Guess or Not? A Brain-Computer Interface Using EEG Signals for Revealing the Secret behind Scores

Tao Xu
School of Software and Microelectronics
Northwestern Polytechnical University
Xi'an, Shaanxi, China
xutao@nwpu.edu.cn

Yun Zhou
School of Education
Shaanxi Normal University
Xi'an, Shaanxi, China
zhouyun@snnu.edu.cn

Yuhan Wang
Zichen Zhao
School of Software and Microelectronics
Northwestern Polytechnical University
Xi'an, Shaanxi, China
xwyh@mail.nwpu.edu.cn
zzcnetwork@gmail.com

Shiqian Li
School of Education
Shaanxi Normal University
Xi'an, Shaanxi, China
qc@snnu.edu.cn

ABSTRACT

Now examinations and scores serve as the main criterion for a student's academic performance. However, students use guessing strategies to improve the chances of choosing the right answer in a test. Therefore, scores do not reflect actual levels of the student's knowledge and skills. In this paper, we propose a brain-computer interface (BCI) to estimate whether a student guesses on a test question or masters it when s/he chooses the right answer in logic reasoning. To build this BCI, we first define

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05.
https://doi.org/10.1145/3290607.3312904

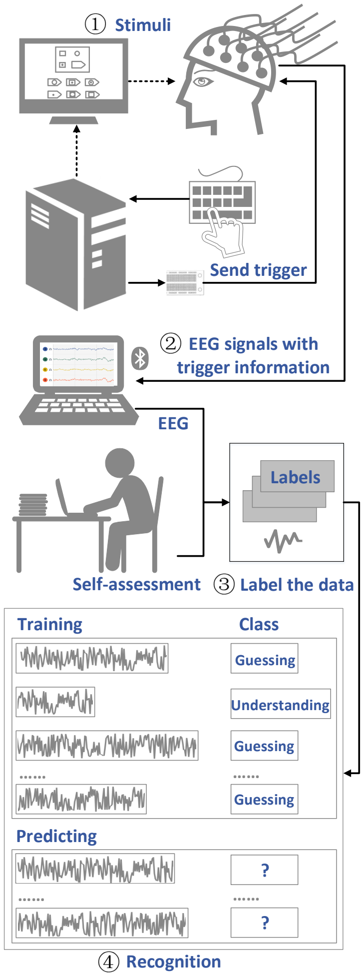

Figure 1: The methodology and BCI design

the "Guessing" and employ Raven's Progressive Matrices as logic tests in the experiment to collect EEG signals, then we propose a sliding time-window with quorum-based voting (STQV) approach to recognize the state of "Guessing" or "Understanding", together with FBCSP and end-to-end ConvNet classification algorithms. Results show that this BCI yields an accuracy of 83.71% and achieves a good performance in distinguishing "Guessing" from "Understanding".

CCS CONCEPTS

• **Human-centered computing → Interaction paradigms**; • **Computing methodologies → Cognitive science**.

KEYWORDS

EEG data; brain-computer interface; learning; guessing; reasoning; cognition

ACM Reference Format:

Tao Xu, Yun Zhou, Yuhan Wang, Zichen Zhao, and Shiqian Li. 2019. Guess or Not? A Brain-Computer Interface Using EEG Signals for Revealing the Secret behind Scores. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3290607.3312904

INTRODUCTION

Guessing is used commonly in examinations, when the student is stumped on a test question and cannot proceed. Some may use strategies to gain on the right answer, like looking for grammatical clues, eliminating outliers, choosing the opposite answer if surrounding answers are the same. Some may randomly choose an answer. These behaviors make it probably to hit the right answer. Therefore, scores do not fully reflect a student's actual levels on mastering knowledge and skills. It is inaccurate to merely use scores of tests to evaluate whether a student masters learning contents.

Logic reasoning distinguishes humans as a species and helps achieve the superior intellectual performance [1]. In most of tests, logic reasoning is involved unavoidably, which plays an important role in the development of knowledge related to STEM (Science, Technology, Engineering, and Mathematics) disciplines. In this work, we propose a brain-computer interface (BCI) to estimate whether a student guesses on a test question or masters it when s/he chooses the right answer in logic reasoning. To build this BCI with a high performance, two key issues are supposed to be addressed. The first issue is that how to define the "Guessing" and guarantee that the "Guessing" can be induced precisely in the experiment. The second one is that how to distinguish the "Guessing" from "Understanding" when the learner chooses the right answer.

The individual would experience confused states when s/he is uncertain about the answer. Therefore, confused state is a key index related to the "Guessing". To solve the first issue, we use confused state

	X = 1: Correct answer
Y = 1: Confused	Y = 1|X = 1: Guessing
Y = 0: Non-confused	Y = 0|X = 1: Understanding

Figure 2: The paradigm to define "Guessing" and "Understanding".

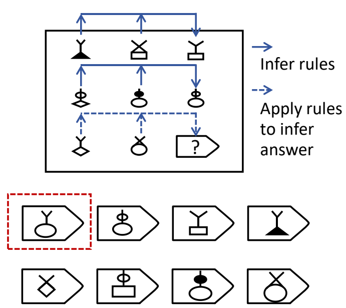

Figure 3: An example test created by us to illustrate the form and rules of RPM. In a test, a matrix of figures is presented with one piece missing, and the task is to select from six to eight given choices according to reasoning. The visual geometric design is mostly in the form of a 3×3 matrix. These tests have little dependency on language abilities, education and culture background. It is suitable for subjects spanning large range of age and profession, and makes them focus on reasoning.

to define the "Guessing" and combine item answers with the subject's self-reported data to label EEG data for classification. We employ the test of Raven's Progressive Matrices [5] (RPM) to guarantee that the "Guessing" can be induced precisely in the experiment. RPM test is a family of standardized and nonverbal intelligence test, which can be typically used in educational settings to measure the taker's abstract reasoning ability.

Previous studies [6][7] show that confused states can be measured by electroencephalogram (EEG). In our experiment, we record EEG signals, which are multi-channeled, temporal, and mixed with noises. Specially, EEG signals collecting from different questions and subjects are of variable length since the answering time of each question and each person is not fixed. Classification algorithms cannot be applied directly to recognize the state of "Guessing" or "Understanding". Therefore, we propose a sliding time-window with quorum-based voting approach (STQV) to process the data of variable length. Besides, we employ both conventional and end-to-end classification algorithms as classifiers. Results show that this BCI with the proposed approaches achieves a good performance in predicting whether an item of the logic test is answered through "Guessing" or "Understanding", which yields an accuracy up to 83.71%.

METHODOLOGY AND BCI DESIGN

As shown in Fig. 1, the methodology of our BCI is composed of four steps, including evoking "Guessing" in the logic test in the experiment, collecting EEG signals with trigger information from acquisition device, labeling EEG signals with self-reported data, preprocessing the data and building classifier to recognize the "Guessing". In this section, we mainly illustrate and discuss the core parts in the methodology, that is, the designed experiment to define and evoke "Guessing", and the proposed STQV approach to recognize "Guessing".

Define and evoke "Guessing"

The learner chooses the right answer in logic reasoning tests, either through: 1) guessing, called "Guessing" in our paper; or 2) mastering related concepts and skills to work out the answer, called "Understanding". Our hypothesis is that the learner's guessing behavior can be identified and distinguished from "Understanding" through our BCI. Fig. 2 demonstrates the paradigm to define the "Guessing". The variable X means that the item is answered either correctly or incorrectly in the experiment, which is noted as 1 or 0. The subject is asked to write down whether s/he feels confused or not for each question in a post questionnaire. The variable Y refers to the confused state corresponding to each item, the value of which is 1 or 0, representing confused or non-confused respectively. Given that X is observed to be the value 1, if Y is equal to 1, we assume that guessing occurs. If Y is equal to 0, we believe that the learner understands and works out the question. Since the guessing behavior of the incorrect answers does not impact the score, in this paper we only focus on the condition that the

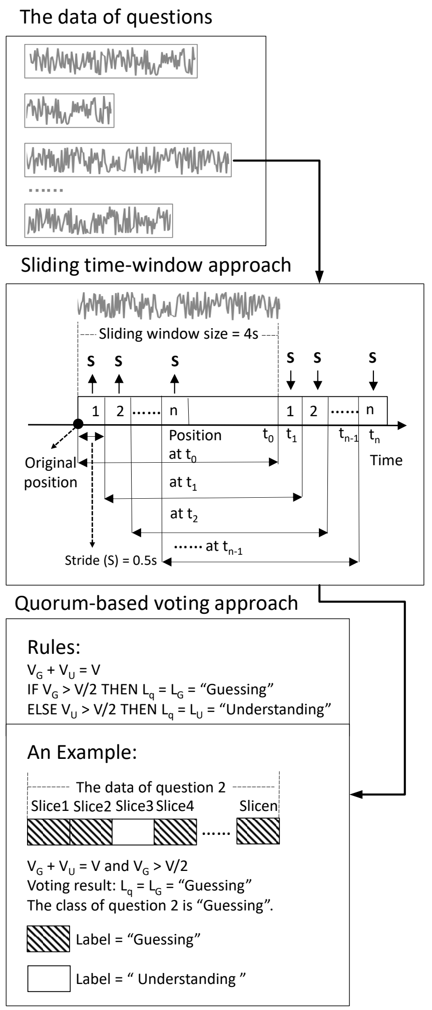

Figure 4: The sliding time-window with quorum-based voting approach.

learner chooses the right answer. In this way, we combine item answers with subjects' self-reported data to label EEG signals for classification.

Then, the difficulty is how to choose an appropriate logic test as stimuli in the experiment. First, this test should be able to apply to everyone, the scores of which should not be influenced by culture and knowledge. Second, the item difficulty index in our experiment is supposed to be difficult to evoke guessing. Item difficulty index is the proportion of students who answered an item correctly [3]. Finally, to reduce maximally the movement cause by interaction, the test should only contain multiple choice questions with the single-answer. Based on this idea, we employ Raven's Progressive Matrices [5] (RPM) of tests to design the experiment (see Fig. 3). In this experiment, we selected 48 items and the answering time of each item is limited to 15 seconds. The item difficulty index ranging from 0 to 0.83 with an average of 0.27, which shows a bias towards higher difficulty.

Recognize "Guessing"

To recognize "Guessing", we propose a sliding time-window with quorum-based voting (STQV) approach (see Fig. 4). We segment EEG data of variable length into fixed-length pieces, predict the class of each piece, and determine the class of each question through voting from pieces.

The sliding time-window (ST) approach is used to mainly overcome the issue of variable length input. Besides, it solves the small sample problem and effectively increases the number of data. EEG signals are segmented into a number of 4-second time window with an overlap of 3.5s between two successive time windows. The stride is 0.5s and each slice is 4s. These slices are used as the input for classification, and each is labeled a class: "Guessing" or "Understanding". In the testing step with ST approach, we acquire the classification result of each slice rather than that of the question item.

The quorum-based voting approach is used to determine the class of the question item. Each slice is assigned a vote, based on its class value. Each question item then has to obtain a Guessing quorum (V_G) or an Understanding quorum (V_U), respectively. A given question item has a total of V votes. To ensure that the value of V is odd, we dispose the last extra small segments away from integral number of seconds. The quorums should obey the rules (see Fig. 4). These rules ensure that a question item does not occur in both states of "Guessing" and "Understanding" concurrently and only one state wins.

CONFIGURATION AND EXPERIMENT

We recruited twenty-three paid subjects, including 11 females and 12 males. Their ages were distributed between 20 and 47 years of age (Mean = 24.48, SD = 6.36). All subjects had normal or corrected vision and were right-handed. There was a bias towards higher education. In the experiment, first, the tester explained this study and guided the subject to read and sign the consent form. Second, the subject wore the OPENBCI headset (Fig. 5) and had a 150-second of brain resting with watching

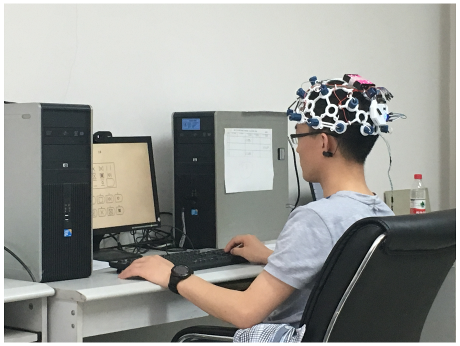

Figure 5: The experiment and OpenBCI device. We adapt OpenBCI Cyton board with the 3D printed headset to acquire raw EEG signals and the data are delivered to the computer via Bluetooth. The neuroheadset feature 8 channels (Fp1, Fp2, C3, C4, T5, T6, O1 and O2) plus 2 references (A1 and A2) based on the 10-20 format. The sampling rate is 250. We develop the trigger function and hardware parts to help segment the data precisely.

Table 1: Decoding accuracies of FBCSP and end-to-end ConvNets.

	FBCSP			
Acc.	**SVM**	**LDA**	**NBPW**	**CNN**
S	81.22%	82.04%	82.04%	86.26%
Q	83.71%	82.62%	80.65%	83.58%

Acc.: Accuracy.
S: Slices. Q: Questions.
CNN: ConvNets.

10 scenery pictures. Then s/he performed the task of watching the stimuli coded by E-Prime 2.0, including 48 Raven's tests. E-Prime also recorded answers of tests. One computer presented stimuli and another laptop recorded EEG signals, with the trigger function synchronizing time stamps. Third, after finishing watching, the subject filled out the questionnaire, including basic information, the self-assessment of confusion levels for each test, and the explanation of his/her choices.

DATA ANALYSIS, RESULTS AND DISCUSSION

We applied STQV approach to recognize the state of "Guessing" or "Understanding", together with FBCSP and end-to-end ConvNet classification algorithms. EEG data were divided by subjects to increase the generalization ability of the classifier in our BCI. The data of 16 subjects were randomly sampled to build the training set, and the data of 7 subjects were used for testing set.

With regard to conventional classification approach, we employed Filter Bank Common Spatial Pattern (FBCSP) approach [2] to process EEG signals, which has been widely used to decode oscillatory EEG data. This approach contains four steps including multiple bandpass filters, spatial filtering using the Common Spatial Pattern (CSP) algorithm, feature selection of the CSP features, and classification of the selected CSP features. We chose three mainstream methods for classification: Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and Naïve Bayesian Parzen Window (NBPW). Each slice was assigned a predicted class, being either "Guessing" or "Understanding" (see Fig. 6). Next, the class of a question was predicted using the quorum-based voting approach.

The end-to-end approach [4] relates to deep learning algorithms, which leverages raw data for classification without considering handcrafted features. It replaces multiple steps with just a single neural network, reducing the process of feature extraction. We employed end-to-end learning with convolutional neural networks (ConvNets). As shown in Fig. 7, through a sliding time-window, we segmented EEG data and obtained slices as the input. In order to avoid the impact of individual differences in EEG, the normalization process is adopted at first, using Z-score standardization method. After normalizing, we used a four-layer convolutional neural networks to predict the class of each slice. Then each question was predicted using the quorum-based voting approach.

We evaluated our STQV approach in two steps. The first step is to test the performance of different classifiers on EEG data slices, which were generated from EEG raw data by the sliding time-window approach. The second one is to measure the accuracy of the predicted class of each question using the data in the testing set.

Table 1 shows results of the FBCSP with the SVM, LDA and NBPW classification algorithms, as well as end-to-end ConvNets. Overall, in predicting the class of EEG slices and questions, all of the four classification algorithms with STQV approach yield inspiring test accuracies, which are above 80%. With regard to predicting the class of EEG slices, the accuracy of the end-to-end approach reaches up to 86.26%, which is higher than that of three other classification approaches (SVM, LDA, and NBPW)

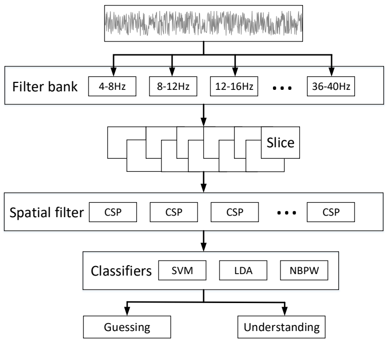

Figure 6: Decoding process of FBCSP machine learning approach.

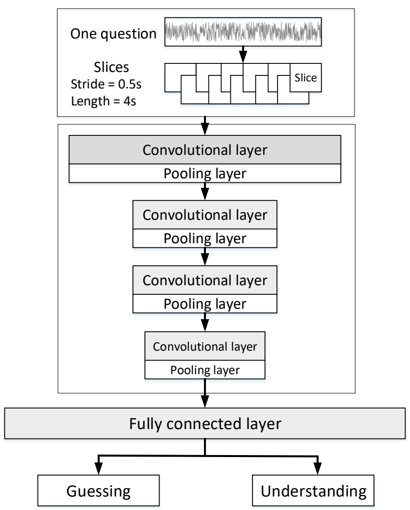

Figure 7: Architecture of ConvNets approach.

with FBSCP. With regard to predicting the class of question, SVM with FBCSP performs best, which is slightly better than the end-to-end approach, yielding an accuracy of 83.71%.

Preliminary experimental results proved our hypothesis. Therefore, our BCI with the proposed STQV approach is capable to distinguish whether the student guesses on a test question or masters it in logic inference tests. Besides, it is worth emphasizing that subjects assigned in testing set are different from subjects in training set, which shows the powerful generalization ability of this BCI.

CONCLUSION AND FUTURE INVESTIGATION

In this work, we proposed a BCI to estimate whether a student guesses on a test question or masters it when s/he chooses the right answer. Addressing the questions as we stated in the introduction, the main contributions of this study are as follows. First, we defined the "Guessing" state and designed an experiment to elicit "Guessing" in logic tests. Second, we proposed a STQV approach to process the variable input and output in length for classification. The performance of this approach combining with FBCSP and end-to-end ConvNets classification algorithms shows powerful recognition of student's guessing in tests. Thus, in summary the approaches and findings described in this study provide a potential mean to explore the examinee's true levels and pave the way for EEG decoding in education applications. In the future, we attempt a profound investigation to epistemic changes in the learning process using BCI and other intelligent user interfaces.

ACKNOWLEDGEMENTS

This work was supported by projects of the National Natural Science Foundation of China (61703259 and 61702417), and the Natural Science Foundation of Shaanxi Province (2017JM6097).

REFERENCES

[1] John R. Anderson. 2000. *Cognitive psychology and its implications.* WH Freeman, New York. 237–259 pages.

[2] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan. 2008. Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence).* 2390–2397. https://doi.org/10.1109/IJCNN.2008.4634130

[3] Ronald Jay Cohen, Mark E. Swerdlik, and Suzanne M. Phillips. 1996. *Psychological testing and assessment: An introduction to tests and measurement, 3rd ed.* Mayfield Publishing Co, Mountain View, CA, US.

[4] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521 (2015), 436–444.

[5] John Raven. 2000. The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology* 41, 1 (2000), 1–48.

[6] Jingkang Yang, Haohan Wang, Jun Zhu, and Eric P. Xing. 2016. SeDMiD for Confusion Detection: Uncovering Mind State from Time Series Brain Wave Data. In *2016 Conference on Neural Information Processing Systems, Time Series Workshop.*

[7] Yun Zhou, Tao Xu, Shiqian Li, and Shaoqi Li. 2018. Confusion State Induction and EEG-based Detection in Learning. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* 3290–3293. https://doi.org/10.1109/EMBC.2018.8512943