ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

Facebook stock has been incredibly rising over the last five years. Hence the algorithm has not much difficulty finding the optimal strategy that is to buy and hold the stock.

6 Conclusion

We show in this article that there are tight connections between SL and RL. PGMs in RL can be cast as cross-entropy minimization problems where true labels are replaced by expected future rewards or advantages while the log term is changed into the log policy term. This analogy takes its root from the minimization problem where we are looking for the parameters that maximize the expected future rewards or advantages. Should this optimization objective changed, we conjecture that we could make other analogies between SL and RL.

References

[1] R. J. Williams. *Simple statistical gradient-following algorithms for connectionist reinforcement learning*, volume 8. Springer, 1992.

[2] Richard S. Sutton, David McAllester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In *NIPS*, pages 1057–1063, Cambridge, MA, USA, 1999. MIT Press.

[3] David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. Deterministic policy gradient algorithms. In *ICML*, number 1 in Proceedings of Machine Learning Research, pages 387–395, Bejing, China, 22-24 Jun 2014. PMLR.

[4] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra. Continuous control with deep reinforcement learning. *ArXiv e-prints*, 2015.

[5] Vijay R. Konda and John N. Tsitsiklis. On actor-critic algorithms. *SIAM J. Control Optim.*, 42(4):1143–1166, April 2003.

[6] J. Peters and S. Schaal. Natural actor-critic. *Neurocomputing*, 71(7-9):1180–1190, March 2008.

[7] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *ICML*, volume 48, pages 1928–1937, NY USA, 20-22 Jun 2016. PMLR.

[8] Alexander L. Strehl. *Associative Reinforcement Learning*, pages 49–51. Springer US, Boston, MA, 2010.

[9] Matthias Rolf and Minoru Asada. Where do goals come from? a generic approach to autonomous goal-system development. *ArXiv*, abs/1410.5557, 2014.

[10] S. Schaal. Learning from demonstration. In *NIPS, MIT Press*, pages 1040–1046. NIPS, MIT Press, 1996.

[11] S. Ross, G. J. Gordon, and D. Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. *In: AISTATS, JMLR.org, JMLR Proceedings*, 15:627–635, 2011.

[12] Todd Hester, Matej Vecerik, Olivier Pietquin, Marc Lanctot, Tom Schaul, Bilal Piot, Andrew Sendonaris, Gabriel Dulac-Arnold, Ian Osband, John Agapiou, Joel Z. Leibo, and Audrunas Gruslys. Learning from demonstrations for real world reinforcement learning. *CoRR*, abs/1704.03732, 2017.

[13] Christopher John Cornish Hellaby Watkins. *Learning from delayed rewards*. PhD thesis, University of Cambridge England, 1989.

[14] R. S. Sutton and A. G. Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition, 1998.

[15] Thomas Degris, Patrick M. Pilarski, and Richard S. Sutton. Model-free reinforcement learning with continuous action in practice. *IEEE In ACC*, 2012:2177–2182, 2012.