Classification of Brain Signals Collected During a Rule Learning Paradigm

Alicia Howell-Munson[1(✉)], Deniz Sonmez Unal[2], Theresa Mowad[3], Catherine Arrington[3], Erin Walker[2], and Erin Solovey[1]

[1] Worcester Polytechnic Institute, Worcester, MA 01609, USA
ahowell4565@gmail.com
[2] University of Pittsburgh, Pittsburgh, PA 15260, USA
[3] Lehigh University, Bethlehem, PA 18015, USA

Abstract. We propose incorporating biophysical data with behavioral data to inform digital learning environments on an individual's current cognitive state and how it relates to their learning. We used a rule learning paradigm drawn from cognitive psychology to define phases of rule learning across multiple domains. This paradigm can simulate an inductive reasoning framework seen during mathematics education while reducing the number of covariates compared to real-world settings. We combined the time series brain data with behavioral and contextual data in machine learning models for prediction of rule learning phases with the aim of developing approaches to incorporate a mixture of behavioral and neural data into digital learning designs.

Keywords: rule learning · inductive reasoning · functional near-infrared spectroscopy · brain-computer interfaces

1 Introduction

In realistic learning environments, students encounter multiple domains where they need to learn rules through inductive reasoning [4]. The mechanisms for inductive reasoning involve gathering information, generalizing it, classifying it, and chunking it, then recalling it. These mechanisms are collectively labeled as induction and refinement, a class of processes that lead to robust learning [6]. We aim to use brain data to improve knowledge modeling in intelligent tutoring systems (ITSs). While knowledge modeling from log data is well-studied in ITSs [8], it is still not possible to detect the precise moment a student has learned a rule [3]. Brain data may supplement behavioral data to provide a fine grained indication of when and to what degree a student has mastered a rule. Our first step is to demonstrate that brain signals can be used to recognize three stages of inductive reasoning as described in a cognitive science rule learning paradigm [5]. These stages form the fundamental mechanisms of induction and refinement processes which are the basis of important features of ITSs. Our goal is to expand upon this work and develop machine learning models that can classify a rule learning state of students and that transfer across task domains.

The National Science Foundation under Grant Nos. (1835307 and 1912474).

N. Wang et al. (Eds.): AIED 2023, CCIS 1831, pp. 625–631, 2023.
https://doi.org/10.1007/978-3-031-36336-8_97

2 Related Literature

2.1 Neural Processes During Rule Learning Tasks. During rule learning, individuals move through phases of rule search, rule discovery, and rule following [5]. These phases can be studied in controlled laboratory tasks that measure behavior across a series of trials in standard rule learning tasks [4]. Stimuli such as numbers or spatial locations appear in a continuous stream and alternate between random non rule steps and rule sequences that follow a set rule or pattern. Participants must indicate with a key press whether the current stimulus matches a rule that they have discovered through inductive reasoning. Rule search is defined as when the participant is responding that they do not think the stimuli are following a pattern. Rule discovery is the first response where the participant indicates the stimuli are following a pattern. Finally, rule following is the continued response that the stimuli are following a rule. Prior work on rule learning has shown bilateral activation in the prefrontal cortex (PFC) with shifts between the medial, inferior, and posterior PFC regions between rule acquisition and rule following phases [4,5]. Similar brain patterns should be observed in learning tasks at points where students are just learning a rule and at points where they have mastered the rule, paralleling the concepts of induction, refinement, and fluency in cognitive states related to learning [4].

2.2 Neuroimaging and Usage of Neural Data with ITSs. Neuroimaging tools provide measures of brain activity while participants engage, in many tasks. Functional near-infrared spectroscopy (fNIRS), unlike more common tools like fMRI, is portable, easy to use, and quick to set up [9]. The fNIRS optodes are arranged on a mesh cap that emits near-infrared light into the head, then measures how much of the light was absorbed by the blood and tissue in the head with a detector. fNIRS typically has a range of 1–3cm into the cortex and the response time for peak signal after an event occurs is 4-7 s [9].

Researchers have used brain imaging data to study the neural processes underlying learning and to inform the design and development of intelligent tutoring systems. For example, Anderson and colleagues showed functional magnetic resonance imaging (fMRI) can be used to identify deep processing behaviors during problem-solving [1]. Others used the electroencephalogram (EEG) for understanding student engagement [10] during use of an ITS.

3 Data Collection and Curation

We conducted a study to build a dataset that can be used to explore whether fNIRS brain signals can be used to recognize stages of inductive reasoning. To do this, we expand on prior fMRI studies [5]. We aim to demonstrate to what extent it is possible to develop machine learning models that can classify the rule learning state of students, that transfer across task domains, and that do not require the student to provide individual data prior to using the model.

3.1 Rule Learning Task. We created isomorphic tasks in *numeric* and *spatial* domains for cross domain comparisons. Stimuli appeared sequentially on the screen with a fixation, '+', between stimuli. The *numeric* task involved the numbers 0 to 99; the *spatial* task involved a filled circle in one of 10 locations positioned in a circle. Stimuli shifted position alternating between sequences of 5-11 trials that followed a rule and 2-4 trials that occurred randomly within displacement between -4 and 4. Easy rules involved one step while hard rules followed two steps. A list of all 16 rules and their features are in Table 1. In the spatial task, subtraction steps moved counterclockwise, while addition steps moved clockwise. The order of the domain tasks was counterbalanced.

Table 1. The selected hard and easy rules all participants saw with examples.

Easy	Rule	+1	+2	+3	+4	- -1	–2	–3	–4
	Example	5, 6, 7	5, 7, 9	5, 8, 11	5, 9, 13	5, 4, 3	5, 3, 1	5, 2, 99	5, 1, 96
Hard	Rule	+1, +2	+2, +3	-2, -1	-3, -2	-2, +1	-2, +3	+2, -1	+2, –3
	Example	5, 6, 8	5, 7, 10	5, 3, 2	5, 2, 99	5, 3, 4	5, 3, 6	5, 7, 6	5, 7, 4

Our data processing approach followed the procedure outlined in [5]. Non-rule trials were removed from the analysis. Whenever a participant responded "f", the trial was coded as *rule search*. For rules that were successfully discovered, the first trial the participant responded "j" was coded as *rule discovery* and all subsequent trials were *rule following* until the rule ended. Rule sequences were coded as discovered if they responded "j" in the last two trials of the rule.

3.2 Equipment. The fNIRS signals were recorded using a NIRx NIRSport2 fNIRS device with a sampling rate of 10.17Hz. The device was configured with a fifty-two channel design using sixteen sensors and fifteen detectors (Fig. 1). A note was left in a visible area of the workstation that stated to press "j" if they thought it was a rule sequence and to press "f" if they thought it was not.

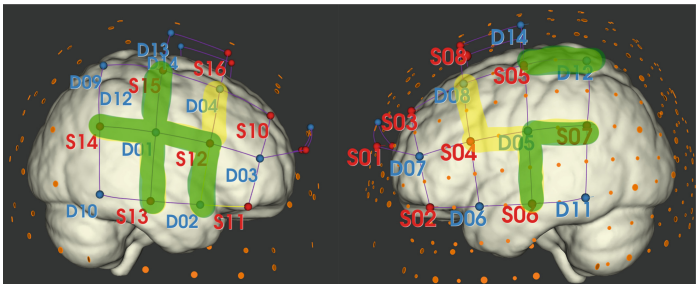

Fig. 1. Sensors (in red) emit light and detectors (in blue) receive light. Highlighted are the channels that correspond to regions of interest for rule acquisition (yellow) and rule following (green) based on previous work [5]. (Color figure online)

3.3 Participants. We recruited 22 university students as participants (5 male, 13 female, and 4 other) between 18 and 23 years old (M = 20.09, SD = 1.6). 71% of participants identified as white, 14% as Asian, 14% as mixed ethnicity, and 6% as black or African American. Participants were compensated with either coursework credit or monetary payment of $15.00. All participants signed an informed consent that described the procedure, compensation, and risks of the study. The informed consent was approved by the university's IRB. Seven participants were excluded from analysis due to errors in data collection, excessively noisy brain data, or failure to complete the task correctly, leaving 15 participants in the final analysis.

3.4 fNIRS Preprocessing. Channels were removed due to noise if they exceeded a 15% coefficient of variance threshold. A bandpass filter was applied to the raw data with a high cutoff frequency of 0.2 Hz and a low cutoff frequency of 0.01 Hz. We used the modified Beer-Lambert Law to convert the raw data into change in micromolar oxygenation with a differential pathlength factor of 7.25 for oxygenated data and 6.38 for deoxygenated data.

3.5 Dataset Overview for Machine Learning. We separated the data from the numeric and spatial domains into separate datasets to compare the performance of each machine learning model on the individual domain, combined domains, and cross-domain training and testing. The goal was to determine if the information domain affected a model's performance Each dataset contained two classes: rule acquisition and rule following. The datasets were marginally unbalanced with the acquisition class containing approximately 6% more trials. We selected fNIRS channels that corresponded to regions of activation for rule acquisition and rule following, as shown in Fig. 1 based upon previous results from the spatial task [5]. This resulted in 3 channels corresponding with rule acquisition and eight channels corresponding with rule following. Each time-series was approximately 5 s long, ensuring time for the hemodynamic response to start to peak, but reducing the influence from multiple trials within a time-series.

3.6 Machine Learning Feature and Classifier Selection. Models included a combination of a behavioral feature (response time), contextual features (rule difficulty, time since the rule started, and the unique rule identifier), and neural data (time series subsequences for each fNIRS channel). We used a logistic regression (LR) on the behavioral and contextual data and we used the time-series forest (TSF) classifier from sk-time on the fNIRS data. In addition, we used an ensemble method where each fNIRS channel was fitted to a model, then the channels would vote to determine the final label. For models that include computer logged features, the label from the LR would be included in the voting.

We fitted five separate models which were: behavioral model (BM) that had response times as its only feature, contextual model (CM) that had rule identity, difficulty, rule domain, and trial position in rule sequence as features; behavioral and contextual model (BCM); neural model (NM) that contained only fNIRS channels as features; neural and behavioral model (NBM); and neural, behavioral, and contextual features (NBCM).

4 Results

4.1 Behavioral Results. To confirme the behavioral results matched previous research, we analyzed the frequency of rule discovery in a 2 (Domain) x 2 (Difficulty) repeated measures ANOVA and the response time (RT) in a 3 (Rule learning phase) by 2 (Domain) x 2 (Difficulty) repeated measures ANOVA. We considered the number of rules discovered out of a maximum of 8 per condition. Rule discovery occurred in 78% of all rule sequences. There was a main effect of difficulty, with participants discovering significantly more easy rules, $F(1,21) = 29.105$, $p < 0.001$, $n_p^2 = 0.581$. RTs sped up as participants moved from rule search to discovery to following, $F(2, 42) = 35.3$, $p < .001$, $n_p^2 = 0.627$. The other significant effect was of domain, $F(1, 21) = 15.8$, $p < .001$, $n_p^2 = 0.429$, with participants responding more quickly on spatial trials. These results aligned with previous research, giving validation to perform neural analysis [4,5].

4.2 Computer Logged Machine Learning Results. The computer logged data result in robust F1 scores for all three behavioral and contextual models. Across both numeric and spatial datasets, as well as the combined data and crossed domain training and testing, behavioral data resulted in the worse performance while the combination of behavioral and context data resulted in the best performance. The results from the mixed domain dataset mirrored those of the spatial dataset. Interestingly, the cross domain testing and training data sets resulted in F1 scores that were in line with models trained and tested within domains, suggesting that rule learning phases may share features across different domains.

Table 2. F1 scores on the models trained with combinations of RT, contextual, and neural features. All models performed better than the dummy classifier.

Dataset	BM	CM	BCM	NM	NBM	NBCM	Dummy
Numeric	0.69	0.76	0.78	0.50	0.57	0.69	0.39
Spatial	0.74	0.79	0.82	0.50	0.61	0.74	0.27
Numeric + Spatial	0.74	0.79	0.82	0.51	0.65	0.76	0.35
Numeric Train, Spatial Test	0.69	0.77	0.81	0.46	0.56	0.71	0.35
Spatial Train, Numeric Test	0.73	0.76	0.79	0.50	0.63	0.74	0.39

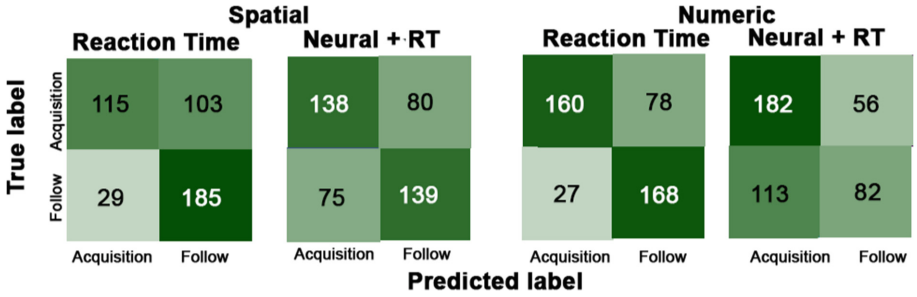

Fig. 2. Confusion matrices of true and predicted labels from the spatial (left) and numeric (right) datasets for BM and NBM. For both, the addition of neural features causes more accurate predictions of the acquisition class.

4.3 fNIRS Machine Learning Results. The classifiers including neural features consistently had a higher F1 score than the dummy classifier (Table 2). Classifiers using brain data are consistently better than random chance. In addition, the neural data improves the misclassification of the rule acquisition phase from the RT features in the spatial dataset (Fig. 2). The neural data similarly correctly predicts more acquisition labels on the numeric dataset; however, in exchange it misclassifies the majority of follow labels. Consistent with the computer logged features, the F1 scores of the cross domain training and testing remained consistent to the datasets of a singular information domain. This continues to support the logic that rule learning phases may share characteristics across task domains, even when incorporating the brain data. These results lead us to predict that brain data, RT, and possible context features would be ideal features for a classifier that works in real-time with an ITS.

5 Discussion

In this paper, we have shown the viability of predicting rule acquisition and rule following phases from fNIRS and behavioral data in a controlled psychological task in both numeric and spatial domains. Computer logged data is the most successful predictor of rule learning phase and should be relied upon when the data is available. However, we have shown that there is promise in using brain data to supplement computer logged data. This supplementation can be applied when computer logged data is either sparse or non-existent to maintain an informed state of the student's learning phase. In this study, we used a group model to predict the rule learning phase of students who were not included in the training data and showed the predictions carried across numeric and spatial domains. We believe that brain data can help fill the gaps during long pauses during ITS use and determine if a long pause is related to engagement [2] or disengagement [7] from the problem. Future work should transition models into real-world applications.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant Nos. (1835307 and 1912474).

References

1. Anderson, J.R., Betts, S., Ferris, J.L., Fincham, J.M.: Cognitive and metacognitive activity in mathematical problem solving: prefrontal and parietal patterns. Cognit. Affect. Behav. Neurosci. **11**(1), 52–67 (2011)
2. Arroyo, I., Mehranian, H., Woolf, B.P.: Effort-based tutoring: An empirical approach to intelligent tutoring. In: Educational Data Mining 2010. Citeseer (2010)
3. Baker, Ryan S. J. D.., Gowda, Sujith M.., Corbett, Albert T.., Ocumpaugh, Jaclyn: Towards automatically detecting whether student learning is shallow. In: Cerri, Stefano A.., Clancey, William J.., Papadourakis, Giorgos, Panourgia, Kitty (eds.) ITS 2012. LNCS, vol. 7315, pp. 444–453. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30950-2_57
4. Cao, B., Li, W., Li, F., Li, H.: Dissociable roles of medial and lateral pfc in rule learning. Brain Behav. **6**(11), e00551 (2016)
5. Crescentini, C., Seyed-Allaei, S., De Pisapia, N., Jovicich, J., Amati, D., Shallice, T.: Mechanisms of rule acquisition and rule following in inductive reasoning. J. Neurosci. **31**(21), 7763–7774 (2011)
6. Koedinger, K.R., Corbett, A.T., Perfetti, C.: The knowledge-learning-instruction (kli) framework: Toward bridging the science-practice chasm to enhance robust student learning. Cognitive Sci. (2010)
7. Muldner, K., Burleson, W., Van de Sande, B., VanLehn, K.: An analysis of students' gaming behaviors in an intelligent tutoring system: Predictors and impacts. User Model. User-Adap. Inter. **21**(1), 99–135 (2011)
8. Pelánek, R.: Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. User Model. User-Adapted Inter., 313–350 (2017). https://doi.org/10.1007/s11257-017-9193-2
9. Solovey, E.T., et al.: Using fnirs brain sensing in realistic hci settings: experiments and guidelines. In: Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology, pp. 157–166 (2009)
10. Stevens, Ronald H.., Galloway, Trysha, Berka, Chris: EEG-related changes in cognitive workload, engagement and distraction as students acquire problem solving skills. In: Conati, Cristina, McCoy, Kathleen, Paliouras, Georgios (eds.) UM 2007. LNCS (LNAI), vol. 4511, pp. 187–196. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73078-1_22