Machine learning predicts responses to conceptual tasks using eye movements

N. Sanjay Rebello,[1,2] Minh Hoai Nguyen,[3] Yang Wang,[3] Tianlong Zu,[1] John Hutson,[4] and Lester C. Loschky[4]

[1]*Department of Physics and Astronomy, Purdue University, 525 Northwestern Ave., West Lafayette, IN, 47907*
[2]*Department of Curriculum and Instruction, Purdue University, 100 N. University St., West Lafayette, IN, 47907*
[3]*Department of Computer Science, Stony Brook University, 100 Nicolls Rd., Stony Brook, NY 11794-2424*
[4]*Department of Psychological Sciences, Kansas State University, 1114 Mid-Campus Dr. N., Manhattan, KS 66506-5302*

Research has shown that students' responses to conceptual questions correlate with their eye movements. However, to what extent is it possible to predict whether a particular learner might answer a question correctly by monitoring their eye movements in real time? To answer this question, we used spatial-temporal eye-movement data from about 400 participants, as well as their responses to four conceptual physics questions with diagrams. Half of these data were used as a training set for a machine learning algorithm (MLA) that would predict the correctness of students' responses to these questions. The other half of the data were used as a test set to determine the performance of the MLA in terms of the accuracy of the prediction. We will discuss the results of our study with specific attention to the prediction accuracy of the MLA under different conditions.

I. INTRODUCTION

Research has shown that short duration visual cues can facilitate students to improve their performance on physics conceptual questions and near transfer tasks [1]. Research has also shown that visual cues can shift students' attention from areas of a diagram associated with incorrect responses to those associated with correct responses [2]. This shift in visual attention can facilitate students to re-represent a task, by activating the relevant domain knowledge thereby enabling them to correctly solve the task [3].

While visual cues have been shown to be effective, an important issue is whether or not all students can benefit from them. Students who may lack the required domain knowledge of the underlying principles may not benefit from visual cues. Further, students who already know how to answer the question clearly do not need visual cues. In fact, based on the expertise reversal effect, providing guidance to these learners may increase their cognitive load [4]. So, it is important to determine when a learner needs guidance.

Oftentimes, online homework systems, either do not provide guidance, or provide guidance to learners on demand. A more personalized alternative -- Attentive User Interfaces (AUIs) -- provide individualized guidance *when* a learner needs it. For instance an AUI will detect difficulties a learner experiences and provide hints or cues that can facilitate the learner to comprehend the material. Therefore it is important to develop systems that can anticipate when a student might need guidance.

The goal of this study is test the proof-of-concept that machine learning can be used to predict student performance on a conceptual physics task, by using data from the current and prior learners' eye movements on the task. We developed and tested a machine learning algorithm (MLA) that predicts, based on real-time eye-movements, whether or not a learner will correctly solve the conceptual physics task.

Our research questions are: 1) Can the MLA discover the thematically relevant and irrelevant areas associated with correct and incorrect answers on each task? 2) Can the MLA accurately predict student correctness on a conceptual task based on her eye movements? 3) If so, how much training data is needed to develop an MLA that achieves greater than 80% prediction accuracy? 4) Likewise, using real-time data, within what fraction of the total response time can the MLA achieve greater than 80% prediction accuracy?

II. BACKGROUND

Attentive user interfaces (AUIs) are individualized attention assistants that use data of prior activity plus real-time data from ongoing activity to predict and influence future activity to assist the learner in achieving their goal [5]. These systems take into account individual differences in cognitive styles, modality preferences, working memory capacities, and prior and current activity from the learner [6]. They use eye movement data [7] to anticipate the learner's next move [8] and adjust the display in real time

[9-10] to adapt to the learner's information processing system [11-12]. Attentive displays can present information in different formats and rates, and adjust difficulty level [13-15]. They can also reduce clutter, guide attention [16-17], and provide momentary information [18] to facilitate reasoning [19] and provide visual cues [20].

This study is part of a larger project to develop an AUI that would facilitate learners to complete physics tasks. AUIs involve three aspects: detection of current attentional states, evaluation of alternative states, and strategies to maintain or change attentional state [6]. In this study we are focused on the first aspect alone.

III. METHOD

We used eye-movement and correctness data from over 400 students collected from solving four different conceptual tasks (Fig. 1) in our previous studies [1, 2, 21].

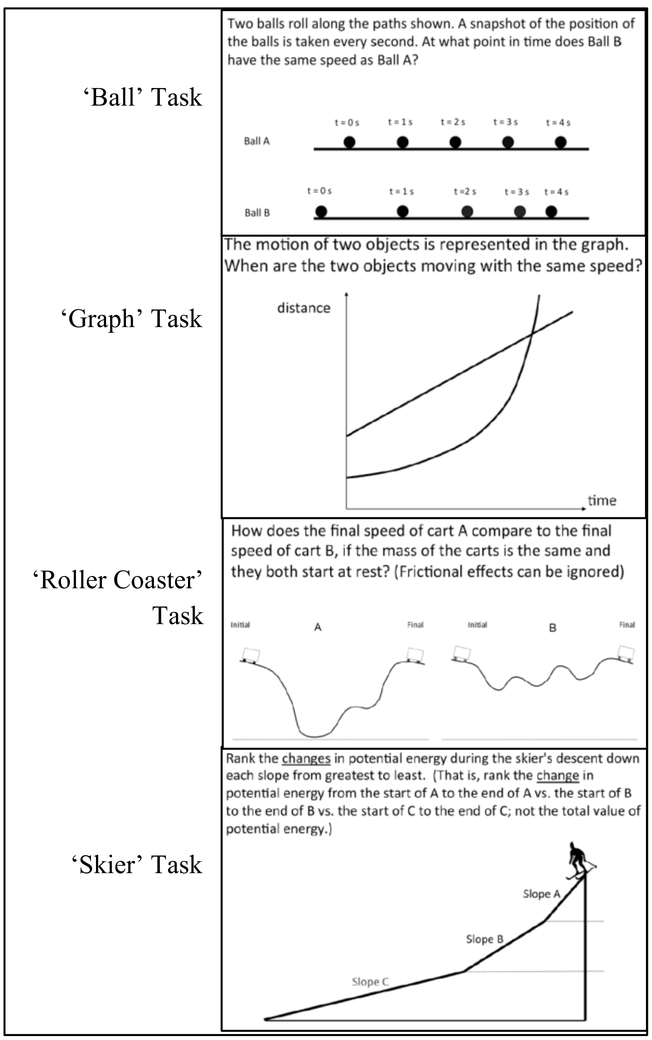


FIG 1. Four conceptual tasks from previous studies

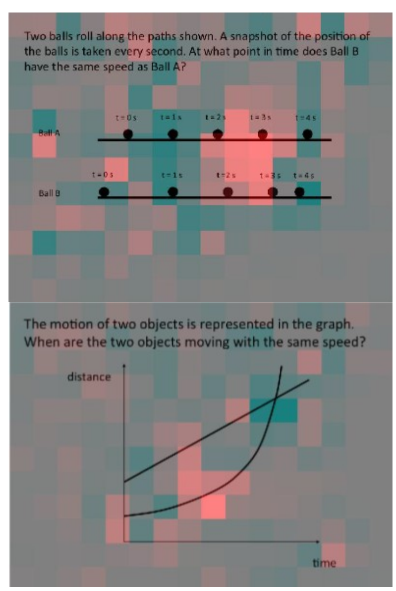
The data from these participants were divided equally into disjoint training and testing sets. A fixation is when eye movements do not exceed a pre-specified spatial and temporal threshold. For each task, each gaze fixation for each student was vectorized using label of +1 if the fixation was made by a student who solved the task correctly or a label of -1 if it was made by a student who solved the task incorrectly. Then we combined data for all participants in the training set to create a sequence of gaze fixations. Given a sequence of gaze fixations for each task, we computed the 12x16 2D histogram of visual attention, and the vectorized histogram was used as the input feature vector for classification (correct versus incorrect answers). The 12x16 division was determined to be the most optimal after trying several other smaller and larger divisions.

We considered two classification methods: Naive Bayes generative model and Least-Squares SVM (Support Vector Machine) discriminative model. We found that the Naive Bayes method does not work as well as the Least-Squares SVM in our experiments. We developed a separate MLA – a Least-Squares SVM classifier -- for each task [22]..

IV. RESULTS

A. Thematically Relevant and Irrelevant Areas

Figure 2 shows the learned classifier weights overlaid on the original task image. The red regions (positive weights) are relevant areas (correct answer), the green regions (negative weights) are irrelevant areas (incorrect answer) This shows that the MLA can *discover* the thematically relevant and irrelevant areas *on its own*.



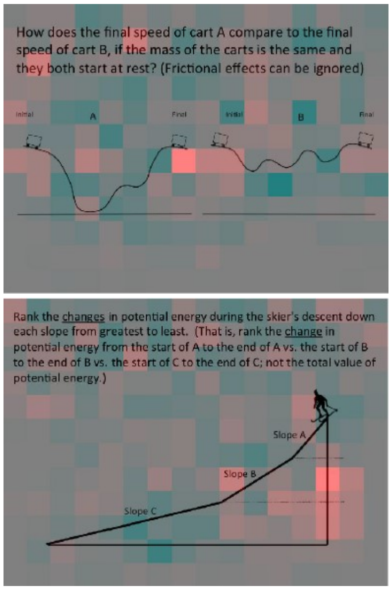


FIG 2. Overlaid classifier weights for the gaze attention map for the tasks in Fig. 1.

B. Classification Performance

To test performance, the MLA (machine learning algorithm) is provided with testing data, which it has to classify as correct or incorrect. The testing data is randomly selected one half of the overall dataset, and is indistinguishable from the training data. For this binary classifier, the performance metric for prediction accuracy that we used is the area under the receiver operating characteristic (ROC) curve (i.e., a graph of true positive rate vs. false positive rate). The mean prediction accuracy for all four tasks is 79.4%. So, the MLA can predict, based on eye movements, whether or not a participant answers a task correctly in about four out of five cases. The results for each task are shown in Table I.

TABLE I. Prediction accuracy of MLA for each task

Conceptual Task	Prediction Accuracy
Ball	73.6%
Graph	84.3%
Roller Coaster	74.6%
Skier	85.0%

C. Number of Training Data & Performance

One of the important considerations when designing an MLA is to determine the number of training data that are needed to achieve a certain level of performance (prediction accuracy). Ideally, the smaller the number of training data needed to achieve an acceptable level of accuracy, the better. The results (Fig. 3) show, that in general, as expected, the performance of the MLA improves slightly with more training data.

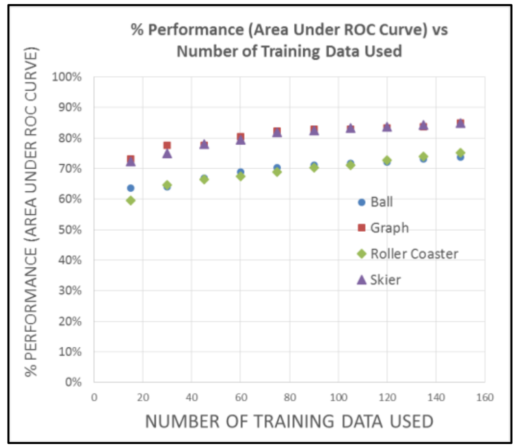


FIG 3. Performance versus number of training data used

On average with about 150 training data, it is possible to achieve about 80% prediction accuracy performance, however the prediction accuracy flattens after about 90 training data. There is a slight difference between the 'ball' and 'roller coaster' tasks on one hand the 'graph' and 'skier' tasks on the other hand. This difference is consistent with results of our prior studies [1,2] which showed that for the ball and roller coaster tasks it is the sequence of eye movements, and not just the areas of attention that are predictive of performance, and likewise, learners showed performance improvement on these tasks only with the use of integration cues (which cued participants to attend to information in a particular sequence) rather than selection cues (which cued participants to attend to the thematically relevant area).

D. Performance & Fraction of Time

The goal of the study is to detect whether you can classify, based on eye movements, the response as correct or incorrect. Ideally, we would want to do this as the person reads the problem *before* they actually complete the task. This is an early classification problem, where we make predictions with observation up to a fraction R of the total response time. If R = 1, the classification problem becomes equivalent to making prediction based on the full observation, which has already been discussed in the previous sections. Results (Fig. 4) shows that in general, as expected, as time passes, the performance accuracy of the classifier improves.

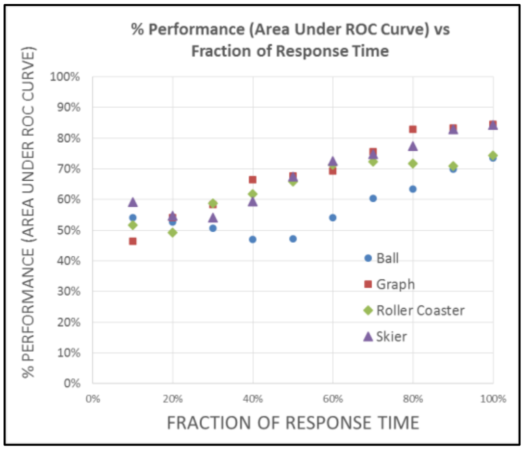


FIG 4. Performance versus fraction of response time

There is a slight dip in performance for the 'ball' task at about 50% of the total response time. This might indicate that the initial (before 50% of response time) eye movements were adequate to classify the solution correctly because learners' eye movements have clear distinctions between those that solve the task correctly versus those that solve it incorrectly. As time progresses, the differences between correct and incorrect solvers tend to diminish and then again increase for the 'ball' task however. The other three tasks showed an almost monotonic increase in performance accuracy with time. In general, it seems that at data from up to 70% of the response time is needed to achieve 70% performance accuracy.

V. SUMMARY

In summary, we found that our MLA achieves about 80% prediction accuracy for physics conceptual tasks. Further the MLA can *discover* areas of interest in the diagram that are associated with correct and incorrect answers. The accuracy of the MLA increases with more training data, and on average the MLA can achieve about 80% performance (prediction accuracy) with about 150 training data. We also find that on average the MLA can make a prediction with an accuracy of about 70% within about 70% of the full response time for the task. This result is particularly significant, because it indicates that we can use such an MLA to predict in advance which student might answer a problem incorrectly/correctly based on their real time eye movements before they have actually answered it. Therefore, it is possible to anticipate, by recording students' real time eye movements while they solve problems, which student might actually benefit from receiving visual cues or other kinds of guidance while they are solving a physics task. This research looks for patterns that associate eye movements with correct or incorrect answers. Understanding *why* the learner is providing the correct or incorrect answers is not the focus of this research study.

ACKNOWLEDGEMENTS

The research is supported in part by U.S. National Science Foundation grant 1348857.

[1] A.R. Rouinfar et. al, in *Proceedings of the Physics Education Research Conference* (2014)
[2] A.R. Rouinfar et. al, *Frontiers in Psych*, **5**, (2014)
[3] S. Ohlsson, *Adv. in Psych of Thinking*, 1-44 (1992)
[4] S. Kalyuga, *Managing Cognitive Load in Adaptive Multimedia Learning*. 58-80 (2009)
[5] R. Vertegaal et. al *Comp. in Hum. Behav.* **22**, 4 (2006)
[6] C. Roda and J. Thomas, *Comp. in Hum. Behav.* **22**, 4 (2006).
[7] L. Rothrock, et. al, *Theor.Issues in Ergo.Sc*i. **3,** 1, (2002)
[8] C. R. Beal, *CyberPsychology & Behavior*, **7**, 6 (2004)
[9] R. Vertegaal, in *Proc. of the symposium on eye tracking research & applications*, p. 23 (2002)
[10] R. Vertegaal, *Comm. of the ACM*, **46**, 3 (2003)
[11] G. W.McConkie & K. Rayner, *Perception & Psychophysics*, **17**, 6 (1975)
[12] E.M. Reingold, et. al, *Human Factors*, **45**, 2 (2003)
[13] L. Copeland, Ph.D. thesis, Australian Natl. Univ. 2016
[14] L. Copeland & T. Gedeon, in *Proceedings of Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on.* (pp. 791-796): IEEE.
[15] L. Copeland & T. Gedeon,. in *Proceedings of the 26th Australian Comp.-Human Inter* (pp. 230-239). ACM
[16] A. Toet, *Computers in Human Behavior*, **22**, 4 2006;
[17] M. Dorr, et. al, in *Proceedings of the fifth workshop on dynamic perception 2004* (pp. 89–94)
[18] I. Starker, and R. A. Bolt in *Proceedings of CHI 1990, ACM*, 3-9
[19] N.H. Narayanan and D. Yoon, in *Proc. of INTERACT 2003: Conf. on Human–Comp. Int.* (pp. 244–251).
[20] D.N. Rapp, *Comp.in Human Behavior*, **22**, 4 (2006)
[21] A. Madsen, et. al, *Phys. Rev. ST-PER* **9**, 020104 (2013)
[22] J. A. Suykens and J. Vandewalle, *Neural Processing Letters*, **9**,3 (1999)