Pre-trained Language Models as Prior Knowledge for Playing Text-based Games

Extended Abstract

Ishika Singh, Gargi Singh, Ashutosh Modi
Department of Computer Science and Engineering
Indian Institute of Technology Kanpur (IITK), India
{ishikas, sgargi}@iitk.ac.in, ashutoshm@cse.iitk.ac.in

ABSTRACT

Recently, text world games have been proposed to enable artificial agents to understand and reason about real-world scenarios. These text-based games are challenging for artificial agents, as it requires an understanding of and interaction using natural language in a partially observable environment. Past approaches have paid less attention to the language understanding capability of the proposed agents. In this paper, we improve the semantic understanding of the agent by proposing a simple *RL with LM* framework where we use transformer-based language models with Deep RL models. Overall, our proposed approach outperforms on 4 games out of the 14 text-based games, while performing comparable to the state-of-the-art models on the remaining games.

KEYWORDS

Interactive Fiction Games; Reinforcement Learning; NLP

ACM Reference Format:
Ishika Singh, Gargi Singh, Ashutosh Modi. 2022. Pre-trained Language Models as Prior Knowledge for Playing Text-based Games: Extended Abstract. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 3 pages.

1 INTRODUCTION

Artificial autonomous agents suffer from a number of challenges during training, such as reward, goal, or task under- or over- specification(s), and generalization. In this scenario, being able to utilize language as an interface between a user and an artificial agent simplifies a number of these challenges. Text-based Interactive Fiction Games (IFG) [4] provide such an environment where the agent learns to consume and produce natural language-based inputs and outputs. IFG present several challenges for artificial agents as these cover the real world settings. It requires an agent to understand the textual state description, handle combinatorial textual action space, and learn a policy in a partially observable environment to maximize the game score. The key challenge is to decipher the long textual observations, extract reward cues from them, and generate a semantically rich representation such that the policy learned on top of it is well informed. Most of the existing works learn textual representations from scratch during the RL training [1, 2, 4, 12–14], except [3] which uses pre-trained GloVe embeddings [7]. Online reinforcement learning is known to be sample-inefficient [11]. The reward-based temporal difference objective used for training is proposed for learning the gameplay, and it does not necessarily reinforce the agent to learn the semantics of the game. Learning textual representations solely from the game-generated text during online learning does not provide rich enough representation to be able to handle such complicated decision-making tasks. In addition, these games also need some prior knowledge about the functioning of the world. We propose the use of a pre-trained Language Model (LM) fine-tuned on game dynamics of RL agents trained on text-based games. It provides three-fold benefits to the RL agent: linguistic priors, world sense priors, and game sense priors, thereby, facilitating the agent to achieve SOTA results on several text-based games.

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online.

2 RELATED WORK

Recently, Hausknecht et al. [4] proposed a learning environment (*Jericho*) that supports a set of 32 human-written IF games. These games are written to be challenging for human players, thereby providing a realistic test-bed for training intelligent agents. Many of the recent modeling approaches [1, 2, 12] use a dynamically updated Knowledge Graph (KG) to represent the current state of the universe (game). KG-A2C [1] is the first such proposal. One of the SOTA methods by Xu et al. [12], in addition to KG-A2C architecture, computes a representation of the game state with attention on sub-components of the graph and on the complete graph. Q*BERT [2] presents an open domain QA method to update the KG with more information and an additional intrinsic motivation reward to enable structured exploration. MPRC-DQN [3] solves the partial observability by an object-centric retrieval of relevant past observations. Yao et al. [13] investigate to what extent semantic information is utilized by the DRRN [5] agent. They show that even in the complete absence of text, the "DRRN + template actions" setup is able to achieve significant scores, indicating the underlying overfitting to the reward system and memorization tendency of DRRN. They use an inverse-dynamics loss function to regularize the DRRN representation space for improving the semantic understanding. CALM [14] generates the next set of possible action by fine-tuning a GPT-2 model [8].

3 METHOD

In this work, we propose a simple approach that performs better than previously proposed complex approaches. We deploy a pre-trained LM with the existing game agents, namely Deep Reinforcement Relevance Network (DRRN) [5] and Template-Deep Q-Network (TDQN) [4], used in the recently proposed Jericho IF game-suite [4]. To the best of our knowledge, we are the first to utilize pre-trained LMs for IF games. We use `DistilBERT`[9] (hereafter

Table 1: Final (raw) and max scores seen by agents (available via prior work) comparing trained DBERT-DRRN with all the current SOTA baselines across a set of games from Jericho game-suite. The missing scores not in the previous work are denoted as '-'. *Max* denotes the maximum possible scores based on human-written optimal walkthroughs for winning the game without the step limit of 100. Note that our model (DBERT-DRRN) is trained for less than half of the training steps used in other baselines. *Bold* scores denote the best score, blue* scores denote the second best. DBERT-DRRN gets an avg. norm of 15.7% on all games, and 21.1% on 6 games reported for INV-DY. *the games with rising learning curve for DBERT-DRRN till the last training step. ^{p}(possible/easy), ^{d}(difficult), ^{e}(extreme) refers to game difficulty level as given in [4].

Game	TDQN [4] raw	DRRN [4] raw / max	MPRC-DQN [3] raw	SHA-KG [12] raw	Q*BERT [2] max	CALM [14] raw	INV-DY [13] raw / max	DBERT-DRRN (ours) raw / max	Max
Inhumane*^{p}	0.7	0	0	5.4	-	25.7*	19.6 / 45	**32.8** / **50**	90
Jewel^{d}	0	1.6	4.46*	1.8	-	0.3	-	**6.5** / 13	90
Library*^{p}	6.3	17.0*	**17.7**	15.8	19	9.0	16.2 / **21**	17.0* / **21**	30
Ludicorp^{d}	6	13.8	**19.7**	17.8*	22.8	10.1	13.5 / **23**	12.5 / 18	150
Omniquest^{p}	**16.8**	5	10.0	-	-	6.9*	5.3 / **10**	4.9 / 5	50
Reverb^{p}	0.3	8.2*	2.0	**10.6**	-	-	-	6.1 / 12	50
Snacktime^{p}	9.7	0 / 0.25	0	-	-	19.4*	-	**20.0** / **20**	50
Spellbrkr*^{e}	18.7	37.8	25	**40**	-	**40**	-	38.2* / 40	600
Spirit^{e}	0.6	0.8	**3.8**	**3.8**	-	1.4	-	2.1* / 8	250
Temple^{p}	7.9*	7.4	**8.0**	7.9*	**8.0**	0	-	**8.0** / 8.0	35
Tryst205*^{e}	0	9.6*	**10.0**	6.9	-	-		9.3 / 17	350
Zork1^{d}	9.9	32.6 / 53	38.3	34.5	41.6	30.4	43.1* / **87**	**44.7** / 55	350
Zork3^{d}	0	0.5	**3.63**	0.7*	-	0.5	0.4 / **4**	0.2 / **4**	7
Yomomma*^{d}	0	0.4	**1.0**	-	-	-	-	0.5* / 1.0	35
Avg. Norm (%)	7.7	10.2	14.2	13.3	-	12.8	18.9	**15.7** (**21.1**)	100

referred to as **DBERT**) as the LM owing to its compact size, thereby computing rich text encoding efficiently. We fine-tune DBERT on an independent set of human gameplay transcripts to add a game sense to the pre-trained LM. This set of transcripts contain games other than those in the evaluation game-suite to test the generalizability of the proposed approach. We use the ClubFloyd dataset [14] for fine-tuning DBERT. It is a collection of human game-play trajectories on 590 games. We pre-process this data to obtain around 217K pairs of observation and action, (o_t, a_t). We test both DBERT-DRRN and DBERT-TDQN setup on Zork1, and additionally, we also evaluate DBERT-DRRN on a set of other games from Jericho to test the generalization capability of the model.

4 RESULTS AND ANALYSIS

Table 1 reports the final scores of our best performing model (trained DBERT with DRRN) in comparison with 7 existing baselines on 14 games. We report raw or maximum or both the scores as given in original papers. Different baselines achieve SOTA scores on different games. Our model achieves SOTA results on 4 games: Zork1, Inhumane, Snacktime, and Jewel, while being second best or comparable on most of the other games. INV-DY [13] uses additional loss objectives inspired from curiosity-driven exploration [6]. While it helps them achieve higher maximum scores on Zork1, but are not able to learn the high score trajectories. On the other hand, our agent efficiently learns the max score trajectories explored by it, thereby indicating that with a better exploration strategy our model has the potential to achieve better scores. None of the other agents, with a max score of 55, are able to stably reach a score as high as our model, that maintains a margin of 6.4 from the best model [3]. Our agent explores higher max score on Inhumane, but more importantly, it is able to learn the best-explored trajectories, thereby plateauing closer to the max scores for many games (Inhumane, Jewel, Omniquest, Zork1), indicating that the trained DBERT is an important learning component, and it also facilitates the exploration to some extent. We also report the normalized score (raw score as a factor of max possible score collected from human-written optimal walk-through) averaged across all games. We get an overall norm of 15.7%, followed by 14.2% achieved by MPRC-DQN [3]. Compared to some of the previous works, our model does not suffer a lot on any game, e.g., the second best (MPRC-DQN) gets 0 scores on Inhumane and Snacktime, and notably higher scores on others (Ludicorp, Spirit, Zork3). It indicates both the generalization tendency and the necessity of the pre-trained LM deployed in our model. Compared to the average norm of 18.9% for INV-DY (evaluated on 6 of the 14 games), we get 21.1%. The average norm is also a measure of human-machine gap for text-based games, indicating that IFG are at best only 15.7% solved. Hence, it is a good benchmark for developing language understanding agents. Moving past trained DBERT-DRRN score will likely require a more intelligent agent with better exploration and learning strategies. We have included other details and more results in the full paper [10].

REFERENCES

[1] Prithviraj Ammanabrolu and Matthew Hausknecht. 2020. Graph Constrained Reinforcement Learning for Natural Language Action Spaces. In *International Conference on Learning Representations*. https://openreview.net/forum?id=B1x6w0EtwH

[2] Prithviraj Ammanabrolu, Ethan Tien, Matthew Hausknecht, and Mark Riedl. 2021. How to Avoid Being Eaten by a Grue: Structured Exploration Strategies for Textual Worlds. (2021). https://openreview.net/forum?id=eYgI3cTPTq9

[3] Xiaoxiao Guo, Mo Yu, Yupeng Gao, Chuang Gan, Murray Campbell, and Shiyu Chang. 2020. Interactive Fiction Game Playing as Multi-Paragraph Reading Comprehension with Reinforcement Learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, 7755–7765. https://doi.org/10.18653/v1/2020.emnlp-main.624

[4] Matthew Hausknecht, Prithviraj Ammanabrolu, Marc-Alexandre CÃťtÃĮ, and Xingdi Yuan. 2020. Interactive Fiction Games: A Colossal Adventure. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 05 (Apr. 2020), 7903–7910. https://doi.org/10.1609/aaai.v34i05.6297

[5] Ji He, Jianshu Chen, Xiaodong He, Jianfeng Gao, Lihong Li, Li Deng, and Mari Ostendorf. 2016. Deep Reinforcement Learning with a Natural Language Action Space. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 1621–1630. https://doi.org/10.18653/v1/P16-1153

[6] Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. 2017. Curiosity-driven Exploration by Self-supervised Prediction. (2017). arXiv:1705.05363 [cs.LG]

[7] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 1532–1543. https://doi.org/10.3115/v1/D14-1162

[8] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[9] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. (2020). arXiv:1910.01108 [cs.CL]

[10] Ishika Singh, Gargi Singh, and Ashutosh Modi. 2021. Pre-trained Language Models as Prior Knowledge for Playing Text-based Games. (2021). arXiv:2107.08408 [cs.CL]

[11] Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.

[12] Yunqiu Xu, Meng Fang, Ling Chen, Yali Du, Joey Tianyi Zhou, and Chengqi Zhang. 2020. Deep Reinforcement Learning with Stacked Hierarchical Attention for Text-based Games. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 16495–16507. https://proceedings.neurips.cc/paper/2020/file/bf65417dcecc7f2b0006e1f5793b7143-Paper.pdf

[13] Shunyu Yao, Karthik Narasimhan, and Matthew Hausknecht. 2021. Reading and Acting while Blindfolded: The Need for Semantics in Text Game Agents. In *North American Association for Computational Linguistics (NAACL)*.

[14] Shunyu Yao, Rohan Rao, Matthew Hausknecht, and Karthik Narasimhan. 2020. Keep CALM and Explore: Language Models for Action Generation in Text-based Games. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Online, 8736–8754. https://doi.org/10.18653/v1/2020.emnlp-main.704