Early Prediction of Cognitive Tool Use in Narrative-Centered Learning Environments

Lucy R. Shores[1], Jonathan P. Rowe[2], and James C. Lester[2]

[1]Department of Curriculum & Instruction, North Carolina State University, Raleigh, NC 27695
[2]Department of Computer Science, North Carolina State University, Raleigh, NC 27695
{lrshores, jprowe, lester}@ncsu.edu

Abstract. Narrative-centered learning environments introduce novel opportunities for supporting student problem solving and learning. By incorporating cognitive tools into plots and character roles, narrative-centered learning environments can promote self-regulated learning in a manner that is transparent to students. In order to adapt narrative plots to explicitly support effective cognitive tool-use, narrative-centered learning environments need to be able to make early predictions about how effectively students will utilize learning resources. This paper presents results from an investigation into machine-learned models for making early predictions about students' use of a specific cognitive tool in the CRYSTAL ISLAND learning environment. Multiple classification models are compared and discussed. Findings suggest that support vector machine and naïve Bayes models offer considerable promise for generating useful predictive models of cognitive tool use in narrative-centered learning environments.

Keywords: Narrative-Centered Learning Environments, Cognitive Tools, Self-Regulated Learning.

1 Introduction

Narrative-centered learning environments have become the subject of increasing attention in the AI in Education community [1,2,3,4,5,6]. By contextualizing learning within narrative settings, narrative-centered learning environments tap into students' innate facilities for crafting and understanding stories [7]. An additional benefit of narrative-centered learning environments is their capacity to discreetly scaffold students' learning processes by tightly integrating pedagogy and narrative elements. For example, narrative-centered learning environments have been developed that teach negotiation skills [3] and foreign languages [2] through conversational interactions with virtual characters. Also, scientific inquiry has been realized in interactive mysteries where students play the roles of detectives [8,9].

A particularly promising opportunity presented by narrative-centered learning environments is supporting self-regulated learning, i.e., students' ability to generate, monitor and control their cognitive, metacognitive, and motivational processes [10]. Students often possess varying degrees of competency in self-regulated learning [11].

Narrative plots and character roles can introduce contextualized *cognitive tools* that discreetly support self-regulation through elements of the story world. However, not all students use cognitive tools equally effectively; tools' effective use may need to be encouraged or guided. In narrative-centered learning, this support is ideally delivered by adapting narrative sequences to encourage effective cognitive tool use. Narrative-centered learning environments should be capable of making early predictions about how a student will use cognitive tools during a narrative-centered learning interaction, and subsequently use these predictions to inform decisions about tailoring the narrative and problem solving support.

This paper focuses on early prediction of students' cognitive tool use in a narrative-centered learning environment. The work extends previous research that identified a particular cognitive tool, a *diagnosis worksheet*, to be associated with significant content learning gains in the CRYSTAL ISLAND environment [18]. Several supervised machine-learning models are compared for early prediction of students' diagnosis worksheet usage, and potential directions for incorporating the predictive models into narrative-centered learning environments are discussed.

2 Related Work

Narrative-centered learning environments are a class of serious games that tightly couple educational content and problem solving with interactive story scenarios. Recent work on narrative-centered learning environments has leveraged a range of techniques for providing effective, engaging learning experiences. FearNot! uses affectively-driven autonomous agents to generate dramatic, educational vignettes about bullying [1]. The Tactical Language and Culture Training System uses a range of AI techniques for speech recognition and virtual human behavior in interactive narrative scenarios for language and culture learning [2,6]. BiLAT is a story-centric serious game that enables students to practice cross-cultural negotiation skills during interactions with virtual characters [3]. BiLAT features a *leader preparation worksheet* that students complete to prepare for virtual negotiations, and it has similarities to the cognitive tool (diagnosis worksheet) that is the focus of this work. However, none of these systems explicitly model students' cognitive tool use during narrative-centered learning interactions to our knowledge.

Cognitive tools [12] are external, compensatory resources for problem solving. They are used to moderate student ability deficits and to maximize the effects of learning experiences. Cognitive tools for supporting self-regulated learning have been incorporated into several intelligent tutoring systems. For example, prompts for self-explanation have been shown to enhance learning during interactions with the Cognitive Tutor and SE Coach systems [13,14]. Similarly, self-regulatory prompts in the Betty's Brain environment have been shown to positively influence student learning and problem-solving behaviors [15]. During interactions with MetaTutor, students receive several forms of self-regulated learning instruction and, as a result, used self-regulation strategies more successfully [16].

Given the benefits of cognitive tools for supporting self-regulatory behaviors, it is critical to provide effective scaffolds for cognitive tool use. Schunk [11] explains that the development of self-regulatory skills occurs socially over time, making a one-size-fits-all approach to self-regulated learning support problematic. Developing predictive models of students' cognitive tool-use can enable intelligent tutoring systems to tailor support for students' self-regulated learning. Narrative-centered learning environments stand to benefit from these predictive models by adapting stories to support cognitive tool-use in a manner that is embedded in plots [17].

3 CRYSTAL ISLAND

CRYSTAL ISLAND is a narrative-centered learning environment built on Valve Software's Source™ engine, the 3D game platform for Half-Life 2. The curriculum underlying CRYSTAL ISLAND's mystery narrative is derived from the North Carolina state standard course of study for eighth-grade microbiology. Students play the role of the protagonist who is attempting to discover the details of an infectious disease plaguing a research station. Several of the team's members have fallen ill, and it is the student's task to discover the cause of the outbreak (for more information, see [8]).

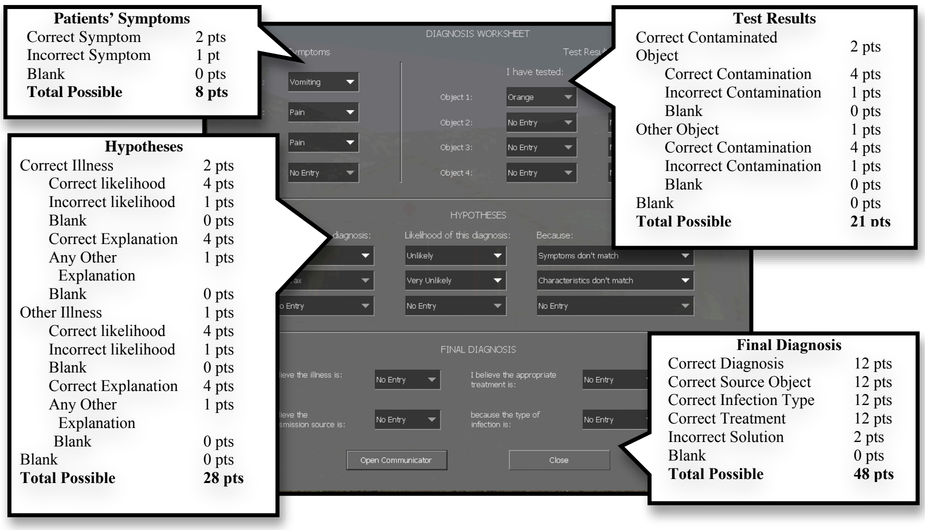

Fig. 1. CRYSTAL ISLAND's diagnosis worksheet and associated scoring rubrics.

An important element throughout CRYSTAL ISLAND's narrative and gameplay is the diagnosis worksheet (Figure 1). The worksheet consists of four sections: the *Patients' Symptoms* area where students record traits of the spreading disease; the *Test Results* area where students record findings from laboratory tests; the *Hypotheses* area where students record their beliefs about the likelihoods of candidate diagnoses; and the *Final Diagnosis* area where students report the identity, source, and treatment of the illness. A scoring scheme was devised to assess students' worksheet completion (see Figure 1). The total worksheet score was calculated by summing the sub-scores for each region (max = 105 points). Regions that involved complex inferences were weighted more heavily than regions that involved rote recording of information.

4 Predicting Cognitive Tool Use

The data used for the current investigation was collected during an experiment involving human participants from the eighth grade of a North Carolina middle school. The primary goal of the experiment was to investigate the impact of different scaffolding techniques on learning and engagement in the CRYSTAL ISLAND narrative-centered learning environment. However, no condition effects were observed for either learning or engagement. This paper's investigation is a secondary analysis of the data and considers data from all conditions as a whole.

4.1 Data Collection

A total of 153 eighth grade students ranging in age from 13 to 15 (M = 13.3, SD = 0.47) interacted with the CRYSTAL ISLAND environment. Eight of the participants were eliminated due to incomplete data and eight participants were removed because they had prior experience with CRYSTAL ISLAND. Among the remaining 137 students (male: 77, female: 60), approximately 3% of the participants were American Indian or Alaska Native, 2% were Asian, 32% were African American, 13% were Hispanic or Latino, and 50% were White. The study was conducted prior to students' exposure to the microbiology curriculum unit of the North Carolina state standard course of study.

Students completed a series of pre-experiment questionnaires one week prior to playing CRYSTAL ISLAND. Post-experiment materials were completed immediately following the learning interaction. In addition to pre- and post-experiment measures, the CRYSTAL ISLAND software logged student actions, locations, and narrative state during gameplay, including the complete state of students' diagnosis worksheets.

4.2 Inductive Framework

A previous investigation indicated that students achieved significant learning gains as a result of their interactions with CRYSTAL ISLAND [8], and maintaining a thorough and accurate diagnosis worksheet was associated with improved learning outcomes, especially for students with low levels of prior microbiology knowledge [18]. Each student was classified as being either a low or high diagnosis worksheet student using a median split on their final diagnosis worksheet score. Students with low prior knowledge who earned high scores on their diagnosis worksheet experienced greater content learning gains than their low-scoring counterparts, and they performed comparably to high prior knowledge students on the microbiology post-test. Significant worksheet differences between the high and low groups began to appear after twenty-five minutes, which was almost halfway through the learning interaction. The current machine learning analysis focuses on early prediction, and it therefore classifies whether students will be high or low diagnosis worksheet users during the first twenty-five minutes of interaction, which is prior to the score divergence.

In order to identify useful predictor features for machine learning, a series of ANOVAs compared the gameplay characteristics of high and low diagnosis worksheet students. *In-game score*, a numerical sum that is based on a student's problem-solving engagement and effectiveness (for full details, see [8]), revealed significant differences between high and low diagnosis worksheet students after one minute of play. *Microbiology manual use,* measured by counting the number of times a student opened the feature on his/her in-game PDA device, was used significantly more by low diagnosis worksheet students around five minutes of elapsed gameplay. *Dialogue moves with non-player characters,* calculated as the total number of conversational turns with virtual characters, was greater among high diagnosis worksheet students after five minutes of play. *Virtual book reading,* calculated as the number of times a student opened in-game virtual books, was found to occur more frequently among high diagnosis worksheet students around ten minutes of gameplay.

A supervised learning approach leveraging the above predictors was taken in order to predict students' diagnosis worksheet group (high/low). All models were induced using the WEKA machine learning toolkit [19]. Naïve Bayes, decision tree, and support vector machine (SVM) classification techniques were compared to a *most frequent category* baseline (in this case, high diagnosis worksheet) for predicting whether a student would end the game as either a high or low diagnosis worksheet student. To enable early classification of student worksheet outcomes, instances of each model were learned for the 10, 12, 15, 18, 20, 22, and 25 minute marks. Predictor feature values were calculated using data up to the relevant time in the logs. In total, 28 models were trained and tested (including baseline). A student-level tenfold cross validation scheme was used to evaluate the performance of each model.

5 Findings

After ten minutes of gameplay, the best performing model (SVM) correctly classified 60.5% of instances, which was found to be significantly better than baseline ($p < .05$). The SVM model maintained significance over baseline for the 12-, 15-, 18-, 20-, 22-, and 25-minute models (see Table 1). The twelve-minute naïve Bayes model was found to significantly out predict baseline, correctly classifying 60.9% of instances ($p < .05$). Again, the naïve Bayes model was found to consistently and significantly outperform baseline for the remaining timestamps. However, decision tree models were not found to be reliable predictors of diagnosis worksheet performance.

Table 1. Accuracy percentages for classification models predicting diagnosis worksheet group with regard to time.

Time (Minutes)	Baseline	Naïve Bayes	Decision Tree	SVM
10	56.9	57.7	52.7	60.6*
12	56.9	60.9*	53.3	62.0**
15	56.9	63.0**	55.6	62.1**
18	56.9	61.1**	55.4	63.6**
20	56.9	61.8**	54.6	61.2*
22	56.9	62.7**	56.0	61.5*
25	56.9	60.9*	55.4	63.8**

Note: * ($p < .05$) and **($p < .01$) indicate significant performance above baseline

The results indicate that SVM models are significantly more accurate than baseline for classifying students' diagnosis worksheet performance after ten minutes of interaction with the CRYSTAL ISLAND environment. Naïve Bayes modeling techniques are effective after twelve minutes have elapsed and tend to sustain higher levels of significance than SVM models. However, a series of ANOVAs found both the twelve-minute SVM and fifteen-minute naïve Bayes models to have higher increased significance over baseline than the ten-minute and twelve-minute models, respectively. The decision tree models' lack of significant improvement over baseline suggests that they may not be well-suited to the current task. The areas under the Receiver Operating Characteristic (ROC) curve for the ten-minute and twelve-minute SVM and twelve-minute and fifteen-minute naïve Bayes models are displayed in Table 2.

Table 2. Areas under the ROC curve for the best performing time-based models.

Class	Ten-Minute SVM	Twelve-Minute SVM	Twelve-Minute Naïve Bayes	Fifteen-Minute Naïve Bayes
High Diagnosis Worksheet	0.55	0.61	0.68	0.74
Low Diagnosis Worksheet	0.54	0.65	0.66	0.73

As previous analyses of the diagnosis worksheet suggest, efficient use of the diagnosis worksheet is particularly advantageous for low prior knowledge students in terms of content learning gains [18]. Although accurately classifying students into four groups, high/low prior knowledge and high/low worksheet, is a more challenging problem than the previous two-group task, this finer-grained classification can better inform real-time, personalized scaffolding, particularly to assist low prior knowledge students in utilizing the diagnosis worksheet. A low diagnosis worksheet/low prior knowledge student might benefit from both tool use and content-related scaffolding; whereas, a low diagnosis worksheet/high prior knowledge student might find content scaffolding to be redundant, running the risk of expertise reversal effects [20]. Additionally, this finer-grained classification opens opportunities for tailoring scaffolding without the need for prior information about students.

In a follow-up analysis, models were created to classify students as high diagnosis worksheet/high prior knowledge, high diagnosis worksheet/low prior knowledge, low diagnosis worksheet/high prior knowledge, or low diagnosis worksheet/low prior. Again, a median split was used to distinguish performance on the microbiology pre-test. The highest performing model (SVM) accurately classified 40.00% of instances after ten minutes of gameplay, which significantly outperformed the baseline model (33.57%; $p < .01$). The SVM model maintained significance over baseline models for all time periods ($p < .01$). Naïve Bayes 10-, 12-, and 15-minute models were found to significantly out predict baseline models ($p < .01$); however, this dominance was not found for later time points after fifteen minutes. Again, the decision tree model was observed to be insufficient for accurately classifying the students.

6 Conclusions

Narrative-centered learning environments offer important opportunities for supporting effective self-regulated learning behaviors. By incorporating cognitive tools into narrative plots and character roles, narrative-centered learning environments can discreetly scaffold complex cognition and metacognition during problem solving. Previous research has indicated that not all students use cognitive tools equally effectively. In order to dynamically adapt narrative plots to support effective cognitive tool use, it is necessary for narrative-centered learning environments to make early predictions about how students will use provided cognitive supports.

Several machine-learning models were trained and evaluated for predicting students' diagnosis worksheet performance in the CRYSTAL ISLAND learning environment. Support vector machine (SVM) and naïve Bayes models were found to achieve promising predictive accuracy as early as ten minutes into a learning interaction. SVM and naïve Bayes models were also found to be a promising method for jointly predicting diagnosis worksheet performance and microbiology prior knowledge, although more work is needed to enhance the accuracy of these fine-grained classifications. Continued investigations of machine-learned models for predicting cognitive tool-use may introduce opportunities for dynamically scaffolding students' self-regulatory behaviors in narrative-centered learning environments.

It should be noted that the machine-learned models were trained using only 137 instances, a relatively small dataset for machine learning purposes. As a consequence, very few predictor features were used for training the models. This may explain the relatively low accuracies, particularly for the decision tree models. Future work will utilize a larger corpus of students with additional predictor features in hopes of improving predictive accuracy. Furthermore, incorporating the models into runtime narrative-centered learning environments will enable further investigations to determine whether model-informed narrative adaptations can lead to more effective use of the diagnosis worksheet, and consequently improved learning outcomes.

Acknowledgments. The authors 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 Source™ SDK. This research was supported by the National Science Foundation under Grants REC-0632450, DRL-0822200, CNS-0540523, and IIS-0812291. This work was also supported by the National Science Foundation through a Graduate Research Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

1. Aylett, R., Louchart, S., Dias, J., Paiva, A., Vala, M.: FearNot! An Experiment in Emergent Narrative. In: 5th International Conference on Intelligent Virtual Agents, pp. 305-316 (2005)

2. Johnson, W.L.: Serious use of a Serious Game for Language Learning. In: 13th International Conference on Artificial Intelligence in Education, pp. 67-74 (2007)
3. Kim, Hill, Durlach, Lane, Forbell, Core, Marsella, Pynadath, Hart.: BiLAT: A Game-Based Environment for Practicing Negotiation in a Cultural Context. Int. J. AIED, 19, 289-308 (2009)
4. Niehaus, J., Riedl, M.: Toward Scenario Adaptation for Learning. In: 14th International Conference on AI in Education, pp. 686-688 (2009)
5. Thomas, J., Young, R. M.: Using Task-Based Modeling to Generate Scaffolding in Narrative-Guided Exploratory Learning Environments. In: 14th International Conference on Artificial Intelligence in Education, pp. 107-114. IOS Press, Amsterdam, The Netherlands (2009)
6. Si, M., Marsella, S., Pynadath, D.: THESPIAN: An Architecture for Interactive Pedagogical Drama. In: 12th International Conference on Artificial Intelligence in Education, pp. 595-602 (2005)
7. Bruner, J.: Acts of Meaning. Harvard University Press, Cambridge, MA (1990)
8. Rowe, J., Shores, L., Mott, B., Lester, J.: Integrating Learning and Engagement in Narrative-Centered Learning Environments. In: 10th International Conference on Intelligent Tutoring Systems, pp. 166-177 (2010)
9. Ketelhut, D.: The Impact of Student Self-Efficacy on Scientific Inquiry Skills: An Exploratory Investigation in River City, a Multi-User Virtual Environment. J. Science Ed. and Tech. 16, 99-111 (2007)
10. Zimmerman, B. J.: Self-Regulated Learning and Academic Achievement: An Overview. Ed. Psychologist, 25, 3-18 (1990)
11. Schunk, D.: Social Cognitive Theory and Self-Regulated Learning. In: Zimmerman & Schunk (eds.) Self-Regulated Learning and Academic Achievement: Theoretical Perspectives (2nd ed.), pp. 125-152. Lawrence Erlbaum Associates, Mahwah (2001)
12. Lajoie, S., Derry, S. J. (eds.) Computers as Cognitive Tools. Lawrence Erlbaum Associates, Hilldale, NJ (1993)
13. Aleven, V., Koedinger, K.: An Effective Metacognitive Strategy: Learning by Doing and Explaining with a Computer-Based Cognitive Tutor. Cog. Sci.: A Multidisciplinary Journal, 26, 147-179 (2002)
14. Conati, C., VanLehn, K. Towards Computer-Based Support of Meta-Cognitive Skills: a Computational Framework to Coach Self-Explanation. Int. J. AIED, 11, 398-415 (2000)
15. Biswas, G., Jeong, H, Roscoe, R., Sulcer, B.: Promoting Motivation and Self-Regulated Learning Skills through Social Interactions in an Agent-Based Learning Environment. In: 2009 AAAI Fall Symposium on Cognitive and Metacognitive Educational Systems (2009)
16. Azevedo, R., Witherspoon, A., Chauncey, A., Burkett, C., Fike, A.: MetaTutor: A Metacognitive Tool for Enhancing Self-Regulated Learning. In: 2009 AAAI Fall Symposium on Cognitive and Metacognitive Educational Systems, pp. 34-39 (2009)
17. Rowe, J., Shores, L., Mott, B., Lester, J.: A Framework for Narrative Adaptation in Interactive Story-Based Learning Environments. In: Workshop on Intelligent Narrative Technologies III at the 5th International Conference on Foundations of Digital Games (2010)
18. Shores, L., Nietfeld, J.: The Role of Compensatory Scaffolds for Inquiry Learning in Narrative-Centered Learning Environments. To appear in: American Educational Research Association (in press 2011)
19. Witten, I., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques (2nd ed.). Morgan Kaufman, San Francisco (2005)
20. Kalyuga, S., Ayers, P., Chandler, P., Sweller, J.: The Expertise Reversal Effect. Ed. Psychologist, 38, 23-31 (2003)