Decomposing Drama Management in Educational Interactive Narrative: A Modular Reinforcement Learning Approach

Pengcheng Wang, Jonathan Rowe, Bradford Mott, James Lester

North Carolina State University, Raleigh, NC 27695
{pwang8, jprowe, bwmott, lester}@ncsu.edu

Abstract. Recent years have seen growing interest in data-driven approaches to personalized interactive narrative generation and drama management. Reinforcement learning (RL) shows particular promise for training policies to dynamically shape interactive narratives based on corpora of player-interaction data. An important open question is how to design reinforcement learning-based drama managers in order to make effective use of player interaction data, which is often expensive to gather and sparse relative to the vast state and action spaces required by drama management. We investigate an offline optimization framework for training modular reinforcement learning-based drama managers in an educational interactive narrative, CRYSTAL ISLAND. We leverage *importance sampling* to evaluate drama manager policies derived from different decompositional representations of the interactive narrative. Empirical results show significant improvements in drama manager quality from adopting an optimized modular RL decomposition compared to competing representations.

Keywords: Intelligent Narrative Technologies • Drama Management • Modular Reinforcement Learning • Educational Interactive Narrative

1 Introduction

Data-driven techniques for interactive narrative generation hold considerable promise for creating story experiences that are both rich and personalized. Training computational models from interactive storytelling data offers the potential to enhance narrative systems' capacity to personalize stories to individual players [1], endow machines with narrative intelligence with reduced programming effort [2], and facilitate interactive story content creation [3]. Recent years have seen growing interest in data-driven methods for drama management. For example, Yu and Riedl [4] employed prefix-based collaborative filtering to personalize interactive stories based on recurring player self-reports. Crowdsourcing-based methods task human users with generating narrative scripts—by writing stories or playing brief interactive narratives—in order to train computational models of interactive narrative generation [3,5]. Supervised machine learning techniques have been employed to induce interactive narrative planners from data gathered in Wizard of Oz studies [2]. These methods

leverage large datasets of interactive narrative log data to train models that complement manually authored approaches to runtime decision-making in interactive storytelling.

Reinforcement learning (RL) methods are the subject of growing interest within the intelligent narrative technologies community because they align naturally with key characteristics of interactive narrative and drama management [6,7,8,9]. RL focuses on training software agents to perform sequential decision-making in uncertain environments with delayed rewards [10]. Despite growing interest in RL, only a small number of RL-based drama managers have been deployed with playable interactive narratives, particularly those involving non-linear storylines and open virtual worlds. These environments yield potentially vast state and action spaces for reinforcement learning and demand large volumes of training data. Because this can be problematic for training models directly from human interaction data, much of the work on RL-based drama management has relied on synthetic training data from simulations [6] or pre-defined plot trajectory distributions [7]. This raises an important open question: how can we formalize drama management to be amenable to RL techniques trained with human interaction data from playable interactive narratives?

In this paper, we investigate an offline optimization framework for modular reinforcement learning-based drama management, with a specific focus on identifying an optimal decomposition of the task in terms of *adaptable event sequences*. We investigate this framework in an educational interactive narrative, CRYSTAL ISLAND, which features a science mystery about a spreading epidemic on a remote island research station. With a corpus of player interaction data from 402 middle school students, we evaluate potential competing structures for decomposing CRYSTAL ISLAND's drama management task by leveraging importance sampling (IS) [11], a statistical off-policy evaluation method. Empirical results suggest that the modular structure of an RL-based drama manager can have a significant influence on the drama manager's effectiveness, and an optimal structure based on *adaptable event sequences* can be identified using the proposed offline framework.

2 Related Work

Two families of technical approaches to the design of interactive narrative systems have been investigated: character-based story generation and plot-based story generation [12]. *Character-based* systems feature plots that emerge through interactions between believable, autonomous characters [13,14]. *Plot-based* approaches typically implement a director agent or drama manager to create, monitor, and adjust narrative event sequences and produce coherent plots [15,16]. In this work, we focus primarily on plot-based approaches, which to date have used a broad range of computational methods, including adversarial search [17], planning [15,16], and machine learning-based techniques [2,4,7].

RL-based drama managers typically model decisions about interactive narrative in terms of Markov decision processes (MDPs). By modeling sequences of events stochastically, MDPs account for the inherent uncertainty in predicting human player

actions. Further, MDPs support encoding *narrative quality* in terms of reward functions, which guide the process of training a drama manager by optimizing narratives' measured "goodness." Nelson et al. utilized temporal-difference methods to train a drama manager for a text-based interactive fiction, Anchorhead, using simulated user data and an author-specified story evaluation function [6]. This approach was extended by Roberts et al. [7], which applied target-trajectory distribution MDPs (TTD-MDPs) to generate interactive narratives according to an author-specified target distribution over possible stories. Thue and Bulitko [8] used MDPs to model player behavior in a player-adaptive interactive narrative by dynamically augmenting story events based on estimates of player gameplay preferences. We build on this foundational work on RL-based drama management to undertake the first systematic investigation of alternate decompositional representations of a drama management task, as well as by training a drama manager exclusively from player interaction data.

Decomposition-based approaches to RL, such as modular reinforcement learning, have long been a subject of interest in the machine learning community [18,19]. Rowe et al. presented the first example of a modular RL framework for interactive narrative adaptation by introducing the concept of *adaptable event sequences* [1,9], a decomposition-focused abstraction for recurring story units within interactive narratives. However, Rowe et al. investigated only one type of decompositional representation based on author-specified adaptable event sequences [9]. We extend this work by systematically exploring alternative decomposition structures for modular RL-based drama management. In addition, we utilize off-line evaluation methods to assess drama management policies induced from player interaction data, enabling policy evaluation without the requirement to collect additional data from new groups of human players [1].

3 Modular RL-Based Drama Management in CRYSTAL ISLAND

To investigate a data-driven optimization framework for modular RL-based drama management, we utilize CRYSTAL ISLAND, an educational interactive narrative that features a science mystery about an infectious outbreak on a remote island research station (Figure 1). The player adopts the role of a medical detective who must determine the source and treatment of the outbreak. The player investigates the illness by conversing with non-player characters, collecting data in a virtual laboratory, reading virtual books and articles, and completing a diagnosis worksheet. The drama manager monitors the player's behavior within the story world and makes recurring decisions about how *adaptable event sequences* (AESs) should unfold during the narrative. As defined in [9], an AES is a recurring series of one or more story events that, once triggered, can unfold in several different ways, leading to potentially different plot trajectories and player experiences. An AES can occur multiple times over the course of an interactive narrative, each time unfolding in a potentially distinct manner.

In this work, we focus on four AESs in CRYSTAL ISLAND[1]:

[1] There are 13 AESs in CRYSTAL ISLAND. In this work, we focus on four AESs, which were chosen because they were the most commonly occurring in our training corpus.

Fig 1. CRYSTAL ISLAND interactive narrative.

- **Teresa Symptoms**. This AES is triggered each time the player initiates a conversation with Teresa, a sick scientist in the camp infirmary. If a player inquires about Teresa's symptoms, the drama manager selects one of three possible conversational responses: providing minimal detail about her symptoms, providing moderate detail about her symptoms, or providing maximal detail about her symptoms.
- **Record Findings Reminder**. Whenever the player uncovers useful information that is relevant to the mystery's solution, such as the result of conducting a laboratory test or information contained in an important (virtual) book, the drama manager determines whether to deliver a hint suggesting the player take in-game notes about the information.
- **Diagnosis Feedback**. When seeking to solve the mystery, if a player submits an incorrect diagnosis to the camp nurse, the drama manager selects one of three options for feedback: minimally detailed feedback, moderately detailed feedback, or maximally detailed feedback.
- **Knowledge Quiz**. When players converse with certain characters, the drama manager optionally delivers an embedded assessment (i.e., quiz) about related microbiology concepts. Each time this occurs, the drama manager decides whether to present the quiz or not to present it.

In order to model AESs computationally, and devise drama management policies using reinforcement learning, we use Markov decision processes (MDPs). An MDP is defined as a quintuple $<S, A, p, r, \gamma>$, in which S is a set of states, A is a set of actions, p is a probabilistic transition function with $p^a_{s,s'}$, representing the probability of taking action a in state s and transitioning to state s', r is a reward function following the form of $r^a_{s,s'}: S \times A \times S \rightarrow \mathbb{R}$, and $\gamma \in (0,1]$ denotes the discount factor, trading off the importance of long-term rewards versus short-term rewards. In CRYSTAL ISLAND, p and r do not have explicit forms due to the inherent uncertainty of the environment, but we can estimate their values from a corpus of player interactions. The solution to an RL problem is a policy π, which generally takes the form of a conditional probability mass function $Pr_\pi(a|s)$, representing the probability of taking action a in state s.

Policies for MDPs can be obtained using reinforcement learning techniques, of which we focus on two broad families: model-based RL techniques and model-free RL techniques [10]. Specifically, we utilize policy iteration and Q-learning to induce drama manager policies under different modular RL representational structures.

To represent AESs in CRYSTAL ISLAND, we utilize five binary features to encode the state representation of each MDP (Table 1). We limit the state representation to these features to reduce potential data sparsity issues. It should be noted that this form of compact state representation is not uncommon in reinforcement learning-based intelligent user interfaces [20]. The first two state features encode key elements of plot state. The third state feature encodes player trait information about prior science knowledge. The final two state features denote AES indices, indicating which subset of AES related actions are available to drama manager when a narrative adaptation is triggered.

Table 1. MDP state features for CRYSTAL ISLAND RL-based drama manager.

State Feature	Bits	Description
Submit Solution	1	Player has tried to submit solution?
Solved Mystery	1	Mystery solved correctly?
Pre-test Score	1	Player's pre-test score above median?
AES	2	AES index

The action sets for the MDPs represent the drama manager's possible actions, e.g., how much information an NPC reveals to the player, or whether a player receives a hint or quiz after an important event. The action set for each MDP is comprised of the possible event sequences in its corresponding AES, which are described above.

The four AESs shared the same reward function. The reward function was derived from a measure of player knowledge acquisition, normalized learning gain, because of the educational focus of CRYSTAL ISLAND. Data on normalized learning gains were obtained by administering a pre-test and post-test to each player about relevant microbiology concepts and calculating the normalized difference between the two test scores. When generating training episodes for reinforcement learning, rewards were assigned when the terminal state of the game was reached, or at the conclusion of gameplay. The reward value was either 100 when the player's normalized learning gain was above median value, or -100 if it was below median.

To induce drama manager policies for the MDPs, we utilized player interaction data from a pair of human subject studies conducted with CRYSTAL ISLAND [9]. All participants utilized the same version of CRYSTAL ISLAND, which was deployed in two public middle schools involving 300 students and 153 students, respectively. Participants played the game until they solved the mystery, or 55 minutes elapsed, whichever occurred first. While using CRYSTAL ISLAND, participants unknowingly encountered AESs several times. At each AES, the drama manager selected a narrative adaptation according to a random policy, uniformly sampling the planning space. By logging these narrative adaptations, as well as participants' subsequent responses, the environment broadly sampled the space of policies for controlling adaptable event sequences. In addition, several questionnaires were administered prior to, and imme-

diately after, participants' interactions with CRYSTAL ISLAND. The questionnaires provided data about participants' individual characteristics, curricular knowledge, and engagement with the environment.

The data collected from both studies were combined into a single corpus. After removing data from participants with incomplete records, there were 402 participants remaining in the data set. Out of the total 402 trials, we observed the relevant AESs with the following frequency: the Teresa Symptoms AES occurred 559 times, the Record Findings Reminder AES occurred 3,435 times, the Diagnosis Feedback AES occurred 804 times, and the Knowledge Quiz AES occurred 1,073 times. Each trial, on average, contained 14.6 occurrences of AESs.

4 Decomposing Drama Management for Modular Reinforcement Learning

We utilize AESs to represent modular units of interactive narrative that can be shaped by a modular RL-based drama manager in CRYSTAL ISLAND. However, it is difficult to anticipate the optimal grain-size for representing events in drama management. A hand-authored AES-based representation is likely to produce less effective policies than an optimized encoding, which may be more fine- or coarse-grained. Prior work on RL-based drama management has typically utilized a coarse-grained representation: a single monolithic MDP, which encodes all possible state features and actions for a drama manager [6,7]. An alternate approach is a modular representation, which clusters subsets of narrative events into AESs, each individually modeled as a separate MDP. This approach emphasizes compact RL sub-tasks, which are more readily solved by training policies with datasets of limited size, as one might expect to obtain from logs of human player interactions. However, determining how to best cluster events into AESs is challenging to do manually. We investigate a data-driven framework for evaluating alternate decompositions.

In CRYSTAL ISLAND, one could model all four AESs in terms of a single MDP. In other words, the Teresa Symptoms AES (denoted as "T"), Record Findings Reminder AES (denoted as "R"), Diagnosis Feedback AES (denoted as "D"), and Knowledge Quiz AES (denoted as "Q") are modeled together (denoted as "TRDQ"), and thus solved as a single RL problem. An alternative representation could involve decomposing the task in terms of two modular sub-tasks (Figure 2), encoding Teresa Symptoms (T) events in one MDP, and the other three types of AESs (denoted as "RDQ") in a separate MDP. In this case, an overall drama management policy would consist of the combined policies from the two modules (denoted as "T_RDQ"), as well as an arbitration procedure for resolving inter-policy conflicts.

When the number of AESs is greater than two, there may be multiple modular representations that can be considered. For example, in CRYSTAL ISLAND, the monolithic model TRDQ can be decomposed into the form T_RDQ by isolating the Teresa Symptoms AES (T) from the other AESs, or it can be decomposed into the form TR_DQ, in which events from Teresa Symptoms (T) and Record Findings Reminder (R) are encoded in one MDP and the other two AESs are modeled by a second MDP.

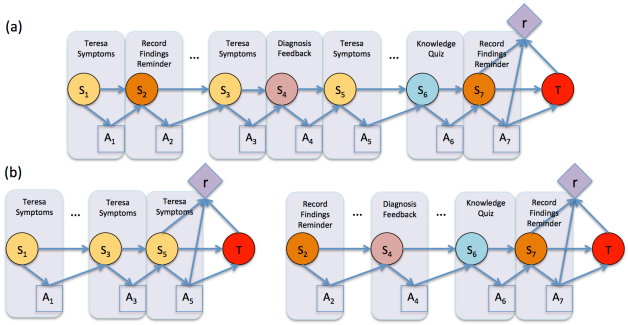

Fig. 2. Illustration of two decompositional representations for a modular RL-based drama manager. AESs are color-coded and denoted with boxes. Events progress in temporal order from left to right. *S* and *A* represent states and actions, *T* represents a terminal state, and *r* denotes a reward. (a) A monolithic structure in which all events and actions are encoded as a single MDP. (b) A modular structure in which one AES is modeled as an MDP (left), and the other three AESs are modeled as a different MDP (right).

The number of possible modular structures is determined by the number of possible combinations of AESs. Because each sub-task's MDP can potentially exploit a smaller state space and action set, this approach addresses the curse of dimensionality inherent in many reinforcement learning tasks and contributes to improved training speed. Further, alternate decompositional representations change the MDP models' transition dynamics by altering the task environment's transition probability distributions. Thus, although both monolithic and decomposed models can yield locally "optimal" solution policies, the policies produced by each model are likely to be different.

5 Offline Policy Evaluation

We consider two primary factors in selecting an offline policy evaluation technique. First, we seek a method that evaluates MDP policies based on sample episodes rather than an explicit environment model. This excludes evaluation metrics such as expected cumulative reward [20]. Second, we seek an evaluation method that accounts for generalizability to unseen situations. Specifically, we utilize cross validation to evaluate policies with corpus data not utilized in model training. To address these two factors, we employ importance sampling [21].

Importance sampling is a statistical evaluation technique that can be used to evaluate a policy π when it is infeasible to draw samples under π [11]. In CRYSTAL ISLAND, trials were collected with a uniformly random policy π'. In order to assess policy π, the following equation can be utilized:

$$v^{IS}_{\pi',h}(\pi) = \frac{1}{N}\sum_{i} R(h_i)\frac{\Pr(h_i \mid \pi)}{\Pr(h_i \mid \pi')} \tag{1}$$

In Equation 1, h is a set of trials, in which each sampled trial is labeled as h_i. $R(h_i)$ is the sum of discounted rewards across the trial h_i:

$$R(h_i) = \sum_{t=1}^{T} \gamma^{t-1} \cdot r_{t,h_i} \tag{2}$$

The ratio of likelihoods of observing the trial h_i following evaluating policy π and sampling policy π' could be simplified as the ratio of likelihoods of making a series of action choices under given policies, as is demonstrated in Equation 3.

$$\frac{\Pr(h_i \mid \theta)}{\Pr(h_i \mid \theta')} = \frac{\prod_{t=1}^{T} \pi_{\theta}(a_t \mid s_t)}{\prod_{t=1}^{T} \pi_{\theta'}(a_t \mid s_t)} \tag{3}$$

Although importance sampling yields an unbiased estimate of policy value, it can introduce a large variance when samples are scarce. A biased estimate with lower variance can be obtained with weighted importance sampling (WIS), as shown in Equation 4. In this work, we apply both IS and WIS metrics to compare policies from different modular structures.

$$v_{\pi',h}^{WIS}(\pi) = \frac{1}{\sum_i \frac{\Pr(h_i \mid \pi)}{\Pr(h_i \mid \pi')}} \sum_i R(h_i) \frac{\Pr(h_i \mid \pi)}{\Pr(h_i \mid \pi')} \tag{4}$$

6 Results

We empirically evaluate drama management policies generated by model-based and model-free RL techniques under several possible decompositional representations. In this work, there are 15 possible decompositions, each comprised of a different combination of the four AESs, ranging from a monolithic model (TRDQ), to a completely decomposed representation (denoted as "T_R_D_Q"). We also compare to a uniform random policy, a baseline model that was utilized to collect the initial RL training data. It should be noted that in CRYSTAL ISLAND, any interactive narrative generated under a random AES policy is guaranteed to still be coherent because the AESs have been designed to never introduce events that might threaten future events in the narrative. Random policy-generated narratives may deviate substantially from highly personalized narratives, but they will still be playable and sensible.

Importance sampling and weighted importance sampling were utilized to evaluate policies generated using policy iteration and Q-learning under five-fold cross validation across 20 runs. During each run, all trials were randomly re-assigned across the 5 folds. We utilized approximate randomization to assess the statistical significance of differences in measured quality between different drama manager policies [22]. Approximate randomization is a statistical test that has been broadly used in the natural language processing community and does not assume that data points from two groups are independently sampled. In our experiment, we set the random shuffling parameter in approximate randomization to 5,000.

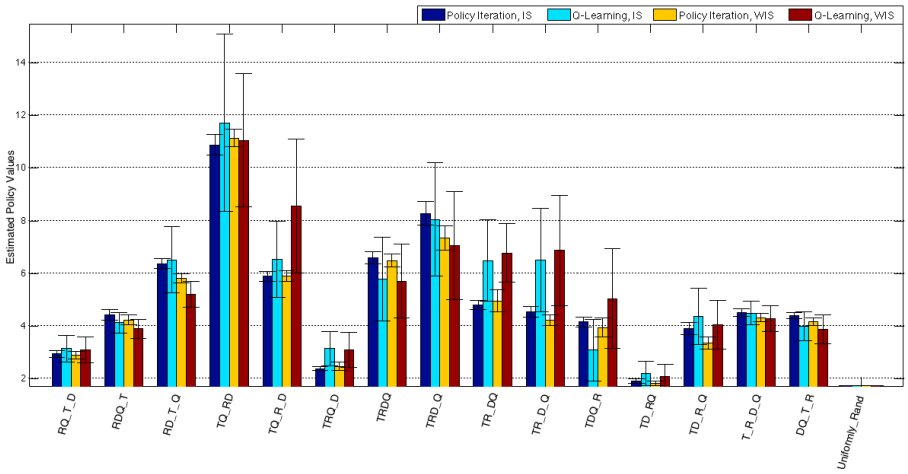

Fig. 3. Importance sampling and weighted importance sampling based policy values with 95% confidence intervals for a discount factor of 1.0.

Our first analysis utilized a discount factor of 1, which treats long-term rewards as equivalent to short-term rewards (Figure 3). These experiments revealed that an RL-based drama manager's decompositional representation has a significant impact on policy quality. The highest-performing model was TQ_RD, which generated policies that were significantly better than most of the other modular structures, including the monolithic model TRDQ and the fine-grained decomposed model T_R_D_Q, $p<0.005$. The TQ_RD decomposition groups the Teresa Symptoms AES (T) and Knowledge Quiz AES (Q) together, and the Record Findings Reminder (R) and Diagnosis Feedback (D) AESs in a separate MDP. In addition, we ran a second set of experiments with a discount factor of 0.9 and obtained similar results (Figure 4).

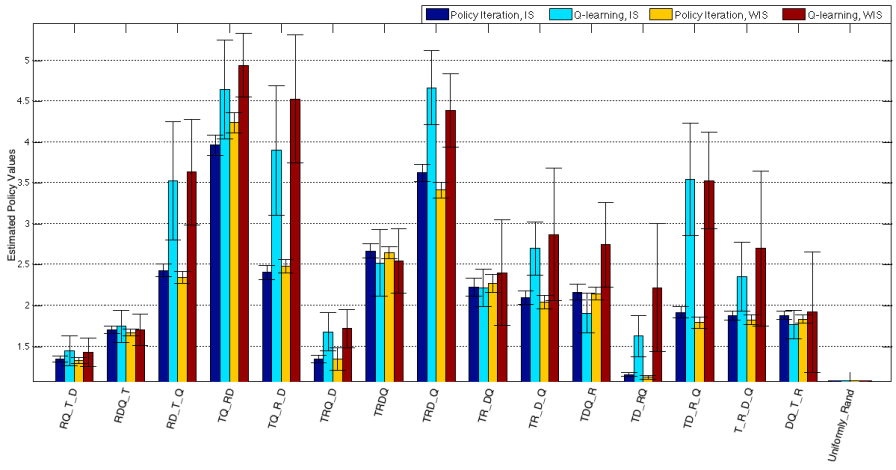

Fig. 4. Importance sampling and weighted importance sampling based policy value estimates with 95% confidence intervals for a discount factor of 0.9.

In Tables 2 and 3, we list the average policy values calculated using importance sampling and weighted importance sampling for the optimized modular RL structure TQ_RD, the monolithic structure TRDQ, and the fully decomposed structure T_R_D_Q, trained by Q-learning. We also add results from a uniform random policy as a baseline. When utilizing a discount factor of 1.0, the policy value for the optimized model yielded a nearly six-fold improvement over the uniform random policy. The optimized model also yielded a policy value that is more than double the monolithic and fully decomposed models. The same trend is observed for weighted importance sampling, as well as when we set the discount factor to 0.9 (Table 3). Notably, for most RL modular structures, we observe no significant differences in policy quality due to the choice of policy iteration or Q-learning as the learning algorithm. When the discount factor is 0.9, several modular structures (e.g., TD_R_Q) yielded higher values under Q-learning. A possible explanation for this is that we utilize an early stopping criterion in Q-learning, which may help reduce overfitting.

Table 2. Average policy values from 20 runs of 5-fold cross validation using different modular RL structures with a discount factor of 1.0. All policies are trained using Q-learning. Policies trained with policy iteration yield similar results.

Modular Structure	Policy Value Based on IS	Policy Value Based on WIS
Optimal Structure (TQ_RD)	**11.73**	**11.07**
Monolithic (TRDQ)	5.78	5.70
Fully Decomposed (T_R_D_Q)	4.49	4.28
Uniformly Rand	1.73	1.73

Note. N=20, $p<.005$

Overall, results suggest that the structure of the decompositional representation in modular RL-based drama management has a significant effect on generated policy quality. However, it is challenging to determine why the best-performing modular structure, TQ_RD, yielded such substantially greater policy values than other competing representations. One possible explanation is that the optimal representation may have benefitted from grouping diagnosis worksheet-related AESs separately from other AESs. However, further analysis is necessary to confirm or refute this explanation. It should be noted that policy value is a reflection of the policy's quality with respect to its associated reward function and RL task environment. In this work, we have utilized a reward function that is based upon normalized learning gains, due to the educational focus of CRYSTAL ISLAND, but any narrative evaluation metric, if properly quantified, can be utilized in the framework (e.g., user engagement, sense of narrative transportation). Because it is difficult to analytically discern how quantitative differences in policy values correspond to differences in user experience or overall narrative quality, a promising direction for future work is to integrate these policies back into a run-time version of CRYSTAL ISLAND and test their effects on actual human players.

Table 3. Average policy values from 20 runs of 5-fold cross validation using different modular RL structures with a discount factor of 0.9. All policies are trained using Q-learning. Policies trained with policy iteration yield similar comparison results.

Modular Structure	Policy Value Based on IS	Policy Value Based on WIS
Optimal Structure (TQ_RD)	**4.65**	**4.94**
Monolithic (TRDQ)	2.52	2.55
Fully Decomposed (T_R_D_Q)	2.36	2.70
Uniformly Rand	1.08	1.08

Note. N=20, $p<.005$

7 Conclusions and Future Work

We have presented a data-driven framework for evaluating alternate decompositional representations of modular RL-based drama management in the educational interactive narrative CRYSTAL ISLAND. Devising an optimal computational representation for RL-based drama management has a significant impact on the quality of policies induced from a training corpus. Decompositional representations help address data sparsity challenges presented by efforts to train drama managers from human player data. Using an offline evaluation framework based on importance sampling, we find that an optimized decompositional representation for RL-based drama management yields superior policies to traditional monolithic or fully decomposed representations by a significant margin.

There are several promising directions for future work. It will be important to investigate automated procedures for devising high-quality decompositional representations for drama management, thereby automating the process of identifying optimal breakdowns of events into AESs. In addition, it will be important to explore the scalability and generalizability of the modular RL-based interactive narrative generation framework and to understand its applicability to other narrative domains and genres.

References

1. Rowe, J., Lester, J.: Improving Student Problem Solving in Narrative-Centered Learning Environments: A Modular Reinforcement Learning Framework. In: 17th International Conference on Artificial Intelligence in Education, pp. 419–428. (2015)
2. 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, 203–215. (2014)
3. Li, B., Lee-Urban, S., Johnston, G., Riedl, M.: Story Generation with Crowdsourced Plot Graphs. In: 27th AAAI Conference on Artificial Intelligence, pp. 598–604. (2013)
4. Yu, H., Riedl, M.: Personalized Interactive Narratives via Sequential Recommendation of Plot Points. IEEE Transactions on Computational Intelligence and AI in Games, 6(2), 174–187. (2014)

5. Orkin, J., Roy, D.: Understanding speech in interactive narratives with crowd sourced data. In: 8th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, pp. 57–62. (2012)
6. Nelson, M., Roberts, D., Isbell Jr, C., Mateas, M.: Reinforcement learning for declarative optimization-based drama management. In: 5th International Conference on Autonomous Agents and Multi-Agent Systems, pp. 775–782. (2006)
7. Roberts, D., Nelson, M., Isbell, C., Mateas, M., Littman, M.: Targeting specific distributions of trajectories in MDPs. In: 21st AAAI Conference on Artificial Intelligence, pp. 1213–1218. (2006)
8. Thue, D., Bulitko, V.: Procedural Game Adaptation: Framing Experience Management as Changing an MDP. In: Fifth Workshop on Intelligent Narrative Technologies, pp. 44–50. (2012)
9. Rowe, J., Mott, B., Lester, J.: Optimizing Player Experience in Interactive Narrative Planning: A Modular Reinforcement Learning Approach. In: 10th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, pp. 160–166. AAAI Press, Menlo Park, CA (2014)
10. Sutton, R., Barto, A.: Reinforcement Learning: An Introduction. MIT Press. (1998)
11. Peshkin, L., Shelton, C.: Learning from Scarce Experience. In: Proceedings of the Nineteenth International Conference on Machine Learning, pp. 498–505. (2002)
12. Riedl, M., Bulitko, V.: Interactive Narrative: An Intelligent Systems Approach. AI Magazine. 34(1), 67–77. (2012)
13. McCoy, J. Treanor, M., Samuel, B., Mateas, M., Wardrip-Fruin, N.: Prom Week: Social Physics as Gameplay. In: 6th International Conference on Foundations of Digital Games, pp. 319–321. (2011)
14. Suttie, N., Louchart, S., Aylett, R., Lim, T. Theoretical Considerations Towards Authoring Emergent Narrative. In: 6th International Conference on Interactive Digital Storytelling, pp. 205–216. (2013)
15. Porteous, J., Lindsay, A., Read, J., Truran, M., Cavazza. M.: Automated Extension of Narrative Planning Domains with Antonymic Operators. In: 14th International Conference on Autonomous Agents and Multiagent Systems, pp. 1547–1555. (2015)
16. Robertson, J., Young, R.: Automated Gameplay Generation from Declarative World Representations. In: 11th AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, pp. 72–78. (2015)
17. Lamstein, A., Mateas, M.: Search-based drama management. In: 2004 AAAI Workshop on Challenges in Game Artificial Intelligence, pp. 103–107. (2004)
18. Singh, S.: Transfer of Learning by Composing Solutions of Elemental Sequential Tasks. Machine Learning. 8, 323–339. (1992)
19. Dietterich, T.: Hierarchical reinforcement learning with the MAXQ value function decomposition. Journal of Artificial Intelligence Research. 13, pp. 227–303. (2000)
20. 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, 137–180. (2011)
21. Mandel, T., Liu, Y., Levine, S., Brunskill, E., Popovic, Z.: Offline Policy Evaluation Across Representations with Applications to Educational Games. In: 13th International Conference on Autonomous Agents and Multi-Agent Systems, pp. 1077–1084. (2014)
22. Riezler, S., Maxwell, J.: On Some Pitfalls in Automatic Evaluation And Significance Testing for MT. In: Proceedings of the ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization. 57–64. (2005)