Early Prediction of Student Self-Regulation Strategies by Combining Multiple Models

Jennifer L. Sabourin
Department of Computer Science
North Carolina State University
Raleigh, North Carolina 27695
jlrobiso@ncsu.edu

Bradford W. Mott
Department of Computer Science
North Carolina State University
Raleigh, North Carolina 27695
bwmott@ncsu.edu

James C. Lester
Department of Computer Science
North Carolina State University
Raleigh, North Carolina 27695
lester@ncsu.edu

ABSTRACT

Self-regulated learning behaviors such as goal setting and monitoring have been found to be crucial to students' success in computer-based learning environments. Consequently, understanding students' self-regulated learning behavior has been the subject of increasing interest. Unfortunately, monitoring these behaviors in real-time has proven challenging. This paper explores a variety of data mining approaches to predicting student self-regulation capabilities. Students are classified into SRL-use categories based on evidence of goal-setting and monitoring activities. Prior work on early prediction of these categories pointed to logistic regression and decision tree models as effective techniques. This paper builds on these findings by exploring techniques by which these models can be combined to improve classification accuracy and early prediction capabilities. By improving classification accuracy, this work can be leveraged in the design of computer-based learning environments to provide adaptive scaffolding of self-regulation behaviors.

Keywords

Self-regulated learning, early prediction, machine learning

1. BACKGROUND

Understanding and facilitating students' self-regulated learning behaviors has been the subject of increasing attention in recent years. This line of investigation is fueled by evidence suggesting the strong role that self-regulatory behaviors play in a student's overall academic success [1]. Self-regulated learning (SRL) can be described as "the process by which students activate and sustain cognitions, behaviors, and affects that are systematically directed toward the attainment of goals" [2]. Unfortunately, students can demonstrate a wide range of fluency in their SRL behaviors [3] with some students lagging behind their peers in their ability to appropriately set and monitor learning goals. Findings that students with low SRL skills are less likely to achieve academic success have prompted efforts to mediate these differences [1,4].

Identifying and scaffolding SRL strategies has also been a focus of much work in the intelligent tutoring systems community. For example, in MetaTutor, a hypermedia environment for learning biology, think-aloud protocols have been used to examine which strategies students use, while analysis of students' navigation through the hypermedia environment helps to identify profiles of self-regulated learners [5]. Similarly, researchers have identified patterns of behavior in the Betty's Brain system that are indicative of low and high levels of self-regulation [6]. Prompting students to use SRL strategies when these patterns of behavior occur has shown promise in improving student learning. For example, Conati et al. have examined the benefits of prompting students to self-explain when learning physics content and how these explanations can be facilitated in a computer-based learning environment [7].

Such work has focused primarily on examining SRL in highly structured problem-solving and learning environments. However, understanding and scaffolding students' SRL behaviors is particularly important in open-ended learning environments where goals may be less clear and students do not necessarily have a clear indicator of their progress [8]. In order to be successful in this type of learning environment, students must actively identify and select their own goals and evaluate their progress accordingly. While the nature of the learning task may have implicit overarching goals such as 'completing the task' or 'learning a lot,' it is important for students to set more specific, concrete and measurable goals [9]. Unfortunately, students do not consistently demonstrate sufficient self-regulatory behaviors during interactions with these environments, which may reduce the educational potential of these systems [10,11]. Consequently, identifying and scaffolding students with low SRL skills is a necessary next step to ensure that these systems can be used as effective learning tools.

This paper reports on an investigation of self-regulatory behaviors of students in a game-based science mystery, CRYSTAL ISLAND. During interactions with the CRYSTAL ISLAND environment, students were prompted to report on their mood and status in a way that is similar to many social networking tools available today. Though students were not explicitly asked about their goals or progress, many students included this information in their short, typed status statements. This data is used to classify students into low, medium, and high self-regulated learning behavior classes. Prior work has pointed to the importance of being able to identify and scaffold the low SRL students [4]. While logistic regression and decision tree models have been found to be effective at early prediction of these classes, this work expands upon these findings by exploring ways in which these models can be combined to improve classification accuracy and early prediction capabilities. Ensemble methods have been found to be effective at a variety of predictive tasks including predicting student knowledge [12]. By improving classification accuracy, this work can be leveraged in future systems to provide adaptive scaffolding of self-regulation behavior early into interaction with the environment, offering the possibility for timely intervention. The implications of these results and areas of future work are then discussed.

2. METHOD

The investigation of students' SRL behaviors was conducted with students from a middle school interacting with CRYSTAL ISLAND, a game-based learning environment being developed for the domain of microbiology that follows the standard course of study for eighth grade science in North Carolina.

2.1 CRYSTAL ISLAND

CRYSTAL ISLAND features a science mystery set on a recently discovered volcanic island. Students play the role of the protagonist, Alex, who is attempting to discover the identity and source of an unknown disease plaguing a newly established research station. The story opens by introducing the student to the island and the members of the research team for which her father serves as the lead scientist. As members of the research team fall ill, it is her task to discover the cause and the specific source of the outbreak. Typical game play involves navigating the island, manipulating objects, taking notes, viewing posters, operating lab equipment, and talking with non-player characters to gather clues about the disease's source. To progress through the mystery, a student must explore the world and interact with other characters while forming questions, generating hypotheses, collecting data, and testing hypotheses.

2.2 Study Procedure

A study with 296 eighth grade students was conducted. After removing instances with incomplete data or logging errors, there were 260 students remaining. Among the remaining students, there were 129 male and 131 female participants varying in age and race. Participants interacted with CRYSTAL ISLAND in their school classroom, although the study was not directly integrated into their regular classroom activities. Pre-study materials were completed during the week prior to interacting with CRYSTAL ISLAND. The pre-study materials included a demographic survey, researcher-generated CRYSTAL ISLAND curriculum test, and several validated instruments. Personality was measured using the Big 5 Personality Questionnaire, which indexes subjects' personalities across five dimensions: openness, conscientiousness, extraversion, agreeableness and neuroticism [12]. Goal orientation was measured using a 2-dimensional taxonomy considering subjects' mastery or performance orientations along with their approach or avoidance tendencies [13]. Subjects' affect regulation tendencies were measured with the Cognitive Emotion Regulation Questionnaire [14] though features from this survey were not included in the current models.

Immediately after solving the mystery, or after 55 minutes of interaction, students moved to a different room in order to complete several post-study questionnaires including the curriculum post-test.

Students' affect data were collected during the learning interactions through self-report prompts. Students were prompted every seven minutes to self-report their current mood and status through an in-game smartphone device. Students selected one emotion from a set of seven options, which included the following: anxious, bored, confused, curious, excited, focused, and frustrated. After selecting an emotion, students were instructed to briefly type a few words about their current status in the game, similarly to how they might update their status in an online social network.

2.3 SRL Classification

The typed status reports were later tagged for SRL evidence using the following four ranked classifications: 1) specific reflection, 2) general reflection, 3) non-reflective statement, or 4) unrelated (Table 1). This ranking is motivated by the observation that setting and reflecting upon goals is a hallmark of self-regulatory behavior and that specific goals are more beneficial than those that are more general [9]. Students were then given an overall SRL score based on the average score of their statements. An even tertiary split was then used to assign the students to a Low, Medium, and High SRL category.

Table 1. Categories of SRL tags

SRL Category	Description	Examples
Specific reflection	Student evaluates progress towards a specific goal or area of knowledge	"I am trying to find the food or drink that caused these people to get sick." "Well...the influenza is looking more and more right. I think I'll try testing for mutagens or pathogens – [I] ruled out carcinogens"
General reflection	Student evaluates progress or knowledge but without referencing a particular goal	"I think I'm getting it" "I don't know what to do"
Non-reflective	Student describes what they are doing or lists a fact without providing an evaluation	"testing food" "in the lab"
Unrelated	Any statement which did not fall into the above three categories is considered unrelated, including non-word or unidentifiable statements	"having fun" "arghhh!"

From the 260 students, a total of 1836 statements were collected, resulting in an average of 7.2 statements per student. All statements were tagged by one member of the research team with a second member of the research team tagging a randomly selected subset (10%) of the statements to assess the validity of the protocol. Inter-rater reliability was measured at $\kappa = 0.77$, which is an acceptable level of agreement. General reflective statements were the most common (37.2%), followed by unrelated (35.6%), specific reflections (18.3%) and finally non-reflective statements (9.0%).

The tertiary split of students into Low, Medium, and High SRL classes has yielded interesting findings in prior work [4]. One important finding is that Medium and High SRL students have both higher prior knowledge and higher learning gains than Low SRL students. This suggests that Low SRL students begin with some disadvantage and that the overall gap in knowledge is increased after interactions with CRYSTAL ISLAND. Though all groups have significant learning gains, Low SRL students are not experiencing the same advantages of interaction with CRYSTAL ISLAND. This finding points to the strong need to provide these students with additional scaffolding to improve the quality of their interaction.

2.4 SRL Prediction

The difference in learning between Low, Medium, and High SRL students has motivated the goal of early prediction of students' SRL skills. Prior work [4] has shown promise in being able to predict SRL class early into the interaction. This work compared the ability of naïve Bayes, neural network, logistic regression,

support vector machine, and decision tree models to predict SRL class at different time intervals. Overall it was found that logistic regression and decision trees offered the best performance, with the best model correctly predicting 57% of students' classes after one-third of their interaction with CRYSTAL ISLAND. Compared with a most-frequent-class baseline of 34%, this offers a significant improvement in the ability to recognize SRL skill. However, while both logistic regression and decision tree models significantly outperformed other modeling techniques, neither of the two best performers consistently outperformed the other. This raised the question of whether some method of combining these two learned models might offer improved or more stable performance.

2.4.1 Original Models

The original logistic regression and decision tree models were trained using 10-fold cross validation with the WEKA machine learning toolkit [15]. For the original models, a total of 49 features were used to train machine-learning models. Of these, 26 features represented personal data collected prior to the student's interaction with CRYSTAL ISLAND. This included demographic information, pre-test score, and scores on the personality, goal orientation, and emotion regulation questionnaires. The remaining 23 features represented a summary of student's interactions in the environments. This included information on how students used each of the curricular resources, how many in-game goals they had completed, as well as evidence of off-task behavior. Additionally, data from the student's self-reports were included, such as the most recent emotion report and the character count of their "status".

In order to examine early prediction of the students' SRL-use categories, these features were calculated at four different points in time resulting in four unique datasets. The first of these (**Initial**) represented information available at the beginning of the student's interaction and consequently only contained the 26 personal attributes. Each of the remaining three datasets (**Report1**-**3**) contained data representing the student's progress at each of the first three emotion self-report instances. These datasets contained the same 26 personal attributes, but the values of the remaining 23 in-game attributes differentially reflected the student's progress up until that point. The first self-report occurred approximately 4 minutes into game play with the second and third reports occurring at 11 minutes and 18 minutes, respectively. The third report occurs after approximately one-third of the total time allotted for interaction has been completed, so it is still fairly early into the interaction time.

2.4.2 Combining Multiple Models

To combine the predictions of multiple models, a variety of different voting schemes were used in which both the predicted class from the original decision tree and logistic regression models were taken into account:

- **Standard**: The prediction from each model is weighted equally.
- **Weighted by Accuracy**: The prediction from each model is weighted by the model's overall predictive accuracy.
- **Weighted by Precision:** The prediction from each model is weighted by its precision at predicting the class for which it is voting.
- **Select Lowest Class**: The model predicting the lowest SRL skill is selected.

The final model of always selecting the lowest level prediction is based on the assumption that we would rather underestimate students' abilities and provide additional scaffolding than overestimate their abilities. Additionally, in all of the above voting schemes, the lower class was chosen in case of a tie.

3. RESULTS

For each time slice, we compared the original models with the combined models by evaluating overall predictive accuracy as well as recall on the Low-SRL class. The first metric represents how well the model does overall at correctly identifying each class, while the latter represents the proportion of Low-SRL students who were correctly identified. This second metric is especially important given the proposed style of intervention. These metrics for each model are shown in Table 2.

The results indicate that the most successful voting model was the **Weighted by Precision** model. It offered statistically significantly ($p < 0.05$) better accuracy than any other model, and better Low-SRL recall than either original model for all time-slices, with the exception of the **Initial** prediction. It also offered improved stability of performance over the original models and other ensemble models, with both accuracy and recall improving as more data became available. The **Select Lowest Class** combined model had the highest recall of the Low-SRL class which is to be expected given its favoritism for low classifications. The **Select Lowest Class** model identified almost exactly half of all students as Low-SRL However, it was able to correctly identify up to 85% of the actual Low-SRL students, making it a promising contender for identifying cases where additional scaffolding would be beneficial.

Table 2. Predictive models and evaluation metrics

	Predictive Accuracy				***Low-SRL Recall***			
	Initial	Report1	Report2	Report3	Initial	Report1	Report2	Report3
Original Models								
Decision Tree	37.7	46.5	51.6	53.4	0.36	0.58	0.63	0.70
Logistic Regression	40.8	55.0	53.1	57.2	0.43	0.65	0.68	0.77
Combined Models								
Standard Voting	38.1	50.0	54.3	54.1	0.45	0.67	0.75	0.79
Weighted by Accuracy	37.1	53.1	55.0	54.5	0.33	0.56	0.65	0.65
Weighted by Precision	40.1	57.3	57.0	59.1	0.44	0.67	0.75	0.80
Select Lowest Class	36.9	51.5	52.3	51.4	0.58	0.81	0.79	0.84

With the exception of the **Weighted by Precision** model, the predictive accuracy of each ensemble model tended to fall somewhere between accuracy of the original decision tree and logistic regression models. This suggests that these models did not have enough additional information in their weighting scheme to offer improvements in performance. It is especially interesting that weighing votes by overall accuracy was not beneficial. This is likely due to the high and mostly equivalent accuracies of both the original models. However, the **Weighted by Precision** model takes into account each model's likelihood of correctness given a particular prediction which varied between models. Specifically, the logistic regression model was generally better at Low and High SRL predictions while the decision tree model was stronger at Medium SRL predictions.

4. CONCLUSION

Predicting students' self-regulated learning skills can form the basis for effective scaffolding strategies. Combining multiple machine learned models can be used for early prediction of students' self-regulated learning skills, as was shown in an investigation with the narrative-centered learning environment, CRYSTAL ISLAND. Results indicate that early prediction of self-regulation skills is feasible and that combining multiple models can offer improvements over individual models alone. Specifically, logistic regression and decision tree models were combined using a variety of voting strategies. Some of these strategies were able to offer significant improvements in both predictive accuracy and Low-SRL recall.

These findings point to several directions for future work. The most prominent of these is developing intervention mechanisms for aiding student self-regulation. Early prediction of SRL skills is not useful unless we are able to act intelligently upon this prediction. Therefore, the development of appropriate and effective scaffolding strategies is an important next step in this line of investigation. These techniques could then be used in conjunction with several of the top-performing models in order to determine which optimizations have the best impacts on students overall learning.

5. ACKNOWLEDGMENTS

The authors wish to thank members of the IntelliMedia Group for their assistance, Omer Sturlovich and Pavel Turzo for use of their 3D model libraries, and Valve Software for access to the SourceTM engine and SDK. This research was supported by the National Science Foundation under Grants REC-0632450, DRL-0822200, IIS-0812291, and CNS-0739216. This material is based upon work supported under a National Science Foundation Graduate Research Fellowship.

6. REFERENCES

[1] Zimmerman B.J. Self-regulated learning and academic achievement: An overview, *Educational Psychologist*, vol. 25, pp. 3-17, 1990.

[2] Schunk, D. H. Attributions as Motivators of Self-Regulated Learning, in *Motivation and Self-Regulated Learning: Theory, Research, and Applications*, D. H. Schunk and B. J. Zimmerman (eds). 2008, pp. 245-266.

[3] Ellis, D. and Zimmerman, B. J. Enhancing self-monitoring during self-regulated learning of speech, H. J. Hartman (ed) Dordrecht, The Netherlands: Kluwer, 2001, pp. 205-228.

[4] Sabourin, J. L., Shores, L.R., Mott, B. W., and Lester, J. C. Predicting Student Self-Regulation Strategies in Game-Based Learning, to appear in *Proceedings of the 11th International Conference on Intelligent Tutoring Systems*, 2012.

[5] Azevedo, R., Johnson, A., Chauncey, A., and Burkett, C. Self-Regulated Learning with MetaTutor: Advancing the Science of Learning with MetaCognitive Tools, in *New Science of Learning: Cognition, Computers and Collaboration in Education*, M. Khine and I. Saleh (eds) New York: Springer, 2010, pp. 225-248.

[6] Biswas, G., Jeong, H, Roscoe, R., and Sulcer, B. Promoting Motivation and Self-Regulated Learning Skills through Social Interactions in Agent-Based Learning Environments, in *2009 AAAI Fall Symposium on Cognitive and Metacognitive Educational Systems*, 2009.

[7] Conati, C. and VanLehn, K. Towards Computer-Based Support of Meta-Cognitive Skills: a Computational Framework to Coach Self-Explanation, *International Journal of Artificial Intelligence in Education*, vol. 11, pp. 398-415, 2000.

[8] Land, S. Cognitive requirements for learning with open-ended learning environments, *Educational Technology Research and Development*, vol. 48, no. 3, pp. 61-78, 2000.

[9] Zimmerman, B. Goal Setting: A Key Proactive Source of Academic Self-Regulation, in *Motivation and Self-Regulated Learning: Theory, Research, and Applications*, D. H. Schunk and B. J. Zimmerman (eds) New York: Routledge, 2008, pp. 267-286.

[10] Kirschner, P. A., Sweller, J. and Clark, R. E. Why Minimal Guidance during instruction does not work: An analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching, *Educational Psychologist*, vol. 41, pp. 75-86, 2006.

[11] Alfieri, L., Brooks, P., Aldrich, N., and Tenenbaum, H. Does Discovery-Based Instruction Enhance Learning, *Journal of Education Psychology*, vol. 103, pp. 1-18, 2011.

[12] Pardos, Z.A., Gowda, S. M., Baker, R. S.J.D., Heffernan, N. T. The Sum is Greater than the Parts: Ensembling Models of Student Knowledge in Educational Software. To appear in *ACM's Knowledge Discovery and Datamining Explorations*, vol. 13, 2012.

[13] McCrae, R. and Costa, P. *Personality in Adulthood: A Five-Factor Theory Perspective*, 2nd ed. New York: Guilford Press, 1993.

[14] Elliot, A. and. McGregor, H.A . A 2 x 2 achievement goal framework, *Journal of Personality and Social Psychology*, vol. 80, pp. 501-519, 2001.

[15] Gernefski, N. and Kraati, V. Cognitive Emotion Regulation Questionnaire: Development of a Short 18-Item Version, *Personality and Individual Differences*, vol. 41, pp. 1045-1053, 2006.

[16] Hall, M., Frank, E., Holmes, G., Pfajhringer, B., Reutmann, P., and Witten, I. The WEKA Data Mining Software: An Update, *SIGKKD Explorations*, vol. 11, no. 1, 2009.