A Generative Policy Gradient Approach for Learning to Play Text-Based Adventure Games

René Raab[*]
Thesis supervisor: Kurt Driessens

Department of Data Science and Knowledge Engineering
Maastricht University, The Netherlands

Keywords: Deep Reinforcement Learning · Text-based Games · Policy Gradients · Representation Learning · Variational Autoencoders

In recent years, improvements in deep reinforcement learning have enabled agents to achieve superhuman performance on a range of tasks [7,10]. These advances are mostly limited to areas that consist of a visual input and require selecting the best of a small number of discrete actions. Applying deep reinforcement learning to text-based games, on the other hand, remains difficult. There are many reasons for this, but the most important differences to successful applications are the large state space and textual observations of text-based games, on the one hand, and the large, quasi-continuous yet at the same time sparse action space on the other hand [3].

Existing work that is concerned with learning to play text-based games is either using specifically engineered solutions [6] or deep Q networks [7] with discrete actions [13,4,8,1,14]. Some of these approaches [13,4,8] artificially reduce the number of actions to only contain two-word actions consisting of a verb and an object. This reduces their ability to work with more realistic games that require more complex actions. This thesis therefore evaluates the idea of moving away from discrete actions and instead modelling text-based game actions as continuous variables. This change enables the application of policy gradient algorithms as an alternative to Q-learning. The idea behind using continuous actions is to enable a more generative approach to action handling in text-based games in the future.

To build a proof of concept reinforcement learning system that has the ability to generate textual actions, we employ a continuous latent space that can be decoded into text as an action space. To construct this latent space, we train a variant of the variational autoencoder (VAE) [2] on a set of pre-defined textual actions. The action set is then defined as $\mathbb{R}^d$ where any action $a' \in \mathbb{R}^d$ can be decoded into a textual action a using one of two methods: (i) By returning the text associated with the nearest known action representation (taken from the

[*] René Raab is a PhD candidate at the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) in Germany. This work constitutes a summary of his Master's thesis at Maastricht University. The thesis can be found at https://reneraab.org/master_thesis.pdf

training set), or (ii) using the decoder part of the VAE to decode the vector a' into text a. The advantage of method (i) is that all generated text is meaningful and interpretable by the game. The advantage of method (ii) is that it treats the latent space as truly continuous, which is expected to benefit the policy gradient algorithm.

For training the agent, we focus on a modification of the simple policy gradient scheme called REINFORCE [12,11] and avoid more advanced policy gradient algorithms to get a clearer understanding of possible issues when applying these kinds of techniques to the domain of text-based games. We use REINFORCE and backpropagation [9] to train a neural network that produces a policy π which is described by a probability distribution over the action space. The network takes the textual state description from the game as an input, uses an LSTM [5] layer to learn a vector-shaped state representation which is then passed through several hidden layers and finally results in the predicted mean and standard deviation of a Gaussian distribution. This distribution π indicates which action(s) are expected to perform best (an actual action can be obtained by sampling from the distribution and then decoding the resulting vector). When updating the neural network we can either use the sampled action or the nearest neighbour (comparable to method (i) and (ii) of action decoding described above) in the update equations of REINFORCE.

We have tested this setup on short games that are generated using the TextWorld framework [3]. These games contain a very sparse room and require the player to perform a small number of actions to win. The described setup shows some promise but it turns out that it was not able to learn how to finish a game that requires more than two correctly generated consecutive actions. We assume that this is due to the simple network structure and training algorithm used in this initial step towards a generative solution. On games that were finished successfully, the nearest neighbour (i) and VAE decoding (ii) options perform similarly well. Furthermore, we have seen that the best update rule was to use the point that was responsible for the resulting text (i.e. the nearest neighbour in the first case and the sampled point in the VAE decoding case). This lets us hope that future work can expand on this generative approach and replace the VAE with a generic language model which is detached from the initial action/training set that was required for this work. We furthermore assume that more advanced learning algorithms will help solve the problem of learning longer games.

In summary, the combination of a policy gradient approach and continuous natural language representations seems to be a viable foundation for agents that can learn to play text-based games. Future research needs to be performed into a better integration of natural language processing, especially in terms of considering textual structure, and more expressive generative language models. This thesis has only focused on one of the many open issues in learning to play text-based games; to succeed at learning to play real and more complex games, future work also needs to tackle the other problems described in [3].

References

1. Ammanabrolu, P., Riedl, M.O.: Playing text-adventure games with graph-based deep reinforcement learning. CoRR **abs/1812.01628** (2018)
2. Bowman, S.R., Vilnis, L., Vinyals, O., Dai, A., Jozefowicz, R., Bengio, S.: Generating sentences from a continuous space. In: Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning. pp. 10–21 (2016)
3. Côté, M., Kádár, Á., Yuan, X., Kybartas, B., Barnes, T., Fine, E., Moore, J., Hausknecht, M.J., Asri, L.E., Adada, M., Tay, W., Trischler, A.: Textworld: A learning environment for text-based games. CoRR **abs/1806.11532** (2018)
4. Fulda, N., Ricks, D., Murdoch, B., Wingate, D.: What can you do with a rock? affordance extraction via word embeddings. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (Aug 2017)
5. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural computation **9**(8), 1735–1780 (1997)
6. Kostka, B., Kwiecieli, J., Kowalski, J., Rychlikowski, P.: Text-based adventures of the golovin ai agent. In: 2017 IEEE Conference on Computational Intelligence and Games (CIG). pp. 181–188. IEEE (2017)
7. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., Hassabis, D.: Human-level control through deep reinforcement learning. Nature **518**(7540), 529–533 (Feb 2015)
8. Narasimhan, K., Kulkarni, T., Barzilay, R.: Language understanding for text-based games using deep reinforcement learning. Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (2015)
9. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**(6088), 533–536 (1986)
10. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al.: Mastering the game of go without human knowledge. Nature **550**(7676), 354 (2017)
11. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT press, 2 edn. (2018)
12. Williams, R.J.: Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine learning **8**(3-4), 229–256 (1992)
13. Yuan, X., Côté, M., Sordoni, A., Laroche, R., des Combes, R.T., Hausknecht, M.J., Trischler, A.: Counting to explore and generalize in text-based games. CoRR **abs/1806.11525** (2018)
14. Zahavy, T., Haroush, M., Merlis, N., Mankowitz, D.J., Mannor, S.: Learn what not to learn: Action elimination with deep reinforcement learning. In: Advances in Neural Information Processing Systems. pp. 3562–3573 (2018)