TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 09/13 pp878–885
DOI: 10.26599/TST.2020.9010052
Volume 26, Number 6, December 2021

Parallel-Data-Based Social Evolution Modeling

Weishan Zhang*, Zhaoxiang Hou, Xiao Wang, Zhidong Xu, Xin Liu, and Fei-Yue Wang

Abstract: Abnormal or drastic changes in the natural environment may lead to unexpected events, such as tsunamis and earthquakes, which are becoming a major threat to national economy. Currently, no effective assessment approach can deduce a situation and determine the optimal response strategy when a natural disaster occurs. In this study, we propose a social evolution modeling approach and construct a deduction model for self-playing, self-learning, and self-upgrading on the basis of the idea of parallel data and reinforcement learning. The proposed approach can evaluate the impact of an event, deduce the situation, and provide optimal strategies for decision-making. Taking the breakage of a submarine cable caused by earthquake as an example, we find that the proposed modeling approach can obtain a higher reward compared with other existing methods.

Key words: parallel data; reinforcement learning; decision-making

1 Introduction

Social life is related not only to economic development but also with social harmony and stability. Furthermore, the security of economic and social activities is threatened by the occurrence of extreme natural disasters. Thus, making wise decisions during disasters occurs is crucial; that is, the situation must be analyzed as soon as possible, the development trend must be deduced, and reasonable maintenance must be performed on the basis of the objective of total loss reduction.

• Weishan Zhang is with China University of Petroleum, Qingdao 266580, China, and also with Qingdao Academy of Intelligent Industry, Qingdao 266111, China. E-mail: zhangws@upc.edu.cn.
• Zhaoxiang Hou, and Xin Liu are with China University of Petroleum, Qingdao 266580, China. E-mail: s19070035@s.upc.edu.cn; lx@upc.edu.cn.
• Xiao Wang and Fei-Yue Wang are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100049, China, and also with Qingdao Academy of Intelligent Industry, Qingdao 266111, China. E-mail: x.wang@ia.ac.cn; feiyue.wang@ia.ac.cn.
• Zhidong Xu is with the Institute of National Security, National Defense University, Beijing 100081, China. E-mail: xuzhd1981@163.com.
∗ To whom correspondence should be addressed.
Manuscript received: 2020-09-20; accepted: 2020-10-09

Some research on the use of big data or artificial intelligence has been conducted to analyze society and nature (e.g., automatic detection of smart cities[1] and natural disaster detection[2, 3]). Most existing machine and deep learning models are trained on the basis of labeled data. However, big data of social evolution are difficult to format. Reinforcement Learning (RL) can achieve unsupervised learning by interacting with the environment. At present, many RL algorithms, including Deep Q-Network (DQN)[4], deep deterministic policy gradient[5], and Monte Carlo tree search[6], can be used for games[7] and automatic control of unmanned aerial vehicles[8]. However, no general method has been effective in combining social data with RL to solve social problems and in analyzing virtual data to evolve and make decisions.

The idea of parallel data provides a new solution for simulating the normal operation and evolution of the society[9]. Therefore, this study presents a parallel-data-based social evolution model to achieve social assessment, evolution, and decision-making. The social evolution model involves (1) an overall evolution framework combining parallel data and RL, (2) an economic and social value model, and (3) an RL-based resource allocation and evolution strategy.

The remainder of this paper is organized as follows: In Section 2, we analyze the current situation of related research. Then, we show our overall architecture and propose a set of models and methods for abnormal event evolution and decision-making in Section 3. We then evaluate the method through experiments in Section 4. Conclusions and future work are presented in Section 5.

2 Related Work

Digital twins form a digital image that can be disassembled, copied, transferred, modified, and repeated. They are now redefined as digital replications of living and nonliving entities[10]. Considerable research on digital twins for economic systems has been conducted; such research includes the use of blockchains and digital twins to study cluster economic systems[11] and the use of digital twins and robots as assistants to improve the quality, productivity, and efficiency of economic development[12]. In the risk assessment of chemical plants, Song et al.[13] proposed a method to evaluate the probability of abnormal events dynamically; this method can analyze the probability of abnormal events and causal factors. Data-driven digital twins will become the core of simulation-based development processes, because they can simplify the development process, solve the problem of diagnosing and predicting the status of production system components[14], and realize intelligent systems[15].

As an important branch of machine learning, RL has been widely used in war games, robot control, Unmanned Air Vehicle (UAV) cooperative scheduling, and other fields due to its unique unsupervised mechanism. RL mainly includes methods based on value function, strategy gradient, as well as search and supervision. DQN[4] represents the value-based deep RL algorithm. Many improved algorithms exist, including double DQN[16] and dueling DQN[17]; however, most DQNs can only handle discrete action sets and cannot express continuous actions. Policy-based[18] RL makes up for the shortcomings of DQN. By modeling strategy functions, the distribution of actions, including actor-critic, Deep Deterministic Policy Gradient (DDPG)[5], A3C[19], and PPO2[20] algorithms, can be output. In 2017, AlphaGo Zero made international headlines with their incredible success, which was achieved from scratch. Learning from the blank state, AlphaGo Zero can quickly perform self-study without any human input. In addition to no sample and self-training, AlphaGo Zero used the Monte Carlo tree search algorithm and brought new development[6].

Parallel data and digital twins can model the social state, provide data support, and achieve parallel evolution. RL can make improved decisions in a complex environment. Currently, RL is not combined with parallel data or digital twins to achieve social state assessment, evolution, and decision-making. For the state interpretation and resource redistribution after the occurrence of social abnormal events, we must comprehensively consider various factors to analyze the development trend of the event and obtain optimal strategies to guide the execution of remedy practices. Therefore, we propose a parallel-data-based social evolution modeling approach to solve the above problems.

3 Overview of Parallel-Data-Based Social Evolution

We build a data-driven social evolution approach, which integrates social, economic, and other factors to form a virtual society that can be simulated and evolved. On the one hand, the model is used to calculate the global loss for deduction. On the other hand, it can be used to calculate the reward of RL and evaluate the strategy. We use RL to model the current environment through the continuous exploration and optimization of agents to find solution; in this manner, the global loss caused by abnormal events is minimized. In turn, the final decision affects the social model. The overall architecture is shown in Fig. 1.

3.1 Social and economic value model

Economic value refers to the proportion of economic benefits between input and output in production activities. In enterprises, the economic benefits constitute the economic value and directly affect the development. The economic value and social responsibility of

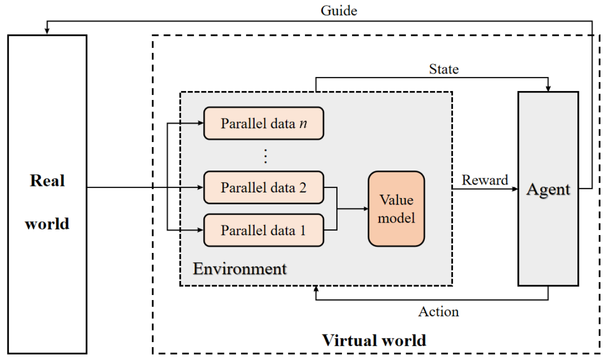


Fig. 1 Architecture of parallel-data-based social evolution.

enterprises exist simultaneously, interact, and influence each other. In the real world, the evaluation of an enterprise should be based on its economic and social value. Based on parallel data theory[9], we establish the economic and social value model by improving the LM3[21] and SROI[22] models.

The economic value model is as follows:

$$V_{\mathrm{E}} = \sum_{t=1}^{y} \frac{(R_t - C_t)}{(1+d)^t} \tag{1}$$

where y represents the year, R_t represents the income in the year t, C_t represents the cost in the year t, and d represents the discount rate.

The social value model is as follows:

$$V_{\mathrm{S}} = \sum_{t=1}^{y} \frac{(G_t + T_t + E_t + P_t + O_t)}{(1+d)^t} \tag{2}$$

where G_t represents the promoting local GDP growth value in the year t, T_t represents the tax value in the year t, E_t represents the saving energy value in the year t, P_t represents the value of environmental pollution control in the year t, and O_t represents other social values in the year t.

The social and economic value of the enterprise is as follows:

$$V = V_{\mathrm{E}} + V_{\mathrm{S}} \tag{3}$$

3.2 RL-based resource allocation

RL is a mapping that agent learns from environment to behavior. Through continuous interaction with the environment, RL can obtain the optimal behavior sequence under the stimulation of reward or punishment given by the environment, thus obtaining the maximum reward.

When an abnormal event occurs, some social resources may be destroyed. How to allocate the remaining resources in accordance with the current state to make the loss as small as possible is a crucial issue. Traditional machine and deep learning methods require substantial labeled data for training, but the social distribution problem is difficult to train due to data with uncertain dimensions. RL can solve the problem of social distribution by using its unique interaction mechanism with the environment.

3.2.1 Problem modeling

RL modeling problems must consider the environment, action space, and reward. We describe the problem from the following three aspects.

(1) Observation space. For the resource allocation task, all data that interact with agents are called environment. Different types of environments can be divided into the following categories: completely observable, partially observable, continuous, and discrete environments. In the process of resource allocation, the resource status information of each node, the economic and social value of each enterprise, and other factors, such as distance, are regarded as the environment's responding to the agent's action. The agent perceives the current environmental state,

$$S_i = (\mathrm{Demand}_i, \mathrm{Value}_i, \mathrm{Resource}_{1i}, \ldots, \mathrm{Resource}_{ni}, \mathrm{Other}_i) \tag{4}$$

$$S_i = (\mathrm{Demand}_{1i}, \mathrm{Value}_{1i}, \ldots, \mathrm{Demand}_{mi}, \mathrm{Value}_{mi}, \mathrm{Resource}_{1i}, \ldots, \mathrm{Resource}_{ni}, \mathrm{Other}_i) \tag{5}$$

Equations (4) and (5) represent the state space of partially and fully observable states, respectively, where m represents the number of enterprises that need resource allocation, and n represents the number of resource nodes. Demand_{ji} and Value_{ji} $(1 \leqslant j \leqslant m)$ represent the resource demand and value of the enterprise j, respectively, $\mathrm{Resource}_{ji}$ $(1 \leqslant j \leqslant n)$ represents the resource surplus of the j node, and Other_i represents other factors that can be considered (e.g., distance).

(2) Action space. The action space can be divided into continuous and discrete action spaces. The length of policy or $Q(s,a)$[4] output by a neural network is consistent with the number of actions that can be taken in the discrete space. However, in a resource allocation problem, resources can also be expressed as continuous variables. For multinode resource allocation, the number of resources is determined after selecting nodes. A stands for action space,

$$A = \mathrm{Number}(1,2,\ldots,n) + \mathrm{Quantity}(1,2,\ldots,m) \tag{6}$$

$$A = \mathrm{Number}(1,2,\ldots,n) + \mathrm{Quantity}[0,1] \tag{7}$$

Equations (6) and (7) represent discrete and continuous action spaces, respectively. Number ( ) is the number of the resource node, Quantity ( ) is the discrete allocation scheme, and [0, 1] represents continuous space, in which each value can represent an allocation scheme.

(3) Reward. The reward function is the most important factor in RL, because it can measure the quality of model prediction and decision-making. Through the reward function, RL can make improved decisions to define and optimize loss function.

Through the investigation, we found that each entity has a minimum resource guarantee, and the required resource is not linear with the value created. Therefore, we designed a value function that considers not only the economic and social value of each entity, but also the different resource needs of each entity,

$$\text{value}(x) = V_i \times \mu(x) \tag{8}$$

$$\mu(x) = \frac{1}{1+\beta \times \mathrm{e}^{-\alpha \times (x-\text{Max}(r))}} \tag{9}$$

where i represents the entity node, x represents the amount of resource allocated by the current node, V_i represents the economic and social value of current enterprises (see Section 3.1), α and β is the coefficient of the curve, and Max(r) represents the maximum resource required by the current entity.

Function curve μ is shown in Fig. 2.

Therefore, in state s, the reward of action (a_{1j}, a_{2k}) can be expressed as

$$\text{reward}(s, a_{1j}, a_{2k}) = \begin{cases} -2, & \text{node}[j]-a_{2k}<0; \\ 0, & a_{2k}=0; \\ \text{value}(\max_i)\times 0.8-\text{distance}_{ij}, & a_{2k}>\max_i; \\ \text{value}(\max_i)-\text{distance}_{ij}, & a_{2k}=\max_i; \\ \text{value}(a_{2k})-\text{distance}_{ij}, & \text{other} \end{cases} \tag{10}$$

where a_{1j}, and a_{2k} represent a_1 takes action j, and a_2 takes action k, respectively, they are different action types. node[j] represents the resource of node j, $\max_i$ represents the maximum number of resources required by node i, and distance_{ij} represents the influence of the distance factor.

3.2.2 Model design

In the allocation scenario, the allocation strategy is a continuous sequence, and each sequence can be regarded as an action of the agent. The action can be divided into continuous and discrete actions.

For the single resource allocation node and discrete action, we design a Parallel-Data-based DQN (PD-DQN) model by combining DQN[4] with parallel data[9] to select a Q_{value} by ϵ greedy policy, with the mapped action as the allocated quantity, which is shown in Fig. 3. On the basis of DQN, the problem of overestimation is eliminated by decoupling the selection of target Q_{value} action and the calculation of target Q_{value}. The Q-network is updated by calculating the loss, as shown in Eq. (11), and the Q′ network updates parameters through soft updates, as shown in Formula (12),

$$\text{Loss} = (r+\gamma \max(Q(s',a',\theta'))-Q(s,a,\theta))^2 \tag{11}$$

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta' \tag{12}$$

where γ is the attenuation coefficient and τ is the renewal coefficient.

In PD-DQN network, parallel data environment is used as the input state s. When action a is taken in state s, the environment transfers to the new state s', and the reward r is obtained. In state s', $\max(Q())$ is selected as the reward, corresponding to action a'. θ and θ' are the DQN network and target network parameters, respectively.

The output of the DQN network is the value of each action; thus, extending to a multiresource node problem is difficult. In this situation, we use the Parallel-Data-based DDPG (PD-DDPG) model by combining DDPG[5] with parallel data[9] to transform the multinode resource allocation problem into a multiobjective optimization problem with nonlinear functions; thus, the strategy is expressed by parameterized functions. The PD-DDPG model consists of actors and critics. In this model, the output of an actor is a set of deterministic strategies (i.e., selecting a fixed resource allocation node and number of allocation), and a critic is responsible for evaluating the actions of the actor output. We also add a layer of linear loss function and use Mean Square Error (MSE) to measure the gap between the action of the model output and the actual resource demand, and thus optimizing the actor and completing the strategy learning. The

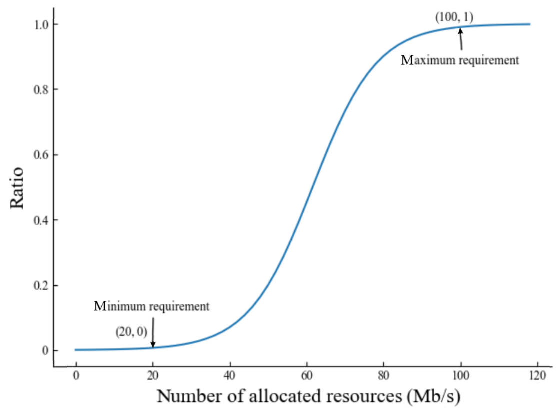


Fig. 2 Resource allocation curve μ. (For example, in bandwidth resource allocation, Max(r) is 100 Mbps, α is 0.12, and β is 0.01.)

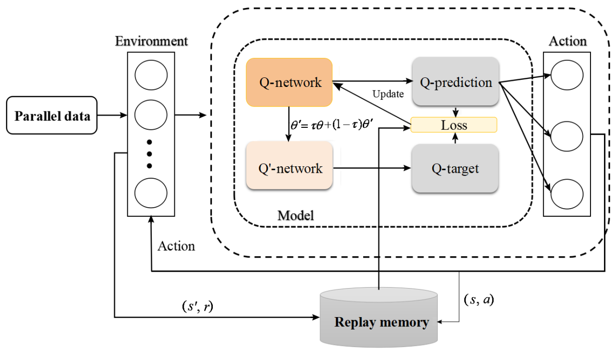


Fig. 3 PD-DQN model structure.

PD-DDPG model structure is shown in Fig. 4. In Fig. 4, DDPG includes four networks: Actor, Actor$'$, Critic, and Critic$'$. The parameters of neural network are θ, θ', ω, and ω'. Action a includes resource node and resource quantity.

In the actor-critic structure of RL, the output of the actor is approximated by a parameter equation (neural network), which is used to obtain a definite action from the current state. The critic model uses the Bellman equation of action to measure the quality of action. The DDPG algorithm adopts an off-policy style; thus, the difference between behavior strategy and evaluation strategy can increase agent exploration. At the same time, the robustness of the model is improved by adding noise to the deterministic behavior strategy. The goal of model training is to maximize the Q_{value} (see Eq. (13)) and minimize the loss of critic network (see Eq. (14)).

$$Q_{\text{value}} = Q_\omega\,(s, f(s,\theta)) - \text{MSE}(\text{size}, \text{max}_{\text{dem}}) \quad (13)$$

$$\text{Loss} = \text{MSE}(Q_\omega\,(s, f(s,\theta)), r + \gamma Q_{\omega'}(s', f(s',\theta'))) \quad (14)$$

where size represents the number of resources allocated, max_{dem} represents maximum resource demand, $Q_\omega(\,)$ is the critic network output, and $f(\,)$ represents the nonlinear transformation of neural network..

In a complex environment, the parameters change greatly due to the influence of learning rate, thus leading to the instability of the model. To resolve these problems, we use the Proximal Policy Optimization 2 (PPO2)[20] algorithm. Based on the actor-critic structure, the PPO2 algorithm introduces the advantage function and clip strategy to ensure the smooth updating of parameters.

4 Evaluation

As the most important means of communication in the world, submarine optical cables are mainly used for long-distance communication. More than 95% of international communication is carried out through submarine optical cables.

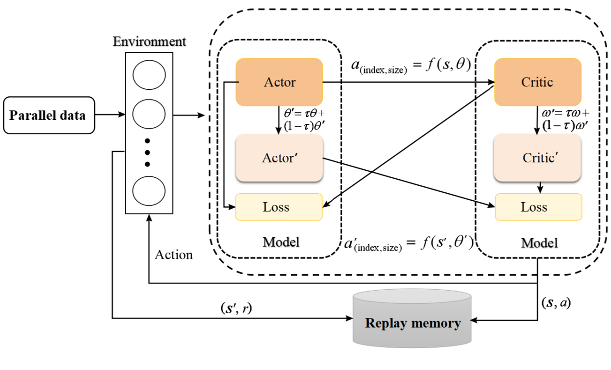


Fig. 4 PD-DDPG model structure.

With deep-sea emergency taken as an application scenario, submarine cable distribution is selected as the deduction object. When the submarine cable breaks unexpectedly in the normal operation process, the parallel data evolution method is used to evaluate the effect of events, and the RL algorithm is used to model the current state. In accordance with the model results, the optimal bandwidth allocation strategy is selected to guide practice.

4.1 Experimental design and results

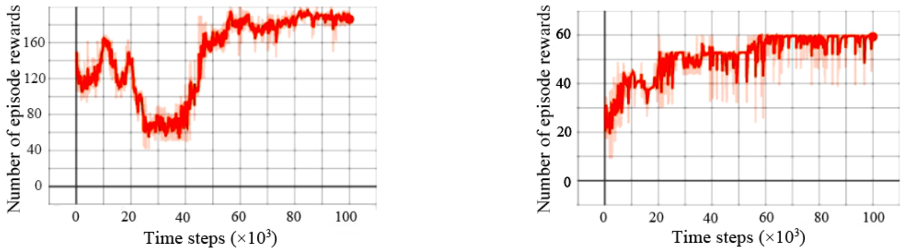
For single-resource and multi resource nodes, as well as continuous and discrete allocation strategies, we use the different models proposed in Section 3.2 to simulate bandwidth allocation and use Formula (10) to calculate the reward. At the same time, we consider the distance between each resource node and an enterprise. The goal of the experiment is to maximize the reward of all enterprises. The experimental steps are as follows:

Data preprocessing. From the National Bureau of Statistics (http://www.stats.gov.cn/), we have collected economic and social data on industries in recent years (e.g., GDP, resources, employment rate, and environment). The economic and social value of each industry is calculated through the value model. The economic and social model is normalized to make the value between [1,10] due to the lack of different data dimensions in various industries.

Data generation. We randomly simulate the missing data. The number of bandwidth required by an industry ranges from 20 Mbps to 100 Mbps, the number bandwidth resources in a single resource node ranges from 100 Mbps to 1000 Mbps, and the distance between industries and different resource nodes is 1–10 km.

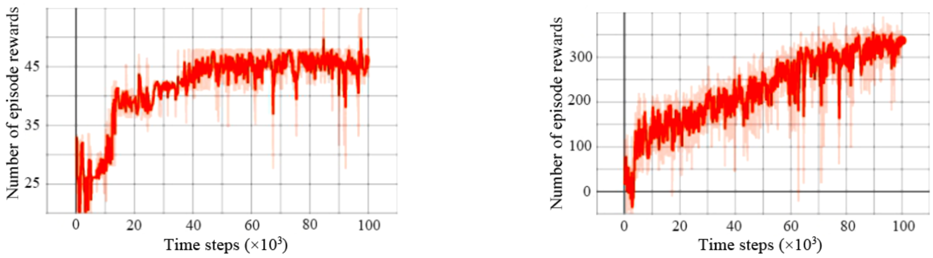
Model allocation simulation. For a single-resource node, the allocation strategy adopts three discrete allocation values, which are 10 Mbps, 50 Mbps, and 100 Mbps, seperately, we use the PD-DQN model to allocate. For multiresource nodes, the allocation strategy adopts the same method of distribution, which are allocated by the PD-DDPG model. For multiresource nodes, the allocation strategy adopts the continuous value with the range of 0–100 Mbps; we allocate through the PPO2 model. Lastly, a comparative study between RL and other methods is conducted.

The experimental results of PD-DQN, PD-DDPG, and PPO2 are shown in Figs. 5–7, respectively. After multiple training, the final reward of each algorithm tends to be stable and can reach the highest value. In



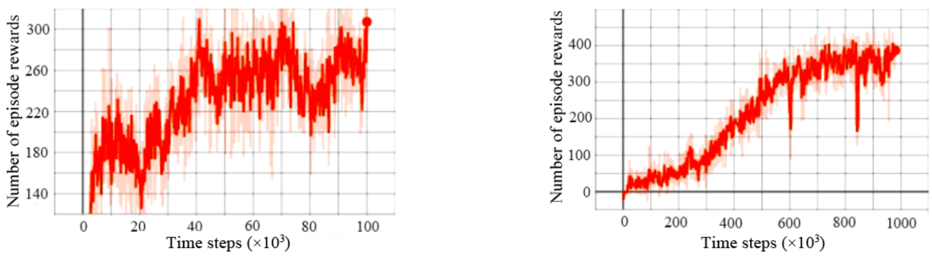
(a) (b)

Fig. 5 PD-DQN model simulates single-node bandwidth allocation adopted discrete allocation strategies. (a) Reward change tendency of 20 enterprise nodes, and (b) reward change tendency of 100 enterprise nodes .



(a) (b)

Fig. 6 PD-DDPG model simulates bandwidth allocation adopted discrete allocation strategies. (a) Reward change tendency of 1 resource allocation node and 100 enterprise nodes, and (b) reward change tendency of 5 resource allocation nodes and 100 enterprise nodes.



(a) (b)

Fig. 7 PPO2 model simulates bandwidth allocation. (a) Reward change tendency of 5 resource allocation nodes and 100 enterprise nodes adopted discrete allocation strategy, and (b) reward change tendency of 5 allocation nodes and 100 enterprise nodes adopted continuous allocation strategy.

the experiment of one resource allocation node and 100 enterprise nodes, PD-DDPG reaches the stable point faster than PD-DQN by using the actor-critic structure. In the experiment of five resource allocation nodes and 100 enterprise nodes, PD-DDPG has fast update and convergence speeds, so the reward changes smoothly. Compared with the discrete allocation strategy, the continuous allocation strategy has a larger action space, so the parameter update is slower and requires more training time steps. Experimental results show that our proposed social evolution modeling approach can adapt to various scenarios.

As shown in Table 1, experiments on RL algorithms (PD-DQN, PPO2), random trials, as well as genetic algorithms are conducted, and the average values of multiple final rewards are compared. For the allocation algorithm that is not suitable for this scenario, the allocation result is represented by "–". In the discrete scenario, at 20 and 100 enterprise nodes, the final reward of the PD-DQN model is higher than that of the random trail and genetic algorithms. In the continuous scenario, with 100 enterprise nodes, the final reward

Table 1 Experimental comparison on reward.

Number of nodes	Number of enterprises	Distribution type	PD-DQN	PPO2	Random trial	Genetic algorithm
1	20	Discrete	190	–	184	161
1	100	Discrete	117	–	–49	93
1	100	Continuous	–	49	–	–101

of the PPO2 model is higher than that of the genetic algorithm. Experiments show that our proposed methods based on RL and parallel data have high reward, good robustness, and effectiveness.

4.2 Discussion

We use three different RL algorithms (i.e., PD-DQN, PD-DDPG, and PPO2) to allocate bandwidths for single-node, multinode, discrete, and continuous allocation strategies. Each model can find a high reward through continuous exploration and simulation. After many trainings, the algorithm gradually converges, and the reward stabilizes. By combining with social evolution methods, the results of each step can be deduced, and the optimal strategy for resource allocation can be found.

In the approach of social evolution modeling, we use the economic and social value model to calculate the reward of RL. To some extent, the reward can reflect the social loss. We can analyze social loss through the reward of RL. If we do not take any action or it is not the optimal strategy, we obtain less rewards. The final strategy obtained by RL can achieve the highest reward, thus reducing social loss.

For the assignment problem, we also attempt to find the optimal value through the genetic algorithm and random trials. The genetic algorithm adopts the "survival of the fittest" evolutionary method between successive generations of individuals to solve the optimal problem. Through reward or fitness, they can eventually find a better value. The essential difference between the genetic algorithm and RL is that RL not only finds the optimal value, but also learns the representation from state to action. Moreover, the genetic algorithm hardly constructs a fitness function for complex problems.

5 Conclusion and Future Work

The occurrence of social abnormal events causes a chain reaction that affects social stability, leading to serious harm and loss. If no effective mechanism can be used to deal with these events, the situation may worsen. Based on the idea of parallel data, we propose a parallel-data-based social evolution modeling approach to deduce the evolution of events. Based on RL, we design decision-making mechanisms for different situations to provide optimal strategies for handling abnormal events. Experiments in different scenarios show that compared with other methods, the proposed approach can find the optimal strategy. The evaluations also show that our approach is versatile, robust, and efficient.

In the future, we will consider forming a more complete digital social system to deduce the state evolution. We commit to improving the RL algorithm and the training speed.

Acknowledgment

This research was supported by the National Natural Science Foundation of China (No. 62072469), the National Key R&D Program of China (No. 2018YFE0116700), the Shandong Provincial Natural Science Foundation (No. ZR2019MF049, Parallel Data Driven Fault Prediction under Online-Offline Combined Cloud Computing Environment), and the Fundamental Research Funds for the Central Universities (No. 2015020031).

References

[1] W. Jiang and L. Zhang, Geospatial data to images: A deep-learning framework for traffic forecasting, *Tsinghua Science and Technology*, doi: CNKI:SUN:QHDY.0.2019-01-006.

[2] H. Qiao, X. Wan, Y. Wan, S. Li, and W, Zhang, A novel change detection method for natural disaster detection and segmentation from video sequence, *Sensors*, vol. 20, no. 18, p. 5076, 2020.

[3] G. Xu, C. Cheng, W. Yang, W. Xie, L. Kong, R. Hang, F. Ma, C. Dong, and J. Yang, Oceanic eddy identification using an AI scheme, *Remote Sensing*, vol. 11, no. 11, p. 1349, 2019.

[4] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, Playing atari with deep reinforcement learning, arXiv preprint arXiv: 1312.5602, 2013.

[5] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, Continuous control with deep reinforcement learning, arXiv preprint arXiv: 1509.02971, 2015.

[6] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, et al., Mastering the game of go without human knowledge, *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[7] Z. Zhang, H. Li, L. Zhang, T. Zheng, T. Zhang, X. Hao, X. Chen, M. Chen, F. Xiao, and W. Zhou, Hierarchical reinforcement learning for multi-agent MOBA game, arXiv preprint arXiv:1901.08004, 2019.

[8] A. Asheralieva and D. Niyato, Hierarchical game-theoretic and reinforcement learning framework for computational offloading in UAV-enabled mobile edge computing networks with multiple service providers, *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8753–8769, 2019.

[9] X. Liu, X. Wang, W. Zhang, J. Wang, and F. Wang, Parallel data: From big data to data intelligence, (in Chinese), *Pattern Recognition and Artificial Intelligence*, vol. 30, no. 8, pp. 673–682, 2017.

[10] A. E. Saddik, Digital twins: The convergence of multimedia technologies, *IEEE multimedia*, vol. 25, no. 2, pp. 87–92, 2018.

[11] S. Obushnyi, R. Kravchenko, and Y. Babichenko, Blockchain as a transaction protocol for guaranteed transfer

of values in cluster economic systems with digital twins, in *Proc. of 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology* (*PIC S&T*), Kyiv, Ukraine, 2019, pp. 241–245.

[12] E. Bryndin, Digital cyclical ecological regional self-sufficient economy, *Journal of Applied Science, Engineering, Technology, and Education*, vol. 2, no. 2, pp. 104–111, 2020.

[13] G. Song, F. Khan, and M. Yang, Probabilistic assessment of integrated safety and security related abnormal events: A case of chemical plants, *Safety Science*, vol. 113, pp. 115–125, 2019.

[14] Y. Tian and X. Liu, A deep adaptive learning method for rolling bearing fault diagnosis using immunity, *Tsinghua Science and Technology*, vol. 24, no. 6, pp. 750–762, 2019.

[15] M. Schluse and J. Rossmann, From simulation to experimentable digital twins: Simulation-based development and operation of complex technical systems, in *Proc. of 2016 IEEE International Symposium on Systems Engineering* (*ISSE*), Edinburgh, UK, 2016, pp. 1–6.

[16] H. Van Hasselt, A. Guez, and D. Silver, Deep reinforcement learning with double Q-learning, arXiv preprint arXiv: 1509.06461, 2015.

[17] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, Dueling network architectures for deep reinforcement learning, in *Proc. of International Conference on Machine Learning*, New York, NY, USA, 2016, pp. 1995–2003.

[18] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, Deterministic policy gradient algorithms, in *Proc. of 31st International Conference on Machine Learning* (*ICML*), Beijing, China, 2014, pp. 387–395.

[19] V. Mnih, A. P. Badia, M. Mirza, A. Graves, T. Lillicrap, T. Harley, D. Silver, and K. Kavukcuoglu, Asynchronous methods for deep reinforcement learning, in *Proc. of International Conference on Machine Learning*, New York, NY, USA, 2016, pp. 1928–1937.

[20] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, Proximal policy optimization algorithms, arXiv preprint arXiv: 1707.06347, 2017.

[21] J. Sacks, *The Money Trail*: *Measuring Your Impact on the Local Economy Using LM3*. Londen, UK: New Economics Foundation, 2002.

[22] R. Millar and K. Hall, Social return on investment (SROI) and performance measurement: The opportunities and barriers for social enterprises in health and social care, *Public Management Review*, vol. 15, no. 6, pp. 923–941, 2013.

Weishan Zhang received the PhD degree from Northwestern Polytechnical University, China in 2001. He is currently a full professor and the deputy head for research at the College of Computer Science and Technology, China University of Petroleum. He has published more than 100 papers. His current H-index according to Google scholar is 17. His current research interests are big data platforms, pervasive cloud computing, and service oriented computing.

Zhaoxiang Hou received the BEng degree from Shandong Agriculture University in 2019. He is now a master student at the College of Computer Science and Technology, China University of Petroleum (East China), Qingdao, China. His main research interests include deep learning and big data processing.

Xiao Wang received the PhD degree from the University of Chinese Academy of Sciences in 2016. She is the executive vice president of Qingdao Institute of Intelligent Industry Technology, and an associate researcher at State Key Laboratory of Management and Control for Complex System, Institute of Automation, Chinese Academy of Sciences. Her main research interests include social computing, intelligent transportation and Internet of vehicles, and artificial intelligence.

Zhidong Xu received the MS degree from National Defense University in 2009, he is now a PhD candidate in Beijing Normal University. He is a researcher at the Institute of National Security, National Defense University. His main research interests include natural language processing and social computing.

Xin Liu received the PhD degree from Nankai University, China in 2012. She is currently an associate professor at the College of Computer Science and Technology, China University of Petroleum, Qingdao, China. Her research interests include social networks, data mining, and big data processing.

Fei-Yue Wang received the PhD degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, New York, USA in 1990. He is the State Specially Appointed Expert and the director of the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His main research interests include methods and applications for parallel systems, social computing, and knowledge automation.