2022 International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM) | 978-1-6654-9858-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/MLCCIM55934.2022.00079

A Counting Method based on Deep Reinforcement Learning Combined with Generative Adversarial Network

1st Zhoubao Sun
School of Engineering Audit,
Nanjing Audit University,
Nanjing, China
270247@nau.edu.cn

2nd Yi Zhu*
School of Information Engineering
Nanjing Audit University,
Nanjing, China
449091513@qq.com

***Abstract*—In the process of function optimization and scheduling of multi-agent system model, the precise step calculation has attracted more and more attention. How to accurately calculate the number of steps has become an important research area, and it is also an important prerequisite for multi-agent cooperation in navigation and collision avoidance. This function can also be extended to intelligent wearable devices. The main drawbacks of the current popular methods are that they have extremely high requirements for noise data filtering, and the parameters in the model are not updated in the calculation process. To solve the above problems, this paper proposes a step counting algorithm of deep reinforcement learning combined with generative adversarial network. The critical contains a generator and a discriminator. The generator is used to learn the value function distribution, the generator and discriminator are learned at the same time with the help of confrontation training, and the generated state action values are used to train the policy network. The algorithm is applied to the step counting, which can further improve the accuracy of the step counting while accelerating the learning rate of the original gradient algorithm.**

Keywords-multi-agent system, deep reinforcement learning, generative adversarial network

I. INTRODUCTION

With the development of Internet of Things technology, some wearable devices and monitoring bracelets have been born, which can better display people's health data, mainly based on people's gait characteristics [1,2]. One is common applications that are simple and easy to install acceleration sensors based on gait recognition system.

For the optimization, updating and self-learning of parameters in the counting process, many methods currently use the method based on deep reinforcement learning [3] to filter noise data, and use return values and incentives to achieve the ultimate goal. However, the biggest drawback is that the updating strategy depends on the current latest return values, which is easy to lead to slow learning speed and excessive variance, reducing the efficiency and effect of the algorithm, which is unfavorable to counting..

As a generation model, generative adaptive net can model the distribution of samples without specifying the specific parameter representation of the distribution, and directly output new sampling samples [4]. The model includes a generator(G) and a discriminator(D). G models the potential distribution of data to generate data; D distinguishes between data from the real world and data from the generator. During the training, the G and the D are trained by opposing each other, and finally a better generation effect is achieved. Therefore, this paper proposes a counting algorithm for deep reinforcement learning combined with GAN(GAN-DL). The network is trained by the generator and discriminator of GAN, and the algorithm is applied to the step counting, which can accelerate the learning rate of the original strategy gradient algorithm and further improve the accuracy of counting.

The work arrangement of this article is as follows. The second part mainly introduces the related work. The third is the core of this article, mainly introducing the model and process of the algorithm. The fourth illustrates the efficiency and effect of the algorithm on the data set, and finally summarizes.

II. RELATED WORK

Many traditional methods have been put forward one after another, such as threshold methods and wave peak detection, the biggest disadvantage of these models is the low efficiency of the computing process due to the use of embedded systems. With the development of intelligent computing software and hardware, the computing capacity and speed have been greatly improved, and many new counting models have been designed.

Xiao [5] found that the interference signal received by the equipment was not completely random during the step counting process. He found out the rule and designed a new encoding method for counting to improve the accuracy; In the reference [6,7], the researchers used the design waveform method to count. According to the phase change law of the received signal, whether it is swinging up and down or other ways, they used this type of feature to design the counting model to improve the counting accuracy of the traditional counter. Of course, the main disadvantage is that the time complexity is too high, there may be uncertainty in the actual application, and the accuracy is prone to jitter changes, Especially in the case of long time movement, the accuracy is easier to reduce. As human beings have higher and higher requirements for intelligent wearable devices, many positive and effective devices have also been produced. In [8], researchers trained and learned by classifying the acceleration data of mobile phones into rhythmic and non rhythmic ones to design a step counting model that is more consistent with the human body. Not all received signals will be considered as a step, and only when they meet the requirements of length, strength and shape of characteristics will they be considered as counting data. In [9], the researchers collected data by calculating the number of zeros during data collection and transmission. Yang et al. [10], in order to improve the accuracy, only extract amplitude features, amplitude difference features, and energy ratio that meet certain conditions when collecting data, and use classification methods to train and distinguish. These methods generally filter the noise

978-1-6654-9858-6/22/$31.00 ©2022 IEEE
DOI 10.1109/MLCCIM55934.2022.00079

generated by random motion in a certain way to calculate the step size, so as to improve the counting effect.

III. THE FRAMEWORK OF DEEP REINFORCEMENT LEARNING COMBINED WITH GENERATIVE ADVERSARIAL NETWORK

GAN is an emerging technology that can handle unsupervised and semi supervised learning. It contains two competitive roles. A Generator G is to distinguish the probability distribution of real data, while D is to identify the difference between real samples and generated samples. G is to maximize the result, while D is to minimize the error. The opposite operation is used for the final effect. The relationship between G and D can be formalized as following:

$$\text{Max}[E[\log D(x)]]+E[\log(1-D(G(z)))] \quad (1)$$

The Generator G can generate enough learning samples when GAN reaches the Nash equilibrium, which means that G can't really distinguish between true samples and generated data samples, so as to really learn the same data. The current research mainly takes GAN as a method to solve data imbalance through oversampling technology caused by real data. They first train the GAN on a specific data set and then use G to generate simulated sample data. Finally, the original real samples are mixed with the generated samples. Although this kind of method has made great progress in counting, it is more complex and discontinuous because it is processed in two segments. Moreover, the learned G cannot be used in subsequent calculations.

For the convenience of training, the loss function is as follows:

$$\text{Loss}_D=-E[\log D(x)]-E[\log (1-D(G(z)))] \quad (2)$$

Where formula (2) is the loss of the discriminator and the generator. In the training process, one network is temporarily fixed each time, and then another network is trained. D network and G network are trained alternately. Combining the GAN with the actor critical algorithm, the generated countermeasure network is used to train the critical network, and then the estimated value of the critical network is used to train the actor network parameters, which is finally applied to step counting. The network structure is shown in Figure 1.

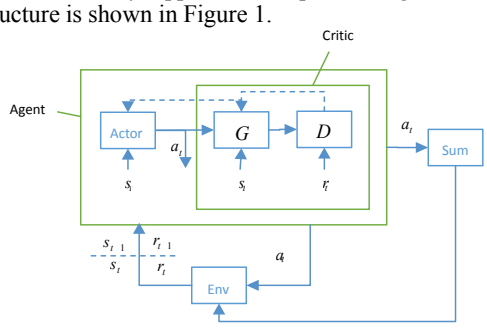

Figure 1. The network structure

At each time t, the agent obtains observations from the environment as input, the actor selects actions according to the acquisition, and outputs them to the environment and critical. After receiving them, the environment executes the actions and transfers them to the next state, and inputs the rewards to the agent; Within critical, critical inputs the observation and reward of the environment, including generator g input, sampling of output value distribution, discriminator D input and the results generated by the generator, and identifying whether the data comes from the environment or from the generator. The specific flow of the step counting algorithm combining the generation of countermeasure network and actor critical is shown in algorithm 1.

Algorithm 1: a step counting algorithm based on GAN

Randomly initialize the parameters of the critical network and the actor network, set the critical update factor *tr*, set the D network update factor *k*, and the total number of rounds *M*

1. for round $i=1,...M$, do
2. get initialization status
3. for time step $t=1,...T$, do
4. actor input status s_t, act according to output a_t
5. input a_i to summation unit for accumulation
6. environment execution action a_t, return to reward and new status s_{t+1}
7. end for
8. obtain the step count sum y' obtained by the summation unit
9. enter y and y' into the reward function to obtain the current reward *r*
10. for time step $t=1,...T$, do
11. calculation G_t
12. end for
13. for times $td=1,...tr$, do
14. for times $tk=1,...k$, do
15. calculate x' according to formula 1 and 2
16. calculate the gradient of the discriminator and update the parameters using the gradient descent method
17. end for

End for

IV. EXPERIMENT ANALYSIS

In this experiment, we use the real data set to validate the effectiveness of the proposed approach. The data set is provided by Cambridge University, The characteristics of the experimental data are shown in the following table 1.

TABLE I. CHARACTERISTIC DESCRIPTION OF EXPERIMENTAL DATASET

Data set	Cambridge University Data set
frequency	100 Hz
Number	27
Sex ratio	2(M):1(F)
Age range	15-29
Height range	150-189(cm)
Acquisition location	Six kinds of locations, such as hands handbags and so on
walking speed	Random to accelerating to decelerating

Ratio of noise duration	47%

In this experiment, we will compare with several commonly used model algorithms to prove the improvement advantages of our proposed model: GAN Q-learning and GAN-AC(Actor-GANCritic). Firstly, the influence of different GAN on the step counting effect is compared, and then the improved method is compared with the corresponding method to further verify the effectiveness of this method.

In this section, the impact of different GAN on the step counting effect is analyzed. The impact of different GANs within critical is shown in Figure 2. The figures show the impact of three different GAN on the step counting effect. The learning rate is 0.00001 to 0.01.

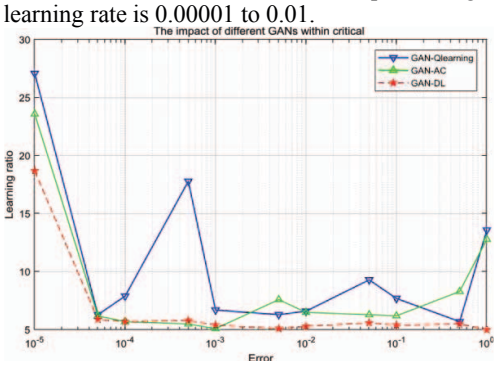

Figure 2. The impact of different GANs within critical on the step counting effect

As can be seen from the above, among the three methods, when the learning rate varies from 0.00001 to 0.01, the counting error is not much different. When the learning rate is between 0.01 and 1, the curve of GANDL is low, and the distance is far away. It can be seen that at this time, the step counting error is much greater than that of the other two methods, and the effect is poor. In order to further compare the performance of the three GANs, the data with a learning rate between 0.00001 and 0.01 are further compared. As can be seen from the figure, when the learning rate is between 0.00005 and 0.01, the GAN-DL curve appears below the GAN-AC curve. However, the distance between the GAN-AC algorithm and the GAN-Qlearning algorithm is relatively close, which needs further analysis. At the learning rates of 0.00005, 0.0001, 0.0005 and 0.005, the performance of both is consistent.

We compare the learning speed of the compared algorithms. The iteration times required for the three algorithms to reach their respective minimum step counting error are shown in Figure 3. As can be seen, when the learning rate is between 0.00001 and 0.001, GAN Q-learning is best. When the learning rate is between 0.001 and 1, GAN-DL algorithm reaches the minimum step counting error earlier than GAN Qlearning algorithm. It is worth noting that when the learning rate is around 0.05, the speed difference of the three algorithms is very small and almost the same. On the whole, the method proposed in this paper improves the learning speed of the algorithm to a certain extent.

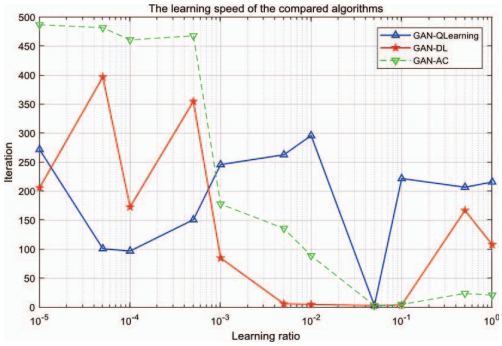

Figure 3. The learning speed of the compared algorithms

V. CONCLUSIONS

In the strategy gradient based algorithm, the update depends on the cumulative return, which has the problems of slow learning and large variance. At the same time, it is necessary to specify the specific distribution form of parameters in the distribution of learning state action values. In this paper, the generation countermeasure network is used to learn the value distribution, and the value distribution is used to sample and generate the state action value to assist the learning of the strategy. Specifically, based on the actor-critical algorithm, the generated countermeasure network is integrated into the critical, the generator is used to learn the distribution of state action values, and the state action values are generated in the form of sampling to update the actor, so as to accelerate the learning of strategies in the actor. At the same time, this paper uses the sampling of the real value distribution as the real input of the discriminator. Finally, combined with the step counting framework based on DRL, the method in this paper is applied to the step counting problem to further improve the step counting effect.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 6200612. The National Key Research and Development Program of China (No. 2019YFB1404602).

REFERENCES

[1] Kjartansdottir I, Arngrimsson S A, Bjarnason R, et al, "Cross-sectional study of randomly selected 18-year-old students showed that body mass index was only associated with sleep duration in girls," Acta Paediatrica, 2018, 107(6).

[2] Yilmaz F T, Aydin H T, "The effect of a regular walking program on dyspnoea severity and quality of life in normal weight, overweight, and obese patients with chronic obstructive pulmonary disease," International Journal of Nursing Practice, 2018, 24(3):e12636.

[3] Zhang H, Yuan W, Shen Q, et al, "A Handheld Inertial Pedestrian Navigation System With Accurate Step Modes and Device Poses Recognition," IEEE Sensors Journal, 2015,15(3):1421-1429.

[4] Arulkumaran K, Deisenroth M P, Brundage M, et al, "Deep Reinforcement Learning: A Brief Survey," IEEE Signal Processing Magazine, 2017, 34(6): 26-38.

[5] F Xiao, "Deep Reinforcement Learning framework for Autonomous Driving," electronic imaging, 2017, 2017(19): 70-76.

[6] an Hasselt H, Guez A, Silver D, "Deep reinforcement learning with double Q-Learning," national conference on artificial intelligence, 2016:2094-2100.

[7] Ioffe S, Szegedy C, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” international conference on machine learning, 2015: 448-456.

[8] Schulman J, Moritz P, Levine S, et al, “High-Dimensional Continuous Control Using Generalized Advantage Estimation,” international conference on learning representations, 2016.

[9] Bellemare M G, Ostrovski G, Guez A, et al, “Increasing the action gap: new operators for reinforcement learning,” national conference on artificial intelligence, 2016: 1476-1483.

[10] Schulman J, Moritz P, Levine S, et al, “High-Dimensional Continuous Control Using Generalized Advantage Estimation,” international conference on learning representations, 2016.