Introduction to Reinforcement learning and its application in anomaly detection

Group 14

Richard Li, Alex Cheung

Simon Fraser University CMPT318

2020/11/29

Introduction

In March 2016, Alpha Dog, developed by DeepMind which is a subsidiary of Google, successfully defeated the world Go champion Lee Se-dol by a score of three to one, and again defeated the world Go champion Ke Jie in May of the following year. So far, the view that Alpha Dog's chess power has surpassed the top level of human professional Go which has become an undeniable fact. It makes people feel the power of artificial intelligence (AI). The Alpha Dog is so magical thanks to working under a principle called "reinforcement learning". Reinforcement learning seems to be very far away from daily life, but it always protects people's property and information security in the era of highly developed information. The Internet has now covered the whole world, greatly increasing the scale of system coverage, and at the same time it has brought about a more dangerous and fragile network security environment than ever before. The complexity and uncertainty of network attacks make the protection mechanism have to be responsive, scalable and self-adaptive. Statistical research shows that 62% of attacks are recognized after they cause major damage to the network system. Therefore, with the continuous acceleration of the information age and the continuous escalation of attack methods, how can the surveillance and defence system grow with the times? This is inseparable from the help of reinforcement learning. This article focuses on the concept of reinforcement learning, through some simple examples to tell what reinforcement learning is composed of, what are the uses of the components, and how they work together. Moreover, will combine multiple examples to show the importance of reinforcement learning as an online machine learning method for network intrusion detection and what specific help it has for the future well-being of mankind.

Why is reinforcement learning needed?

Before specifically explaining reinforcement learning, two related terminologies are needed to be introduced, supervised learning and unsupervised learning. And show differences compared with reinforcement learning. First of all, these three words are the most important branch of the field called machine learning(ML). Supervised learning has a label, which indicates to the algorithm about what kind of input should be for what kind of output, in the opposite the unsupervised learning does not have a label. It can be seen that the overall supervised learning or the relatively rigid input-output corresponding model cannot provide the flexibility which is necessarily needed to deal with various scenarios. However, the main difference between supervised learning and reinforcement learning is whether the feedback received is evaluative or instructive. Instructive feedback is to provide methods to achieve goals, and supervised learning is to solve problems based on instructive feedback. However, reinforcement learning uses evaluative feedback, which provides information based on how far the goal will be achieved. For example, for a system that controls oxygen content, the operator cannot tell the algorithm what the correct setting of each part is at any given time. Such instructive feedback does not seem to have any effect. However, the staff can easily obtain the various instrument data generated in a specific period. This kind of evaluative feedback idea is more intuitive and easier to realize. To sum up, reinforcement learning is a flexible learning method in machine learning. Due to the four factors behind it, which can cope with more complex and changeable environments.

What is reinforcement learning?

Reinforcement learning is considered to be the closest form of human learning because it can learn based on its own experience by exploring unknown environments. It includes four main factors: agent, status, reward, and action. To facilitate understanding, take the robot walking a maze as a scene to illustrate what agent, status and reward factors represent. An agent is a hypothetical entity, usually an object trained in a specific environment to make correct decisions. In the example,

the agent can be understood as a robot, what it has to do is try to out of the maze without any collision. The state defines the current real-time state information of the Agent, such as the robot's position in the maze, movement posture, current movement speed and body distortion angle. The state information of the agent entirely depends on the method of solving the problem. When an Agent performs a specific action or task, it will receive a real-time feedback called a reward, which will be treated as a scalar. Based on the execution of the behaviour in the current environment, the reward can be divided into positive rewards and negative rewards. For example, as the absolute distance of the robot from the exit is greater, the score will drop, (Bad reward) vice versa. At this point, having the four elements of the agent, state, action, and reward can form a simple Markov decision process, which is also the main theory behind reinforcement learning.

The fundamental principles of the Markov Decision Process

Since many previous states need to be taken into account in the transition of the real environment, which is so complicated and it is difficult to model. This can be simplified by assuming that the state is converted to Markov property (Markov property refers to the probability of transition to the next state Only related to the current state, not related to any previous state). The dynamic process of MDP is as follows. The initial state of an agent is S_0, and an action a_0 is selected from the actions for execution. After execution, it randomly transfers to the next state S_1 according to the state transition probability PS_0, Pa_0. The following figure shows one of the processes sketch.

$$\begin{matrix} & a_0 & & a_1 & & a_2 & \\ S_0 & \text{-----}> & S_1 & \text{-----}> & S_2 & \text{-----}> & S_3 \ \ldots \end{matrix}$$

Figure 1.1 MDP process

After each decision-making will receive the corresponding reward, and the goal of reinforcement learning is how to gradually form an expectation of the stimulus under the incentives or punishments given by the environment, to find the optimal strategy to maximize the long-term future reward. The following two pictures are the goal of reinforcement learning by David using the maze as an example. (As just mentioned, when the absolute distance of the robot from the exit is greater, the score will be -1).

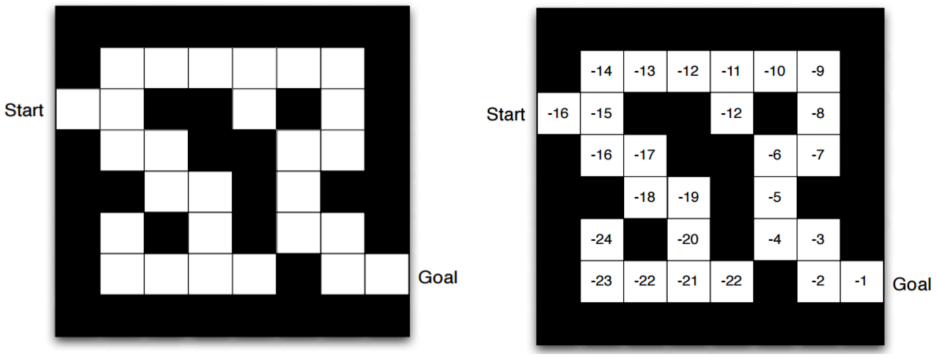

Figure 1.2 Example of RL (Value Iteration)

By illustrating some basic examples, we already know the basic principles of reinforcement learning. Although these jobs can be done by humans even faster than computers can run. So what difficult tasks can be done by reinforcement learning instead of humans is a question. In other words, what benefits reinforcement learning can bring to human well-being needs to be explored.

Reinforcement application in Real Life

Reinforcement learning can also be used in the self-driving car industry. Some of the autonomous driving tasks, such as trajectory optimization, motion planning, dynamic pathing, controller optimization, and scenario-based learning policies for highways, can all be achieved by installing cameras on cars and applying reinforcement learning. Based on the camera captures of states of cars speed, driving lanes, and location, the system makes actions and gets corresponding rewards for tasks. As more reinforcement learning is processed, autopilot systems will be better able to use faster than the human brain to cope with more complex road conditions. This can improve the efficiency, intelligence and safety of self-driving cars. The more varieties of qualified disabled people can drive cars as normal people do. Reinforcement learning can also be applied in automated medical diagnosis, chronic disease and critical care systems. The system will evaluate the patient's states of diseases and give a stable and reasonable solution for the specific states of patients. The use of RL in healthcare enables the improvement of long-term outcomes by factoring in the delayed effects of treatments. For example, [1]KenSci uses reinforcement learning to predict patients' dynamic changes and help practitioners to find treatments at patients' early stages.

Reinforcement learning applied in anomaly detection

As the volume of data surges rapidly when the network size is enlarged. Using manual network data anomaly detection is impossible. Reinforcement learning plays a huge role in finding anomalies for network intrusion detection in the network traffic flow. A network intrusion detection system is a software or hardware platform installed on network equipment to detect and report to the administrator abnormal or malicious activities by analyzing the audit data. The architecture can be implemented by setting network flow parameters as the state variables, such as the port number, packet size and transmission rate. The action can either flag or not flag an anomaly detection warning based on Q(s, a) values given by the output of the neural network in the previous state and an ε-greedy manner. The following image is the network intrusion detection architecture.

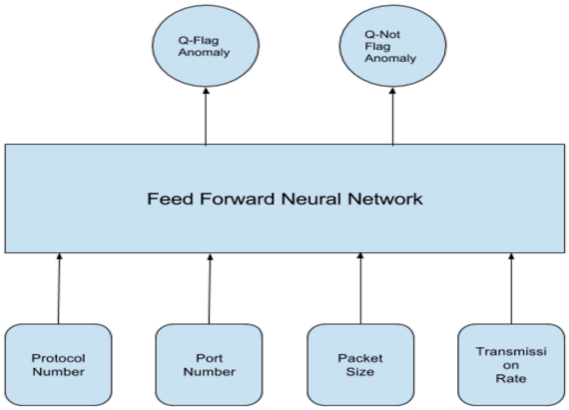

Figure 1.3 Network Intrusion Detection Architecture.

The number of reinforcement learning processes for network intrusion detection could be generalized into 4 parts of the back propagation algorithm. [2] The first is to initialize all the weights in the deep neural network with random values and Initialize the total accumulated reward to zero. And then to get an initial state from the environment. A third part is a finite number of recursive steps which contains starting with the state obtained in the previous step; performing a feed-forward of the current state using deep neural network, and getting the predicted Q(s, a) values; Taking action of either flag or not flag from the current state, based on the Q(s, a) values given by the output of the deep neural network in the previous state and an ε-greedy manner; Getting the reward and moving to state; Passing the new state into the deep neural network and calculate the Q(s, a) values by Bellman's equation; Performing a

training of the deep neural network by back propagation of the error of prediction, where the difference between target Q(s, a) and predicted Q(s, a) is taken as the error of prediction. The last part is computing the new cumulative total reward. Through this advanced and comprehensive network anomaly detection reinforcement learning process, hackers can be identified, such as Denial-of-Service (DoS) attack, Remote-to-Local (R2L) attack, User-to-Root (U2R) attack and the Probing attack. This can prevent people from being stolen and being accessed private data unauthorized which guarantees to access the network under a private and trustworthy environment for everyone.

Conclusions

In summary, reinforcement learning is a concept of learning while obtaining examples based on Markov Decision. The learning system is formed through agent, reward, state and action to find a way to maximize future returns. Reinforcement learning technology has been widely used in many fields, such as telecommunications, manufacturing, power management, healthcare, government, and even entertainment. With the advent of the technological age, everything is closely related to the Internet. As Thanh Thi Nguyen and Vijay Janapa Reddi said, "The scale of Internet-connected systems has increased considerably, and these systems are being exposed to cyber-attacks more than ever." (Nguyen & Reddi, 2019, p. 1). For network security, traditional supervised learning (learning through data labels) is far from sufficient to resist network attacks. The extremely fast adaptability of reinforcement learning can perfectly cope with the complex and changeable characteristics of network systems. Even so, the speed of network update is amazing, and the variety of attack methods and the degree of concealment have also increased. Developers must continuously improve the functionality and anomaly awareness of reinforcement learning to deal with various threats.

References

AlphaGo. (n.d.). Retrieved December 1, 2020, from Deepmind.com website: https://deepmind.com/research/case-studies/alphago-the-story-so-far

Great Learning Team. (2020, February 17). Use of Reinforcement Learning (RL) in healthcare. Retrieved December 1, 2020, from Mygreatlearning.com website: https://www.mygreatlearning.com/blog/reinforcement-learning-in-healthcare/

Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., Sallab, A. A. A., Yogamani, S., & Pérez, P. (2020). Deep reinforcement learning for autonomous driving: A survey. Retrieved from http://arxiv.org/abs/2002.00444

Koduvely, D. H. (2018, January 19). Anomaly detection through reinforcement learning. Retrieved November 30, 2020, from Zighra.com website: https://zighra.com/blogs/anomaly-detection-through-reinforcement-learning/

Nguyen, T. T., & Reddi, V. J. (2019). Deep reinforcement learning for cybersecurity. Retrieved from http://arxiv.org/abs/1906.05799

Rungta, K. (2020, January 1). Reinforcement Learning: What is, Algorithms, Applications, Example. Retrieved December 1, 2020, from Guru99.com website: https://www.guru99.com/reinforcement-learning-tutorial.html

Sharma, A., Kalbarczyk, Z., Barlow, J., & Iyer, R. (2011). Analysis of security data from a large computing organization. 2011 IEEE/IFIP 41st International Conference on Dependable Systems & Networks (DSN). IEEE.

Vincent, J. (2019, November 27). Former Go champion beaten by DeepMind retires after declaring AI invincible. Retrieved December 1, 2020, from The Verge website: https://www.theverge.com/2019/11/27/20985260/ai-go-alphago-lee-se-dol-retired-deepmind-defeat