0 Introduction

Deep Learning became popular since 2012, It has been in the computer vision, voice recognition, translation, games and Go and other areas have made revolutionary breakthroughs, especially AlphaGo, so that people around the world know the great power of deep learning The However, we all know that the success of deep learning depends entirely on massive amounts of data and powerful computing resources. In the face of a new task, we have to re-train again, very time-consuming and laborious. For AlphaGo, many people have raised the question: Will the checkerboard be bigger? AlphaGo can do it? The current method is clearly not, AlphaGo will immediately become a fool. And humans have much more, every minute can adapt to the new board. At present, artificial intelligence does not have human fast learning ability. And then give an example of face recognition, we can often only one side can remember and identify, and now the depth of learning but need tens of thousands of pictures to do. Therefore, how to make Artificial Intelligence to have the ability to quickly learn to become a cutting-edge research problem.

What can AI learn fast?

1 Quick Learning is very important for AI

AI's ability to learn quickly will be a revolution in the revolution in artificial intelligence. Why do you say that? Deep learning is a revolution in artificial intelligence, but because of the depth of learning is currently unable to quickly learn, making the application of deep learning by a lot of limitations. We often encounter in the real life of many new tasks, and now the depth of learning because they can’t quickly adapt to new tasks, there is no way to replace the human work. Take robots as an example, we hope that robots will one day be able to walk into the tens of thousands of households. But everyone is not the same as the use of robots, each family's environment is not the same, how can the robot to adapt to a variety of requirements? If you can’t, the robot can’t be popular.

Therefore, to let the robot into the tens of thousands of households, we need the robot to real-time learning, continuous learning, quickly learning, even in the face of a new similar task, but also quickly grasp. Such a robot will be very powerful, can really handle a variety of tasks!

3 Meta Learning can make AI achieve Quick Learning

Think about why human can learn fast? Because we can use the past experience to learn! Really is very simple truth.

Why is the depth of learning now can’t learn quickly? Because we do not understand the depth of learning to use the experience of the past!

In most cases we can only start from scratch. Using Finetune to learn a new task is often ineffective. So, to let the depth of learning fast learning, we must study how to make the neural network can make good use of the previous knowledge, making the neural network can be adjusted according to the new task. Meta Learning, one of the quick ways to learn!

Meta Learning, in fact, also known as Learning to Learn. What is learning to learn? Is the ability to have learning.

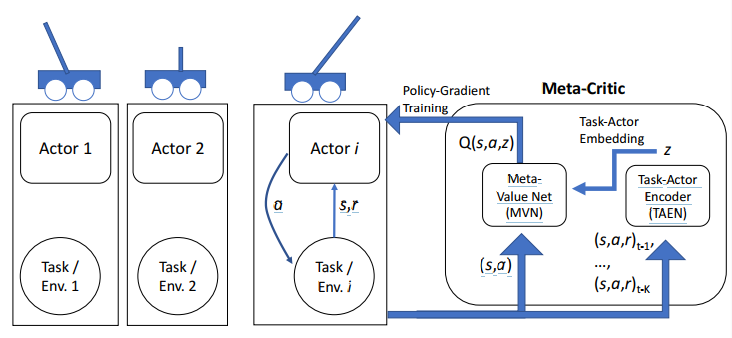
You can learn more about Meta Learning in this link: <https://arxiv.org/abs/1706.09529>.

4. What is Meta Learning?

We are based on value-driven animals. What we don’t do is because there are a steelyard brain thinking about which is more important. Even if sometimes very emotional, but also because the emotional time to do that thing the value of the maximized. Does we use this values to drive AI fast learning?

The answer, of course, can be, and what this paper is doing. The method is simple:

Let AI learn a variety of tasks after the formation of a core value of the network, so the face of new tasks, you can use the existing core value network to accelerate AI learning speed!

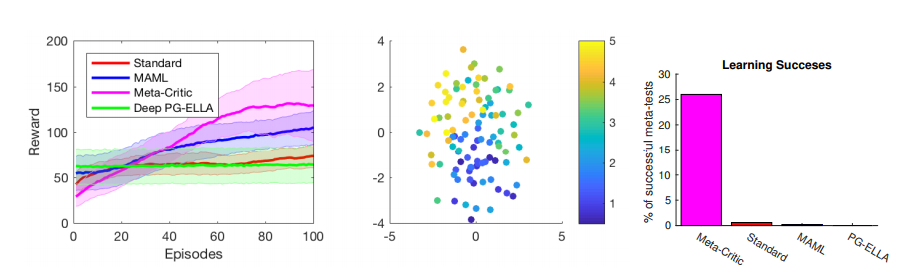


*The picture above shows the basic diagram of the Meta-Critic Network. We do this with CartPole to keep the balance of the task to do the analysis. In our case, the length of the rod is arbitrary, we hope that the AI in the study of the length of the rod after the task, the face of a new length of the bar, to quickly learn to master the balance of the handle to keep the knot.*

How to do it?

Each training task we construct a Actor Network, but we only have Meta-Critic Network, the network consists of two parts: one is the core value of the Meta Value Network, the other is Task-Actor Encoder. We train this Meta Critic Network with multiple tasks at the same time. The training method can be a common Actor-Critic. Training is the most critical Task-Actor Encoder, we enter the historical experience of the task (including state, action, and reward), and then get a task that information z, z and general value of the network input (state and action), and enter it into the Meta Value Network.

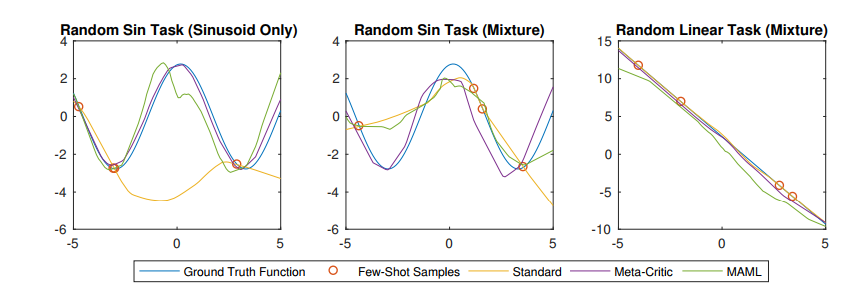
In this way, we can train a Meta Critic Network. In the face of the new task (that is, the length of the pole changed), we create a new mobile network Actor Network, but keep the Meta Critic Network unchanged, and then use the same Actor-Critic method of training. This time, the effect came out, we can learn very fast:



*Look at the first graph of the purple learning curve, the reward is very fast, standard is complete Actor-Critic training, basically still flat (usually for CartPole task need to train thousands of times to converge to the 195 score through the task) The Then look at the third graph on the right, after only 100 bar training, Meta-Critic method can achieve 25% through the success rate of the task, and other methods are still early yet. In fact, paper did not show a result is based on Meta Critic Network training 300 steps can make the task through the basic rate of 100%. This result is very promising!*

So what do we care about Task-Actor Encoder? So we extracted the different tasks z with t-SNE display as shown in the middle of the figure. Then we were surprised to find that the distribution of z is directly related to the length of the CartPole bar, which means that the task behavior encoder can actually use the previous experience to understand the configuration information of a task.

In addition to applying the Meta-Critic Network to the enhanced learning area, we can also apply it to supervise learning. The specific method here is not analyzed, we look at the results:



We use the basic function to fit the ability to see fast learning. The above figure is the result of training with only four samples. We get two tasks: the first is to fit the sin function, and the second to fit sin or linear function. We can see that the second task is very different. The first figure on the left is the first task on the left, we can see the use of Meta-Critic fitting effect is very good, and our general supervision and training (yellow line that) did not fit the basic success! MAML is currently a new study of Meta-Learning, but the effect is different from ours. Then look at the second task, the difficulty becomes larger, we can see the second and third graph, Meta-Critic for sin and linear function are doing well, but MAML effect is poor. MAML's idea is to build a good initial network and then finetune, obviously it is difficult to adapt to different types of tasks, and Meta-Critic due to the existence of Task Behavior Encoder (Task-Actor Encoder), can face a variety of different types of tasks。

6 Summary

Meta-Critic Network, as a new Meta Learning method, has a great potential by training out a core network (that is, core values) that can guide the rapid learning of new tasks. In the future work, we will use Meta-Critic Network to more complex tasks, to achieve better application!