Deep Reinforcement Learning Agents Revealing Uncertainties in Blockchain Systems

Contact

Önder GÜRCAN, e-mail: onder.gurcan@cea.fr, desk phone: +33 (0) 1 69 08 00 07

Beginning date: September 2020

Host Laboratory

CEA is a public multidisciplinary research organization whose research fields range from nuclear industry to biosciences and fundamental physics. It is made of several research centers located in France. CEA represents 15024 employees, 2.7 B Euros budget, 1689 patents registered or active, and 1300 contracts signed with industry. More than 80 new companies have been created since 1984 in high technologies sectors and 9 research centers (http://www.cea.fr/english). In HORIZON 2020, the EU Framework Programme for Research and Innovation, CEA is already involved in more than 100 projects.

The CEA LIST (Laboratory for Integration of Systems and Technology) institute is part of CEA TECH, the CEA Technological Research Division. CEA LIST combines basic research and industrial R&D and is primarily concerned with the development of technologies that combine software and hardware to form highly integrated complex systems. The research activities are structured into three major themes: embedded systems, interactive systems and sensors, and signal processing. CEA LIST focuses on methods and tools for the design of embedded systems with appropriate architectures, software, and an optimal level of safety.

The successful candidate will join a research team in distributed systems at Laboratory for Trustworthy, Smart and Self-Organizing Information Systems (LICIA) at CEA Paris-Saclay campus (Paris area). LICIA is in charge of the development of techniques, methods and tools for the formalization and management of large distributed, open and collaborative information systems. In this context, LICIA develops methods, algorithms and tools for formal analysis and agent-based engineering for:

- Modeling, formalization and implementation of enforceable contracts in collaborative systems
- Development and implementation of multi-agent algorithms for self-governance in distributed collaborative environments.
- The development of techniques based on game theory for resilience analysis.

This research thread, while still quite recent, is already very active and well integrated in the related European research area. Thanks to the various outstanding academic partnerships we have developed, the coming years are very promising, with several National and European collaborative projects.

Context

Since its genesis in late 2008 [11], Bitcoin had a rapid growth in terms of participation, number of transactions and market value. This success is mostly due to innovative use of existing technologies for building a trusted ledger called blockchain. A blockchain system allows its participants (agents) to collectively build a distributed economic, social and technical system where anyone can join (or leave) and perform transactions in-between without needing to trust each other, having a trusted third party and having a global view of the system. It does so by maintaining a public, immutable and ordered log of transactions, which provides an auditable trusted ledger accessible by anyone.

Problematic

Technically speaking, all agents store unconfirmed transactions in their memory pools and confirmed transactions in their blockchains. Users agents create transactions with a fee and then broadcast them across the blockchain network to be confirmed (i.e., totally ordered and cryptographically linked to the block-chain). After receiving a certain number of transactions, block creator agents try to confirm them as a block by using a consensus algorithm (e.g., a hash-based proof-of-work by solving a computational puzzle of pre-defined difficulty or a practical byzantine fault-tolerance protocol). The successful block creator agent(s) broadcast(s) the next block to the network to be chained to the blockchain. However, this process is not trivial because a block-chain system is open and dynamic. Hence, both types of agents have to take into account uncertain constraints (e.g., the global merit, the transaction confirmation times, the transaction fees, the delays in the network, and the topology of the network) during their decision-making process for carefully balancing their objectives, otherwise this can lead to important consequences; where for instance, a trend on a growing number of unconfirmed transactions may create a service degradation and may result decreased participation of users agents [8], which in result may make no user agent to stay in the system, and thus make block creator agents to have no transactions to confirm and eventually make the whole system to be confined to end. This is a challenging task and has not been covered by formal studies conducted so far [6, 5, 14, 3, 12].

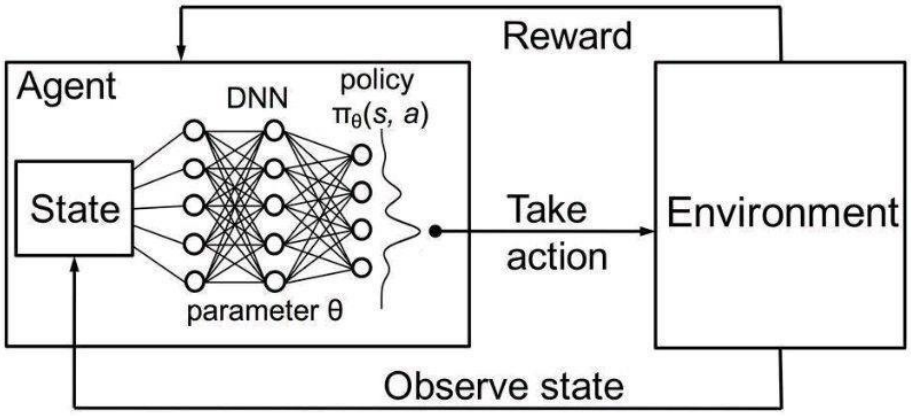

Figure 1: Representation of a Deep Reinforcement Learning Agent.

A promising approach to tackle such kind of problems is reinforcement learning [16], where agents learn how to behave in an unknown environment by performing actions and seeing the results, in order to maximize their individual cumulative returns (rewards) [2]. While reinforcement learning agents have achieved some successes in a variety of domains [13, 4, 17], their applicability has previously been limited to domains in which useful features can be handcrafted, or to domains with fully observed, low-dimensional state spaces [10]. In other words, traditional reinforcement learning algorithms have difficulty with domains featuring

- high-dimensional continuous state spaces,
- high-dimensional parameterized-continuous action spaces,
- partial observability, and
- multiple independent learning agents.

Blockchain systems, however, are environments that are too complex for humans to predetermine the correct actions using hand-designed solutions. Furthermore, the agents performing in these systems have limited observability, and the state and parameter spaces are vast and changing dynamically. Consequently, agents that can learn to tackle such complex real-world domains are needed.

Based on this observation, we hypothesize that deep reinforcement learning [7], where each agent is with a deep neural network; hold the key to scaling reinforcement learning towards complex tasks for agents acting in blockchain systems. Deep reinforcement learning had a great growth during the last decade [15, 9]. However, most of its successes have been in single agent domains, where behavior of

other agents is not so relevant. In blockchain systems, on the other hand, the interaction between multiple agents, which can cooperate or compete, is critical. Few studies focused on this topic so far [18,19].

Subject

This thesis seeks to answer the following two research questions:

1. How can the power of deep reinforcement learning be used for blockchain systems (i.e. complex environments featuring partial observability, high-dimensional parameterized-continuous state and action spaces, and sparse rewards)?
2. How can multiple deep reinforcement learning agents learn to cooperate in a multi-agent setting in blockchain systems and continuously learn the uncertain constraints?

Concretely, the objective of this thesis is to investigate the uncertain constraints of blockchain systems and to propose a deep reinforcement learning decision-making approach based on utility and rewards for both user and block creator agents.

The thesis will also contribute to develop and extend the agent-based simulation platform Multi-Agent eXperimenter (MAX) of LICIA.

References

[1] S. Bandini, S. Manzoni, and G. Vizzari. Agent based modeling and simulation: An informatics perspective. JASSS, 12(4), 2009. cited By 75.

[2] L. Bu soniu, R. Babuska, and B. De Schutter. Multi-agent Reinforcement Learning: An Overview, pages 183–221. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[3] M. Carlsten, H. Kalodner, S. M. Weinberg, and A. Narayanan. On the instability of bitcoin without the block reward. In Proceedings of the 2016ACM SIGSAC Conference on Computer and Communications Security, pages 154–167. ACM, 2016.

[4] C. Diuk, A. Cohen, and M. L. Littman. An object-oriented representation for efficient reinforcement learning. In Proceedings of the 25th international conference on Machine learning, pages 240–247. ACM, 2008.

[5] I. Eyal and E. G. Sirer. Majority is not enough: Bitcoin mining is vulnerable. In International Conference on Financial Cryptography and Data Security, pages 436–454. Springer, 2014.

[6] J. Garay, A. Kiayias, and N. Leonardos. The Bitcoin Backbone Proto-col: Analysis and Applications, pages 281–310. Springer Berlin Heidelberg, Berlin, Heidelberg, 2015.

[7] J. K. Gupta, M. Egorov, and M. Kochenderfer. Cooperative multi-agent control using deep reinforcement learning. In G. Sukthankar and J. A.Rodriguez-Aguilar, editors, Autonomous Agents and Multiagent Systems, pages 66–83, Cham, 2017. Springer International Publishing.

[8] Ö. Gürcan, A. Del Pozzo, and S. Tucci-Piergiovanni. On the bitcoin limitations to deliver fairness to users. In H. Panetto, C. Debruyne, W. Gaaloul, M. Papazoglou, A. Paschke, C. A. Ardagna, and R. Meersman, editors, On the Move to Meaningful Internet Systems. OTM 2017 Conferences, pages589–606, Cham, 2017. Springer International Publishing.

[9] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. A. Riedmiller. Playing atari with deep reinforcement learning. CoRR, abs/1312.5602, 2013.

[10] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Belle-mare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra,S. Legg, and D. Hassabis. Human-level control through deep reinforcement learning. Nature, 518:529 EP –, Feb 2015.

[11] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008.https://bitcoin.org/bitcoin.pdf.

[12] R. Pass, L. Seeman, and A. Shelat. Analysis of the blockchain protocol in asynchronous networks. IACR Cryptology ePrint Archive, 2016:454, 2016.

[13] M. Riedmiller, T. Gabel, R. Hafner, and S. Lange. Reinforcement learning for robot soccer. Autonomous Robots, 27(1):55–73, 2009.

[14] A. Sapirshtein, Y. Sompolinsky, and A. Zohar. Optimal selfish mining strategies in bitcoin. In International Conference on Financial Cryptography and Data Security, pages 515–532. Springer, 2016.

[15] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lilli-crap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis. Mastering the game of go with deep neural networks and tree search. Nature, 529:484EP –, Jan 2016. Article.

[16] M. Tan. Multi-agent reinforcement learning: Independent vs. cooperative agents. In Machine Learning Proceedings 1993, pages 330 – 337. Morgan Kaufmann, San Francisco (CA), 1993.

[17] G. Tesauro. Temporal difference learning and td-gammon. Communications of the ACM, 38(3):58–68, 1995.

[18] C. Hou, M. Zhou, Y. Ji, P. Daian, F. Tramer, G. Fanti and A. Juels. SquirRL: Automating Attack Discovery on Blockchain Incentive Mechanisms with Deep Reinforcement Learning. ArXiv 1912.01798, https://arxiv.org/abs/1912.01798, 2019.

[19] T. Wang, S. C. Liew, S. Zhang. When Blockchain Meets AI: Optimal Mining Strategy Achieved By Machine Learning. ArXiv 1911.12942, https://arxiv.org/abs/1911.12942, 2019.