Optimizing Tutorial Planning in Educational Games: A Modular Reinforcement Learning Approach

Pengcheng Wang PWANG8@NCSU.EDU
Jonathan Rowe JPROWE@NCSU.EDU
Bradford Mott BWMOTT@NCSU.EDU
James Lester LESTER@NCSU.EDU
North Carolina State University, Raleigh, NC 27695 USA

Abstract

Recent years have seen a growing interest in educational games, which integrate the engaging features of digital games with the personalized learning functionalities of intelligent tutoring systems. A key challenge in creating educational games, particularly those supported with interactive narrative, is devising narrative-centered tutorial planners, which dynamically adapt gameplay events to individual students to enhance learning. Reinforcement learning (RL) techniques show considerable promise for creating tutorial planners from corpora of student log data. In this paper, we investigate a modular reinforcement learning framework for tutorial planning in narrative-centered educational games, with a focus on exploring multiple modular structures for modeling tutorial planning sub-tasks. We utilize offline policy evaluation methods to investigate the quality of alternate tutorial planner representations for a narrative-centered educational game for middle school science, CRYSTAL ISLAND. Results show significant improvements in policy quality from adopting a data-driven optimized modular structure compared to a traditional monolithic MDP model structure, particularly when training data is limited.

1. Introduction

There is growing interest in narrative-centered educational games, which adaptively tailor students' learning experiences within engaging virtual environments. Reinforcement learning has shown considerable promise for inducing strategies that dynamically regulate learning processes in advanced learning environments, such as intelligent tutoring systems and digital games (Chi et al., 2011; Rowe et al., 2014; Mandel et al., 2014). However, educational games, particularly those with rich interactive narratives, present significant challenges for modeling tutorial planning as a reinforcement learning task. Specifically, we seek to embed tutorial actions in interactive storylines discreetly, making them highly context-sensitive; however, this can yield vast state and action spaces for reinforcement learning. An important open question is how we can identify optimal representation structures for RL-based tutorial planning that cope with the challenges posed by game-based learning. In this work, we investigate an offline optimization framework for modular reinforcement learning-based tutorial planning, with a specific focus on identifying the optimal modular structure for representing adaptable event sequences. We leverage *importance sampling* (IS) (Precup, 2000; Peshkin & Shelton, 2002) to evaluate narrative-centered tutorial planning policies offline in the context of a narrative-centered educational game for middle school science, CRYSTAL ISLAND.

2. Tutorial Planning in a Narrative-Centered Educational Game

CRYSTAL ISLAND is a narrative-centered educational game that features a science mystery about an outbreak afflicting a team of scientists on a remote island research outpost (Rowe et al., 2014). Students adopt the role of a medical detective who must investigate the outbreak and identify the cause and treatment of the disease. Four types of narrative events are organized as *adaptable event sequences* (Rowe et al., 2014), which connect narrative adjustments chronologically. These four types of events include adapting details of information revealed to the player about an NPC's symptoms, the appearance of a reminder to players prompting them to record their findings, the details of feedback given to players for submitting their diagnosis worksheet, and the appearance of an in-game knowledge quiz. Within each adaptation, a proper action fitting to a certain event type would be opted by the narrative-centered tutorial planner to interact with players.

We use Markov decision processes (MDPs) to model adaptable event sequences in CRYSTAL ISLAND and investigate both model-based (policy iteration) and model-free (Q-learning) RL techniques in generating tutorial planning policies.

Five binary features covering narrative features and player individual difference features are extracted into the RL state space. Players' game processes, such as whether a player has tried to submit a solution for the mystery, and players' knowledge of related subjects, such as her content knowledge scores, have been encoded into the state features. Action spaces in MDP models represent narrative-centered tutorial planner's options, e.g., how much information an NPC reveals to players when they ask the character about her symptoms. The reward function reflects a student's learning gain calculated from pre- and post-test scores. Because all tutorial adaptation of trials in a training set is assessed by students' normalized learning gain, a single reward is given to the tutorial planner only when the terminal state of the game is reached. This reward could be either 100 when student's normalized learning gain is above average, or -100 if it is below average. The corpus contains 402 students' playing records, with narratives in all games directed by a uniformly random tutorial adaptation policy. Each trial, on average, contains 14.60 adaptable events from the four event types.

3. Optimizing Tutorial Planner Modular Structure

The core research problem of this paper is investigating modular representation structure in determining policy's quality. A monolithic MDP model that describes adaptable event sequences of all types in one model could be decomposed into multiple using the modular MDP approach, with each modular model concentrating only on events from certain types of adaptable event sequences, as it is demonstrated in Figure 1.

When the number of event types is greater than two, there could be multiple modular structures to decompose the monolithic MDP model into. Because a modular MDP exploits a smaller state set and action set, it would help to avoid the curse of dimensionality, as well as improve training speed. At the same time, decomposition changes the MDP model's transition dynamics by altering transition probability distribution. Thus, although both monolithic MDP models and decomposed modular MDPs can generate optimal policies, optimal policies from different models are probably different. In cases in which the environment can not be accurately modeled, an empirical approach is called for to identify which modular representation fits the real environment best. In the CRYSTAL ISLAND environment, because the player triggers each adaptable event, there is no need to design an arbitration module in the modular MDP model.

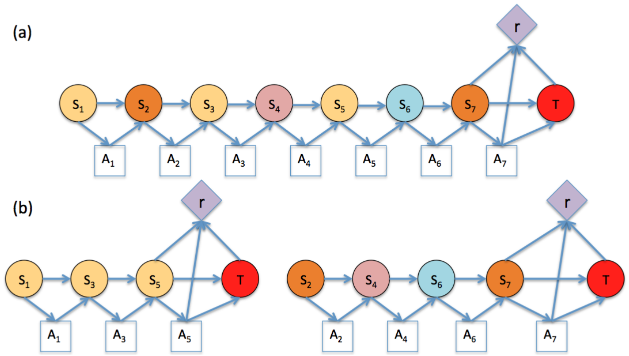

Figure 1. Tutorial planning MDP decomposition according to event types in narrative-centered tutorial adaptation: (a) shows a monolithic MDP structure, and (b) demonstrates one modular structure in which one event type is modeled in an MDP (left), and the other three types are in another MDP (right). Event types are color-coded. S and A represent state and action, with subscript indicating time step. T represents terminal state, and r denotes reward.

4. Offline Policy Evaluation

Two factors affect the choice of a proper offline policy evaluation method in our problem setting. First, because MDP decomposition changes the environment's transition dynamics, a policy evaluation method should be selected based on samples instead of methods explicitly requiring an environment model, e.g., *expected cumulative reward* (Chi et al., 2011; Tetreault & Litman, 2008). Second, to verify the generalizability of each model based on limited amount of samples, cross validation is needed. To address these two factors, we employ importance sampling (IS), which estimates unbiased policy values from samples. IS is an off-policy evaluation method, so evaluating trials do not have to be sampled by evaluated policy. Estimated policy value is calculated from trials, meaning it could fit into cross validation. A variance of IS—weighted IS (WIS)—is also used in our experiment for comparing policy's quality.

5. Empirical Results

To investigate the effectiveness of optimal reinforcement learning modular structure in deriving narrative-centered tutorial planner in educational games, a series of experiments have been conducted on the CRYSTAL ISLAND cor-

Table 1. Averaged policy values of 20 runs 5-fold cross validation using different modular structures when discount factor is 1.0. Numbers in parentheses are policy value improvement of optimal structural model over model in that row. All policies used here are trained with policy iteration. Policies trained with Q-learning show the same trend and very similar results.

Modular Structure	Policy Value Based on IS	Policy Value Based on WIS
Optimal Structure	1.8029	1.7982
Monolithic	1.7390(3.67%)	1.7375(3.49%)
Fully Decomposed	1.7875(0.86%)	1.7828(0.87%)
Uniformly Random	1.7274(4.38%)	1.7274(4.10%)

Note. N=20, p<.001

Table 2. Averaged policy values of 20 runs 5-fold cross validation using different modular structures when discount factor is 0.9. Numbers in parentheses are policy value improvement of optimal structural model over model in that row. All policies used here are trained with policy iteration. Policies trained with Q-learning show the same trend and very similar results.

Modular Structure	Policy Value Based on IS	Policy Value Based on WIS
Optimal Structure	1.1060	1.1050
Monolithic	1.0884(1.62%)	1.0880(1.56%)
Fully Decomposed	1.1018(0.38%)	1.1008(0.38%)
Uniformly Random	1.0816(2.26%)	1.0816(2.17%)

Note. N=20, p<.001

pus. Of the four types of events, we compare 15 different modular structures (including monolithic structure). A discount factor of 0.9 and 1 were used to test both situations where game playing time would be cared or not. Policy iteration and Q-learning were applied in solving the MDP problems. Five-fold cross validation was ran for 20 times, whose results were analyzed using approximate randomization (Riezler & Maxwell, 2005) to test their statistical significance.

According to evaluation from both IS and WIS, the optimal modular structure for our problem is one modeling the NPC's revelation and knowledge quiz prompting together, while modeling the other two types of events with the other MDP. We compare it with other commonly used structures, like monolithic and fully decomposed structures, using a single or four MDPs respectively.

From the results shown in Table 1 and 2, we can see in both situations when the discount factor is equal to or smaller than 1, the optimal modular structure is able to train a policy that is statistically significantly better than policies from either monolithic structure or fully decomposed structure, which were more commonly used. The optimal modular structure MDP generating policy also statistically significantly outperforms policies from most of the other structural models.

6. Conclusions

In this work, we investigated the effect of optimal modular structure in MDPs in solving tutorial adaptation problems in narrative-centered educational games. Empirical results from the offline evaluation method, importance sampling, demonstrate that we can optimize modular reinforcement learning structures to generate a tutorial planning policy that is significantly better than those derived from other commonly used structures.

References

Chi, Min, VanLehn, Kurt, Litman, Diane, and Jordan, Pamela. 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.

Mandel, Travis, Liu, Yun-En, Levine, Sergey, Brunskill, Emma, and Popovic, Zoran. 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. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

Peshkin, Leonid and Shelton, Christian R. Learning from Scarce Experience. In *Proceedings of the Nineteenth International Conference on Machine Learning*, pp. 498–505. Morgan Kaufmann Publishers Inc., 2002.

Precup, Doina. Eligibility traces for off-policy policy evaluation. *Computer Science Department Faculty Publication Series*, pp. 80, 2000.

Riezler, Stefan and Maxwell, John T. 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*, pp. 57–64, 2005.

Rowe, Jonathan, Mott, Bradford, and Lester, James. Optimizing player experience in interactive narrative planning: a modular reinforcement learning approach. In *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2014.

Tetreault, Joel R and Litman, Diane J. A reinforcement learning approach to evaluating state representations in spoken dialogue systems. *Speech Communication*, 50 (8):683–696, 2008.