International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

Reinforcement Learning for Real - World Applications - A Comprehensive Review

Ankur Mahida

Java Developer, Barclays

Abstract: ***Reinforcement learning (RL) has been transformed from theory to real - life application in the domains of healthcare fashion and many more. This decision paper highlights the history of RL in robotics, autonomous vehicles, business operations, business healthcare, and more. We kickstart RL in sequential decision - making where little human engineering is required. Major algorithms, including profound Q - network, actor - critic methods, and distributional RL, have been implemented into the processes of finding human - level policies to solve complex tasks. Simulation innovations, hierarchical RL, as well as a more efficient transfer from a simulator to the real world have increased the applicability. For instance, the robotics used in manufacturing processes, self - driving cars, supply chain management, automated trading, medical assistance systems, and gaming AI are all examples of their application. RL is responsible for full autonomy in making the best dynamic decisions, which, in turn, leads to increased automation and productivity. The long - term efforts of overcoming the bottlenecks of sample efficiency, safety, sim - to - real transfer, and stability are in the process of being actively investigated. With the development of parallelism and transfer learning, the effect of RL on real - world problems will be more striking.***

Keywords: reinforcement learning, deep RL, real - world applications, sequential decision - making

1. Introduction

The reinforcement learning (RL) approach has been gaining a lot of popularity in recent years for its capability to train intelligent agents to make the best decisions and policies in sequential environments. The agent in RL interacts with the surroundings through actions and feedback, rewards, and penalties. While playing by trial - and - error, the agent seeks to maximize its future reward, and over time, the agent's learning ability will be enhanced. This is in contrast with supervised learning from pre - set data sets. RL coupled with deep neural networks as universal function approximators have achieved human - level performance in high - dimensional/complex games and simulations. This success has consequently created zest in replicating the approach to real - world problems in a multitude of domains, including robotics, transport systems, business operations, finance, health, and so much more. While some problems, such as data efficiency, sim - to - real transfer, safety, and stability, remain, considerable progress has been made in the field.

2. Problem Statement

Frequently, we encounter real - life situations that require us to make long - term sequential decisions with uncertainty with the goal of attaining the highest cumulative rewards. Examples of these are – automated vehicles driving on the roads, robots building items in the manufacturing field, and artificial intelligence trading stocks or managing business operations. In these types of problems, the best action to take at each timestep depends on both the current state of the environment and the expected future states. However, the future states are not known in hindsight. Also, the environment can be very complex in a way that handcoding of the optimal policies for every possible scenario is infeasible.

Reinforcement learning (RL) gives a solution to this problem as it helps intelligent agents to learn efficient policies by themselves which are from experience. The agent interacts with the environment by taking some actions and getting feedback (rewards or penalties) [1]. The advantage of these trial - and - error RL algorithms is that they can develop policies that are able to beat the ones that have been previously hand - crafted. While practical implementation of RL is still an issue, it is still a popular choice. Vital ongoing challenges include:

- Sample efficiency - Real - world RL agents may require a considerable number of training hours, which can be expensive to acquire in the real world [2]. More cost - effective solutions are required.
- Sim - to - real transfer - The behaviors and methods learned from the simulation may translate less robustly into the complexities of the real world. [3] There is a necessity for more sophisticated simulation and transfer techniques.
- Security - This can happen when an RL agent is optimized by maximizing the accumulated rewards and may attempt unsafe actions. The safety mechanisms necessary during training could be more manageable [4].
- Stability - RL algorithms can be unstable or not converge at all when tackling complex state/action space, partial - observability, and noisy data [5]. To resolve this issue, more robust algorithms should be developed.

3. Solutions

Deep reinforcement learning (deep RL) has gained ground as it is a powerful technique for solving the challenging problems of the real world. Deep RL utilizes deep neural networks as a function approximator in addition to reinforcement learning algorithms [6, 7]. This makes space for the agent's policy and value functions to have an even higher representational capacity to model tasks with a high - dimensional state and action space. Several critical innovations in deep RL have led to human - level performance on complex game environments and expanded applicability to real - world domains: Several key innovations in deep RL have led to human - level performance on complex game environments and expanded applicability to real - world domains:

- Experience Replay - This involves the storing of agent experiences in a replay buffer from which samples are drawn, which reduces the correlation in the training data, leading to more stable learning. This supports sample efficient online action - based learning.
- Secondary networks - A secondary network of high - speed and stable target networks helps the algorithm analyze its response and make accurate decisions.
- Actor - critic methods - Actor - critic approaches directly learn both policy and value functions (actor and critic), which results in stable and sample - efficient learning compared to what Q - learning is able to achieve alone.
- Distributional RL – The selection of the value distribution instead of just the expected value gives better training signals that develop policy learning.
- Advances in simulators - The recent rise of powerful simulators, which are low - cost and safe, enables agents to be trained on massive scales. Such expanded applicability means that this presents not only challenges but also a chance for sim - to - real transfer.
- Hierarchical RL - Hierarchical agents that operate at a variety of temporal levels of abstraction are just as capable of learning quickly on complex tasks.
- Imitation learning - Human demonstrations can be utilized to cut the sample complexity in RL by using them to infer agent policies.
- Sim - to - real transfer - Advances in domain shuffling, system models, and noise injection make the transport of policies from simulations to the real world a reality.

4. Uses

a) *Robotics*

Reinforcement learning is making great strides in robotics in pulling out specific capabilities such as manipulation, grasping, navigation, and legged locomotion [7, 9]. RL enables robots to learn complex sensorimotor skills from experience rather than purely exploiting hand - crafted control policies. In particular, RL is able to train robot limbs to manipulate and sort objects in a complex, cluttered environment. In navigation, RL allows the use of robots in factories and warehouses for optimum path [planning and] collision avoidance. RL exhibits potential for such abilities that will make robots more dynamic and enable them to undertake actions such as walking, running, and recovering from falls. Rather than using input - output experiences for dynamics modeling and controller design, Reinforcement Learning learns locomotion policies directly from the experience. Although the challenges have remained, the main issues now are real - to - sim transfer and training vision - based policies. While RL can be considered a viable approach for robot learning, it is also the most promising one.

b) *Autonomous Vehicles*

RL (reinforcement learning) is used to train and control autonomous vehicles on roads so that they maneuver smoothly, safely, and efficiently in complex road environments [6, 8]. Unlike rule - based approaches that would be difficult to cover all scenarios, RL can enable self driving - cars to learn driving policies from driving experience directly. RL societies use road states, traffic patterns, and maneuvers as rewards and penalties to make di experience decisions. RL - trained models are capable of maintaining intelligent speed matching, lane changing and overtaking, and intersection and roundabout navigations. Nonetheless, essential issues remain about the ability to convert simulation policy into real - life action and the safety of personnel during the training session. However, RL provides a means of demonstrating conforming rules and planning and operating driverless cars.

c) *Logistics*

RL is minimizing logistics operation costs in provisioning, fleet management, supply chain, and logistics. RL can achieve this by having more sophisticated inventory and order fulfillment policies that are optimally tuned to minimize cost and maximize services [9]. At the delivery of the last mile, RL enables dynamic routing and scheduling with the traffic, demand, and vehicle positions in consideration. In fleet management, RL optimizes vehicle dispatch for better service to the customers and to increase the margins. For networks, the RL models the network dynamics to avert disruptions and errors. Logistics planning involving reinforcement learning produces scale economies manifesting themselves in the form of higher productivity, lower cost, improved service quality, and better resilience when compared to the traditional optimization of transportation networks.

d) *Finance*

The scope of RL in finance is applied to automated trading, portfolio optimization, risk management, and so forth. The market - order nature of markets makes RL effective. RL - based algorithmic trading can predict profitable short - term trading strategies by analyzing price patterns and trends. In portfolio management, RL dynamically shifts assets to gain the maximum return with risk objectives and constraints [10]. Banks can take advantage of RL not only in risk modeling and hedging but also in strategic decision - making. The issues are sample market efficiency and precise incentive design. RL provides adaptive and self - duplicative learning to maximize trading returns and portfolio performance beyond the best - case heuristic techniques.

e) *Healthcare*

RL has been improving healthcare that is able to provide personalized treatments, enhance medical imaging, and manage hospital resources. Precision medicine applies RL to patient treatment regimens by incorporating markers of disease response. For medical imaging, RL enhances reconstruction quality at the same time scanning time, and radiation exposure is decreased. At hospital business operations, RL approaches optimize scheduling, logistics, and utilization of resources to achieve service efficiency improvement [11]. However, this is continued by data availability. Still, the promising clinical use proves that RL has a strong advantage in healthcare.

f) *Game AI*

RL agents use excessive on - game self - play, allowing them to overcome human - level performance on complex video games such as Dota 2, StarCraft, and Go. These grid worlds serve as valuable benchmarks within the training and test procedures. Teamwork of RL - powered virtual characters is applied even in the latest competitive evolution. The bots for the RL games game aim at beating humans; however, they also act as a stepping stone in creating intelligent teammates and competitors with fairness and spiced - up gameplay in video games.

g) Business Operations

RL drives the adoption of automation across the production line, revenue management, pricing, and supply chain [12]. The ML can find a dynamic way to schedule machines and route workflows to minimize bottlenecks and delays in manufacturing. With pricing, RL aims at having a good margin and meeting customer demand. In revenue management, RL minimizes overbooking and adjusts availability, targeting maximum profitability. For a complicated supply chain, the algorithm helps to model the network dynamics so as to minimize and handle uncertainties. They show how RL is more effective than the predefined syntax in solving complex tasks at the business level.

5. Impact

Reinforcement learning makes it possible for automated systems to make multiple decisions in sophisticated real - world situations unprompted [13]. Regression algorithms get trained on the data and build policies that maximize return on investment. This is leading to increases in automation, productivity, and optimized outcomes across many sectors: This is leading to increases in automation, productivity, and optimized outcomes across many sectors:

In robotics, the capability of RL includes the robots' learning of multiple skills, such as the ability to manipulate objects, motor skills, and navigation [11, 14]. It gives the system increased flexibility and allows it to save human efforts on every motion programming. Industrial robots schooled by RL can work with the system and thus flexibly optimize the production rate on a production line. In logistics, RL increases the warehousing storage density, optimizes delivery routing, and lowers the fleet operating cost relative to traditional optimization methods.

A reinforcement learning algorithm trains self - driving cars to have a smooth and safe driving experience even in complicated situations. This led to a learning - as - planning approach of complementing rule - based self - driving algorithms. In finance, RL creates trading strategies that are profitable and re - balances investment portfolios automatically to achieve the best returns. This is coupled with the fact that it is faster than traditional quantitative trading systems.

Other RL applications that are resulting in the growth of automation and system performance improvement include revenue management, supply chain optimization, manufacturing scheduling, precision medicine, and automatic game - playing.

The main merit of RL in these areas is its propensity to learn and improve its policies by gaining new experiences and adapting to changing environments. Machine - generated policies are designed to work efficiently under some static conditions, but they need to gain the common sense, dynamic conditions, and creativity of humans. The intelligent agents, being trained via RL, can react in a timely manner to increase the end outcomes such as profitability, cost minimization, safety, treatment effectiveness, and service to customers.

As RL is taking off and addressing the current issues of sample efficiency and sim - to - real transfer, its influence in automating complex decision-making tasks across different sectors is expected to grow aggressively. RL is a paradigm shift from devising in a human - like way to individual lifelong autonomous learning.

6. Scope

While RL has achieved breakthrough results in recent years, there remain the same research challenges to be addressed to increase real - world capabilities and create long - term impact. Key directions include:

a) Stability & Data - efficiency

Contemporary RL algorithms may be unstable or data inefficient when encountering noise and high - dimensional state/action spaces. Recent developments in distributional RL, new optimizers as well as network architectures are directed toward optimizing stability and reducing sample complexity. Besides learning from human demonstrations, task decomposition also presents their applications.

b) Safety

Learning RL over natural physical objects poses some safety risks during the training process and the deployment stage. Research into techniques such as adversarial RL, reward shaping, and safe interruption are the most important for real - world implementation. Also, such methods as formal verification are required.

c) Sim - 2 - Real Transfer

The transition of policies from the simulator to the complexities of the natural world still draws a difficult pass. The update of the simulator fidelity through the system identification and the physical/rendering development will help. Moreover, domain randomization and noise injection are proving to work successfully for sim - to - real transfer.

d) Multi - Agent RL

Most of the real - world problems require human - to - human collaborative or competitive interactions, and multi - agent RL provides an adequate model for such scenarios. Open problems such as coordination, generalization, and scaling up to large agent populations are among some of the difficult ones.

e) Hierarchical RL

Architectures based on hierarchy will give significant benefits to handling big problems but make them require new algorithms and frameworks. A promising direction here is the hierarchical integration of model - based RL.

There is more to be enhanced in the extensive and very sparse reward settings. RL's connection with neuroevolution and population - based technologies will tackle the challenges. With the challenges notwithstanding, the vast array of innovations in RL continues to point towards hype as its impact on practical, real - world applications should continue driving fast.

7. Conclusion

Reinforcement learning technology has come a long way from the initial years of laying theoretical foundations to now developing beneficial outcomes for complex decision - making problems in robotics, autonomy, healthcare, finance, and many other areas. RL facilitates a framework that can be employed to actuate automation of complex decision - making without the extensive coding of rules. As the function approximators (the deep neural networks), RL can cope with huge state and action spaces. The constant progress in research is projected to widen the applicability and impact of RL by working on the sample - efficiency, stability, safety, and sim - to - real transfer. Since the RL innovations continue to increase, its role as an optimizer in making sequential decisions to improve the real world without requiring resource - exhausting features and policy engineering will be more significant.

References

[1] Y. Li, "Deep Reinforcement Learning: An Overview, " *arXiv (Cornell University),* Jan.2017, doi: https: //doi. org/10.48550/arxiv.1701.07274.

[2] A. R. Mahmood, D. Korenkevych, G. Vasan, W. Ma, and J. Bergstra, "Benchmarking Reinforcement Learning Algorithms on Real - World Robots, " *proceedings. mlr. press*, Oct.23, 2018. https: //proceedings. mlr. press/v87/mahmood18a. html

[3] N. Sünderhauf *et al.,* "The limits and potentials of deep learning for robotics, " *The International Journal of Robotics Research*, vol.37, no.4–5, pp.405–420, Apr.2018, doi: https: //doi. org/10.1177/0278364918770733.

[4] R. Leo, R. S. Milton, and A Kaviya, "Multi agent reinforcement learning based distributed optimization of solar microgrid, " Dec.2014, doi: https: //doi. org/10.1109/iccic.2014.7238438.

[5] M. J. Hausknecht, "Cooperation and communication in multiagent deep reinforcement learning, " *repositories. lib. utexas. edu*, Dec.2016. Available: http: //hdl. handle. net/2152/45681

[6] F. P. Such, V. Madhavan, E. Conti, J. Lehman, K. O. Stanley, and J. Clune, "Deep Neuroevolution: Genetic Algorithms Are a Competitive Alternative for Training Deep Neural Networks for Reinforcement Learning, " *arXiv: 1712.06567 [cs]*, Apr.2018, Available: https: //arxiv. org/abs/1712.06567

[7] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey, " *The International Journal of Robotics Research*, vol.32, no.11, pp.1238–1274, Aug.2013, doi: https: //doi. org/10.1177/0278364913495721.

[8] S. Shalev - Shwartz, S. Shammah, and A. Shashua, "Safe, Multi - Agent, Reinforcement Learning for Autonomous Driving, " *arXiv: 1610.03295 [cs, stat]*, Oct.2016, Available: https: //arxiv. org/abs/1610.03295

[9] A. De Meyer, D. Cattrysse, J. Rasinmäki, and J. Van Orshoven, "Methods to optimise the design and management of biomass - for - bioenergy supply chains: A review, " *Renewable and Sustainable Energy Reviews*, vol.31, pp.657–670, Mar.2014, doi: https: //doi. org/10.1016/j. rser.2013.12.036.

[10] T. G. Fischer, "Reinforcement learning in financial markets - a survey, " *www.econstor. eu*, 2018. http: //hdl. handle. net/10419/183139

[11] "Using simulation - based optimization in the context of IT service management change process, " *Decision Support Systems*, vol.112, pp.35–47, Aug.2018, doi: https: //doi. org/10.1016/j. dss.2018.06.004.

[12] N. Syam and A. Sharma, "Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice, " *Industrial Marketing Management*, vol.69, no.0019 - 8501, pp.135–146, Feb.2018, doi: https: //doi. org/10.1016/j. indmarman.2017.12.019.

[13] Y. Acar, M. Backes, S. Fahl, D. Kim, M. L. Mazurek, and C. Stransky, "You Get Where You're Looking for: The Impact of Information Sources on Code Security, " *2016 IEEE Symposium on Security and Privacy (SP),* May 2016, doi: https: //doi. org/10.1109/sp.2016.25.

[14] P. Kormushev, S. Calinon, and D. Caldwell, "Reinforcement Learning in Robotics: Applications and Real - World Challenges, " *Robotics*, vol.2, no.3, pp.122–148, Jul.2013, doi: https: //doi. org/10.3390/robotics2030122.