An Exploration of Two Methods for using fMRI to Identify Student Problem Solving Strategies

Caitlin Tenison[1], John R. Anderson[1]

Psychology Department[1]
Carnegie Mellon University, Pittsburgh PA 15213, USA
ctenison@andrew.cmu.edu and ja0s@andrew.cmu.edu

Abstract. Using a math-learning paradigm, we explore two potential uses for fMRI when modeling problem-solving strategies. First, we use fMRI as an additional data source for our model. Second, we employ fMRI as a method of testing and understanding our behavioral models. We evaluate each method and consider which gains the greatest benefit from the inclusion of fMRI data.

1 Introduction

Intelligent tutoring systems (ITSs) emerged from the idea that we can model the cognitive learning processes of students to better inform instruction. Early ITS modeling relied heavily on latency and student productions as responses of underlying cognition; however, with the improved ability to record neural responses, methods like functional magnetic resonance imaging (fMRI) are increasingly used to advance research. In their study of an Algebra ITS, Anderson et al. [1] found that a model using both fMRI and keystroke information better predicted when a student was problem solving than one that relied solely on a single data source. They attribute the models success to the merging information from the brain with behavioral measures. Predicting whether or not a student is problem solving, however, is a simpler task than predicting how a student is problem solving. Additionally, when predicting problem solving strategy there is more variety in the type of behavioral data that can be collected. In this paper, we identify two potential functions of fMRI use in modeling problem solving states. The first is as an additional data source to build better student models, and the second, is as a method for testing and understanding behavioral models. To investigate this question we use a simple math-learning paradigm. As students gain practice problem solving, the strategies that they use change from calculation to retreival. In a previous study we found students also employ intermediary strategies containing a mixture of both retreival and calculation as well [2]. Rather than predicting if a student is problem solving, our models will predict what strategy they are using. We will build two models, one that uses behavioral indicators of strategy use, the other which uses a combination of both behavioral and neurological indicators. We will compare these two methods to assess the value of incorporating fMRI as an additional data source. Finally, we will consider the use of brain data to better understand the behavioral model.

2 Materials and Methods

2.1 Participants

Twenty right-handed university students (9 females; mean age 22; SD 2.3) participated in the study. Participants gave informed written consent and received monetary compensation for their participation.

2.2 Stimuli and Experimental Design

To investigate the change in strategy that occurs when learning a new type of operation, we trained participants on a novel operation. Participants learned how to calculate the value of a pyramid expression b$n by adding n decreasing numbers starting with b [2]. For instance, 11$4=11+10+9+8=38. Participants used the keypad to type out the answers to these problems and to indicate the problem solving strategies that were used. After answering a problem, participants chose from a list of strategies the option that best matched the one used to solve the problem. We compiled the strategy options from the reported strategies in a previous study [2]. Students chose from 4 options: "Retrieve" was defined as remembering the answer; "calculate" was defined as using arithmetic to find the answer; "partial" was described as partially calculating and partially remembering the problem; the "other" strategy was used for anything else but only one participant indicated use of this strategy.

2.3 Scanning Procedure

Participants completed 6 blocks of fMRI scans. Participants completed a concurrent assessment on the 2nd, 4th and 6th scans. The alternating of scans featuring an assessment allowed us to check its reactivity (no reaction was found). Overall there were 3 practiced problems that were repeated 36 times over the course of the experiment and 18 novel problems that were repeated twice. Pyramid problems were presented on the screen following a 2 second fixation period. Once the problem appeared on the screen, the participant was allowed a maximum of 30 seconds to indicate knowledge of a solution by pressing the return key on the numeric keypad. After pressing return, participants input a solution using the keypad and were given correctness feedback. Problem solving time was defined as the time between the appearance of the math problem and the point at which the participant indicated a readiness to input the answer. fMRI images were aquired using gradient echo-echo planar image aquisition on a Siemens 3T Verio Scanner using a 32 channel RF head coil, with 2 s. repetition time (TR). More detailed data aquisition and processing steps is described in Tenison et al. [2].

2.4 fMRI Analysis

To create a single measure of strategy use from the fMRI data we used a classification analysis to quantify how similar a given trial was to other retrieval trials.

Without a direct report of retrieval for all problems, we trained our classifier on the distinction between practiced and novel problems, since we knew novel problem could not be solved by retrieval, whereas many practiced problems would be solved by retrieval. For the purposes of this paper, we will briefly summarize the processing steps applied to our data (again see Tenison et al. [2] for detail on a similar analysis). First, we subdivided the brain into 4x4x4 voxel cubes (a voxel is 3.2 x 3.125 x 3.125mm) over 32 slices of the 64x64 acquisition matrix to create an initial 408 mega-voxel regions of interest (ROIs) [1]. The second step was to eliminate regions that had highly variable fMRI signals. ROIs containing more than 15 TRs across all participants that fluctuated more than 15% during a block were eliminated. The reduced sample comprised 288, 4x4x4 voxel regions of raw data. For the 288 regions, we estimated the activity during problem-solving for each trial and calculated the z-scores of this measure. Normalizing allowed for comparison across subjects. To eliminate fluctuations in the blood-oxygen-level-dependent signal that were physiologically implausible, z-scores were Winsorized such that scores greater than 5 or less than -5 were changed to 5 or -5 respectively. As a third step, we performed dimensionality reduction using Principle Components Analysis (PCA), which creates a set of uncorrelated variables from linear combinations of the ROI activity. We then preformed a linear discriminate analysis (LDA) on the first 50 factors extracted from the PCA to identify which of these factors contributed to distinguishing between practiced and novel problems. We used a leave-one-out cross-validation method where we predicted each subjects from the results of the other subjects. Besides returning a predicted category for each item, an LDA generates a continuously varying evidence measure for category membership and a posterior probability that an item is from a category. These measures were used in subsequent analysis.

3 Results

3.1 Effects of Practice

A repeated measures ANOVA run on the latency data revealed a significant main effect of problem group (practiced vs. novel), $F(1,18)=69.28$, $p<0.0001$, scan block, $F(5,90)=18.66$, $p<0.0001$, and a significant problem by scan block interaction, $F(5,90)=14.95$, $p<0.0001$. The time it took participants to solve practiced problems decreased, but the time to solve novel problems remained constant. Additionally, there was an increase in reports of retrieval with practice, $F(2,38)=42.04$, $p<0.0001$, and a decrease of reports of both computation, $F(2,38)=8.396$, $p=0.001$, and partial strategies, $F(2,38)=18.598$, $p<0.0001$. Novel problems showed no changes in reported strategy use. Averaged across all assessed trials practiced problems were reported as retrieved 81.7% of the time and novel problems were reported as calculated 89% of the time. We took this as evidence that a LDA classifier trained to distinguish between practiced and novel problems would use information similar to what would be used to distinguish between retrieval and calculation (Figure 1). In cross-subject tests, the classifier predicted all subjects better than chance. The average d-prime measure

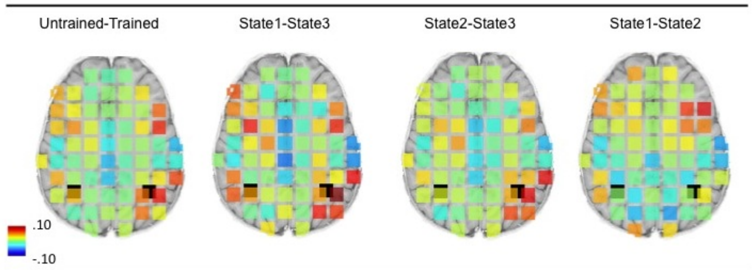

Fig. 1. Four classification analyses are represented here in m-n format. Warm voxels are more active for m problems, cool voxels are more active for n problems. Locus of the HIPS represented by the black square.

of performance in this analysis was 1.71, $t(19) = 14.5$, $p<0.001$, with a hit rate of 60% and a false alarm rate of 11%. The major contribution of this classifier to this study is to label each trial with the probability that it was retrieved. We will use this evidence score as one source of information about strategy use.

3.2 Results from Two Hidden Markov Models

Our first aim of this study is to assess the value of fMRI as an additional data source for modeling strategy changes. We used a Hidden Markov Model (HMM) to study the three practiced problems over the course of the experiment. We ran two HMMs, the first HMM used only behavioral data (the reports, the latency and the accuracy information). Each state was associated with three measures: probability of the three reported strategies, latency (the normalized log latency was modeled as a Gaussian), and the probability of correct solution. We calculated the probability of a student being in a specific strategy-use state after having observed that problem a given number of times using a forward-backwards algorithm. We fit HMMs for 1-10 states, using BIC to penalize for added parameters. The best model had 3 strategy states. Table 1 shows the mean parameter value in each state for the behavioral HMM. The second HMM we used combined the latency and fMRI evidence data along with the accuracy and report data. Since latency and fMRI are highly correlated, we orthogonalize these two measures by use of a PCA. The first component of the PCA proved to carry all the information accounting for 88.7% of the variance. This first component can be taken as a general strength measure and was used, in combination with the other measures, to train the HMM. The right side of Table 1 indicates the mean parameter values of each state for the combination brain behavioral HMM. The HMM generates likelihood of state belonging, we assign each problem to the most likely state. Using this discrete assignment we can compare the two HMMs by looking at the similarity of state assignments. We ran a Cohens Kappa calculation to show the agreement between the two HMMs is above chance ($\kappa =$

Table 1. Average parameter values for the two models

Models	Behavioral			Brain and Behavioral		
	State1	State2	State3	State1	State2	State3
Accuracy (%)	84.2	95.3	98.4	85.7	95.7	98.7
Latency (s)	3.6	1.4	.72	3.3	1.3	.69
fMRI Evidence	.50	-.25	-1.3	.46	-.29	-1.44
Percent Strategy Use						
Calculation	35.7	9.4	1.1	35	6.4	1.1
Partial	52	20.6	.78	56.4	13.3	.74
Retrieval	12.2	69.9	98.1	8.54	80.3	98.1

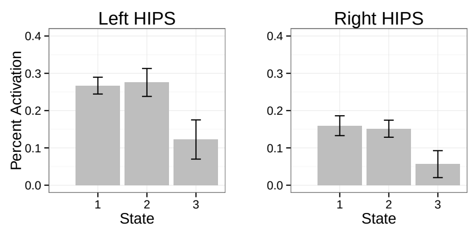

Fig. 2. Mean activation of the Bilateral HIPS. Error bars represent standard error.

0.80) There are no cases in which a problem is assigned as State 1 by one model and State 3 by the other. This evidence suggests the addition of the brain data brings little additional benefit to the state estimation.

3.3 Understanding the States

Our second aim of this study is to explore the potential of fMRI data as a means for understanding the states identified by behavioral models. We ran an LDA similar to the one described in Section 2.4 but this time we classified state assignments from the behavioral HMM. We mapped the weights from the classifier back to the brain in order to observe the areas associated with the classification of the different states (Figure 1) Among the regions used by the classifier to distinguish between states, we identified the horizontal intraparietal sulcus (HIPS), an area used in calculation and numerical cognition. We used the coordinates from a meta-analysis of numerical cognition [3] (Maxima TC: -31, -50, 45) to investigate if there were significant differences in bilateral HIPS activity for the three states. We found significant bilateral differences between State 1 and 3 (Left: $t(19) = 2.6$, $p<0.05$ Right: $t(19) = 2.7$, $p<0.05$) and State 2 and 3 (Left: $t(19) = 4.0$, $p<0.05$ Right: : $t(19) = 3.6$, $p<0.05$) and no differences between State 1 and 2 (Figure 2).

4 Discussion

In this study we put forth two possible methods for using fMRI to inform how we model student strategy use. The first method was to use brain data as a source to incorporate in our models. To this end, the fMRI data did little to change the classification we obtained just from behavioral data. The second method was to use brain data to interpret the latent states identified when modeling behavioral data. Our results bring some insight as to the nature of the three states identified by our model. It is clear from both the behavioral signatures and the brain data that State 1 is a calculation state and State 3 is a retrieval state. The nature of State 2 is less clear, and according to the participant reports State 2 is a retrieval state. However, the brain data suggests that State 2 is more similar to a calculation than a retrieval state. The classification of the brain data associated with the states showed the HIPS, an area used in math calculation, helped distinguish the 3 states. A further ROI analysis of this region verified that the bilateral HIPS showed significantly more activation in State 1 and 2 than State 3, and no difference between States 1 and 2. The direction of the HIPS activation is supporting evidence that States 1 and 2 involve more number processing than State 3 [3]. Future studies could use exploratory analyses, comparisons with the untrained problems, or techniques such as representational similarity analysis to build a more detailed picture of the states. Our goal in this study was to consider how we use fMRI in modeling cognitive states. Previous work found that models using both fMRI and keystroke information, better predicted when students were problem solving [1], however it was unclear if the benefit of fMRI held when identifying shorter cognitive states or using additional behavioral information. Our findings suggest that the high spatial and low temporal resolution of the fMRI is better suited for understanding models rather than building models of brief cognitive states. Future studies could explore how systems with better temporal resolution such as MEG or EEG perform in such scenarios.

Acknowledgments

This work was supported in part by NSF grant DRL-1007945 and by CMU's Program in Interdisciplinary Education Research funded by IES grant R305B090023.

References

1. Anderson, J.R., Betts, S., Ferris, J.L., Fincham, J.M.: Neural imaging to track mental states while using an intelligent tutoring system. Proceedings of the National Academy of Sciences **107**(15) (April 2010) 7018–23
2. Tenison, C., Fincham, J.M., Anderson, J.R.: Detecting math problem solving strategies: an investigation into the use of retrospective self-reports, latency and fMRI data. Neuropsychologia **54** (February 2014) 41–52
3. Cohen Kadosh, R., Lammertyn, J., Izard, V.: Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. Progress in neurobiology **84**(2) (February 2008) 132–47