Balancing Learning and Engagement in Game-Based Learning Environments with Multi-Objective Reinforcement Learning

Robert Sawyer, Jonathan Rowe, James Lester

North Carolina State University, Raleigh, NC 27695
{rssawyer, jprowe, lester}@ncsu.edu

Abstract. Game-based learning environments create rich learning experiences that are both effective and engaging. Recent years have seen growing interest in data-driven techniques for tutorial planning, which dynamically personalize learning experiences by providing hints, feedback, and problem scenarios at run-time. In game-based learning environments, tutorial planners are designed to adapt gameplay events in order to achieve multiple objectives, such as enhancing student learning or student engagement, which may be complementary or competing aims. In this paper, we introduce a multi-objective reinforcement learning framework for inducing game-based tutorial planners that balance between improving learning and engagement in game-based learning environments. We investigate a model-based, linear-scalarized multi-policy algorithm, Convex Hull Value Iteration, to induce a tutorial planner from a corpus of student interactions with a game-based learning environment for middle school science education. Results indicate that multi-objective reinforcement learning creates policies that are more effective at balancing multiple reward sources than single-objective techniques. A qualitative analysis of select policies and multi-objective preference vectors shows how a multi-objective reinforcement learning framework shapes the selection of tutorial actions during students' game-based learning experiences to effectively achieve targeted learning and engagement outcomes.

Keywords: Tutorial Planning, Multi-Objective Reinforcement Learning, Game-Based Learning Environments, Narrative Centered Learning

1 Introduction

Game-based learning environments enable students to engage in rich problem-solving scenarios that enhance student learning. There is compelling evidence that game-based learning environments improve student learning outcomes compared to traditional instructional methods [14, 15]. A key advantage of game-based learning environments is their potential to foster student engagement through features such as 3D virtual worlds and believable characters [6]. However, important questions have been raised about whether specific features of digital games that foster engagement, such as narratives, are beneficial for learning [1]. A one-size-fits-all approach to designing game-based learning environments has significant limitations in terms of balancing effectively between learning and engagement for all students. Recent years have seen growing interest in *tutorial planners* for game-based learning environments, which

personalize game elements to individual students at runtime [4, 16]. Reinforcement learning (RL) techniques have shown particular promise for devising tutorial planners from logs of student interactions with a virtual learning environment [3, 10].

RL-based tutorial planners are often tasked with making personalization decisions that impact both student learning and engagement. Yet, there has been little systematic investigation of *multi-objective RL* techniques for tutorial planning. Multi-objective techniques are particularly relevant to game-based learning environments because there may be tradeoffs between game elements designed to foster learning and game elements designed to foster engagement. Prior work on RL-based planners has typically focused on single-objective reward models [3, 9] and weighted sum-based evaluation functions with author-specified weights [5]. Single-objective RL techniques provide no guarantees about generating policies that balance across multiple objectives. A tutorial planner that is effective for one objective (e.g., learning) may be ineffective for a secondary objective of comparable importance (e.g., engagement). Further, the weight preferences between objectives for a particular game-based learning environment may not be known *a priori*, as they may be dependent upon the educational setting in which a game-based learning environment will be deployed. For example, a tutorial planner intended to support classroom practice before end-of-grade tests might prioritize content learning gains, whereas a game utilized in an after-school setting might optimize engagement and interest in the subject matter.

In this paper, we present a multi-objective RL framework for tutorial planning in game-based learning environments. Using game interaction log data from over four hundred students, we induce a tutorial planner for a game-based learning environment for middle school microbiology education, CRYSTAL ISLAND.

2 Related Work

Data-driven methods for tutorial planning have been the subject of growing interest in recent years. RL techniques have shown particular promise, potentially reducing the need for labor-intensive knowledge engineering processes and large datasets of human demonstrations [3, 5, 10]. Many RL techniques formalize tutorial planning in terms of Markov decision processes, which encode sequential decision-making tasks with stochastic environments and delayed rewards. Chi et al. [3] utilized model-based RL to induce models of pedagogical micro-tactics in a tutorial dialogue system for physics education. More recently, Mandel et al. [16] investigated techniques for offline RL policy evaluation to examine alternate tutorial planning models in the educational game Refraction. Rowe et al. [9] investigated a modular reinforcement learning framework for tutorial planning in educational interactive narratives. Their model, which was evaluated in a classroom study, was found to yield improved student learning behaviors relative to a baseline system [10]. Each of these systems utilized single-objective reward functions to guide RL techniques for inducing tutorial planning models.

In related work on user-adaptive games, Nelson et al. [5] proposed an RL framework for experience management that leveraged a hand-authored evaluation function to personalize events in interactive fiction games. Notably, the evaluation function utilized by Nelson et al. adopted the form of a linear scalarization function with weight preferences. This approach required the system designer to specify weights among objectives prior to training the experience manager. This approach is intuitive, but it is

unlikely to generalize effectively across different deployment settings with distinct priorities for users' gameplay experiences.

Multi-objective RL techniques consist of methods for solving a wide array of multi-objective Markov decision processes, with solutions consisting of a single policy or multiple policies depending on the problem context [7]. Recent work by Wiering, Withagen, and Drugan [12] presented a model-based approach for solving deterministic multi-objective Markov decision processes yielding the set of Pareto optimal policies for a given task. Barrett and Narayanan [2] devised a method for calculating all optimal policies for any weight preference vector used in linear scalarization. Their approach enables a system designer to defer specifying weight preferences for each objective until the RL model is deployed, when a specific policy is extracted at run-time by utilizing properties of convex hulls. Multi-objective RL has been applied successfully in a variety of domains, including traffic light control [18] and water reservoir control [17], but to date there has been little work investigating multi-objective RL techniques for educational software.

3 CRYSTAL ISLAND Game-Based Learning Environment

To investigate multi-objective RL for tutorial planning, we utilize a game-based learning environment for middle school microbiology education as a testbed application, CRYSTAL ISLAND. In CRYSTAL ISLAND, students adopt the role of a medical field agent, who has been tasked with investigating a mysterious epidemic on a remote island. The student must determine the source and identity of the illness by interviewing virtual characters, gathering clues, and running tests in a virtual laboratory. As students solve the mystery, they learn relevant microbiology concepts and utilize the scientific method to complete the science problem-solving scenario. CRYSTAL ISLAND has been used by over 4,000 students in middle school classrooms across the United States.

Tutorial planning in CRYSTAL ISLAND encompasses a broad range of possible decisions about scaffolding student learning and tailoring different elements of the game environment. We seek to induce tutorial planning policies directly from a corpus of student interaction data off-line. To address issues of data sparsity, we decompose tutorial planning in terms of several distinct sub-problems, denoted as *adaptable event sequences* (AESs). An AES is an abstraction for one or more recurring tutorial decision-making events that center on a particular facet of the game-based learning environment, such as the behavior of a non-player character, the properties of a virtual object, or the delivery of a scaffolding-related message. We model CRYSTAL ISLAND's tutorial planner with a set of 12 AESs, each separately encoding a series of sequential game events, which interleave with one another and collectively span the game's problem scenario (Figure 1).

In our multi-objective RL framework, each AES is modeled as a *multi-objective Markov decision process* (MOMDP) with its own state representation, action set, state transition model, and reward model. Every occurrence of an AES corresponds to a decision point for the MOMDP. The possible gameplay adaptations that can be performed by the tutorial planner represent the sets of actions for the MOMDPs. In order to collect a corpus of student interaction data for off-line RL, we deployed CRYSTAL ISLAND to students using a version of the tutorial planner that controls AESs according to a uniform random policy, deliberately sampling the manager's state-action

space. As long as each possible combination of gameplay adaptations produces a coherent user experience, we can collect a corpus of student responses to the tutorial planner's decisions for off-line, model-based RL.

Data for inducing tutorial planning policies from student interactions with CRYSTAL ISLAND were collected from two studies. The first study involved 300 students from a middle school and the second study involved 153 students from a different middle school. Students interacted with the game until they solved the mystery, or 55 minutes elapsed, whichever occurred first. Students completed pre- and post-tests one-week before, and immediately after using the game, respectively. These tests gathered data on students' learning gains, prior gameplay experience, and perceptions of *presence* (i.e., the sense of "being there" in the virtual environment) experienced in the game.

Each student's trace of in-game problem-solving actions was logged, including which AESs they encountered, what actions were performed by the tutorial planner (according to a uniform random policy), and timestamps for all game events. After removing data from participants with incomplete or inconsistent records, the resulting data set consisted of 10,057 instances of tutorial planner decisions, corresponding to approximately 25 gameplay adaptations per player.

Each MOMDP shared the same state representation, which consisted of 8 binary features drawn from three categories: narrative state, gameplay behavior, and player traits. We limited the state representation to 8 binary features to mitigate potential data sparsity issues. The first four features were narrative-focused. Each feature was associated with a salient plot point from CRYSTAL ISLAND's storyline and indicated whether the plot point had been completed thus far. The next two features were computed from a median split on players' microbiology pre-test scores and previous video game experience. The final two features were computed from players' observed gameplay behaviors. Specifically, we computed running median splits on the frequency of students' laboratory testing and book-reading behaviors within CRYSTAL ISLAND.

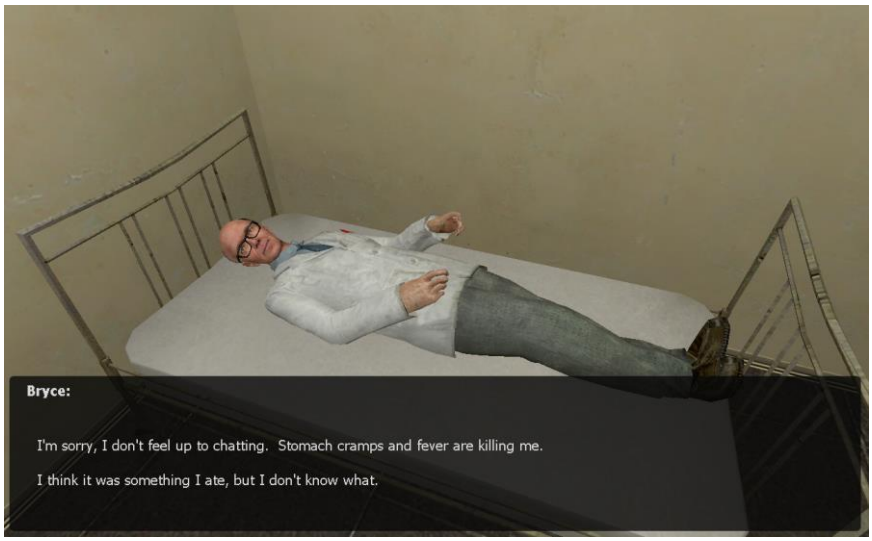

Fig 1. Screenshot of Bryce Symptoms AES in CRYSTAL ISLAND

The action sets for the 12 MOMDPs corresponded to the range of gameplay personalization decisions for the associated AESs. The action sets' cardinalities ranged from binary to 6-way decisions.

Table 1. Summary of AESs by type, name, and number of possible actions. R refers to recurring AESs and O refers to AESs that occur once per episode. Asterisks denote policies selected for additional qualitative analysis in the results section below.

AES Type	AES Name	Cardinality	AES Frequency
Scaffolding	Direct Goal	2	R
	Increase Urgency	2	R
	Knowledge Quiz*	2	R
	Record Findings	2	R
	Reflection Prompt	2	R
Information Availability	Bryce Reveal	2	O
	Bryce Symptoms	2	R
	Quentin Reveal	2	O
	Teresa Symptoms*	3	R
Problem Specification	Mystery Solution	6	O
	Test Count*	3	O
	Worksheet Level	3	R

The AESs ranged broadly in terms of how they affected student gameplay, as well as their frequency of occurring during a typical gameplay episode. Detailed information regarding each AES is provided in [8], and these groupings are summarized in Table 1. If the entire tutorial planning task were modeled as a single MOMDP, it would require encoding approximately 1,644,000 parameters to populate the entire state transition model (256 states × 25 distinct actions × 257 states, including the terminal state), although not all state transitions were possible.

Two distinct reward sources were computed using data from the corpus described above to induce RL-based tutorial planning policies. Each MOMDP utilized the same set of two reward models, which were based upon: (1) participants' normalized learning gains, and (2) self-reported presence after gameplay. Both of these reward sources were calculated using data collected from the pre- and post-tests; no incremental rewards were assigned during gameplay.

The first reward source, normalized learning gain, was selected to obtain a tutorial planner that maximized student learning on microbiology content. Normalized learning gain (NLG) is the normalized difference between pre- and post-game science content knowledge test scores, assessed using a 19-item multiple-choice test. We use NLG because it provides a singular metric for student learning that accounts for individual differences in students' prior knowledge, in contrast to alternative metrics like post-test score or un-normalized learning gain. The reward values for NLG were determined by calculating the NLG for each participant at the conclusion of their gameplay episode.

The second reward source was based upon players' self-reported perceptions of presence, as measured by the Presence Questionnaire [13]. Presence refers to a participant's perception of transportation into a virtual environment. We use it here as a proxy indicator for user engagement in the game. Participants completed the Presence Questionnaire after using CRYSTAL ISLAND. The presence reward function was determined by the student's total Presence Questionnaire score divided by the maximum observed score in the corpus. This normalized the presence reward to be in the interval [0,1] for each student. This objective is important to maximize because fostering engagement is a key motivation of game-based learning environments. These two reward sources reflect each side of the tradeoff between learning and engagement in interactive narrative.

The MOMDPs, one for each AES in CRYSTAL ISLAND, were implemented with a reinforcement learning library written in Python by the first two authors. Policies were induced using a discount rate of 0.9. To encode multiple reward sources for MORL, a vector containing each of the two reward sources was utilized.

4 Multi-Objective Reinforcement Learning for Tutorial Planning

Several multi-objective policies were induced for each AES from the corpus of student interaction data using both the NLG and Presence reward sources. A certainty-equivalence model of the environment was created from the state-action transition counts and observed rewards in the training corpus. This is done with the maximum likelihood model of the MOMDP as in [12].

We derive multiple policies per MOMDP using Convex Hull Value Iteration [2]. This method learns the set of all optimal policies for an MOMDP given a model of the environment through operations on convex hulls similar to the classical dynamic programming method of *value iteration* [11]. In Convex Hull Value Iteration, each Q-value is replaced with a set of possible expected reward vectors. If this set is a convex hull, then each possible vector is optimal under some set of preferences over the reward sources, defined as a weight preference vector where the components sum to one. Given a weight preference vector, the best linear scalarized reward *Q* can be extracted according to the following equation:

$$Q_{\vec{w}}(s,a) = \max_{\vec{q} \in \dot{Q}(s,a)} \vec{w} \cdot \vec{q} \qquad (1)$$

where $\vec{w}$ represents the weight preference vector, $\dot{Q}(s,a)$ is the convex set of optimal reward vectors for a state-action pair, and $Q_{\vec{w}}(s,a)$ is the resulting linear scalarized Q-value. Once the Q-values have been scalarized by a weight preference vector, a policy can be obtained greedily by selecting the best action per state, because Q-values take expected discounted future rewards into account. The weight vector is constrained to consist of positive real numbers that sum to one.

Since CRYSTAL ISLAND can be used in many different educational settings (e.g. classrooms, home, after-school clubs), the tutorial planner requires a weight preference vector defined at run-time, which is contingent on the particular educational priorities

of the deployment setting. This results in the need for a multi-policy approach that can learn all optimal policies regardless of the preference weight vector that will be utilized at run-time.

In order to evaluate the policies derived from Convex Hull Value Iteration, we used the extraction method from Equation 1 to generate all distinct policies for each MOMDP. This was performed by generating the convex sets of Q-values for each MOMDP, running a grid search over weight preference vectors to extract their corresponding Q-values, and utilizing greedy selection to derive distinct policies for each MOMDP. Multiple policies were derived for each MOMDP because optimal mappings between states and actions may be dependent on the weight preference vector. Every policy induced with this method is optimal under some subset of the possible weight preference vectors. In the case of tutorial planning in CRYSTAL ISLAND, we considered two reward sources—NLG and Presence—that together sum to 1. In other words, if NLG is the primary reward source, then the secondary objective Presence is assigned a weight of 1 – NLG in the weight preference vector.

5 Evaluation

The multi-objective RL framework yields multiple policies for each AES because a weight preference vector is not specified prior to training the model. Thus, for different specifications of the weight preference vector, different optimal policies can be obtained. The number of distinct policies generated for a single AES from the multi-objective RL procedure varied from a minimum of 3 (Mystery Solution AES) to a maximum of 11 (Reflection Prompt AES), with a median of 7 distinct policies per AES.

In order to evaluate the quality of the policies induced using multi-objective RL, we conducted an analysis of the policies' expected cumulative rewards for each reward component. *Expected cumulative reward* (ECR) is a measure of the average anticipated reward produced by a policy across all possible gameplay episodes and start states [11]. ECR is calculated by taking the product of the expected discounted reward for each start state with the probability of starting in that state. We compare ECR results calculated by each reward source between each set of induced policies. The convex hull of the MOMDP can be visualized by plotting the expected cumulative reward vector for each distinct policy induced for that MOMDP.

Due to space constraints, we focus on presenting results from 3 of the 12 AES convex hulls in this section. These 3 AESs were chosen as representative examples of each of the three AES categories: Scaffolding, Information Availability, and Problem Specification. They serve as two examples of recurring AESs and one example of an AES that occurs once per episode. The Knowledge Quiz AES, a recurring, scaffolding AES, specifies whether to provide a student with an in-game microbiology quiz or not at several specific points in the problem scenario. The Test Count AES, a single-occurrence problem specification AES, determines whether the student is allotted three, five, or ten initial "scans" with the hypothesis testing equipment in the game's virtual laboratory. The Teresa Symptoms AES, a recurring information availability AES, determines whether a particular non-player character will provide minimum, moderate,

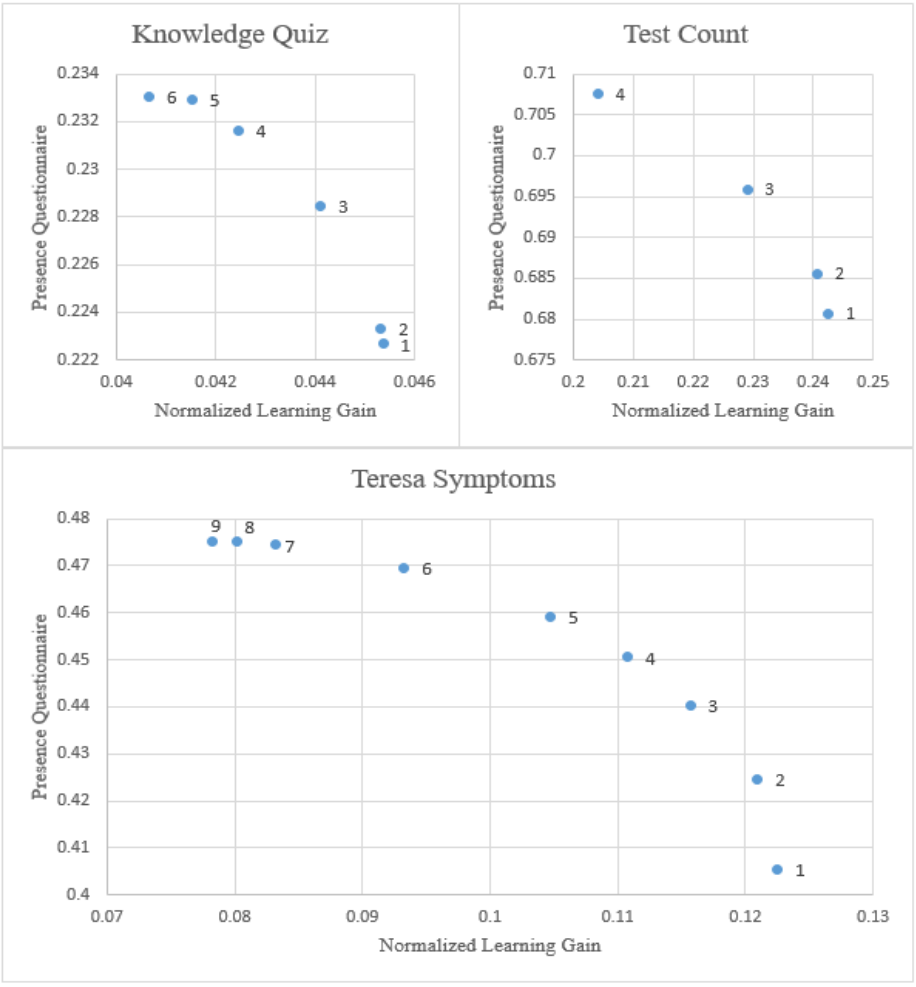

Fig 2. Scatter plot of ECR vectors for select AESs. X-axis denotes NLG reward values, and y-axis denotes Presence reward values.

or maximum detail regarding her symptoms during a branching conversation with the student.

Figure 2 shows the ECR vectors of distinct policies induced by the multi-objective RL framework for the three selected AESs. The x-axis denotes the NLG ECR value of a policy, and the y-axis denotes the Presence ECR value. A qualitative analysis of policies for each AES reveals how changing the weight preference vector affects action choices for the tutorial planner.

In the Knowledge Quiz AES, as the NLG weight decreases, the induced policies tend to give fewer quizzes to students who have read a higher number of books and have higher prior content knowledge. Because this change of policy comes from decreasing the NLG weight (and therefore increasing the Presence weight), this

indicates that presenting the knowledge quizzes may reduce engagement in students who are already familiar with microbiology content, or who are now more knowledgeable from reading the virtual microbiology books. Conversely, this indicates that in-game quizzes may help learning but diminish engagement; it is plausible that quizzes disrupt the flow of gameplay and reduce perceptions of presence in the virtual environment.

In the Test Count AES, the policies induced by weight preference vectors that de-prioritize NLG tend to allot more initial "scans" to students with high number of books read. This indicates that letting students that have already gathered information from reading perform more tests may help engage the students at the cost of decreased learning gains. This may be a way of keeping students engaged by allowing students who have spent time gathering information to form hypotheses continue the problem-solving process by thoroughly testing their hypotheses.

In the Teresa Symptoms AES, policies induced with weight preference vectors that prioritize Presence tended to provide fewer details when students had high prior content knowledge and more detail when students had high prior gameplay experience. This indicates that giving less information to students with high prior content knowledge may help keep them engaged, and it may have also helped engage students who were performing a high number of scans. The lack of information given to a student with high prior content knowledge effectively increases the scenario's difficulty, which may lead to a more appropriate challenge level for a high knowledge student.

In summary, tutorial planning policies are noticeably influenced by the weight assignments in the multi-objective preference vector. In general, increased weight for the NLG reward source corresponds to increased learning support from the tutorial planner. This trend can be observed for both the Knowledge Quiz AES (i.e., more quizzes are given) and Teresa Symptoms AES (i.e., more detailed information is given) with higher NLG weights. The Test Count AES is an exception, where allotting an increased number of tests—an indirect form of learning support—corresponds to a reduction in NLG weight. However, students "earn" additional tests by completing in-game quizzes, so it may be the case that students with fewer allotted tests complete more remedial quizzes, which could be associated with higher learning gains. It should

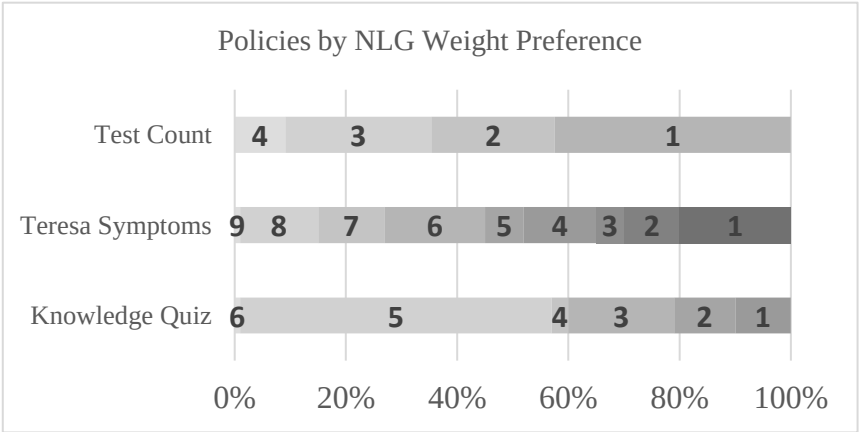

Fig 3. Policies preferred by various values of NLG weight (with Presence = 1 – NLG) for three different AESs corresponding to policies in Figure 2.

be noted that this trend is only observed for students with a strong tendency toward book reading. This would be consistent with a tutorial planner that seeks to limit guessing behavior to encourage learning among students that have already read the relevant content.

As noted above, each of the policies induced is optimal over some subset of possible weight preference vectors. In Figure 3, the subsets of weight preference vectors associated with each optimal policy (from the three AESs shown in Figure 2) are visually represented. The policy numbers corresponding to the hulls from Figure 2 are centered on the ranges of NLG weights that make those policies optimal. For example, Policy 1 in each AES is the policy that favors NLG most and Presence least. In the Test Count AES, this policy is optimal under all weight preference vectors from NLG = 0.58 to NLG = 1.0 (with the corresponding Presence = 0.42 to Presence = 0.0). This image also shows that Policy 2 for Test Count, Policy 5 for Teresa Symptoms, and Policy 5 for Knowledge Quiz are optimal under a weight preference vector that gives even preference to NLG and Presence, i.e. NLG = Presence = 0.5.

Next, we statistically compared policies induced for different weight preference vectors using the multi-objective RL framework. To perform this comparison, we conducted a series of paired t-tests, where each pair consisted of the reward-specific ECR values for two different policies associated with a single AES. Each weight preference vector corresponds to a set of policies from the convex hull; the set is comprised of one policy for each AES. Thus, for two distinct weight preference vectors, there are 12 pairs of policies. We calculate the 12 differences between policy ECRs and average (and take the standard deviation of) these ECR differences to compare two distinct preference weight vectors. These tests investigated whether the ECR value for a particular reward source was statistically different across policies induced by two distinct weight preference vectors.

For example, consider the Teresa Symptoms AES and its induced policies: Policy 1 (induced by $\mathbf{w}$ = [1.0, 0.0]) and Policy 2 (induced by $\mathbf{w}$ = [0.75, 0.25]). We want to compare NLG ECR values between the two policies. From the data in Figure 2, we see that NLG ECR of Policy 1 is 0.122 and the NLG ECR of Policy 2 is 0.121, yielding a pairwise NLG difference of 0.001. This difference is averaged with differences between other AESs, providing the mean NLG ECR difference between policies induced by two weight preference vectors.

Table 2. Paired t-tests comparing policies from different weight preference vectors with differences averaged across all AESs.

Weight Vector One	Weight Vector Two	Reward Source	Mean Difference	SD	t (p-value)
[1.0, 0.0]	[0.75, 0.25]	NLG	0.00421	0.00372	3.76 (0.003)**
[1.0, 0.0]	[0.75, 0.25]	Presence	-0.0209	0.0214	-3.25 (0.008)**
[1.0, 0.0]	[0.5, 0.5]	NLG	0.0132	0.00818	4.99 (< 0.001)***
[1.0, 0.0]	[0.5, 0.5]	Presence	-0.0313	0.0258	-4.02 (0.002)**
[0.5, 0.5]	[0.0, 1.0]	NLG	0.0194	0.0248	2.601 (0.025)*
[0.5, 0.5]	[0.0, 1.0]	Presence	-0.00501	0.00554	-3.00 (0.012)*
[0.25, 0.75]	[0.0, 1.0]	NLG	0.00879	0.0182	1.60 (0.138)
[0.25, 0.75]	[0.0, 1.0]	Presence	-0.0001	0.002	-1.53 (0.153)

It should be noted that policies induced using linear scalarization with weights [1.0, 0.0] and [0.0, 1.0] are equivalent to single-objective policies, enabling a statistical comparison between single-objective and multi-objective policies. The results from these paired t-tests are shown in Table 2.

Table 2 indicates that policies induced with different weight preference vectors have significant differences in ECR across both reward sources when paired by AES. A negative Mean Difference represents the case when policies induced by Weight Vector Two are greater than the policies induced by Weight Vector One for that reward source. Results show that the equal-preference policy given by $\mathbf{w}$ = [0.5, 0.5] outperforms single-objective policies in the secondary objective, but it does not perform as well on the primary objective.

6 Conclusion

Dynamically balancing between multiple objectives is a key functionality of tutorial planners for a broad range of interactive learning environments ranging from intelligent tutoring systems to game-based learning environments. We have presented a multi-objective reinforcement learning framework for tutorial planning in game-based learning environments that addresses the problem of incorporating multiple reward sources, such as learning and engagement, into a data-driven framework for tutorial planning. Our multi-objective RL framework has been investigated in the context of a game-based learning environment for middle school microbiology education, and it was trained using a corpus of student interaction data from classroom studies involving over 400 participants. Multiple reward sources (i.e., content learning, engagement) were used to define an MOMDP for the game-based tutorial planner. These reward sources were chosen because they typify the educational objectives often discussed in the design of game-based learning environments. An analysis of different tutorial planning policies induced using multi-objective RL indicated that tutorial planners utilizing these policies provide a more balanced expected cumulative reward on multiple objectives compared to single-objective policies. We generated an approximate convex hull of optimal policies for several AESs, yielding sets of tutorial planning policies that optimize multiple dimensions of students' game-based learning experiences. These policies can be selected at deployment time by specifying a weighted preference vector tailored to a particular educational setting.

In future work, it will be important to investigate alternate representations for multi-objective policies using complementary evaluation methods, such as importance sampling. In addition, we plan to explore techniques for incorporating multi-objective tutorial planners into the run-time decision cycles of a range of learning environments, investigating how best to dynamically create personalized learning experiences that are simultaneously effective for learning and engagement. In this work, we have utilized ECR as a preliminary evaluation metric to assess multi-objective tutorial policies. This lays the foundation for conducting follow on studies with human subjects to investigate multi-object tutorial planning in laboratory and classroom settings.

References

1. Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., Wainess, R.: Narrative Games for Learning: Testing the Discovery and Narrative Hypotheses. Journal of Educational Psychology. 104(1), 235–249 (2012).
2. Barrett, L., Narayanan, S.: Learning All Optimal Policies with Multiple Criteria. In: Proceedings of the 25th Int. Conference on Machine Learning. pp.41-47. ACM (2008).
3. Chi, M., VanLehn, K., Litman, D., Jordan, P.: Empirically evaluating the application of reinforcement learning to the induction of effective and adaptive pedagogical strategies. User Modeling and User-Adapted Interaction 21(1-2), 137-180 (2011)
4. Lee, S., Rowe, J., Mott, B., Lester, J. A Supervised Learning Framework for Modeling Director Agent Strategies in Educational Interactive Narrative. IEEE Transactions on Computational Intelligence and AI in Games 6(2), 203–215 (2014).
5. Nelson, M., Roberts, D., Isbell, C., Mateas, M.: Reinforcement Learning for Declarative Optimization-Based Drama Management. In: Proceedings of the 5th Int. Conference on Autonomous Agents and Multiagent Systems. pp. 775– 782, ACM, Japan (2006).
6. Prensky, M.: Digital Game-based Learning. McGraw-Hill, New York (2001)
7. Roijers, D., Vamplew, P., Whiteson, S., Dazeley, R.: A Survey of Multi-Objective Sequential Decision-Making. J. of Artificial Intelligence Research 48, 67-113 (2013).
8. Rowe, J.: Narrative-Centered Tutorial Planning with Concurrent Markov Decision Processes. Ph.D. diss., Dept. of Computer Science, North Carolina State University (2013).
9. Rowe, J., Mott, B., Lester, J.: Optimizing Player Experience in Interactive Narrative Planning: A Modular Reinforcement Learning Approach. In: Proceedings of the 10th Artificial Intelligence and Interactive Digital Entertainment Conference, pp. 160-166. Raleigh, NC. (2014)
10. Rowe, J., Lester, J.: Improving Student Problem Solving in Narrative-Centered Learning Environments: A Modular Framework. In: Proceedings of the 17th International Conference on Artificial Intelligence in Education, pp. 419-428. Madrid, Spain, (2015).
11. Sutton, R., Barto, A.: Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA (1998).
12. Wiering, M., Withagen, M., Drugan, M.: Model-based Multi-Objective Reinforcement Learning. IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning. pp. 1-6 (2014).
13. Witmer, B., Singer, M:. Measuring Presence in Virtual Environments: A Presence Questionnaire. Presence: Teleoperators and virtual environments 7(3), 225-240 (1998).
14. Clark, D. B., Tanner-Smith, E., Killingsworth, S.: Digital games, design, and learning: A systematic review and meta-analysis. Review of Educational Research 86(1), 79-122 (2015).
15. Wouters, P., van Nimwegen, C., van Oostendorp, H., van der Spek, E.D.: A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology 105, 249-265 (2013).
16. Mandel, T., Liu, Y., Levine, S., Brunskill, E., Popovic, Z., Offline policy evaluation across representations with applications to educational games. In: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems, pp. 1077-1084. Richland, SC (2014).
17. Castelletti, A., Pianosi, F., Restelli, M.: A Multiobjective Reinforcement Learning Approach to Water Resources Systems Operation: Pareto Frontier Approximation in a single run. Water Resources Research 49(6), 3476-3486 (2013).
18. Houli, D., Zhiheng, L., Yi, Z.: Multiobjective reinforcement learning for traffic signal control using vehicular ad hoc network. EURASIP Journal on Advances in Signal Processing 1, (2010).