An intelligent routing optimization strategy based on deep reinforcement learning

Zhou Xiong1 and Guo Hongyu1*

1 North China Institute of Computing Technology
*Corresponding author’s e-mail: 3220190869@bit.edu.cn

Abstract. Finding the optimal strategy in network routing has always been a NP hard problem. Due to the complexity and dynamics of network traffic, the existing intelligent routing algorithms have poor generalization ability. Therefore, this paper proposes an intelligent routing strategy based on deep reinforcement learning, and with the help of SDN control can dynamically collect network traffic distribution information, can dynamically adjust the routing strategy. Compared with traffic engineering algorithms such as TCMP and DRL-TE, the end-to-end delay is optimized under different throughput.

1. Introduction
Due to the huge fluctuation of network traffic, simple physical equipment expansion is faced with the problems of low resource utilization and long update cycle, which can not achieve better network transmission effect and network service quality with limited cost [1]. Through the routing control of network traffic, network traffic can be more reasonably distributed on different network devices, so as to reduce the unbalanced use of network resources and improve the average utilization of network resources. Traditional network routing algorithms are based on distributed routing algorithms running on different routers. Distributed routing algorithms usually calculate routing according to the principle of best effort, but the accuracy of traffic control is limited, which can not achieve good network quality of service. With the emergence of software defined network (SDN) technology, the traffic control strategy based on centralized network view is realized. By collecting the information of the whole network, SDN controller formulates the network strategy according to the macro control information of the network state, so as to provide more refined service strategy for the network. However, the optimal network routing strategy based on global network information is a NP hard problem [2]. In order to ensure the dynamic and effective routing strategy generation, the existing routing optimization algorithms are usually based on artificial heuristic algorithm, which can not guarantee the quality of routing strategy. To solve this problem, the industry will focus on the rapid development of artificial intelligence technology in recent years. The advantages of routing scheme based on artificial neural network (ANN) and deep reinforcement learning (DRL) are more and more obvious. Compared with other machine learning schemes, the intelligent routing scheme based on deep reinforcement learning has the advantages of autonomous training, strong adaptability, and no need to label a large number of data manually, which stands out among many machine learning based schemes. However, the current DRL based intelligent routing schemes are mostly based on feedforward neural network (FNN), recurrent neural network (RNN) and other neural network structures. The network structure is fixed, and the generalization ability is poor, so it is difficult to adapt to the dynamic changes of network topology.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI. Published under licence by IOP Publishing Ltd
To solve the above problems, this paper proposes an intelligent routing adjustment strategy based on graph neural network (GNN). This strategy is based on the DRL framework for the GNN parameter optimization process, and automatically searches for the optimal routing strategy of the network by constantly updating the strategy parameters in the environment.

2. Related works

At present, intelligent routing schemes based on machine learning algorithm are usually implemented based on SDN framework because they need global network view information. According to the characteristics of machine learning algorithm, it can be divided into three kinds of intelligent routing implementation schemes based on supervised learning, unsupervised learning and reinforcement learning.

The scheme based on supervised learning is mainly realized by deep neural network (DNN). Huang et al. [3] designed a traffic classification scheme based on DNN, and adapted different static routing schemes based on different traffic categories; Literature [4,5] also proposes a traffic classification scheme based on DNN, which realizes different routing algorithms through different traffic categories to complete network traffic scheduling. Mao et al. [6] proposed an intelligent routing scheme based on deep belief network (DBN), which realizes the routing adjustment of edge nodes and core nodes through different traffic characteristics. Generally speaking, the intelligent routing scheme based on supervised learning needs to label a large number of network traffic characteristics manually, so it is difficult to achieve universality. The scheme based on unsupervised learning usually implements network routing strategy based on the analysis of network traffic characteristics. For example, literature [7,8] proposed using k-means algorithm to achieve network traffic classification, and literature [9] used principal content analysis (PCA) to achieve network feature extraction. After analyzing the characteristics of network traffic, the above scheme still needs to design the corresponding routing strategy manually, and because the algorithm accuracy of unsupervised learning is usually difficult to guarantee, it is difficult for the intelligent routing scheme based on unsupervised learning to achieve better performance. The intelligent routing scheme based on reinforcement learning algorithm realizes intelligent routing through reinforcement learning algorithm (including DRL algorithm). Different from the above two kinds of intelligent routing schemes, the intelligent routing scheme based on reinforcement learning algorithm can directly realize the mapping from network input information (such as traffic distribution view) to routing strategy, and can realize self-tuning in the training environment, so its routing scheme can usually achieve better performance. Boyan et al. [10] proposed a routing scheme based on Q-learning algorithm, which can intelligently adjust the next choice of the route; DRL-TE [11] optimizes the performance of multi-path routing by adjusting different path split ratios of multi-path routing through DRL algorithm; References [12,13] use DRL algorithm to adjust the routing weight in the network topology, so as to realize the dynamic adjustment of the whole network routing calculation results.

Generally speaking, reinforcement learning based intelligent routing scheme has become the mainstream of intelligent routing scheme. However, in the above intelligent routing scheme based on reinforcement learning, traditional neural network structures such as feedforward neural network and recurrent neural network are mainly used in neural network. This kind of neural network structure can only deal with European input data structure (such as one-dimensional text data or two-dimensional image data with fixed dimensions), and has strict deterministic restrictions on input data format; In practical application, the input data format of neural network will change after the network topology changes, and the traditional neural network will not be able to generalize the training experience to the new network topology. Therefore, the existing intelligent routing schemes based on reinforcement learning have over fitting for the network topology used in the training process, and can not be flexibly applied to different network topologies, and can not deal with the passive network topology changes such as network link failure.
3. Experimental design

This section mainly introduces the intelligent routing strategy based on SDN framework combined with GNN and DRL, and the experimental design of the proposed intelligent routing strategy compared with traffic engineering algorithms such as TCMP and DRL-TE. The function of the intelligent strategy is mainly realized by the programmable switch in the data plane, the SDN controller in the control plane and the DRL algorithm running on the controller. The programmable switch is responsible for the statistics of network traffic information (including real-time traffic distribution, traffic transmission performance and other indicators); SDN controller is responsible for collecting and counting traffic information and updating forwarding table of switch; DRL algorithm is responsible for generating the output action according to the state information in the network, and the output action value is directly used for routing policy generation. Among them, DRL algorithm combines with graph neural network for dynamic analysis of network input data, and takes the generated output action as link weight. Based on the dynamically adjusted link weight, SDN controller can flexibly adjust network routing through weighted shortest path algorithm, so as to realize intelligent control of routing.

DRL algorithm is extended from the basic framework of reinforcement learning. Reinforcement learning optimizes the parameters of reinforcement learning algorithm through the information interaction between reinforcement learning algorithm and its running environment, so as to realize the algorithm training. In the basic model of reinforcement learning, the network is modeled as a Markov process, the main interactive elements of the algorithm include \( s_t, a_t \) and \( r_t \). \( s_t \) represents the state information of the environment at time \( t \), \( a_t \) represents the output action of the algorithm at time \( t \), and \( r_t \) represents the reward of the result generated by the action content of the algorithm at time \( t \). The output action generated from the state information can be realized by looking up the table, function calculation and other different ways. Deep reinforcement learning uses deep neural network to complete this process, so it is called deep reinforcement learning. The goal of deep reinforcement learning is to maximize the cumulative discount return by adding discount factors in one stage, which is

\[
\max \sum_{t=0}^{T} \gamma^t r_t
\]

Deep reinforcement learning realizes the adjustment of neural network parameters by reward value \( r_t \). According to the different implementation methods, it can be divided into value optimization, strategy gradient optimization and other algorithm frameworks. The deep reinforcement learning framework used in this paper is strategy gradient optimization. The strategy gradient optimization method can be expressed by formula (1):

\[
\nabla_{\theta} \mathbb{E}\left[ \sum_{k=0}^{T} r_k \right] = \mathbb{E}\left[ \sum_{k=0}^{T} \nabla_{\theta} \log \pi(s_k, a_k) Q(s_k, a_k) \right]
\]

Among them, \( Q(s_k, a_k) \) is the evaluation value of the current state action content, represents the current action generation strategy. At each time, the update gradient of the neural network can be calculated by formula (1). In the process of updating the parameters of neural network, this paper adopts the implementation scheme of reference [14], that is, it can calculate the updating gradient value of neural network through the deformation of Q value.

The structure of traffic distribution depends on the network topology. In order to better deal with the network topology information and improve the generalization ability of neural network for different topology data, this paper uses GNN as the neural network implementation of DRL algorithm framework. Specifically, the GNN type used in this paper is information transfer network (MPNN) [15]. Specifically, MPNN mainly completes the network information transmission and calculation through the interaction of node information in its graph, and we can adjust the network routing by adjusting the link weight. The element we adjust is the link information in the network topology. The link of communication network is mapped to the node of MPNN. Through the transmission and calculation of node information in MPNN, the information update and calculation of link in information communication network are
completed. Among them, the link information in the information communication network includes the real-time bandwidth occupation information of the link. Therefore, the interface of DRL status, action and reward is as follows:

**Status:** the status information of DRL algorithm is the link utilization and link delay information. The above information of each link is input in a certain format as the input information of the corresponding node of MPNN.

**Action:** the action content of DRL algorithm is link weight. After MPNN calculation, the output value of each node corresponds to all link weight information in the network. According to the link weight information, the SDN controller can calculate the end-to-end weighted shortest path of the network.

**Reward:** the design of reward determines the independent optimization direction of DRL algorithm, that is, the optimization direction of network performance. Common performance indicators can be combined with the overall network throughput, average end-to-end transmission delay and other performance indicators. In this paper, the average end-to-end delay of the network is used as the reward.

The hardware platform of this experiment has an i7 10750h CPU, 16GB DDR4 memory and a 2060 graphics card. The number of iterations of DRL algorithm is set to 80000, and the number of iterations of MPNN is set to 10. The neural network parameters are optimized by the random gradient descent method based on nesterov momentum, and the learning rate is set to 0.00001. In the hardware experimental platform, it takes about 4 hours to complete an algorithm training. After the training, the scheme can be deployed to different network topologies to achieve fast routing calculation. In the simulation environment, the OS3E topology selected in topology zoo is adopted, and the link bandwidth capacity of each topology is set to 100Mbps. The traffic generation in the network is based on the combination of random traffic and periodic traffic. The traffic characteristics are changed by adjusting the proportion of random traffic and periodic traffic. Each network node can initiate communication demand with any other network node, and its average traffic is set as a specific proportion of the total network throughput (30%, 40%, 50%).

The proposed algorithm is compared with traffic engineering algorithms such as TCMP and DRL-TE. DRL-TE: realize the network end-to-end communication through multi-path (three backup paths are set for each end-to-end communication), sense the traffic distribution in the network through DRL algorithm, so as to adjust the diversion ratio of each data flow on the three backup paths, so as to realize traffic engineering. ECMP: ECMP transmits network traffic through equivalent multipath, which is one of the typical traffic engineering schemes in network.

### 4. Experimental Results

This paper compares the generalization ability of the proposed intelligent strategy with TCMP and DRL-TE in end-to-end delay and routing strategy.

End to end delay comparison: this experiment mainly uses the average end-to-end delay in the network as a measure of network performance. In this experiment, three groups of different network traffic characteristics (i.e. different periodic traffic / random traffic ratio) are set. Figures 1 to 3 show the average end-to-end delay of different traffic characteristics in OS3E topology under different traffic intensities. From the overall trend of figure 1 to figure 3, it can be seen that with the increase of the total traffic in the network, the average end-to-end delay of each scheme shows an upward trend, which is in line with the intuitive inference. The abscissa in figure. 1 ~ figure. 3 shows different traffic composition. Taking figure. 1 as an example (figure. 2 and figure. 3 have the same rule), it can be seen that with the increase of random traffic, the average end-to-end delay of tide shows an obvious growth trend, while the average end-to-end delay of ECMP and the intelligent strategy proposed in this paper has a relatively stable value, Only the delay variance of traffic increases significantly. In general, the delay of intelligent strategy under different traffic characteristics is mostly better than other schemes, and the average end-to-end delay can be saved by 5.8% compared with the current optimal scheme under the best situation.
Figure 1  Average end-to-end delay under the 30% throughput

Figure 2  Average end-to-end delay under the 40% throughput
Figure 3  Average end-to-end delay under the 50% throughput

Generalization ability of routing strategy: the common artificial neural network usually has the disadvantage of over fitting, that is, the input data format needs to remain unchanged during the training process, and it is difficult for the neural network to adapt to different input data formats after the training. DRL algorithm mainly depends on neural network generation strategy, so the DRL algorithm using recurrent neural network also has the problem of over fitting. In this experimental scenario, after completing the training of different routing schemes in OS3E topology, the obtained routing algorithm is applied to other networks to test the generalization ability of different schemes. The results are shown in figure 4. Among them, the abscissa represents the change of the number of nodes (that is, the number of increased nodes) in different test topologies based on the OS3E topology, and the ordinate represents the change of end-to-end delay relative to the original OS3E topology (a positive value indicates that the delay performance is optimized, and a negative value indicates that the delay is increased). As can be seen from the data in figure 4, as the network topology changes more and more (more and more nodes are added than the original training topology OS3E), the traffic transmission of DRL-TE shows an obvious increasing trend of delay, while the performance of ECMP and intelligent strategy shows no obvious downward trend. Among them, the delay fluctuation of ECMP and DRL-TE in different topologies is mainly caused by the nature of the topology itself.
5. Conclusion
This paper proposes an intelligent routing strategy based on GNN and DRL technology. The strategy uses DRL algorithm to realize the independent training of scheme parameters, and can search the optimal strategy through the interaction with the network without relying on human experience. At the same time, by using GNN to calculate the network traffic distribution data based on the topology, the generalization ability of routing strategies under different network topologies can be realized. In the future work, we will consider optimizing the design structure of neural network to further improve the performance of intelligent routing strategy.

References
[1] CAIDA. The cooperative associate for Internet data analysis (CAIDA). [EB/OL].(2019-11-01)[2020-10-11].http://www.caida.org/data.
[2] CUI Yong, WU Jianping, XU Ke, et al. A survey on quality-of-service routing algorithms for the Internet[J]. Journal of Software, 2002, 13(11): 2065-2075 (in Chinese).
[3] HUANG Wenhao, SONG Guojie, HONG Haikun, and XIE Kunqing. Deep Architecture for Traffic Flow Prediction: Deep Belief Networks with Multitask Learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(5): 2191–2201.
[4] KATO N, FADLULLAH Z. M, MAO B, et al. The deep learning vision for heterogeneous network traffic control: proposal, challenges, and future perspective[J]. IEEE Wireless Communications, 2017, 24(3): 146–153.
[5] PARSAEI M. R, SOBOUTI M. J, KHAYAMI S. R, and JAVIDAN R. Network traffic classification using machine learning techniques over software defined networks[J]. International Journal of Advanced Computer Science and Applications, 2017, 8(7): 220–225.
[6] MAO Bomin, Fadlullah Z M, TANG Fengxiao, et al. Routing or computing? The paradigm shift towards intelligent computer network packet transmission based on deep learning [J]. IEEE Transactions on Computers, 2017, 66(11): 1946-1960.

[7] ERMAN J, ARLITT M, and MAHANTI A. Traffic classification using clustering algorithms [C]// Proceedings of the 2006 SIGCOMM workshop on Mining network data. New York, USA: ACM Press, 2006: 281–286.

[8] LIU Yingqi, LI Wwei, and LI Yunchun. Network traffic classification using k-means clustering [C]// International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007). LA, USA: IEEE press, 2007: 360–365.

[9] YAN Ruoyu and LIU Ran. Principal component analysis-based network traffic classification [J]. Journal of Computers, 2014, 9(5): 1234–1240.

[10] BOYAN J A, LITTMAN M L. Packet routing in dynamically changing networks: A reinforcement learning approach. [C]// Proc of Advances in Neural Information Processing System. San Francisco, CA: Morgan Kaufmann press, 1994: 671–678.

[11] XU Zhiyuan, TANG Jian, MENG Jingsong, et al.—Experience-driven networking: A deep reinforcement learning based approach. [C]// IEEE INFOCOM 2018-IEEE Conference on Computer Communications. Hawaii, USA: IEEE press, 2018: 1871–1879.

[12] SUN Penghao, HU Yuexiang, LAN Julong, et al. TIDE: Time-relevant deep reinforcement learning for routing optimization [J]. Future Generation Computer Systems, 2019, 11(3): 134–144.

[13] Sun Penghao, Li Junfei, Guo Zehua, Xu Yang, et al. SINET: Enabling Scalable Network Routing with Deep Reinforcement Learning on Partial Nodes. [C]// SIGCOMM Posters and Demos. Beijing CN: ACM press, 2019: 88–89.

[14] RONALD J. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning [J]. Machine Learning, 1992, 3(4): 229–256.

[15] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, OriolVinyals, and George E Dahl. 2017. Neural message passing for quantum chemistry. [C]// Proceedings of the 34th International Conference on Machine Learning. New York, USA: ACM press, 2017: 1263–1272.

[16] Léon Bottou. 2010. Large-scale machine learning with stochastic gradient descent. [C]// Proceedings of COMPSTAT. Berlin, GER: Springer press, 2010: 177–186.

[17] SIMON K, HUNG X, NICKOLAS F, et al. The internet topology zoo [J]. IEEE Journal on Selected Areas in Communications 2011, 29(9): 1765–1775.