Enhancing Engagement Modeling in Game-Based Learning Environments with Student-Agent Discourse Analysis

Alex Goslen[1], Nathan Henderson[1], Jonathan Rowe[1], Jiayi Zhang[4], Stephen Hutt[2], Jaclyn Ocumpaugh[4], Eric Wiebe[1], Kristy Elizabeth Boyer[3], Bradford Mott[1], and James Lester[1]

[1] North Carolina State University, Raleigh, NC, USA
amgoslen@ncsu.edu
[2]University of Denver, Denver, CO, USA
[3] University of Florida, Gainesville, FL, USA
[4] University of Pennsylvania, Philadelphia, PA, USA

Abstract. Pedagogical agents offer significant promise for engaging students in learning. In this paper, we investigate students' conversational interactions with a pedagogical agent in a game-based learning environment for middle school science education. We utilize word embeddings of student-agent conversations along with features distilled from students' in-game actions to induce predictive models of student engagement. An evaluation of the models' accuracy and early prediction performance indicates that features derived from students' conversations with the pedagogical agent yield the highest accuracy for predicting student engagement. Results also show that combining student problem-solving features and conversation features yields higher performance than a problem solving-only feature set. Overall, the findings suggest that student-agent conversations can greatly enhance student models for game-based learning environments.

Keywords: Student engagement, Game-based learning, Discourse analysis.

1 Introduction

Student engagement plays a central role in effective learning across a wide range of educational settings [7]. Students who disengage often develop a superficial understanding of the material [5]. Pedagogical agents have shown potential in enhancing student engagement through discursive interaction with students [8] and improving student learning [18]. Additionally, positive interactions with pedagogical agents have been shown to help learners feel more engaged [1].

In game-based learning environments, understanding student engagement is multifaceted and may involve measuring students' degree of attention to stimuli in the game (involvement) and their affective response to those stimuli (situational interest) [4]. Pedagogical agents offer the potential to provide such insights for measuring and modeling student engagement [9]. Although prior work has investigated how dialogue with conversational pedagogical agents impacts learners [17], little is known about the relationship between students' dialogue and their overall engagement.

This paper analyzes student discourse with a conversational pedagogical agent in a game-based learning environment. Specifically, we examine six machine learning techniques using neural word embeddings of student-agent discourse and in-game problem solving behavior as input features to predict student engagement. We evaluate the models in terms of accuracy and early prediction performance, and we examine implications of the results for the design of game-based learning environments.

2 Related Work

There is growing literature on utilizing discourse for learning analytics and predicting student engagement in learning environments. Modeling student engagement is an important step in developing adaptive learning environments that can mitigate issues like disengagement or mind-wandering [2]. Emerson et al. [4] investigated features related to student interactions with non-player characters (NPC) in a game-based learning environment and observed that students who interacted more frequently with NPCs showed lower overall interest in the game-based learning environment. This suggests that using students' problem-solving actions and conversational behavior to predict student engagement may help inform adaptive responses for pedagogical agents.

Advances in natural language processing allow for improvements in analyzing unstructured text in dialogue-based learning environments [10]. Analysis of such has included manual annotation and bag-of-words methods to derive meaning from students' text-based utterances [11] and demonstrated the ability of pre-trained neural embeddings to detect student engagement in reflection tasks [6]. More recently, BERT [3] has proven useful for modeling different types of conversational strategies [16].

3 Game-Based Learning Environment

CRYSTAL ISLAND is a game-based learning environment designed to support middle school students learning microbiology through the narrative of an illness outbreak on a remote island research station. A text-based conversational pedagogical agent, Alisha, was integrated into the game [15]. Alisha provides students with an opportunity to share updates about their problem-solving progress and receive pedagogical support by prompting in-game actions like exploring the island or talking to an NPC. Alisha's dialogue was controlled by a finite state machine based on students' dialogue acts [13].

4 Method

4.1 Study Procedure

We use data from 77 middle school students who played CRYSTAL ISLAND over a 3-day study at a public, urban middle school in North Carolina. These students completed both pre- and post-game surveys (32 male, 38 female, and 7 who did not report gender). Following gameplay, students were asked to complete three surveys that were used as a proxy for engagement-related constructs. These were drawn from the original version of the User Engagement Scale (UES) [14] and a revised version (UESz) [19].

4.2 Feature Generation for Predictive Models of Student Engagement

This study utilizes three subscales most relevant to understanding engagement in game-based learning environments: the novelty subscale (NO) and felt involvement (FI) subscale from the original UES and the focused attention (FAz) subscale from the UESz [20]. Novelty (NO) captures a basic measure of situational interest. Focused attention (FAz) captures a retrospective rating of flow-like experience. Felt involvement (FI) can be described as the enjoyment at the intersection of the two previous engagement constructs. We binarized this data using a median split to indicate low and high levels of FAz (*med*=24, *SD*=7.36), FI (*med*=11, *SD*=3.05), and NO (*med*=10, *SD*=3.01).

Student-Agent Conversation Features. There was a total of 2,634 chat messages sent throughout gameplay, with 1,523 messages originating from Alisha and 1,111 messages originating from students. On average, students sent 14.4 messages (*SD*=15.8) and Alisha sent 14.5 messages (*SD*=14.7) across the entirety of gameplay. To generate representations of the students' utterances for use by the pre-trained BERT model [3], tokens were generated for the separate words for each utterance, with stop words being retained due to their contextual relevance. Utterance-level feature vectors were produced by averaging across each token sequence. We constructed both sequential and non-sequential embedding input representations. Sequential models utilized a two-dimensional vector representation, where each row represented a single utterance and each column represented a conversation feature. By summing across sequences, a one-dimensional feature representation was generated for the non-sequential models.

In-Game Problem-Solving Features. To create predictive models of student engagement, we derived problem-solving features from students' gameplay interactions with CRYSTAL ISLAND captured by trace logs: action type, action argument, and location [12]. Action type features represent students' actions during gameplay, action arguments provide details about students' problem-solving actions, and location features denote where the gameplay action took place. We extracted these features from students' gameplay data and converted each action into a binary one-hot vector. Sequential and non-sequential features were generated similarly to the conversation features.

Combined In-Game Problem-Solving and Student-Agent Conversation Features. Finally, we combined in-game features and student-agent conversation features into a single combined representation via sequence-level concatenation.

4.3 Models and Evaluation

The six supervised machine learning techniques used were support vector machines (SVMs), random forest (RF), logistic regression (LR), Naïve Bayes (NB), multilayer perceptron (MLP), and long short-term memory networks (LSTMs). LSTM models utilized sequential input representations and the other models utilized non-sequential input representations. We used student-level cross-validation splits to eliminate data leakage.

Given variability in students' total gameplay time (*M*=75.41 min, *SD*=29.41 min), and a limited number of conversations per student, we split the data cumulatively using five-minute increments and stopped early prediction models after 10 increments.

The models were evaluated across gameplay time intervals in terms of accuracy as well as two early prediction metrics: convergence rate and standardized convergence point. Convergence rate measures the percentage of early predictions that have an accurate final prediction value. Standardized convergence point measures the point at which a model only makes correct predictions from there [19]. Higher convergence rates and lower convergence points correspond to improved predictive performance.

5 Results

Table 2 shows all results from the three representations. All six machine learning models yielded engagement models that improved on a majority-class baseline in terms of accuracy for FAz using conversation-only features. SVM, NB and LR achieved the highest predictive performance in terms of convergence rate and standardized convergence point for FAz, with MLP and LSTM performing well for FI. LSTMs had the lowest convergence rate and highest convergence point of all methods tested, likely due to the small size of the student-agent conversation dataset. Most models failed to converge for NO, implying that novelty might be more difficult to predict in this case.

Table 2. Model results for all experiments. C represents student-agent conversation input, G represents gameplay only input, and C + G represent the combined input representation.

		FAz			FI			NO		
		Acc	CR	CP	Acc	CR	CP	Acc	CR	CP
	Majority	55.84	N/A	N/A	57.14	N/A	N/A	55.84	N/A	N/A
C	SVM	**63.90**	74.00	87.92	55.90	64.00	**94.14**	52.57	58.00	101.50
	RF	61.72	62.00	94.10	58.55	58.00	101.20	**58.16**	60.00	99.12
	NB	59.95	78.00	**86.96**	48.79	26.00	104.60	49.90	28.00	106.16
	LR	61.21	**84.00**	88.98	58.69	**72.00**	96.94	57.71	**80.00**	**98.96**
	MLP	58.62	72.00	96.04	**62.70**	70.00	94.32	55.81	58.00	101.82
	LSTM	58.59	62.00	95.92	62.16	54.00	99.16	59.50	58.00	96.40
G	SVM	52.69	**74.00**	**94.86**	56.01	66.00	**86.30**	51.91	60.00	93.70
	RF	54.39	72.00	96.74	**58.53**	**70.00**	92.10	**59.53**	58.00	97.54
	NB	**56.64**	56.00	97.94	51.25	44.00	99.78	56.50	42.00	97.30
	LR	52.15	68.00	97.42	47.66	54.00	96.28	53.00	68.00	94.24
	MLP	53.63	62.00	97.20	49.87	58.00	95.34	52.98	56.00	97.38
	LSTM	53.97	92.00	85.50	55.37	90.00	79.24	53.84	**88.00**	**82.10**
C + G	SVM	53.06	82.00	88.44	53.80	**60.00**	**87.62**	53.58	58.00	94.22
	RF	54.61	80.00	94.92	**57.01**	**60.00**	91.96	**57.51**	66.00	92.86
	NB	**56.83**	**92.00**	**82.64**	51.28	50.00	93.94	55.36	**86.00**	**83.94**
	LR	48.68	66.00	97.46	48.70	46.00	99.42	51.92	56.00	97.08
	MLP	50.05	56.00	98.10	47.09	52.00	97.96	52.57	54.00	96.74
	LSTM	52.75	90.00	83.48	54.29	92.00	90.70	55.66	86.00	81.04

Results for early prediction models constructed with in-game problem-solving features showed lower accuracy for focused attention and felt involvement and higher accuracy for novelty. For all three engagement measures, most models performed worse than the majority baseline. Overall, models constructed with only in-game problem-solving features performed worse than models using only student-agent conversation features in terms of convergence rate and convergence point with the exclusion of LSTM. These results show promise for using information about student-agent conversational behavior for predicting engagement in game-based learning environments.

All models follow similar trends to the gameplay-only results shown in part G of the table. Notably, the accuracy values for the combined feature representation models (C + G) were not higher than the conversation-only feature models (C). However, the combined representation yields much better predictive performance for FAz and NO, meaning the combined representation predictions converge at a higher rate and generate more accurate predictions. This finding implies that the combined feature representation might be optimal for real-time predictions. All three evaluation metrics appear to improve from the combined feature representation relative to the in-game problem-solving feature representation for focused attention prediction, which suggests that the addition of student-agent conversation features improved predictive model performance.

6 Discussion and Conclusion

Student conversations with pedagogical agents in game-based learning environments can inform predictive models of engagement. Analyses of student-agent conversational behavior in a game-based learning environment for science problem-solving reveal that the conversation features performed best in terms of predictive accuracy for focused attention and felt involvement and improved on a majority baseline. We also found that accuracy, convergence rate, and standardized convergence point improved when adding student-agent conversation features to in-game problem-solving features, suggesting student-agent conversation is an important indicator for engagement modeling.

A limitation of the study is that the surveys were collected retrospectively, immediately after students' completion of the game. For a better understanding of student engagement, intermittent engagement reporting data could be introduced throughout gameplay. Understanding changes in engagement over the course of the student's interaction with the game-based learning environment might impact the observed relationship between student discourse patterns and engagement. Additional measurements of engagement could be also explored since novelty and felt involvement are more summative in. Analyzing student-agent conversations as a means of predicting student engagement shows significant promise and could contribute to cultivating student interest and engagement during science problem-solving in game-based learning environments. Future work should focus on investigating additional text analytic techniques, such as sentiment analysis, to enhance automated analysis of student discourse. Furthermore, mapping trajectories of student engagement over time could prove useful to instructors for guiding pedagogical interventions. Further analysis could also be done to see how patterns of student conversational engagement relate to student learning gains, and real-time tracking of these features could inform adaptive scaffolding for supporting engagement.

References

1. Burgoon, J., et al.: Application of expectancy violations theory to communication with and judgments about embodied agents during a decision-making task. International Journal of Human-Computer Studies, 91, 24-36. (2016).
2. Dermouche, S., Pelachaud, C.: Engagement modeling in dyadic interaction. In: Proceedings of the 2019 International Conference on Multimodal Interaction, pp. 440-445. (2019).
3. Devlin, J., et al.: Bert: Pre-training of deep bidirectional transformers for language understanding. North American Association for Computational Linguistics (NAACL). (2018).
4. Emerson, A., et al.: Multimodal learning analytics for game-based learning. British Journal of Educational Technology, 51(5), 1505-1526 (2020).
5. Forbes-Riley, K., Litman, D.: Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. Speech Communication, 53(9-10), 1115-1136 (2011).
6. Geden, M., et al.: Predictive student modeling in game-based learning environments with word embedding representations of reflection. International Journal of Artificial Intelligence in Education, 31(1), 1-23 (2021).
7. Gobert, J., Baker, R., & Wixon, M.: Operationalizing and detecting disengagement within online science microworlds. Educational Psychologist, 50(1), 43-57 (2015).
8. Graesser, A.: Conversations with AutoTutor help students learn. International Journal of Artificial Intelligence in Education, 26(1), 124-132 (2016).
9. Graesser, A., et al.: Advancing the Science of Collaborative Problem Solving. Psychological Science in the Public Interest, 19(2), 59–92 (2018).
10. Hirschberg, J., Manning, C.: Advances in natural language processing. Science, 349(6245), 261-266 (2015).
11. Lin, Z., et al.: Mintl: Minimalist transfer learning for task-oriented dialogue systems. arXiv preprint arXiv:2009.12005. (2020).
12. Min, W., et al.: Multimodal goal recognition in open-world digital games. In: 13th Artificial Intelligence and Interactive Digital Entertainment Conference, pp. 80-86. (2017).
13. Min, W., et al.: Predicting Dialogue Acts for Intelligent Virtual Agents with Multimodal Student Interaction Data. International Educational Data Mining Society (2016).
14. O'Brien, H., Toms, E.: The development and evaluation of a survey to measure user engagement. Journal of the American Society for Information Science and Technology, 61(1), 50-69 (2010).
15. Pezzullo, L., et al.: "Thanks Alisha, keep in touch": gender effects and engagement with virtual learning companions. In Proceedings of the 18th International Conference on AIED, pp. 299-310, Wuhan, China, Springer International Publishing (2017).
16. Pugh, S., et al.: Do Speech-Based Collaboration Analytics Generalize Across Task Contexts?. In: 12th International LAK Conference, pp. 208-218. (2022).
17. Sikström, P., et al.: How pedagogical agents communicate with students: A two-phase systematic review. Computers & Education, 104564 (2022).
18. Tegos, S., et al.: Conversational agents for academically productive talk: A comparison of directed and undirected agent interventions. International Journal of Computer-Supported Collaborative Learning, 11(4), 417-440 (2016).
19. Wiebe, E., et al.: Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. Computers in Human Behavior, 32, 123-132 (2014).
20. Zhang, J., et al.: Investigating Student Interest and Engagement in Game-Based Learning Environments. In: Proceedings of the 23rd International Conference AIED, pp. 711-716, Durham, Cham: Springer International Publishing (2022, July).