ABSTRACT

JUNG, GYUHUN. Exploring Batch Deep Reinforcement Leaning and Multi-Task Learning across Intelligent Tutoring Systems: Lessons Learned. (Under the direction of Dr. Min Chi).

Nowadays, Intelligent Tutoring Systems(ITSs) have been vastly utilized as a means of adaptive education and training for many learners and Reinforcement Learning (RL) has been widely used to decide an appropriate teaching strategy in ITS. RL is a promising method for the decision-making process in online games, robotics, the military, and so on. Among the variations of the RL algorithm, Deep Reinforcement Learning (DRL) has shown remarkable performance by combining deep neural networks with reinforcement learning. In this work, we focused on two issues for improving students' performance by deploying DRL-induced policies across two ITSs. First, we explored **1) how effective the DRL-induced policy is** in a probability ITS. and 2) the potency of DRL with Multi-Task Learning (MTL) in a logic tutor. Here MTL is about training several similar tasks simultaneously to enhance the model performance and the generalization ability for each task.

For the first issue, we induced a Batch DRL policy based on different input settings . The policy is trained by a stable DRL algorithm, Double DQN combined with Long Short Term Memory (LSTM). Our empirical results show that the offline DRL policy significantly improves students' learning performance compared to the Expert hierarchical RL policy. For the second issue, we used the pre-developed model-based DRL framework, Dreamer. Also, the bisimulation metrics, the distance measuring how much two states are behaviorally similar, are combined with this framework for making a generalized policy for performing two different tasks in ITSs. Following the empirical result, although we did not get higher student learning performance than the control group, we analyzed the reasons for this and the future works for overcoming the limitations of MTL in real environments.

Exploring Batch Deep Reinforcement Leaning and Multi-Task Learning across Intelligent Tutoring Systems: Lessons Learned

by
GYUHUN JUNG

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Computer Science

Raleigh, North Carolina
2023

APPROVED BY:

Dr. Matsuda Noboru	Dr. James Lester

Dr. Min Chi
Chair of Advisory Committee

DEDICATION

I dedicate this thesis to the Republic of Korea Army for allowing me to study abroad and to my family for their support, especially my wife, Shinwon Kim, and my daughter, Yiseo Jung.

BIOGRAPHY

Gyuhun Jung was born in Seoul, South Korea on October 28th, 1991. He graduated from Korea Military Academy in 2015, where he majored in physics. After graduating, he served as an officer in the Republic of Korea Army, where he was stationed in multiple infantry divisions and the Joint Security Area. In the Fall of 2021, he began pursuing his Master's degree in Computer Science in the Department of Engineering at North Carolina State University, located in Raleigh, NC. Soon after, he joined Dr. Min Chi's lab to further his studies.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Dr. Min Chi for her invaluable mentorship and dedication in helping me complete my thesis. Her insights, constructive feedback, and guidance have been instrumental in shaping my academic and personal growth, and have helped me overcome many obstacles and achieve my goals.

TABLE OF CONTENTS

List of Tables . . . . **vii**

List of Figures . . . . **viii**

Chapter 1 Introduction **1**
1.1 DRL for Pyrenees 3
1.2 Multi-Task DRL for Deep Thought 3

Chapter 2 Background **5**
2.1 Reinforcement Learning 6
2.1.1 Model-based and Model-free RL 7
2.1.2 Online and Offline RL 8
2.1.3 Deep Reinforcement Learning 8
2.2 Multi-Task Learning 9
2.3 Pedagogical policy induction in Intelligent Tutoring System (ITS) 10
2.3.1 Reinforcement learning 11
2.3.2 Multi-task learning 12

I Deep Reinforcement Learning (DRL) for Pyrenees 13

Chapter 3 Deep Reinforcement Learning (DRL) **14**
3.1 Markov Decision Processes(MDP) 14
3.2 Q-value based DRL 16

Chapter 4 Probability tutor: Pyrenees **19**
4.1 Tutor Interface 19
4.2 Procedure 20

Chapter 5 DRL policy induction **23**
5.1 Model Architecture 23
5.1.1 Fully Connected and LSTM Layer 24
5.2 Model Training 27
5.2.1 Training Corpus 27
5.3 Experiment 1: Pyrenees Fall 2022 31
5.3.1 Experiment Setup 31
5.3.2 Experiment Results 31
5.4 Conclusion & Discussion 33

II Multi Task RL (MTRL) for Deep Thought 35

Chapter 6 Related Work . . . 36
6.1 Multi-Task Reinforcement Learning . . . 36
6.2 Latent representation learning . . . 38

Chapter 7 Logic tutor: Deep Thought . . . 40
7.1 Tutor Interface . . . 40
7.2 Procedure . . . 42

Chapter 8 MTRL policy induction . . . 45
8.1 Model architecture . . . 45
8.1.1 Latent representation learning . . . 46
8.1.2 Shared latent representation learning . . . 47
8.2 Model Training . . . 51
8.2.1 Training Corpus . . . 51
8.3 Experiment 2: Deep Thought Spring 2023 . . . 52
8.3.1 Experiment Setup . . . 52
8.3.2 Experiment Results . . . 53
8.4 Conclusion & Discussion . . . 53

Chapter 9 Conclusions and Future work . . . 56

References . . . 58

APPENDIX . . . 64
Appendix A RSSM architecture . . . 65

LIST OF TABLES

Table 4.1 Problem-solving process example in Pyrenees 21

Table 5.1 PDIS values for $k = 3$ and 4. Electric blue color cells indicate the input settings with a bigger PDIS value. 29

Table 5.2 F22 Pyrenees experiment result. 32

Table 8.1 PDIS values for multi-task reinforcement learning model and DQN-based models. Electric blue color cells indicate the input settings with a bigger PDIS value. 52

Table 8.2 F23 Deep Thought experiment result. 53

LIST OF FIGURES

Figure 2.1 Learning procedure of online and offline(Batch) RL. The picture at the top is the process of online RL, and the offline RL drawing is at the bottom. 9

Figure 4.1 Tutor Interface of Pyrenees 20

Figure 5.1 LSTM cell structure. In this figure, σ(Sigmoid) and *tanh* are activation functions. 25
Figure 5.2 Model architecture for our DRL model. 26
Figure 5.3 ECR evolution of Double DQN+LSTM with $k = 3$ or 4 30
Figure 5.4 Comparison for Normalized Learning Gains in F22 study 32
Figure 5.5 Comparison for Normalized Learning Gains(Isomorphic) in F22 study . . 33

Figure 7.1 Tutor Interface of Deep Thought(Problem Solving) 41
Figure 7.2 Tutor Interface of Deep Thought(Worked Example) 42
Figure 7.3 Problem solving process example in Deep Thought 43

Figure 8.1 Model architecture generating latent space model for one ITS domain . 47
Figure 8.2 Model architecture generating shared latent space model for two ITSs domain. 50
Figure 8.3 Loss function growth for Multi-Task Reinforcement Learning model. . . 54

Figure A.1 Inner architecture of RSSM 66

CHAPTER

1

INTRODUCTION

Today, the Republic of Korea(ROK) military is looking to improve its military capabilities by expanding the application of AI in the education and training of its soldiers. Providing personalized learning for a large number of soldiers is one of the most significant challenges because every male in ROK is required to serve in the military for about two years, so while there are a large number of soldiers, there are relatively few officers and NCOs(Non-Commissioned Officers) available to teach them. Therefore, a system that can teach military knowledge according to a soldier's rank and level is a significant challenge for the South Korean military. The ROK military already uses computer-based training (CBT) to teach basic tactics and weapon operation, but the system is not efficient enough to improve the learning gain. This is because most of the CBT just visualizes the military manual and randomly issues questions from a large pool of questions to evaluate the learner's learning outcomes.

One of the ways to solve the above is to apply an intelligent tutoring system (ITS), an interactive E-learning system without human intervention. The main advantage of ITS is providing tailored instruction and guidance to each learner by leveraging their ability to adapt to the learning process and individual characteristics of the learner[1]. To maximize this advantage, ITS must provide efficient decision-making processes to adapt to each learner's needs. For that, Reinforcement Learning (RL), a machine learning technique for deciding an effective policy by interacting with the environment, is one of the best approaches because the primary purpose of the RL agent is how to make the best decision maximizing the cumulative future reward provided by the environment. RL has also become a widely used method for decision-making in ITS, allowing the system to make appropriate decisions for enhancing each student's learning gain. Among the variations of RL algorithms, Deep Reinforcement Learning (DRL)[19] has shown outstanding performance by combining deep neural networks with traditional reinforcement learning. Using the advantage of neural networks, DRL can approximate the properties in RL without complex feature engineering. DRL has been successfully applied in various fields, such as online games, robotics, and military applications. In the context of ITS, DRL can help personalize the learning experience by adapting to the individual learning process of each learner.

Following our motivation and the advantages of DRL, this work aims to explore the impact of DRL on ITS and learn lessons about an effective way to apply DRL in ITS. Although it is an appropriate way to use ITS for military tasks, it is currently unavailable because of military security. Thus, we used the other effective ITSs, which are widely applied in previous works. We address **two critical issues** for our purpose. First, we explore the effectiveness of DRL-induced policies in ITS. Second, we study the potency of DRL with Multi-Task Learning(MTL)[8], another critical learning paradigm in RL. We mainly used the same procedures by [3], but our results for both issues show different results and analyses.

1.1 DRL for Pyrenees

We applied our DRL-induced policy in Pyrenees, the ITS teaching probabilistic principles to students. Our DRL algorithm is based on Double Deep Q Network[63], which reduces the over-estimation problem of Deep Q Network[51]. Also, our neural network architecture consists of Long Short Term Memory (LSTM)[30] for handling the long time distance of student learning trajectories data. In addition, we compared the performance of two policies trained by two different input settings per problem. In terms of state s and given time t, the first input setting is to use the most recent three state observations $s_t, s_{t-1}, s_{t-2}(k=3)$ and the second is to use four state observations $s_t, s_{t-1}, s_{t-2}, s_{t-3}(k=4)$. We found that two policies with different input settings perform differently depending on the problems, and we deployed our DRL policies in real ITS based on this result. Our empirical result shows that our DRL policies provided better learning gains than the random but reasonable policy.

1.2 Multi-Task DRL for Deep Thought

The main drawback of DRL is that it needs a massive amount of data for training effective policy under the interaction with the environment. Although it is easy to collect lots of experiences in a simple domain like an Atari game, it is not the case in the real world. Here, Multi-Task Learning (MTL) can tackle this problem. MTL is a method of learning similar tasks simultaneously[8] while enhancing the ability to perform each task by sharing knowledge from each domain. This method is promising for alleviating the above problem because MTL uses multiple task-related data for training, so a task with a limited amount of data can also be effectively learned. For this advantage, we explored the performance of Multi-Task DRL in Deep Thought, the ITS teaching propositional logic rules. Here, we used the state-of-the-art model-based DRL algorithm named Dreamer[27], which shows high performance by predicting the RL environment in latent state space. In addition, for generating a shared latent state between two environments, we applied bisimulation metrics[18], which can measure the behavioral similarity between two states. Unfortunately, our induced Multi-Task DRL cannot get the expected performance in our empirical result. Therefore, we analyzed the potential reasons for this result.

In the following, we will briefly explain RL, MTL, and pedagogical policy induction in ITS in Chapter 2. Then, our main works will be divided into two main parts: Part I) DRL for Pyrenees; Part II) Multi-Task DRL for Deep Thought. Each part has a similar structure as follows. First, we will introduce backgrounds and related work to understanding DRL and MTL. Next, we will describe the ITS used in each part. And then, we will present the model architecture and how we trained our model to induce an effective policy for enhancing students learning outcomes. Finally, we will describe the design of the experiments we carried out and the analysis of the empirical results.

CHAPTER

2

BACKGROUND

This chapter will describe general concepts and information to understand our work. First, we introduce the fundamental notions about Reinforcement Learning (RL), the subarea of Machine Learning (ML) for decision-making. Unlike other ML algorithms, RL is aimed to induce optimal policy for an agent to obtain the highest cumulative reward by interacting with the environment. We will explain the main concepts in RL, the classification of RL, and Deep Reinforcement Learning (DRL), which is directly related to the first methodology in our work. In the second part, we will briefly review the background of Multi-Task Learning (MTL) which is related to our second experiment. MTL is an actively researched branch in the ML world. The objective of multi-task learning is training multiple tasks simultaneously and finding a better model than a single-task trained model. We will cover the concepts of multi-task learning and review the importance of MTL from the perspective of RL. Lastly, we will describe the pedagogical policy induction using RL and MTL in ITS.

2.1 Reinforcement Learning

With outstanding performance in problem-solving, predicting, and analyzing, Artificial Intelligence (AI) is applied in various areas such as education, E-commerce, military, etc. One rapidly growing area in AI is ML. ML is the process of an agent (machine) learning specific tasks by providing task-related data. Thanks to technological advances providing massive and elaborated data, ML has become the core technology in Robotics, Computer Vision, Natural Language Processing, and so on. ML can be divided into three main branches: supervised, unsupervised, and RL. Here, we mainly focus on RL.

RL is a subfield of ML. It is the most promising approach for solving decision-making problems nowadays. Intuitively, RL is a learning process of what behavior is the most beneficial in a given situation by interacting with the environment. For that, it is required for four main components in RL: *policy*, *state*, *action*, and *reward*. In detail, the RL agent aims to induce the *policy*, which can select the best *action* in a given *state*. And here, the best *action* should achieve the maximum future cumulative *reward*. To capture the action providing the maximum numerical reward, the RL agent repeats trial-and-error throughout the interaction with the environment and learns the best action.

Because of these properties, there exist huge differences between RL and those mentioned two main categories of traditional ML types: supervised learning and unsupervised learning. Supervised learning is the learning process with a labeled training dataset. The supervised learning algorithms target to map the function $\hat{y} = f(x)$. In the function, x is the input data, such as features in the data point, and $\hat{y}$ is the output data which should be trained to be equal to labeled data y. On the other hand, unsupervised learning does not need labeled data. As there is no ground truth in the data, an unsupervised learning algorithm finds out the hidden patterns or