Using EEG to Improve Massive Open Online Courses Feedback Interaction

Haohan Wang, Yiwei Li, Xiaobo Hu, Yucong Yang, Zhu Meng, Kai-min Chang

Language Technologies Institute
School of Computer Science
Carnegie Mellon University

Abstract. Unlike classroom education, immediate feedback from the student is less accessible in Massive Open Online Courses (MOOC). A new type of sensor for detecting students' mental states is a single-channel EEG headset simple enough to use in MOOC. Using its signal from adults watching MOOC video clips in a pilot study, we trained and tested classifiers to detect when the student is confused while watching the course material. We found weak but above-chance performance for using EEG to distinguish when a student is confused or not. The classifier performed comparably to the human observers who monitored student body language and rated the students' confusion levels. This pilot study shows promise for MOOC-deployable EEG devices being able to capture tutor relevant information.

Keywords: MOOC, EEG, confuse, feedback, machine learning

1 Introduction

In recent years, there is an increasing trend towards the use of Massive Open Online Courses (MOOC), and it is likely to continue [1]. MOOC can serve millions of students at the same time, but it has its own shortcomings. In [2], Thompson studied post-secondary students who had negative attitudes toward correspondence-based distance education programs. The results indicate that lack of immediate feedback and interaction are two problems with long-distance education. Current MOOC can offer interactive forums and feedback quizzes to help improve the communication between students and professors, but the impact of the absence of a classroom is still being hotly debated. As also discussed in [3], indicates the lack of feedback is one of the main problems for student-teacher long distance communication.

There are many gaps between online education and in-class education [4] and we will focus on one of them: detecting students' confusion level. Unlike in-class education, where a teacher can judge if the students understand the materials by verbal inquiries or noticing their body language (e.g., furrowed brow, head scratching, etc.), immediate feedback from the student is less accessible in long distance education. We address this limitation by using electroencephalography (EEG) input from a commercially available device as evidence of students' mental states.

The EEG signal is a voltage signal that can be measured on the surface of the scalp, arising from large areas of coordinated neural activity manifested as synchronization (groups of neurons firing at the same rate) [5]. This neural activity varies as a function of development, mental state, and cognitive activity, and the EEG signal can measurably detect such variation. Rhythmic fluctuations in the EEG signal occur within several particular frequency bands, and the relative level of activity within each frequency band has been associated with brain states such as focused attentional processing, engagement, and frustration [6-8], which in turn are important for and predictive of learning [9].

The recent availability of simple, low-cost, portable EEG monitoring devices now makes it feasible to take this technology from the lab into schools. The NeuroSky "MindSet," for example, is an audio headset equipped with a single-channel EEG sensor [10]. It measures the voltage between an electrode that rests on the forehead and electrodes in contact with the ear. Unlike the multi-channel electrode nets worn in labs, the sensor requires no gel or saline for recording and therefore requires much less expertise to position. Even with the limitations of recording from only a single sensor and working with untrained users, a previous study [11] found that the MindSet distinguished two fairly similar mental states (neutral and attentive) with 86% accuracy. MindSet has been used to detect reading difficulty [12] and human emotional responses [13] in the domain of intelligent tutoring systems.

A single-channel EEG device headset currently costs around $99-149 USD, which would be a cost deterant to the free service of MOOC. We suggest that MOOC providers (e.g., Coursera, edX) supply EEG devices to a select group of students. In return, MOOC providers would get feedback on students' EEG brain activity or confusion levels while students watch the course materials. These objective EEG brain activities can be aggregated and augment subjective rating of course materials to provide a simulation of real world classroom responses, such as when a teacher is given feedback from an entire class. Then teachers can improve video clips based on these impressions. Moreover, even though an EEG headset is a luxury device at the moment, the increasing popularity of consumer-friendly EEG devices may one day make it a house-hold accessory like audio headsets, keyboards and mice. Thus, we are hopeful of seeing our proposed solution come to fruition as the market for MOOC grows and the importance of course quality and student feedback increases.

To assess the feasibility of collecting useful information about cognitive processing and mental states using a portable EEG monitoring device, we conducted a pilot study with college students watching MOOC video clips. We wanted to know if EEG data can help distinguish among mental states relevant to confusion. If we can do so by better than chance, then these data may contain relevant information that can be decoded more accurately in the future. Thus, we address two questions:

1. Can EEG detect confusion?
2. Can EEG detect confusion better than human observers?

The rest of this paper is organized as follows. Section 2 describes the experiment design. Section 3 and 4 answers the two research questions, respectively. Finally, Section 5 concludes and suggests future work.

2 Experiment Design

In a pilot study, we collected EEG signal data from 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 introduction of basic algebra or geometry. We also prepare videos that are assumed to confuse a normal college student if a student is not familiar with the video topics like Quantum Mechanics, and Stem Cell Research[1]. 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.

We collected data from 10 students. One student was removed because of missing data due to technical difficulties. An experiment with a student consisted of 10 sessions. We randomly picked five videos of each category and randomized the presentation sequence so that the student could not guess the predefined confusion level. In each session, the student was first instructed to relax their mind for 30 seconds. Then, a video clip was shown to the student where he/she was instructed to try to learn as much as possible from the video. After each session, the student rated his/her confusion level on a scale of 1-7, where 1 corresponded to the least confusing and 7 corresponded to the most confusing. Additionally, there were three student observers watching the body-language of the student. Each observer rated the confusion level of the student in each session on a scale of 1-7. The conventional scale of 1-7 was used. Four observers were asked to observe 1-8 students each, so that there was not an effect of observers just studying one student.

The students wore a wireless single-channel 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. More precisely, the position on the forehead is Fp_1 (somewhere between left eye brow and the hairline), as defined by the International 10-20 system [14]. We used NeuroSky's API to collect the EEG data.

3 Can EEG detect confusion?

We trained Gaussian Naïve Bayes classifiers to estimate, based on EEG data, the probability that a given session was confusing rather than not confusing. We chose this method (rather than, say, logistic regression) because it is generally best for problems with sparse (and noisy) training data [15].

To characterize the overall values of the EEG signals while the students watch the 2 minute video, we computed their means over the interval. To characterize the temporal profile of the EEG signal, we computed several features, some of them typically used to measure the shape of statistical distributions rather than of time series: minimum, maximum, variance, skewness, and kurtosis. However, due to the small number of data points (100 data points for 10 subjects, each watching 10 videos), inclusion of

[1] http://open.163.com/

those features tends to overfit the training data and results in poor classifier performance. As a result, we used the means as the classifier features for the main analysis. **Table 1** shows the classifier features.

Table 1. Classifier features

Features	**Description**	**Sampling rate**	**Statistic**
Attention	Proprietary measure of mental focus	1 Hz	Mean
Meditation	Proprietary measure of calmness	1 Hz	Mean
Raw	Raw EEG signal	512 Hz	Mean
Delta	1-3 Hz of power spectrum	8 Hz	Mean
Theta	4-7 Hz of power spectrum	8 Hz	Mean
Alpha1	Lower 8-11 Hz of power spectrum	8 Hz	Mean
Alpha 2	Higher 8-11 Hz of power spectrum	8 Hz	Mean
Beta1	Lower 12-29 Hz of power spectrum	8 Hz	Mean
Beta 2	Higher 12-29 Hz of power spectrum	8 Hz	Mean
Gamma1	Lower 30-100 Hz of power spectrum	8 Hz	Mean
Gamma2	Higher 30-100 Hz of power spectrum	8 Hz	Mean

To avoid overfitting, we used cross validation to evaluate classifier performance. We trained student-*specific* classifiers on a single student's data from all but one stimulus block (e.g., one video), tested on the held-out block (e.g., all other videos), performed this procedure for each block, and averaged the results to cross-validate accuracy within reader. We trained *student-independent* classifiers on the data from all but one student, tested on the held-out student, performed this procedure for each student, and averaged the resulting accuracies to cross-validate across students.

We use two ways to label the mental states we wish to predict. One way is the *pre-defined* confusion level according to the experiment design. Another way is the *user-defined* confusion level according to each user's subjective rating.

Detect pre-defined confusion level. We trained and tested classifiers for pre-defined confusion. Student-specific classifiers achieve a classification accuracy of 67% and a kappa statistic of 0.34, whereas student-independent classifiers achieve a classification accuracy of 57% and a kappa statistic of 0.15. Both classifier performances were statistically significant better than a chance level of 0.5 ($p < 0.05$). **Fig. 1a)** plots the classifier accuracy for each student. **Fig. 1a)** shows that both student-specific classifiers and student-independent classifiers performed significantly above chance in 6 out of 9 students.

Detect user-defined confusion level. We also trained and tested classifiers for student-defined confusion. Since students have different sense of confusing, we mapped the seven scale self-rated confusion level into a binary label, with roughly equal number of cases in the two classes. A middle split is accomplished by mapping scores less than or equal to the median to "not confusing" and the scores greater than the median are mapped to "confusing". Furthermore, we used random undersampling of the larger class(es) to balance the classes in the training data. We performed the sampling 10 times to limit the influence of particularly good or bad runs and obtain a stable measure of classifier performance.

Student-specific classifiers achieve a classification accuracy of 57% and a kappa statistic of 0.13, whereas student-independent classifiers achieve a classification accuracy of 51% and a kappa statistic of -0.04. The student-specific classifier performance was statistically significant and better than a chance level of 0.5 ($p < 0.05$), but not the student-independent classifier. **Fig. 1b)** plots the accuracy for each student. **Fig. 1b)** shows that the student-specific classifier performed significantly above chance for 5 out of 9 students and student-independent classifier performed significantly above chance for 2 out of 9 students.

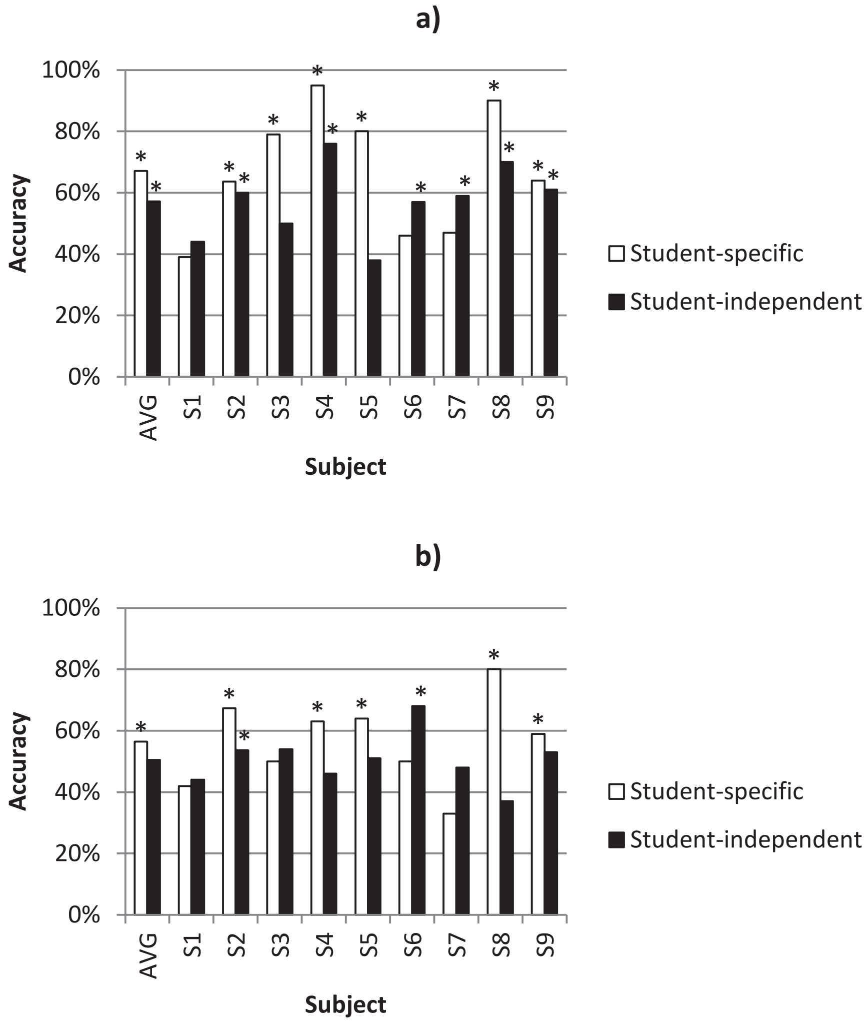

Fig. 1. Detect a) predefined, and b) user-defined confusion level

4 Can EEG detect confusion better than human observers?

To determine if EEG can detect confusion better than human observers of body language, we compared the scores from the observers, the classifier, and the students, with the label of videos. For each student, we used the average scores of the observers as the ‘observer rating’. We used the classifier trained in Section 3 to predict predefined confusion level and linearly mapped the classifier’s estimate of class probability (0-100%) to a scale of 1-7 and labeled it as the ‘classifier rating’.

Fig. 2 shows the scatter plot of a) student vs. observer rating, and b) student vs. classifier rating. The classifier rating had a low, but positive correlation (0.17) with the students’ rating, while the observer rating had a low, but positive correlation of (0.17) with the students’ rating. This shows that the classifier performed comparably to the human observers who monitored student body language and rated the students’ confusion levels.

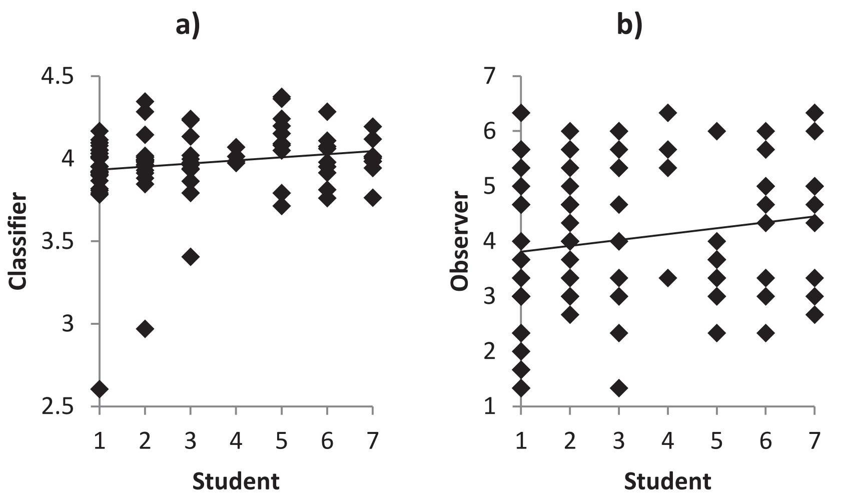

Fig. 2. Scatter plot of a) classifier vs. student rating, and b) observer vs. student rating

5 Conclusions and Future Work

In this paper, we described a pilot study, where we collected students’ EEG brain activity while they watched MOOC video clips. We trained and tested classifiers to detect when a student was confused. We found weak but above-chance performance for using EEG to distinguish whether a student is confused. The classifier performed comparably to the human observers who monitored student body language and rated the students’ confusion levels.

Since the experiment was based on a class project run by a group of graduate students, there were many limitations to the experiment. We now discuss the major limitations and how we plan to address them in future work.

One of the most critical limitations is the definition of experimental construct. Specifically, our pre-defined "confusing" videos could be confounded. For example, a student may not find a video clip on Stem Cell to be confusing when the instructor clearly explains the topic. Also, the predefined confusion level may be confounded with increased mental effort / concentration. To explore this issue, we examined the relationship between the predefined confusion level and the subjective user-defined confusion level. The students' subjective evaluation of the confusion level and our predefined label has a modest correlation of 0.30. Next, we performed a feature selection experiment among all combinations of 11 features; we used cross validation through all the experiments and sorted the combinations according to accuracy. Then we found that the user-specific model Theta signal played an important role in all the leading combinations. Theta signal corresponds to errors, correct responses and feedback, suggesting the experimental construct is indeed related to confusion.

Another limitation is due to the lack of psychological professionalism. For example, the observers in our experiment were not formally trained. As a result, the current scheme allowed each observer to interpret a student's confusion level based on his/her own interpretation. A precise labeling scheme would yield more details that could be compared among raters and, thereby, improve our rating procedure.

Another limitation is the scale of our experiment as we only performed the experiments with 10 students, and each student only watched 10 two-minute video clips. The limited amount of data points prevents us from drawing any strong conclusions about the study. We hope to scale up the experiment and collect more data.

Finally, this pilot study shows positive, but weak classifier performance in detecting confusion. The weak classifier performance may have many false-alarms and thereby frustrate a student. In addition, a student may not be willing to share their brain activity data due to privacy concerns. We are hopeful that the classifier accuracy can be improved once we conduct a more rigorous experiment, by increasing the study size, and improve the classifier with better feature selection and by applying denoising techniques to improve signal-to-noise ratio. Lastly, the classifiers are supposed to help students, so the students should be able to choose not to use EEG if they think the device is hindering.

Acknowledgments. This work was supported by the National Science Foundation under Cyberlearning Grant IIS1124240. The opinions expressed are those of the authors and do not necessarily represent the views of the Institute, or the National Science Foundation. We thank Jessica Nelson for help with experiment design, Donna Gates for preparation of the manuscript, and the students, educators, and LISTENers who helped generate, collect, and analyze our data.

Reference

1. Allen, I.E., Seaman, J., *Going the Distance: Online Education in the United States, 2011*, 2011.
2. Thompson, G., *How Can Correspondence-Based Distance Education be Improved?: A Survey of Attitudes of Students Who Are Not Well Disposed toward Correspondence Study.* The Journal of Distance Education, 1990. **5**(1): p. 53-65.
3. Shute, V., et al. *Assessment and learning in intelligent educational systems: A peek into the future*. in *Proceedings of the 14th International Conference on Artificial Intelligence in Education Workshop on Intelligent Educational Games*. 2009. Brighton, UK.
4. Vardi, M.Y., *Will MOOCs destroy academia?*, in *Communications of the ACM*2012. p. 5.
5. Niedermeyer, E., Fernando H. Lopes da Silva, F. H., *Electroencephalography: basic principles, clinical applications, and related fields*2005: Lippincott Williams & Wilkins.
6. Marosi, E., et al., *Narrow-band spectral measurements of EEG during emotional tasks.* International Journal of Neuroscience, 2002. **112**(7): p. 871-891.
7. Lutsyuk, N.V., E.V. Éismont, and V.B. Pavlenko, *Correlation of the characteristics of EEG potentials with the indices of attention in 12- to 13-year-old children.* Neurophysiology, 2006. **38**(3): p. 209-216.
8. Berka, C., et al., *EEG correlates of task engagement and mental workload in vigilance, learning , and memory tasks.* Aviation, Space, and Environmental Medicine, 2007. **78 (Supp 1)**: p. B231-244.
9. Baker, R., et al., *Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments.* International Journal of Human-Computer Studies, 2010. **68**(4): p. 223-241.
10. NeuroSky, *Brain wave signal (EEG)*, 2009, Neurosky, Inc.
11. NeuroSky, *NeuroSky's eSense™ meters and dtection of mntal sate*, 2009, Neurosky, Inc.
12. Mostow, J., K.M. Chang, and J. Nelson. *Toward exploiting EEG input in a Reading Tutor*. in *15th International Conference on Artificial Intelligence in Education*. 2011. Auckland, New Zealand: Lecture Notes in Computer Science.
13. Crowley, K., et al., *Evaluating a brain-computer interface to categorise human emotional response* in *10th IEEE International Conference on Advanced Learning Technologies*2010: Sousse, Tunisia. p. 276-278.
14. Jasper, H.H., *The ten-twenty electrode system of the International Federation.* Electroencephalography and Clinical Neurophysiology, 1958. **10**: p. 371-375.
15. Ng, A.Y. and M.I. Jordan. *On Discriminative vs. Generative Classifiers: A comparison of logistic regression and naive Bayes* in *Advances in Neural Information Processing Systems* 2002. MIT Press.