The advancements and applications of deep reinforcement learning in Go

Xutao Zheng*

Computer Science and Technology, Harbin Institute of Technology (Shenzhen), 518000 Shenzhen, China

Abstract. Combining Deep Learning's perceptual skills with Reinforcement Learning's decision-making abilities, Deep Reinforcement Learning (DRL) represents a significant breakthrough in Artificial Intelligence (AI). This paper examines the evolution and uses of Deep Reinforcement Learning (DRL), emphasizing both the theoretical underpinnings and the noteworthy real-world applications—like AlphaGo's triumph over elite Go players—of the technology. DRL systems learn optimal policies through interactions with their environment, maximizing long-term cumulative rewards. DRL has achieved remarkable results in complex decision-making tasks through the combination of deep learning models like Convolutional Neural Networks (CNN) and reinforcement learning techniques. DRL's potential to transform AI applications is demonstrated by its success in a number of industries, including robotics, autonomous driving, and video games. AlphaGo's success, leveraging DRL and Monte Carlo Tree Search (MCTS), exemplifies the impact of this method on game theory and strategic decision-making. This paper aims to explore the key concepts of DRL, its historical evolution, and its future prospects in advanced AI research.

1 Introduction

By combining Deep Learning's perceptual ability and Reinforcement Learning's decision-making power, which is directly regulated based on the input image, Deep Reinforcement Learning is an artificial intelligence technique that approaches human cognitive processes more closely. The remarkable achievements in both theory and practice have been achieved by deep reinforcement learning since its debut. An unprecedented record of 4:1 victory over world champion Lee Sedol was achieved by AlphaGo, a Go computer game developed by Google's DeepMind team using deep reinforcement learning techniques, in March 2016 [1].

The discipline of reinforcement learning in machine learning is designed to explore how machines can learn from their interactions with the environment to achieve optimal long-term cumulative rewards from feedback provided by the environment. The environment in RL consists of a description of the state, a transfer of the state, and an immediate reward that is fed back to the machine when a state transfer occurs. The environment is perceived by the machine, which then uses its policy to decide what action to take in the present. After taking

* Corresponding author: 220110402@stu.hit.edu.cn

an action, the machine gets an immediate reward from the environment and logs the cumulative reward. The machine interacts with the environment many times, generating a sequence of "state-action-reward-state" data. The strategy function that generates actions for the state is taught to RL from such data, and it is expected that the strategy will maximize the long-term cumulative reward.

DeepMind, the artificial intelligence research division of Google, in recent years announced two remarkable research results: Atari video game-based deep reinforcement learning algorithms [2] and AlphaGo [3]. Traditional academic designs of human-like learning algorithms can be overpowered by using deep reinforcement learning algorithms that combine deep learning with perceptual abilities and reinforcement learning with decision-making capabilities. The performance of these algorithms has greatly exceeded expectations, much to the amazement of the academic and social worlds.

Deep reinforcement learning has been a popular topic in advanced artificial intelligence thanks to two studies published in Nature by the DeepMind team. The deep Q-network (DQT), which was a breakthrough in Atari video games, was discussed in January 2015 article [2]. With the game screen serving as input and the game score acting as a reinforcement signal, the deep Q-network simulates the movements of a human player in a game. In 49 video games, the converged algorithm outperformed expert human players, as determined by the researchers' testing results. In a January 2016 publication, the Deep Intelligence team further suggested a computerized Go first game number based on this [3]. The technique combines Monte Carlo tree search with deep reinforcement learning, significantly reducing the computational load during the search phase while increasing the precision of game estimation. The first game was played and won 5:0 against the European Go champion, Fan Hui. In the well-known game, AlphaGo faced off against Lee Sedol, the best professional 9-dan player in the world. in March of 2016, and AlphaGo came out on top 4:1. It also means that a new machine learning technique called deep reinforcement learning has been able to match human performance in very hard chess games.

As a result, a thorough understanding of deep reinforcement learning is crucial for the advancement of artificial intelligence and its use in many sectors. The purpose of this work is to present deep reinforcement learning methods and a summary of the history of computer Go development. The following constitutes the main framework of this paper: First, the paper gives a summary of the history of development and important methods in deep learning, reinforcement learning, and deep reinforcement. The benefits of the AlphaGo principle are emphasized, and then possible applications of deep reinforcement learning are examined and concluded. The AlphaGo principle's advantages are highlighted, followed by an exploration of the potential uses of deep reinforcement learning and a final conclusion.

2 Method

2.1 Reinforcement learning

Reinforcement learning is inspired by the efficiency with which species adapt to their environment. They do this by interacting with it through a process of trial and error and determining the optimum line of action by maximizing the cumulative reward. Four fundamental components are needed for a reinforcement learning system: state s, action a, reward signal r, and state transfer probability P. The policy $\pi : S \rightarrow A$ is the mapping between state and action space. With an advance probability of P to the next state s', an intelligent entity chooses an action in the current state s, executes it, and gets input from the environment in the form of a reward r. It does all of this while adhering to the policy π. By modifying the

policy, reinforcement learning seeks to maximize the cumulative reward. Value functions are widely employed in estimating the degree of merit, π, of a plan.

Research on RL has a long history. In 1992, Tesauro successfully used reinforcement learning to achieve master level in backgammon [4]; Sutton wrote the first systematic introduction to reinforcement learning [5]; Kearns proved for the first time that a reinforcement learning problem can be solved to an approximate optimal solution with a small amount of experience [6]. In 2006, Kocsis proposed the confidence upper bound tree algorithm, which revolutionized the field of chess reinforcement learning. Kearns had previously shown that a reinforcement learning problem can be solved to an approximate optimal solution with minimal experience [7]. In 1992, Tesauro employed reinforcement learning to advance to the master level in backgammon. AlphaGo is said to have been created by Kocsis, who in 2006 [7] revolutionized the use of reinforcement learning in Go. In 2015, Littman examined reinforcement learning in Nature [8]. The most popular reinforcement learning approaches these days include policy gradient, adaptive dynamic programming, Q-learning, SARSA learning, TD learning, Monte Carlo, and Q-learning.

2.2 Deep learning

Artificial neural networks were the inspiration for DL. The complex layered network structure of the cerebral cortex was replicated by researchers in the 1990s to perform data processing and inference. Additionally, they suggested the back-propagation algorithm for multilayer neural network optimization. However, neural network research has not advanced much because of the problem of gradient dispersion and the lack of hardware resources. Hinton (2006) suggested automatically extracting the hierarchical feature representation of the original data in order to create a sophisticated function mapping relationship between input and output data. Hinton discovered a basic principle for training deep neural networks in [9], which is to rigorously pre-train the neural network's intermediate levels, layer by layer, using unsupervised methods before fine-tuning the network as a whole using supervised strategies. Better initial parameters for deep neural networks are provided by the pre-training approach, which also lessens the difficulty of deep neural network optimization. In the first few years, the development of deep learning focused on pre-training, and various methods were proposed. In the years following 2010, deep learning has achieved notable strides in artificial intelligence, encompassing speech recognition, visual object recognition and detection, etc., thanks to the advancement of computational resources and pre-training approaches [10–12]. In 2012, Microsoft researchers developed the first deep neural network-hidden Markov hybrid model, which they successfully applied to a large-vocabulary speech recognition system. When this was compared to the traditional Gaussian-hidden Markov model, the speech recognition error rate decreased by about 30% [13]. Krizhevsky's initial use of deep convolutional neural networks (CNNs) on ImageNet resulted in an enormous improvement over previous methods, with the error rate of image recognition being decreased to 37.5% [14]. The 'Google Brain' project, headed by Andrew Ng, utilized an unsupervised learning technique [15] to gain abstract concepts from YouTube videos like 'Google Cat'. Long short-term memory (LSTM) combined with recurrent neural network (RNN) is more efficient than regular recurrent neural network in voice processing, as proved by Graves in 2013 [16]. Deep learning has advanced significantly since 2014, leading to the development of several models, such as attention [17], RNN-CNN [18], and deep residual networks [19]. LeCun summarized the fundamental ideas and key benefits of deep learning in a review that was published in Nature in 2015 [20]. Deep belief networks (DBN), recurrent neural networks, stacked auto-encoders (SAE), and convolutional neural networks are some of the most popular deep learning models available today.

2.3 Deep reinforcement learning

Intelligence is measured by perception and decision-making skills in advanced artificial intelligence studies. The long-term challenge of reward learning is to directly guide intelligence through the acquisition of high-dimensional perceptual inputs (speech, pictures, etc.). Significant breakthroughs have been gained in the theory and methodology of policy selection using reinforcement learning. Most successful reinforcement learning systems involve artificial feature selection, albeit the outcome is heavily influenced by how well this process works [21]. The ability to directly extract high-level features from raw data has been made possible by recent advances in deep learning. Deep learning has exceptional perceptual skills but lacks the ability to make decisions, while reinforcement learning can make decisions but cannot handle problems with perception. The perceptual decision-making issue in complex systems can be solved by integrating the two due to their complementary benefits.

Many scholars are now collaborating with reinforcement learning to address perceptual decision-making tasks in image data due to the inherent advantages of convolutional neural networks in image processing. Convolutional neural networks and Q-learning are combined in the deep Q network (DQN), which was introduced in [22] and connected with experience playback technology [22]. Experience playback lowers the correlation between data and samples past data more frequently, increasing the efficiency of data consumption.

The problems faced by deep reinforcement learning are often very time-dependent, while recurrent neural networks are suitable for problems connected to time series. Cuccu suggested that recurrent neural networks be trained with pre-compressors as a means of utilizing the neural evolution approach for vision-based reinforcement learning. In the vision-based mountain car climbing task, good control results are obtained by dimensionalizing the captured image data using a recurrent neural network and feeding it into reinforcement learning for decision making [23]. Narasimhan proposed a deep network architecture that combines long and short-term memory networks with reinforcement learning to process text games. Semantic information about the game state is obtained by mapping textual data into a vector representation space [24].

3 Typical applications of deep reinforcement learning—AlphaGo

Computerized go, which was introduced in the 1960s, has been recognized as a challenge in the artificial intelligence field and is a valuable testing environment for clever learning algorithms. Computerized Go analyzes and selects move positions by calculating an optimality function on a search tree that comprises a series of one move scenarios, where b is the search's breadth and d is its depth. Unlike chess, which has a limited search space, the computational complexity of Go is about. If the traditional brute force search method is used, it is far from being able to solve the Go problem according to the current computational power [25]. Through the use of fuzzy matching and expert systems, early computers Go decreased the search space and computational intensity. However, because of hardware and computational resource limitations, the real result was not optimal. The year 2006 saw the introduction of Monte Carlo tree search in computer Go, ushering in a new era. The Monte Carlo tree-based optimization search is the primary strategy used in computer go nowadays.

In an inventive move, AlphaGo combines Monte Carlo tree search with deep reinforcement learning. It does this by evaluating the position using a value network to minimize search depth and narrowing the search by using a policy network to increase search efficiency and more precisely estimate the win rate. AlphaGo is now on par with the best players.

One cannot divorce deep neural networks from AlphaGo's success. Much like learning a Go game by heart, traditional rule-based computational Go approaches are limited to

recognizing predefined plays. The capacity to learn the game is much improved by the deep learning-based Hatsune Mikado, which automatically extracts and efficiently integrates the elements of the game positions.Second, Hatsune Mikoto's success is also dependent on the role assessment. Since they are better able to handle the uncertainty of the opponent's next move, the complementarity between the value network and the quick move network in position evaluation is crucial to obtaining more accurate evaluation findings. Furthermore, a significant improvement in hardware configuration has been made. Using asynchronous multi-threaded search, AlphaGo uses the CPU to simulate and the GPU to compute value networks and strategies. While the distributed version of AlphaGo needed 1202 CPUs and 176 GPUs, the final standalone version used 48 CPUs and 8 GPUs. AlphaGo's high performance is a result of this computer hardware's support [3].

Artificial intelligence professionals are familiar with the game of Go due to its intricate move selection and expansive search space. Using a strategy network and a value network based on deep convolutional neural networks, AlphaGo narrows the search space. It then creatively blends the Monte Carlo tree search technique and mixes supervised and reinforcement learning during training. An important development in the realm of artificial intelligence is AlphaGo.

4 Conclusion

This paper provides a comprehensive overview related to Deep Reinforcement Learning and AlphaGo. Deep reinforcement learning, which blends deep learning and reinforcement learning, is a powerful artificial intelligence technique. Apart from deep Q-networks for video game play and AlphaGo for Go, deep reinforcement learning can also be very useful in other domains like robotics and autonomous driving. DRL is still in its early stages as an algorithm for advanced artificial intelligence. It is evident from the launch of Deep Reinforcement Learning and AlphaGo that artificial intelligence is nothing more than algorithmic software that was meticulously created by humans. A huge amount of data samples, computer gear, and human intelligence are necessary for its success. More comprehensive and in-depth basic research on state-of-the-art AI theories and algorithms is desperately needed, as is applied research that integrates hardware and software for a range of applications or jobs.

References

1. Theguardian. AlphaGo seals 4-1 victory over Go grandmaster Lee Sedol, The Guardian, 2016. Available: https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol
2. V. Mnih, K. Kavukcuoglu, D. Silver, et al., Human-level control through deep reinforcement learning. Nature, 518(7540), 529–533 (2015).
3. D. Silver, A. Huang, C. Maddison, et al., Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484–489 (2016).
4. G. Tesauro, TD-Gammon, a self-teaching backgammon program, achieves master-level play. Neural Computation, 6(2), 215–219 (1994).
5. R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction. Cambridge MA: MIT Press (1998).
6. M. Kearns, S. Singh, Near-optimal reinforcement learning in polynomial time. Machine Learning, 49(2/3), 209–232 (2002).

7. L. Kocsis, C. Szepesvari, Bandit based Monte-Carlo planning. In Proceedings of the European Conference on Machine Learning, Springer, Berlin, 282–293 (2006).
8. M. L. Littman, Reinforcement learning improves behaviour from evaluative feedback. Nature, 521(7553), 445–451 (2015).
9. G. E. Hinton, S. Osindero, Y. W. Teh, A fast learning algorithm for deep belief nets. Neural Computation, 18(7), 1527–1554 (2006).
10. O. Abdel-Hamid, A. Mohamed, H. Jiang, et al., Convolutional neural networks for speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(10), 1533–1545 (2014).
11. B. A. Carlson, M. A. Clements, A projection-based likelihood measure for speech recognition in noise. IEEE Transactions on Speech and Audio Processing, 2(1), 97–102 (1994).
12. W. Ouyang, X. Zeng, X. Wang, Learning mutual visibility relationship for pedestrian detection with a deep model. International Journal of Computer Vision, 2016, DOI: 10.1007/s11263-016-0890-9.
13. G. E. Dahl, D. Yu, L. Deng, et al., Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Transactions on Audio, Speech, and Language Processing, 20(1), 30–42 (2012).
14. A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems, Lake Tahoe: MIT Press, 1097–1105 (2012).
15. Q. V. Le, Building high-level features using large scale unsupervised learning. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver: IEEE, 8595–8598 (2013).
16. A. Graves, A. Mohamed, G. Hinton, Speech recognition with deep recurrent neural networks. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver: IEEE, 6645–6649 (2013).
17. K. Xu, J. Ba, R. Kiros, et al., Show, attend and tell: neural image caption generation with visual attention. In Proceedings of the 32nd International Conference on Machine Learning, Lille: ACM, 2048–2057 (2015).
18. P. Pinheiro, R. Collobert, Recurrent convolutional neural networks for scene labeling. In Proceedings of the 31st International Conference on Machine Learning, Beijing: ACM, 82–90 (2014).
19. K. M. He, X. Zhang, S. Ren, et al., Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas: IEEE (2016).
20. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature, 521(7553), 436–444 (2015).
21. V. Mnih, K. Kavukcuoglu, D. Silver, et al., Playing Atari with deep reinforcement learning. In Proceedings of the NIPS Workshop on Deep Learning, Lake Tahoe: MIT Press (2013).
22. L. J. Lin, Reinforcement learning for robots using neural networks. Pittsburgh: Carnegie Mellon University (1993).
23. G. Cuccu, M. Luciw, J. Schmidhuber, et al., Intrinsically motivated neuroevolution for vision-based reinforcement learning. In Proceedings of the IEEE International Conference on Development and Learning, Trondheim: IEEE, vol. 2, 1–7 (2011).

24. K. Narasimhan, T. Kulkarni, R. Barzilay, Language understanding for text-based games using deep reinforcement learning. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Lisbon: ACL (2015).
25. X. Cai, D. C. Wunsch II, Computer Go: a grand challenge to AI. In Challenges for Computational Intelligence, Berlin Heidelberg: Springer, 443–465 (2007).