UNIVERSITY OF TECHNOLOGY SYDNEY

Deep Reinforcement Learning Conditioned on the Natural Language

by

Yunqiu Xu

Thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy
under the supervision of Ling Chen and Chengqi Zhang

Australian Artificial Intelligence Institute
Faculty of Engineering and Information Technology
University of Technology Sydney

May 2022

Certification of Original Authorship

I, Yunqiu Xu, declare that this thesis is submitted in fulfilment of the requirements for the award of degree of Doctor of Philosophy, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Signed:

Production Note:
Signature removed prior to publication.

Date: 13/01/2022

Abstract

Language-conditional reinforcement learning refers to the reinforcement learning task where the language information serves as essential components in the problem formulation. In recent years, the advances of deep reinforcement learning and language representation learning lead to increasing research interest in this cross-domain topic, which brings benefits to the studies in both language learning and reinforcement learning. However, challenges arise along with the premises, hindering the language-conditional reinforcement learning from being applied in the real world. In this research, we aim at designing language-conditional RL agent that is capable of handling the major challenges.

We first address the challenges in state representation learning under partial observability. Motivated by the premises of the transformer architecture in natural language processing, we design an adaptable transformer-based state representation generator featured with reordered layer normalization, weight sharing and block-wise aggregation. We empirically validate our method on both synthetic and man-made text-based games with different settings. The proposed method show higher sample efficiency in solving single synthetic games, better generalizability in solving unseen synthetic games, and better performance in solving complex man-made games.

Secondly, we study the reasoning process in language-conditional reinforcement learning. The reasoning ability enables the agent to generate the actions with the support of an explainable inference procedure. To achieve this ability, we propose an agent featured with the stacked hierarchical attention mechanism. Through exploiting the structure of the knowledge graph, this agent is able to explicitly model the reasoning process. Our agent demonstrates effectiveness on a range of man-made text-based games.

Thirdly, we study the generalization problem in language-conditional RL. We consider the knowledge graph-based observation, and address this challenge by designing a two-level hierarchical RL agent. In the high level, we use a meta-policy for task decomposition and subtask selection. Then, in the low level, we use a sub-policy for subtask-conditioned action selection. In a series of 8 game sets with different generalization types and game difficulty levels, our proposed agent enjoys generalizability and yields favorable performance.

Finally, we provide solutions to the challenges of low sample efficiency and large action space. We introduce the world-perceiving modules, which automatically decompose tasks and prune actions by answering questions about the environment. We then propose a two-phase training framework to decouple language learning from reinforcement learning, which further improves the sample efficiency. We empirically demonstrate that the proposed method not only achieves improved performance with high sample efficiency, but also exhibits robustness against compound error and limited pre-training data.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Prof. Ling Chen and Prof. Chengqi Zhang for their continuous and unconditional assistance throughout the journey in pursuing this degree. They supported me at every stage of my PhD studies, such as leading me to the area of language-conditional reinforcement learning, encouraging me to develop not only the in-depth knowledge in my research topic but also broad research interests in machine learning, helping me to establish connections with academic and industrial bodies, and providing me with valuable career & life advice. Without your patience, encouragement and persistent help, I will never grow into a professional researcher.

Besides my supervisors, I would like to offer my special thanks to Dr. Meng Fang, for providing me with hand in hand guidance to deliver high quality research. Thank you for aiding me to formulate a project management-like research schedule. I benefit from the close and inspiring discussions, which cover almost every aspect of details about how to conduct a research project, such as the problem formulation, model development, experiment design, and the manuscript preparation. Without your help, I probably waste a lot of time on detours.

I would like to acknowledge Dr. Yali Du, Dr. Gangyan Xu, Dr. Joey Tianyi Zhou, Dr. Yang Wang, Dr. Binbin Huang, for providing helpful suggestions throughout our collaboration. I would also like to mention those I met at UTS, specifically Dr. Wei Wu, Dr. Hong Yang, Dr. Yaqiong Li, Dr. Jiamiao Wang, Dr. Jun Li, Mr. Yu Liu, Mr. Shaoshen Wang, Mr. Yayong Li, Ms. Yang Zhang, and Mr. Zihan Zhang. It's my pleasure to work with all of you. In particular, I would like to thank Prof. Yi Yang, for introducing me to Prof. Ling Chen.

I am proud of my parents and grandparents. Thank you for raising me up, providing me with a joyful childhood, and inspiring my curiosity in science and engineering. Thank you for expressing your endless love and support across the ocean. This thesis is dedicated to my grandfather, a veteran passed away in the last year of my PhD study. Thank you, for fostering me a rigorous attitude, and teaching me to be tough to survive hardships.

Last but not the least, I would like to thank my wife, Ms. Yang Liu, for accompanying me through thick and thin.

Contents

List of Figures

List of Tables