



A Fast Estimation Network Model Based on Process Compression and an Optimized Parameter Search Algorithm for Q-Learning

Shudong Zhang^(✉)

Wuhan FiberHome Information Integration Technologies Co., Ltd., Wuhan, China
zhangsd@fiberhome.com

Abstract. The research content and main work of this paper are as follows: Analyzing the characteristics of the network traffic data set, the key index vacancy rate parameters in the process of cleaning the data set. This paper proposes a fast estimation network model based on process compression and an optimized parameter search algorithm for Q-Learning (QV-QL). The model starts from a predictive model based on deep learning, and on the basis of ensuring the functionality and certain accuracy of the model. Through the compression process and the introduction of mixed-precision calculations, the speed of searching for optimal parameters has been greatly improved.

Keywords: Reinforcement Learning · Network Traffic forecast · Q-Learning

1 Introduction

The bandwidth demand of network traffic appears to be unevenly distributed at the moment. In some enterprises where high strength and stability are required to maintain a high-speed network, excessive network load pressure may be experienced, while in some companies, such high-speed traffic is not needed, resulting in resource waste. It is for this reason that there are great challenges facing us in the current situation. To solve the load capacity limit of network traffic quickly, on-line evaluation of static stability of the network system is a basic requirement for on-line evaluation of static stability. In order to maintain the stable operation of high-demand network operators as well as give full play to the scheduling of network resources, accurate network traffic prediction is an important step towards providing optimal service to high-demand network operators. Additionally, in the field of network traffic data prediction, artificial intelligence based on evolutionary algorithms of deep learning [1] has taken the lead in recent years, aiming to overcome some disadvantages of traditional prediction methods in the field of network traffic data. There is still an increase in applications and development of algorithms and prediction models that are being used in network traffic prediction, but the imbalance between development and planning will lead to an imbalance in resources and an invisible pressure on the field of prediction as a result of the imbalance in resources. Therefore, the specific improvement and optimization still need the continuous exploration and

research of relevant experts and scholars to select a better algorithm prediction model that continues to improve and optimize. Despite the fact that these prediction methods are still under development, an imbalance in resources and unseen pressures may result from a lack of balanced planning and development. The refinement and selection of superior algorithmic prediction models require continuous expert exploration.

The main focus of the preprocessing of data used in the development of network traffic prediction models has been on improving the data set as well as handling outliers and null values, but this focus is easily divorced from the reality, which results in a research trend that does not correspond to reality. If it is really applied in the life level, the prediction results are not satisfactory, forming the situation of application limitation. To make data preprocessing more practical, there are new strategies that are being used to make it more practical. When designing a prediction model, the model data indicators tend to be too serious, in order to ensure the integrity of the data, the dimensions of the data and the number of data points are sufficient, leading to too much in cases where the dimensions of the data are not large enough. As an example, combining algorithms with neural networks is often hard to get accurate predictions. Today, a multitude of algorithms have emerged, especially in the current era of artificial intelligence, and these include evolution algorithms that combine evolutionary strategies [4, 5], genetic programming, or algorithms based on physical properties, such as central force optimization (CFO), artificial chemical reaction optimization algorithms (ACROA), and black hole algorithms (BH), have become widely used. Generally, there are a number of algorithms that are used to make animal search decisions such as cuckoo search (CS) [9], Firefly algorithm (FA) [10], Artificial Bee Colony algorithm (ABC) [11], Antlion optimization algorithm (ALO) [12], etc. However, these algorithms are still prone to some deviations and contingencies. Additionally, when it comes to large scale and high latitude data, the data standardization process is difficult and time-consuming. Therefore, it is important to design a system that ensures efficiency and accuracy in prediction, as well as meets the requirements of large volumes of data at high altitudes. It is at this point that the key research content of this paper is focused on network traffic prediction in the context of the problem to be solved.

As part of this interdisciplinary area of research referred to as reinforcement learning [13–15], many aspects of the theory and algorithm have gained great scientific significance during the 1990s, and they have also made significant advances in psychology, intelligence computing, operations research, and control theory over the years. As a result of these achievements, reinforcement learning has greatly expanded in theory. The application of scheduling decisions has proved to be quite successful in artificial intelligence and optimization.

2 Parameter Optimization Algorithm Based on Fast Estimation Network Model and Improved Q-Learning

In traffic prediction, there are a lot of null values and outliers in data due to non-standard operation during data collection, failure of data collection equipment, data system upgrade and other reasons. Poor data quality poses a serious challenge to data preprocessing.

In the data cleaning stage, we can neither blindly remove all the data lines containing null values, nor allow the data lines containing a large number of null values to enter the data enhancement stage or model training stage. In data cleaning process, the non-null rate parameter of valid data is usually used as one of the key indicators to balance data quality.

This parameter determines which rows are retained and moved to the next stage and which are culled from the dataset by controlling the ratio between the number of non-null values contained in the valid data and the total number of values. When the non-null value rate is in the optimal state, data cleaning can improve the overall non-null value rate of the data set while preserving as much data as possible, that is, reducing the number of noisy data in the data set while preserving as much data diversity as possible. Non-null rate index is a percentage, which will not be given automatically. It is generally given by field experts with network engineering and computer background based on experience, and has problems such as difficult to popularize, lack of interpretability and non-optimality. At the same time, if the exhaustive strategy is used to search parameters, there are some problems such as high computational complexity and time consuming. In view of these problems, this chapter proposes an optimization search algorithm combining the fast valuation network model and improved Q-Learning, which can automatically complete the task of parameter optimization under the premise of reasonable computational complexity.

Fast valuation network model based on process compression.

Generally null value for data, abnormal value, will adopt the method of data cleaning and enhance the data cleaning operations, the data can improve forecasting precision of the model within a certain scope, but the introduction of these additional operations will increase the burden of the model and the GPU in the process of running too much memory footprint and resource consumption, Valuations so this paper proposes a fast network, shown in the Fig. 3 below you can see the difference, with a quick first look for value network model parameters, through the selection of design of experiment, a new strategy, every five percent for one iteration, omitted to data processing operations of the zero and outliers, ten generations iterative processing directly, when to find the optimal parameters, Then, data cleaning and enhancement processing were carried out,

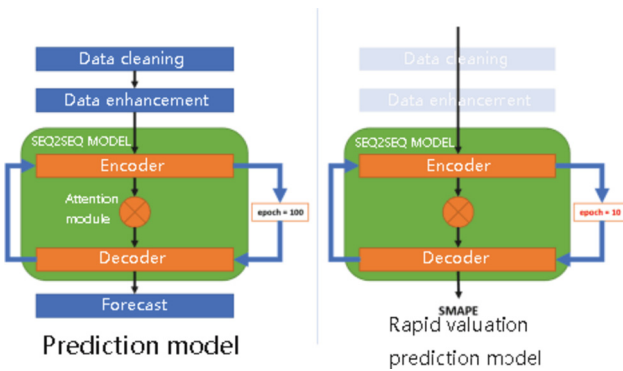


Fig. 1. Comparison of data processing models

so as to omit a lot of useless time. The specific reason why epoch = 10 was selected will be analyzed and explained later.

In order to explain why epoch = 10 is chosen, the change can be seen from Fig. 2. It can be seen that before epoch = 10, the overall loss of baseline has been in a process of rapid decline, that is rapid convergence stage. The increase of epoch at this time has a substantial effect on data processing. However, after epoch = 10, the whole convergence stage entered the long tail stage. It can be seen that, although there was still a slight decrease with the increase of epoch, the cost performance ratio at this time was very low. Therefore, considering the cost performance ratio and reality, the epoch = 10 with the highest cost performance was selected for the universality of application fields.

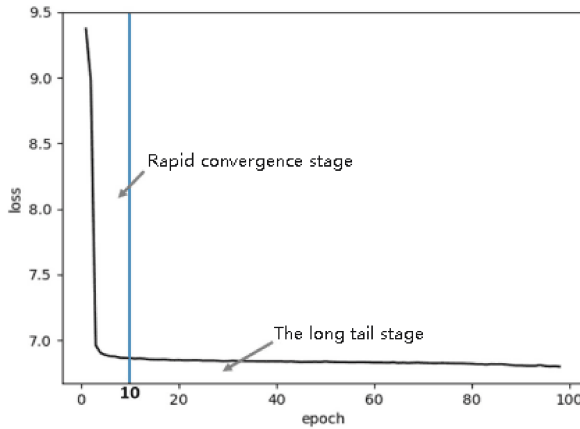


Fig. 2. Cause analysis of epoch = 10

After the fast valuation network model is defined, in order to see that there will not be too much difference before and after the network model changes, the reward index in Q-Learning strategy is used to feedback the network and learn the expected value. From the current step to all subsequent steps, the total expectation gets the maximum value(Q value and Value). Action determines the optimal policy for each state in the Q value function. In each state, the Action with the highest Q value is selected. And the network does not depend on the environment model. The current Action is rewarded, add the next step to get the maximum expected value, the current status action reward, plus the maximum expected value of the next status action The learning rate determines the rate of information acquired before the coverage rate of newly acquired sample information. Usually set a small value to ensure the stability of the learning process and the final convergence. Q-Learning requires an initial value Q. By defining a relatively high initial value, the model is encouraged to explore more. This kind of network may have a certain loss in accuracy, but the speed can be greatly improved. It can be seen from Fig. 5 that the overall accuracy is consistent with the change trend of the processed data.

As shown in Fig. 3, the SMAPE value on the ordinate is the symmetric mean absolute percentage error, which is an accuracy measure based on percentage (or relative) error and can be calculated as shown in formula (2). Indicators used to measure the quality of the network model, the lower the SMAPE value is, the better. SMAPE is a correction index for MAPE problems, which can better avoid the problem that the calculation result of MAPE is too large due to the small real value.

$$SMAPE = \frac{100\%}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{\frac{(|A_t| + |F_t|)}{2}} \quad (1)$$

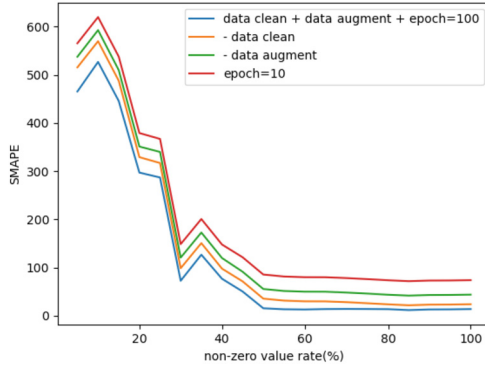


Fig. 3. Comparison of SMAPE values of the rapid valuation network with other treatments

Where A_t is the true value and F_t is the predicted value.

Can see from Fig. 5, although after data cleaning and data fill of the network is still in the overall effect is the result of the optimal, but fast valuations in network epoch = 10, after the overall trends and processing of network is almost the same, change rule is no discrepancy, and maintained a high level in the accuracy. The iteration of 100 generation to 10 generation is reduced and the speed is greatly improved, so the fast valuation network adopted in this paper has better wide applicability in practicability.

Through the proposed fast valuation network for data storage and transportation experiments, it is obvious that the speed optimization before and after is improved, and the accuracy is not far behind. In order to further demonstrate the feasibility of this fast valuation network, the comparison of actual time saving can be shown in Fig. 4.

It can be seen from experimental Fig. 4 above that data cleaning takes 32 s (using non-null value rate of 93% as standard). Data enhancement time (Laplace algorithm is adopted for convolution kernel length of 5): 64 s to 56 s, because if data cleaning is carried out, the data enhancement speed will be improved after removing part of the data. Among them, the training process (epoch = 100) took 118 s, while the training process (epoch = 10) took 27 s. The speed is significantly improved, which is about six times the speed, which is quite remarkable. So whether the data cleaning process and enhance the filling processing of data in the process of iteration tremendous force and resource consumption, the cost of training time, the consumption of the first three figure

is very high, and valuation in the network consume almost negligible, can confirm this fast valuation on the speed of network has the absolute advantage.

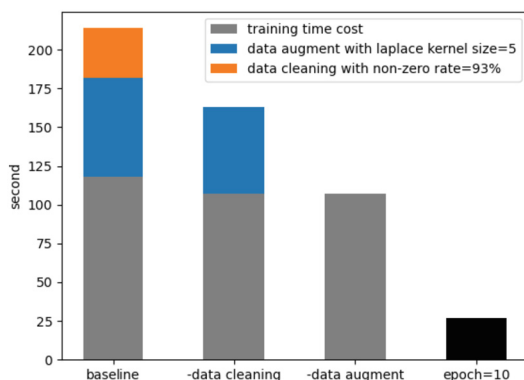


Fig. 4. Comparison of training time for different data processing

2.1 Model Training Process Based on Mixed Accuracy

For circulation data type on the computer, is the most common type of floating-point number, commonly used have double-precision and single-precision floating point, but because of the increasing amount of data now and latitude change big wide, so someone put forward a kind of semi-precision data, double-precision is a 64-bit data, single-precision is 32-bit, semi-precision can reach 16 low storage usage. As the research, the double-precision and single precision are used for calculation, semi-precision is in order to reduce the cost of data transmission and storage, because in many application scenarios, deep learning field, this paper studies the prediction model. For example, with semi-precision data, compared to single-precision can save half of the data transmission cost and resource consumption, In addition, in the field of deep learning, hundreds of millions of parameters are selected for data, so semi-precision transmission is of great significance for research. Figure 7 shows the differences between double-precision, single-precision, and semi-precision floating-point numbers:

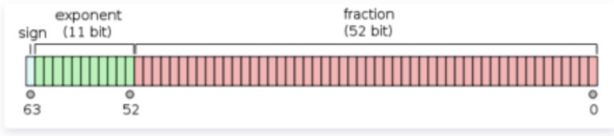
float16, a semi-precision floating-point number consists of one sign bit, five exponent bits, and ten mantissa bits.

float32, a single-precision floating-point number consists of one sign bit, eight exponent bits and 23 mantissa bits.

float64, a double - precision floating-point number consists of one sign bit, 11 exponent bits, and 52 mantissa bits.

As you can see three different precision floating point numbers, is divided into three parts, respectively is the sign bit, index and mantissa, the different precision is only the length of the exponent and the mantissa bits is different, so while keeping on the accuracy of the data at the same time, can very good saves space on the space and memory resources consumption, through the study of the compression of data accuracy,

- Double - precision floating - point number:



- Float-point number with single precision:



- Semi-precision floating point number:

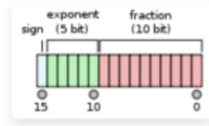


Fig. 5. Feature comparison of three precision data

pre-processing of model algorithms and data can reduce the cost of consumption. The single-pass comparison diagram of the two networks in normal network data processing and fast valuation network can be shown in Fig. 6 below:

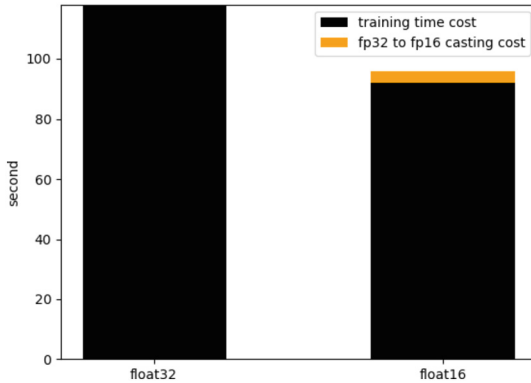


Fig. 6. Comparison of time loss between single precision data and semi-precision data

Can be seen from the comparison above Fig. 6, for data storage and operation process, the data from single precision floating point 32 into half precision 16, get some save time consumption, can probably thirty percent increased performance, further improve the effectiveness of the data preprocessing, illustrates the feasibility of this experiment ideas.

In order to further demonstrate the feasibility of converting from single-precision data to semi-precision data, this paper will verify the comparison between the process of loss value reduction and the effect of amplification in the long-tail stage by high definition comparison in the following two figures. It can be easily seen from the following two figures: The first Fig. 9 shows the comparison of loss values between single-precision data and semi-precision data. The smaller the loss value is, the better. The first figure shows that the overall trend is roughly the same. It can be seen from Fig. 10 that the amplification is about 60 times. It can be seen that although the semi-precision float16 data is not stable in the process of loss decrease, the overall downward trend of loss value is the same as that of float32 single precision data. Therefore, in consideration of the experimental speed, we can use float16 semi-precision data to replace float32 single-precision to realize an optimal data processing.

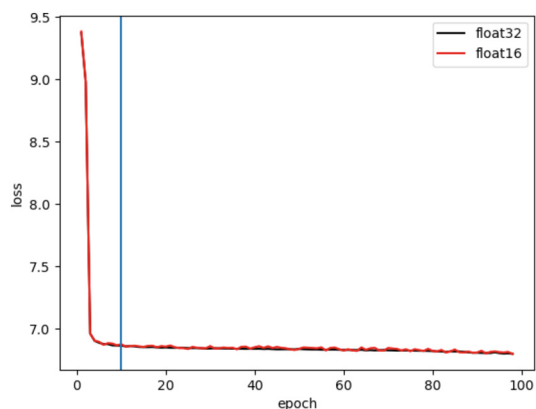


Fig. 7. Comparison of Loss values during training with two kinds of precision

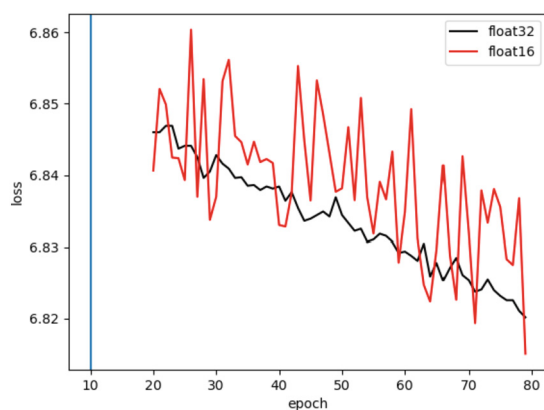


Fig. 8. Comparison of loss values during training with two kinds of accuracy after 60 times magnification

3 Experiment and Result Analysis

3.1 Introduction and Analysis of Data Sets

Because this article research content of the data taken from the records of daily network traffic in and out of an enterprise, is more close to real life, and as a result of the data records have long, data itself exists some days appear more null values, or when business is busy to many outliers, but also there will be a lot of complete data, according to the characteristics of the data, It can be roughly shown as Fig. 11 below:

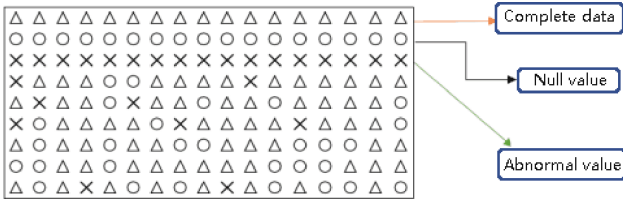


Fig. 9. An introduction to the data set used in this paper

It can be seen from the figure that there are certain null values and outliers in the data. The triangle represents complete data, the circle represents null values, and the cross indicates that there are abnormal values. Because of the difference of these data, the subsequent prediction model will be affected to some extent, therefore, before entering the training, the data should be improved according to the experimental purpose, and the Q-Learning enhanced learning preprocessing strategy should be carried out to pave the way for the subsequent prediction.

3.2 Optimization Parameter Search Experiment Based on Improved Q-Learning

In order to test the difference between the improved Q-Learning and the traditional brute force exhaustive mechanism in search efficiency, this paper carries out an experiment of searching optimization parameters on a traffic prediction system based on deep learning. The basis of the experiment is realized through the improved QV-QL algorithm, in order to better understand the process of parameter search in the experiment, Table 1 shows the operation process of combining the fast valuation network with the improved Q-Learning algorithm in the form of pseudo-code:

Table 1. Quick Valuation Q-Learning algorithm

Algorithm 1: QV-QL (Quick Valuation Q-Learning)

Initialize the reward function r: fast valuation model

Initialize Q(s, a), s: 0.5

Initialize Maximum episode value

Repeat execution (maximum episode detection) :

Repeat execution (knee and elbow detection or reaching the boundary of S):

Starting from initial S, all actions a_n in the action set are performed simultaneously, observe r_n

Update Q table with maximum r value:

$$Q(s, a) \leftarrow Q(s, a) + [\max(r) + \max_{a'} Q(s', a') - Q(s, a)]$$

End of the cycle

End of the cycle

Knee and elbow test

If the action function a' before is equal to the action function a now:

Returns (True), interrupts the loop and jumps out of the current episode

Otherwise:

Return no (False) , to continue the current episode iteration

The following experimental figure is used to show the search process. First, Fig. 12 shows the SAMPE full solution space obtained by exhaustive method as the baseline. On the basis of the baseline, the improved Q-Learning algorithm proposed in this paper is used to search optimization parameters on the same system, and finally the whole search process is annotated to the baseline manually.

Figure 11 shows the first generation episode of optimization. Starting from the non-null rate of 50%, the left and right actions of the action set are respectively performed. In this experiment, there are two actions, namely left and right search. Search in both directions for a knee and an elbow. After comparing the return values of these two points, namely the SMAPE value, the point of the best return value is taken as the starting point of the next episode, and the current episode ends.

Figure 12 shows the second generation episode of optimization search. Starting from the point of the best return value given by the previous generation episode, search both sides until the non-null rate reaches the boundary and the search ends. Compare the return value of the boundary point with the current best return value and return the better one as the best advantage sought.

Through experiments, it is concluded that the improved Q-Learning comparison and exhaust strategy can save 36 times of calculation of return value, which is of great significance for scenarios with high computational complexity of return function caused by deep learning model.

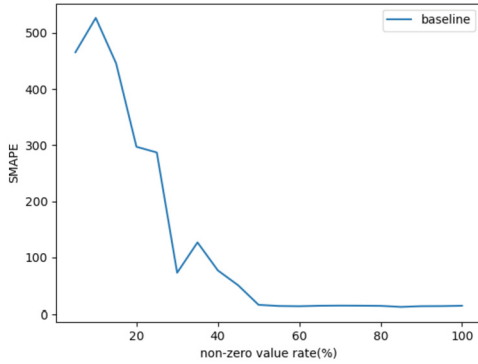


Fig. 10. SMAPE values at different non-null value rates obtained by exhaustive search

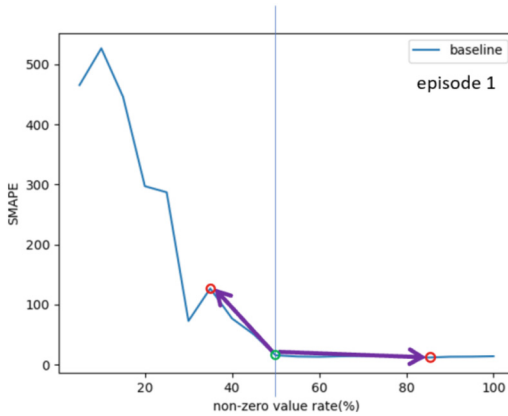


Fig. 11. First generation episode, starting from 50% looking for knee and elbow points

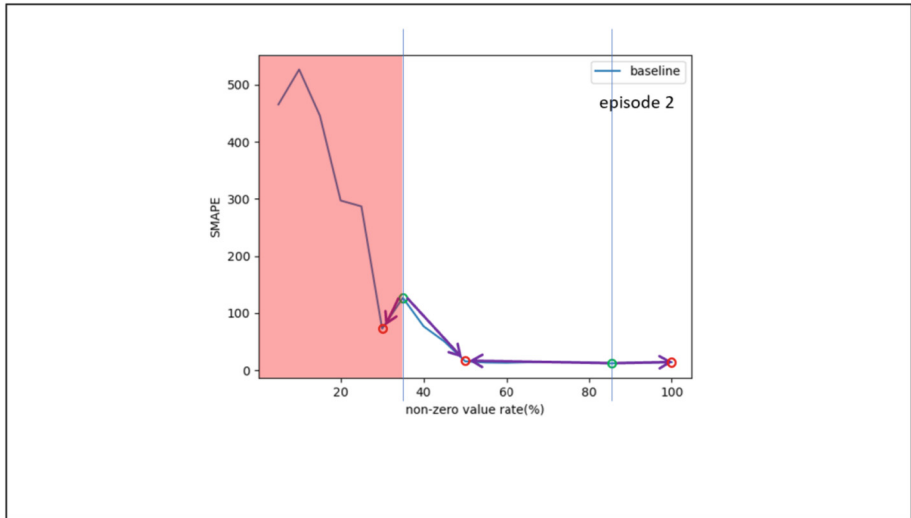


Fig. 12. The second generation episode, starting from the found knee elbow point, search for the next knee elbow point

4 Summary

In this paper, the classic Q-Learning algorithm is introduced in detail and the relevant characteristics are analyzed and summarized, and the good adaptability of the algorithm in the field of traffic prediction is expounded. Then a fast estimation network model based on flow compression is proposed., based on the flow prediction model, a network model that can be used to quickly estimate the return value is constructed by omitting the pre-processing steps of the original model and reducing the training algebra of the prediction model. At the same time, in order to further accelerate the calculation of the return value and reduce the memory consumption of the algorithm, this chapter then proposes a model training process based on mixed precision to accelerate the computational performance of the algorithm by compressing the data tail. It is proved by experiments that the introduction of fast estimation network model based on mixing accuracy has better effect on improving the calculation of return value.

Acknowledgment. This work is funded by the National Natural Science Foundation of China under Grant No. 61772180, the Key R & D plan of Hubei Province (2020BHB004, 2020BAB012).

References

1. Wei, Z.: A summary of research and application of deep learning. *Int. Core J. Eng.* **5**(9) (2019)
2. Cheng, X.F., Yan, Y.J.: Collaborative path planning for unmanned aerial vehicle based on MAXQ hierarchical reinforcement learning. *Res. Inf. Technol.* **46**(01), 13–19 (2020)
3. Ma, P.G., Xie, W., Sun, W.J.: A review of reinforcement Learning. *Command Control Simul.* **40**(06), 68–72 (2018)

4. Jun, S., Kochan, O., Kochan, V., Wang, C.: Development and investigation of the method for compensating thermoelectric inhomogeneity error. *Int. J. Thermophys.* **37**(1), 1–14 (2016)
5. Ding, Q.F., Yin, X.Y.: A review of differential evolution algorithms. *J. Intell. Syst. Intell. Syst.* **12**(04), 431–442 (2017)
6. Talal, T.M., Attiya, G., Metwalli, M.R., et al.: Satellite image fusion based on modified central force optimization. *Multimedia Tools Appl.* **79**, 21129–21154 (2020)
7. Singh, H., Kumar, Y.: Hybrid artificial chemical reaction optimization algorithm for cluster analysis. *Procedia Comput. Sci.* **167**, 531–540 (2020)
8. Singh, P., Arya, R., Titare, L.S., et al.: Optimal load shedding to avoid risks of voltage collapse using black hole algorithm. *J. Inst. Eng. (India) Ser. B* **102**, 261–276 (2021)
9. Yingjie, X., Ning, C., Xi, S., et al.: Proposal and experimental case study on building ventilating fan fault diagnosis based on cuckoo search algorithm optimized extreme learning machine. *Sustain. Energy Technol. Assess.* **45**, 100975 (2021)
10. Vilas, K.J., Asutkar, V.G.: A novel optimal firefly algorithm based gain scheduling proportional integral derivative controller for rotor spinning machine speed control. *Int. J. Dyn. Control* **9**(4), 1730–1745 (2021)
11. Wang, Y., Szeto, W.Y.: An enhanced artificial bee colony algorithm for the green bike repositioning problem with broken bikes. *Transp. Res. Part C* **125**, 102895 (2021)
12. Zhang, J.H., Xu, X.H.: A new evolutionary algorithm – ant colony algorithm. *Syst. Eng.-Theory Pract.* (03), 85–88+110 (1999)
13. Shikang, X., Israel, K., Mani, K.C.: Adaptive workload adjustment for cyber-physical systems using deep reinforcement learning. *Sustain. Comput. Inform. Syst.* **30**, 100525 (2021)
14. Song, W., Beshley, M., Przystupa, K., et al.: A software deep packet inspection system for network traffic analysis and anomaly detection. *Sensors* **20**(6), 1637 (2020)
15. Jun, S., Przystupa, K., Beshley, M., Kochan, O., et al.: A cost-efficient software based router and traffic generator for simulation and testing of IP network. *Electronics* **9**(1), 40 (2019)