

Research Article

Prediction of Soil Heavy Metal Content Based on Deep Reinforcement Learning

Yongqi Zhao,¹ Zhangdong Wei ,¹ and Jing Wen²

¹Miami College of Henan University, Kaifeng 475002, Henan, China

²Hunan Career Technical College Nonferrous Metal, Zhuzhou, Hunan 412006, China

Correspondence should be addressed to Zhangdong Wei; 10380012@vip.henu.edu.cn

Received 4 March 2022; Revised 24 March 2022; Accepted 24 March 2022; Published 15 April 2022

Academic Editor: Hangjun Che

Copyright © 2022 Yongqi Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Since the prediction accuracy of heavy metal content in soil by common spatial prediction algorithms is not ideal, a prediction model based on the improved deep Q network is proposed. The state value reuse is used to accelerate the learning speed of training samples for agents in deep Q network, and the convergence speed of model is improved. At the same time, adaptive fuzzy membership factor is introduced to change the sensitivity of agent to environmental feedback value in different training periods and improve the stability of the model after convergence. Finally, an adaptive inverse distance interpolation method is adopted to predict observed values of interpolation points, which improves the prediction accuracy of the model. The simulation results show that, compared with random forest regression model (RFR) and inverse distance weighted prediction model (IDW), the prediction accuracy of soil heavy metal content of proposed model is higher by 13.03% and 7.47%, respectively.

1. Introduction

Soil is an important resource for human survival and development, as well as the lifeline of the whole ecosystem. However, with the improvement of production level and the rapid development of economics, the problem of soil pollution has become more and more serious. And heavy metal pollution is one of the most difficult pollutants among all soil pollution sources, which is difficult to be degraded by microorganisms. It not only affects the growth of crops and leads to the decline of crop yield, but also may enter the human body through eating and other ways, so as to threaten human life and health. Liu Xingmei et al. believed that there were both noncarcinogenic and carcinogenic potential risks to human health by eating vegetables contaminated by heavy metals. Therefore, it is necessary to study heavy metals in soil. In recent years, with people's attention to soil heavy metal pollution, more and more relevant researches are carried out, which are more and more in-depth [1]. Lan et al. used pinealone-biochar to stably passivate Pb, Cu, Zn, Cr, and As in soil. As can be seen, the addition of pinealone-biochar and the coexistence of indigenous microorganisms

can effectively reduce biological activity of heavy metals and accelerate passivation of heavy metals [2]. Khan Imran and Yang Dong et al. found that silicon in soil has a certain detoxification mechanism for heavy metals, which provides a certain theoretical basis for reducing toxicity of heavy metals in soil [3, 4]. Guo Xujing and Liu Xingmei et al. used spectroscopy combined with parallel factor analysis and two-dimensional correlation spectra to study the complexing characteristics of heavy metals Cr (III) and Cu (II) in soil with biochar source WEOM [1, 5]. Using corn cobs as raw materials, biochar-derived water extractable organic matter can be obtained under low temperature (300°C) pyrolysis conditions, which can be used for soil heavy metal remediation. Rana Anuj and Zhang Jiachun et al. studied the biological activities of heavy metals (Cd and Cr) in crops and believed that heavy metal pollution in soil could be dealt with by reducing the biological activities of heavy metals [6, 7]. The above research results show that the current research on heavy metals in soil mainly focuses on the treatment, and there are few researches on the prediction of contents, while the content prediction of soil heavy metals is the prerequisite for treating heavy metals in soil [8]. Thus a

prediction model based on deep reinforcement learning is proposed.

2. Basic Methods

2.1. Deep Q Networks. Deep Q network is a representative algorithm of deep reinforcement learning. Combining the perception ability of deep learning with the decision-making ability of reinforcement learning, the spatial coverage problem of state-action in Q table can be solved [9]. The calculation of target value of deep Q network can be solved by state value function, which is shown in formula (1) [10]:

$$y = V^s(s; \theta') = r + \gamma \max_{a'} Q^s(s', a'; \theta'), \quad (1)$$

where θ is online value network; θ' is target value network; S is current state; a is action in s state; r is reward of agent for a ; s' means the next state reached by agent in s state when action a is taken.

Figure 1 shows the operation process of deep Q network.

There is the problem of overestimation existing in the combination of deep Q network with neural network and reinforcement learning, which leads to the estimation error of prediction output value of the model and cannot truly reflect the actual value [11]. In addition, the convergence speed of deep Q network is slow, and the stability after model convergence is poor [12]. Therefore, to solve these problems, state value reuse and fuzzy membership factor are utilized to improve deep Q network. At the same time, adaptive inverse distance weighted method is used to adjust the hyperparameters to improve the prediction accuracy.

2.2. Improvement of Deep Q Network

2.2.1. State Value Reuse. State value reuse is to combine the partial output of value function with the obtained reward value to form total reward value, which can replace the original environment reward value to train the agent, and make the total reward value participate in the weight update of Q network. After each round of training, the network error is calculated, and the weight is updated. As can be seen, the calculation method of reward value in the deep Q network model of state value reuse is shown in formula (2) [13]:

$$R(s, a, p) = r(s, a, p) + \lambda V(s; \theta, \beta), \quad (2)$$

where s is current state; d is the action performed in s state. P is state probability after the execution of a , and $r(s, a, p)$ is reward value of environment to action. $V(s; \theta, \beta)$ represents partial output value; λ is regulating factor, which is responsible for determining the dominant position of reward value returned by environment in the total reward value, so as to avoid the influence of the size of reward value returned by environment on the model convergence.

2.2.2. Dynamic Fuzzy Membership Factor. As can be seen the deep Q network can be optimized by state value reuse. The environmental feedback reward value and state value of

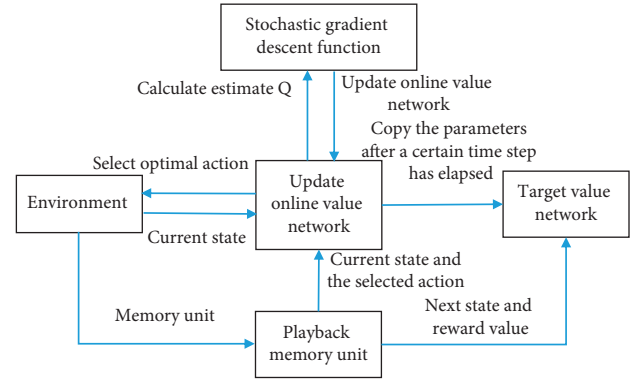


FIGURE 1: Operation process of deep Q network.

Q network are combined in a certain proportion. Moreover, the combination mode remains unchanged in the whole network model training. In practical application, Q network is not sensitive to environmental feedback reward at the initial stage of training, so it cannot accurately judge the advantages and disadvantages of current environment. Therefore, it is necessary to reduce the proportion of state value to improve agent's sensory ability to environment. In the middle of training, parameters of Q network move to the optimal solution, and the network performance is getting higher and higher. The regulatory factor should be appropriately increased to enhance the reward or punishment of environment to agent performing actions. At the later stage of training, parameters of Q network basically remain stable, and the maximum value of regulatory factor should be basically maintained to improve the model convergence rate [14, 15]. Thus in the training process of Q network, the proportion of state value and environmental feedback reward value in total reward value should change dynamically. In addition, dynamic fuzzy membership factor δ is introduced in this paper, which is shown in formula (3) [16]:

$$\delta = \frac{1}{2} - \frac{1}{2} \cos\left(\left(\frac{\pi}{n_{\text{total}}}\right) * n\right), \quad (3)$$

where n is the number of current training steps; n_{total} is the total number of predicted training steps. δ changes with the change of n . When n is small, δ tends to 0. When n is large, δ increases gradually. When $n = n_{\text{total}}$, δ approaches 1.

To sum up, the total reward value calculation method of the improved deep Q network model is as follows [17]:

$$R(s, a, p) = r(s, a, p) + \delta V(s; \theta, \beta), \quad (4)$$

where s is current state; a is the action performed in s state. P is probability of environment transferring to next state, and $r(s, a, p)$ is the reward value of environment to action. $V(s; \theta, \beta)$ represents partial output value of value function in Q network; δ is regulatory factor.

2.3. Adaptive Inverse Distance Weighted Method. Deep Q network model determines the reward value of training environment by the observing interpolation points, which is shown in formula (5). However, the observed value of

interpolation point is an unknown value, which is usually predicted by inverse distance weighted method. According to [18], the inverse distance weighted method has poor interpolation effect because it cannot adapt to complex terrain structure. To solve this problem, an adaptive inverse distance weighted method is proposed.

$$r(s, a) = \sum_{i=1}^m w(i) \left\{ \left[\gamma^{\#}(s, h_i) - \gamma^*(h_i) \right]^2 - \left[\gamma^{\#}(s', h_i) - \gamma^*(h_i) \right]^2 \right\}, \quad (5)$$

where s is current state of agent; a is action performed in s state; s' is the entered next state where agent performs action a in s state; $\gamma^{\#}(s, h_i)$ is fitting value of h_i on mutation function curve corresponding to s ; $\gamma^{\#}(s', h_i)$ is the fitting values of h_i corresponding to s' ; $\gamma^*(h_i)$ is discrete points of mutation function of h_i ; r is reward value with environment to agent carrying out action a in s state.

In the adaptive inverse distance weighted method, hyperparameters of each known point in the model are learned, and the nearest adjacent statistics of each point are calculated. Furthermore, the multidimensional spatial discrete points are formed, and spatial modeling is done by Kriging interpolation method [19, 20]. Finally, the corresponding coordinates of interpolation points to be predicted are input into the spatial model, so as to obtain the corresponding hyperparameters of interpolation points. Thus the final predicted values can be obtained by using this hyperparameter to inversely weight the interpolation point. The adjacent distance is calculated as follows [21]:

$$d_{\text{avg}} = \frac{1}{2(N/A)^{0.5}}, \quad (6)$$

where N is the total number of sample points in research area; A is the area of study area.

The nearest adjacent statistic can be calculated by formula (7) [22]:

$$M = \frac{d_n}{d_{\text{avg}}}, \quad (7)$$

where d_n is the nearest expected distance of prediction point; d_{avg} is the expected nearest distance of study area.

3. Prediction of Soil Heavy Metal Content Based on Improved Deep Q Network

The prediction process of soil heavy metal content is designed as follows:

- (1) Preprocess and divide the collected and sorted original soil heavy metal content into sample point data set and interpolation point data set. The sample point data is the sample data set with known observation value, and the interpolation point data set is the sample data set with unknown observation value.
- (2) Use adaptive deep Q network to train the sample point data set and record the inverse distance weighted optimal hyperparameter of each point.

- (3) Compose all optimal hyperparameters into a new data set, and calculate spatial discrete points of mutation function of the data set to obtain the mutation function model, which is shown in formula (8) [23].

$$\gamma(\vec{h}) = \frac{1}{V} \iiint_V [Z(s + \vec{h}) - Z(s)]^2 dV, \quad (8)$$

where s is a point in support set V in random field z , and \vec{h} is any two-point vector in V . When mutation function is in a second-order stable process, the above equation can be rewritten as

$$\begin{aligned} \gamma(s_1, s_2) &= \text{var}[Z(s_2) - Z(s_1)] \\ &= E[Z(s_2) - Z(s_1) + \mu(s_2) - \mu(s_1)]^2, \end{aligned} \quad (9)$$

where E and var are mathematical expectation and variance operations, and μ is the mathematical expectation of specific point in random field. Since the covariance is related to \vec{h} , formula (9) can be expressed as

$$\begin{aligned} \gamma(s_1, s_2) &= E[Z(s_2) - Z(s_1)]^2, \\ \vec{h} &= s_1 - s_2. \end{aligned} \quad (10)$$

Considering that the covariance is related to Euclidean distance of the two spatial points and has nothing to do with the direction, formula (10) can be expressed as

$$\gamma(s_1, s_2) = \gamma_s(|\vec{h}|). \quad (11)$$

- (4) Use the weight coefficient, as shown in formula (13), establish the fitting standard, as shown in formula (14), and estimate the parameters of mutation function model to obtain the optimal mutation function model and its optimal parameters:

$$w(i) = \frac{N(h_i)}{\sum_{j=1}^m N(h_j)}, \quad (12)$$

$$C_w = \min \left\{ \sum_{i=1}^m w(i) \left[\gamma^{\#}(h_i) - \gamma^*(h_i) \right]^2 \right\}, \quad (13)$$

where $\gamma^{\#}(h_i)$ is fitting value of point h_i on the mutation function curve, $\gamma^*(h_i)$ is discrete point of mutation function of h_i , and C_w is the objective function of model parameter optimization.

- (5) Adopt Kriging method to model the new data set to obtain the hyperparameter distribution model.
- (6) Input the data of interpolation points into hyperparameter distribution model, and introduce the obtained hyperparameters and corresponding interpolation point data into the inverse distance weighted algorithm, so as to obtain the final predicted value of soil heavy metal content.

The above process can be illustrated in Figure 2.

4. Simulation Experiment

4.1. Data Sources and Preprocessing. The soil heavy metal content data set of suburban farmland of Changsha is selected as the experimental data set. Considering that there are geographic data including latitude and longitude in the data set, it is not suitable for direct input into the model. Therefore, the geographic data including latitude and longitude are converted into data in Cartesian coordinate system before the experiment. Meanwhile, to reduce magnitude of geographical data, the converted data coordinates are shifted to the origin as a whole. In addition, considering the existence of missing values in the data set and the different ranges of data values corresponding to different features, mean interpolation or deletion is carried out for the data, and z -score is standardized and preprocessed, which are shown as follows:

$$y_i = \frac{x_i - \bar{x}}{s},$$

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (14)$$

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

4.2. Evaluation Indicators. The mean square error (MSE), root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) are selected as evaluation indicators in this experiment, and the calculation methods are as follows [24]:

$$\text{MSE} = \frac{\sum_{i=1}^n (X_i - x_i)^2}{N},$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (X_i - x_i)^2}{N}}, \quad (15)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right|,$$

$$\text{MAE}(X, h) = \frac{1}{m} \sum_{i=1}^m |h(x_i) - y_i|.$$

4.3. Parameter Setting. The initial parameters of all models are set the same, and the specific settings are as follows: The maximum number of training rounds is 5000; the number of training samples in each round is 64; the random seed is 1; the maximum memory size is 500; the initial state of agent is [2, 14]; the difference of training rounds between weight updates of Q network is 200; the learning rate of

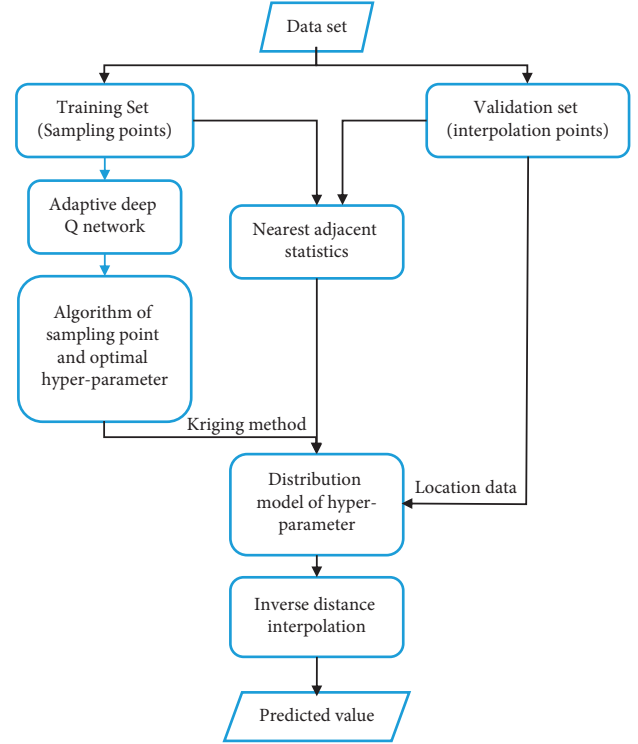


FIGURE 2: Model prediction process.

convolutional neural network is 0.001; the probability factor of e-greedy algorithm is 0.9.

4.4. Experiment Results

4.4.1. Estimation of Model Parameters. To verify the effectiveness of proposed model, the proposed model, deep Q network model, double deep Q network model, and competing deep Q network model are used to learn and train the parameter estimation of inverse distance weighted algorithm, and the training of different models on the data set is recorded. The results are shown in following figures.

As can be seen, the abscissa is the number of training rounds, and the ordinate is the error between predicted value and observed value (mg/kg). To reflect change trend of prediction error more clearly, smoothing processing is carried out on the basis of original graph. Here the initial points of different models are different. The reason is that the agent performs several random decisions before the model training, resulting in different agent states, and then the initial state and initial points of the models are different. However, different initial states of agents do not affect the model performance. For example, on the Ni data set, the initial states of deep Q network and competing Q network model are slightly lower than the proposed model. The convergence speed of proposed model is higher than deep Q network and competing Q network model, which indicates that the initial states of agents do not affect the model performance.

Figure 3 and Figure 4 show that the prediction error of model is not monotone decreasing when all models are

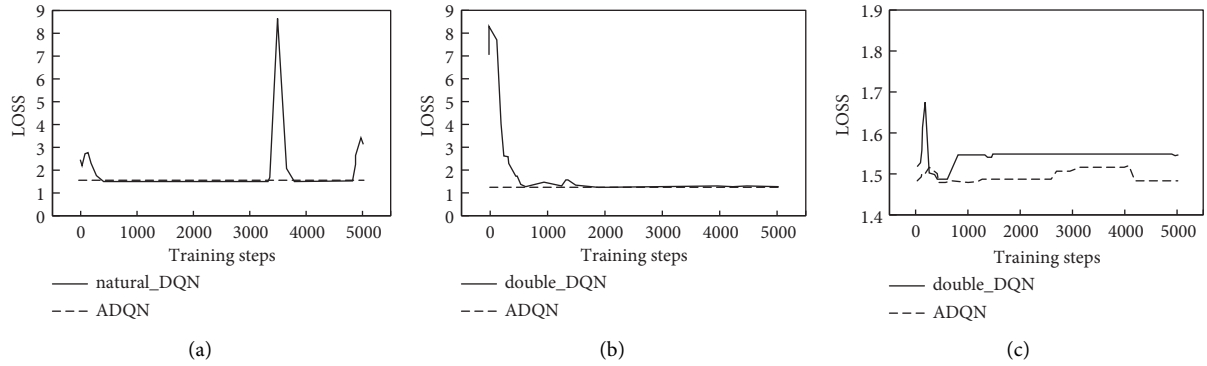


FIGURE 3: Training of different models on Cd data set.

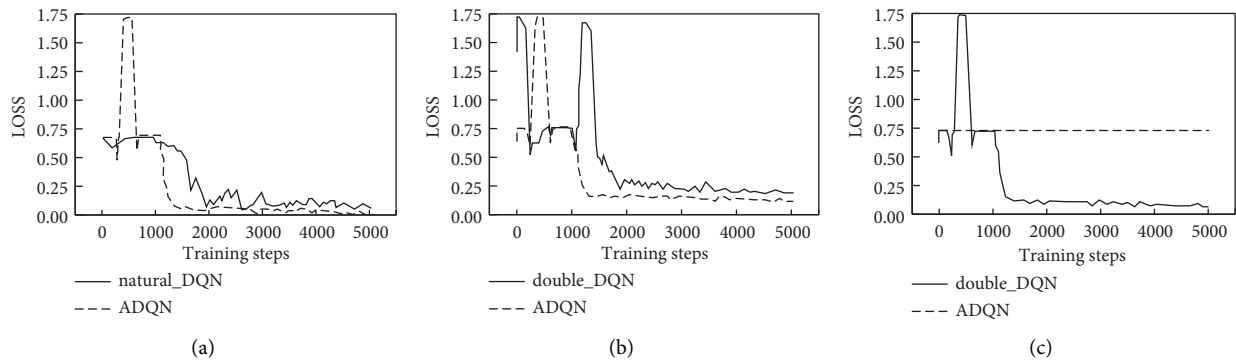


FIGURE 4: Training of different models on Cr data set.

trained, but there is a situation where the minimum error is reached and then moves towards a larger error, and the reason is that the model has reached local optimal state in this learning stage. However, the proposed model can jump out of the local optimal state through adaptive dynamic fuzzy membership factor and converge to the global optimal value; thus the proposed model has certain superiority. Figures 3–5 show that after agent reaches optimal state, it will return to poor state. The reason is that when agent adopts e-greedy strategy to make action decision, there are some unnecessary actions performed, so as to reduce the convergence rate of model and make model return to poor state after reaching optimal state. Figures 3–6 show that, compared with contrast models, the proposed model has faster convergence speed and more stable performance, and it can avoid falling into local optimum.

The convergence time of different models when performing parameter estimation on different soil heavy metal data sets is shown in Table 1. As can be seen, min is minimum convergence time of model repeating 10 times experiment; mean is the average convergence time of model repeating 10 times experiment; >> represents that the model still does not converge after reaching the maximum number of training rounds. According to the table, the convergence time of the same model on different data sets is different, and the convergence time of different models on the same data set is also different. The difference between minimum convergence time and average convergence time of all

models is small, which indicates that each model is stable and the experiment results are reliable. The competing deep Q network model does not converge after reaching the maximum number of training rounds on Cr and Ni data sets. In addition, minimum convergence time and average convergence time of the proposed adaptive deep Q network model on each data set are smaller than those of contrast models; thus the proposed model has better performance and certain advantages. In summary, the convergence speed of proposed model is better than that of contrast models, the performance is better, and the expected effect can be achieved.

4.4.2. Prediction Results of the Model. To verify the prediction effect of proposed model on soil heavy metal content, the prediction effect between proposed and random forest regression model (RFR) and inverse distance weighted model (IDW) is compared. The results are shown in Figures 7~10, and the comparison between predicted value and actual value on test set is shown in Figure 11. In Figures 7~10, the abscissa and ordinate are sampling point and predicted value (mg/kg), respectively. In Figure 11, *a* is the verification result on Cd data set; *b* is the verification result on Cr data set; *c* is the verification result on Ni data set; and *d* is the verification result on Pb data set.

As can be seen from Figures 7(a) and 7(c), for IDW model, the error between predicted value and actual value is

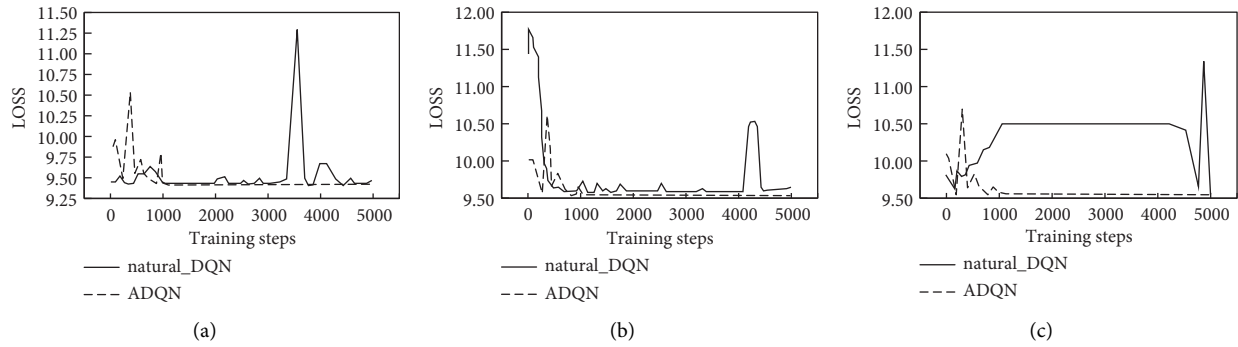


FIGURE 5: Training of different models on Ni data set.

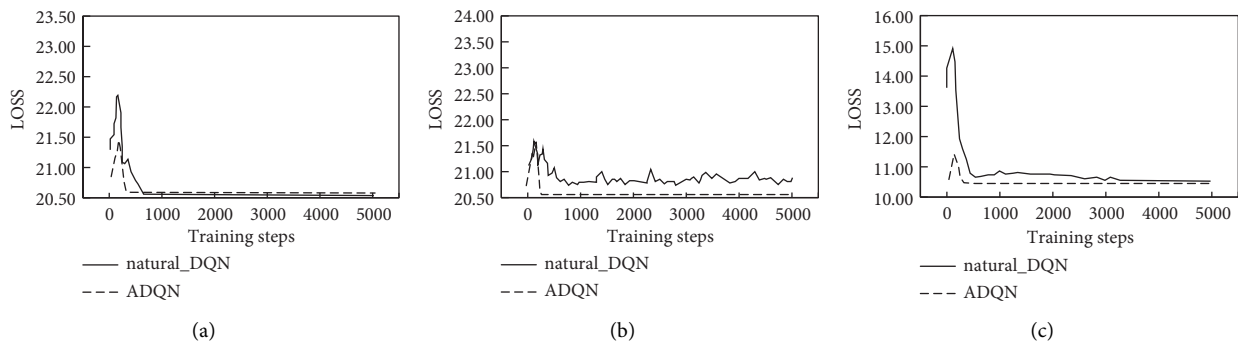


FIGURE 6: Training of different models on Pb data set.

TABLE 1: Convergence time of different models on different data sets (s).

Model —	Cd		Cr		Ni		Pb	
	Min	Mean	Min	Mean	Min	Mean	Min	Mean
DQN	435	4.68	19.50	20.97	6.60	7.10	6.77	7.28
DDQN	4.42	4.76	17.79	19.13	6.39	6.88	9.33	10.04
DuDQN	5.57	5.99	>>	>>	>>	>>	7.45	8.02
ADQN	3.56	3.83	15.22	1637	3.67	3.95	4.59	4.94

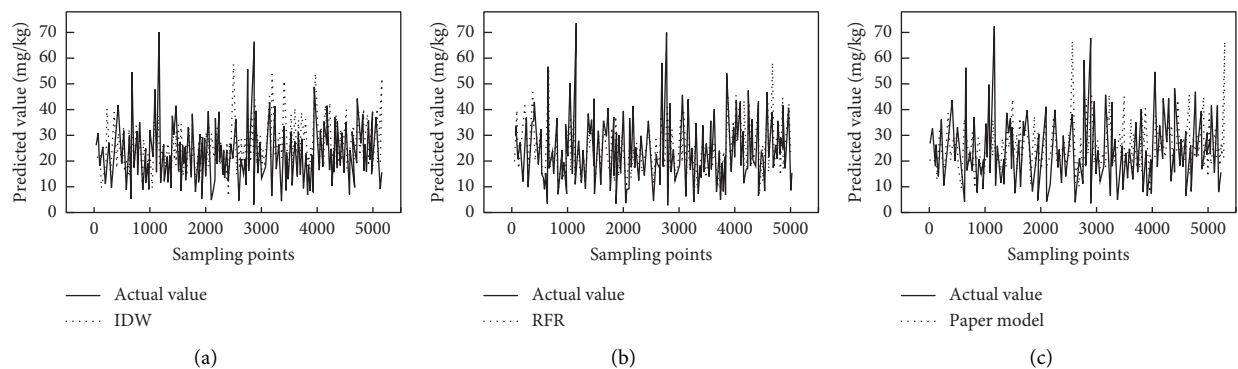


FIGURE 7: Predicted value of different models on Cd data set.

obvious. The proposed model has a large error between predicted value and actual value, and the predicted values of most sampling points basically coincide with the actual

values, thus the hyperparameters of model are adaptively adjusted, and the prediction accuracy of model to spatial data is improved. Moreover, Figures 7(b), 8(b), 9(b), and

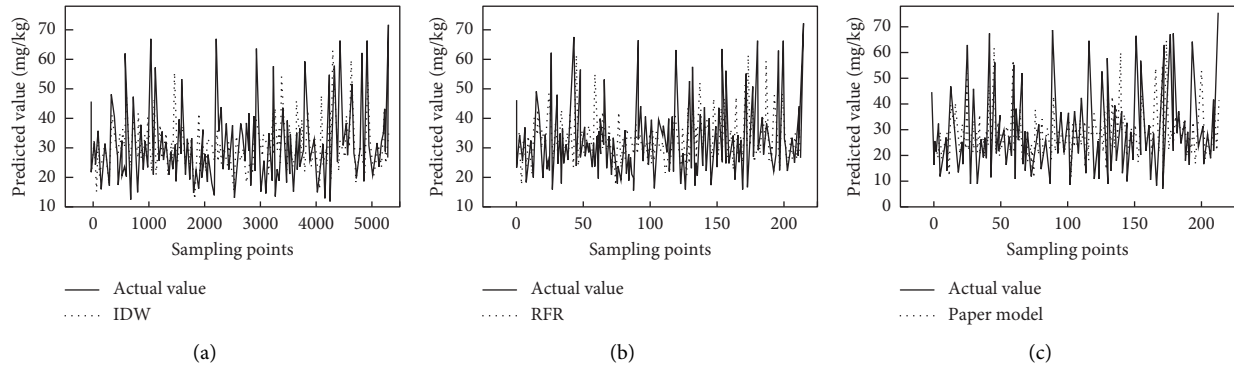


FIGURE 8: Predicted value of different models on Cr data set.

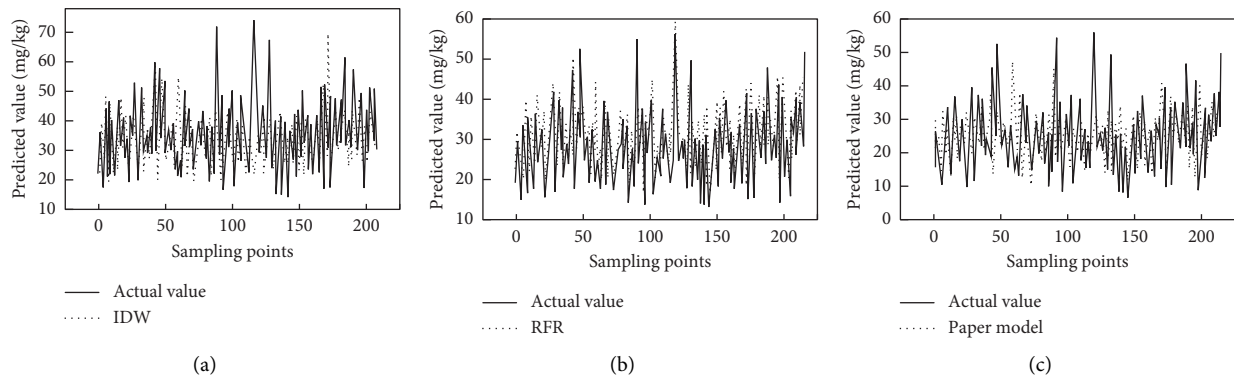


FIGURE 9: Predicted value of different models on Ni data set.

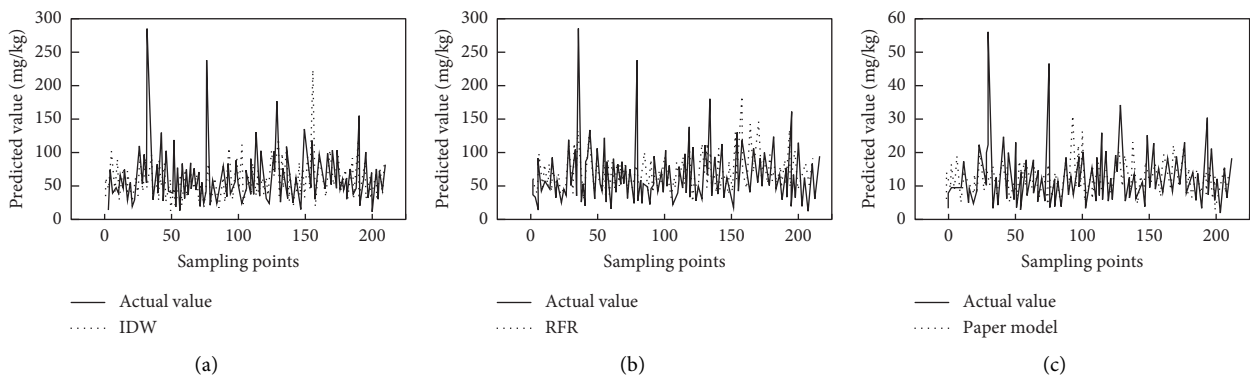


FIGURE 10: Predicted value of different models on Pb data set.

10(b) show that the error between predicted value and actual value of RFR model on training set is the smallest, and the prediction effect is the best. The reason is that RFR model has high feature extraction ability for nonlinear data, and it can adapt to high-dimensional data. Figures 7–10 show that RFR model has the best spatial prediction performance on the training data set. The prediction performance of proposed model and IDW model is poor at a few sampling points, but the error is within the acceptable range.

Figure 11 shows the predicted effect of different models on test data set, and the overall trend is basically the same. For RFR model, the error between predicted

value in test set and actual value is large, which is significantly higher than IDW model and proposed model. The reason may be that there is overfitting in RFR model during training process. What is more, the meaningless features are learned, so as to lead to poor prediction performance. Compared with IDW model, the predicted value of proposed model is closer to the actual value, and there is no abnormal predicted value, which indicates that the proposed model has more stable spatial prediction performance. In conclusion, the prediction performance of proposed model is superior to RFR model and IDW model.

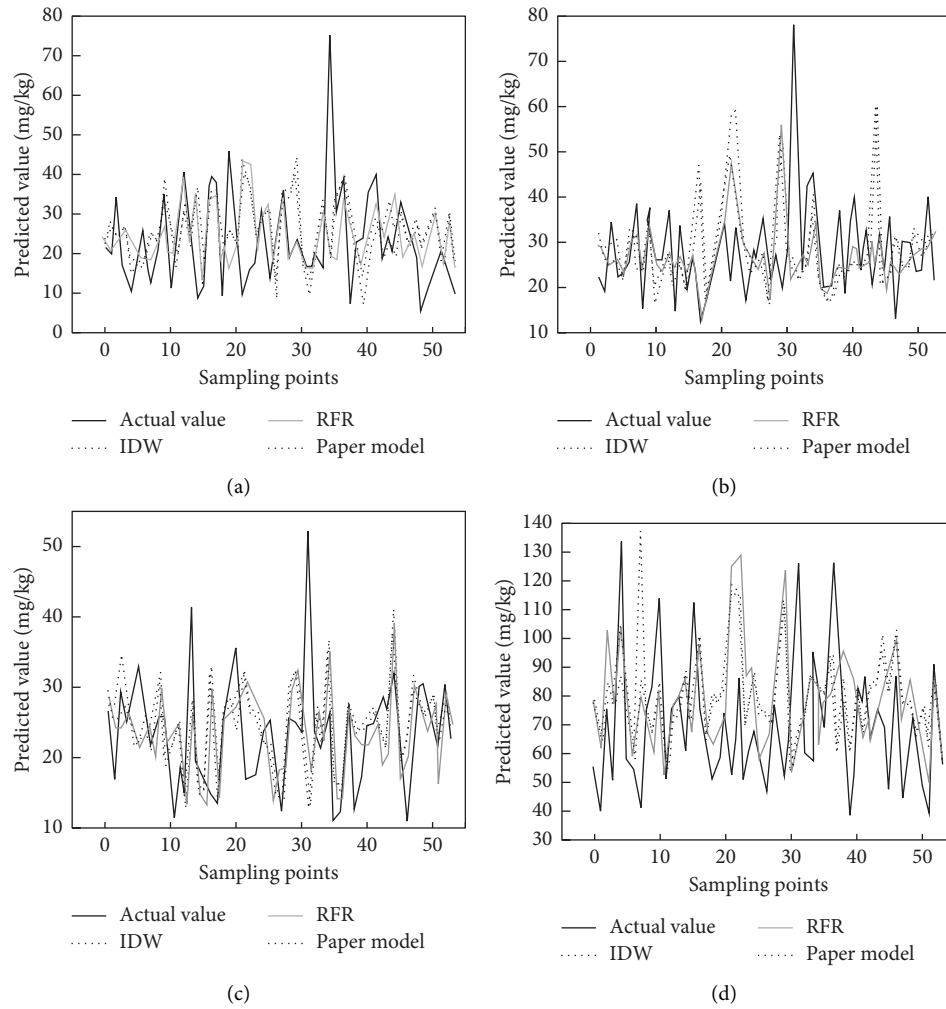


FIGURE 11: Prediction result of different models on different test sets.

TABLE 2: Prediction error of different models on different data sets

Models	Training				Validation				
	MSE	MAE	RMSE	MAPE (%)	MSE	MAE	RMSE	MAPE (%)	
<i>Cd</i>	SAIDW	138.67	8.46	11.77	63.68	203.55	10.11	14.27	84.96
	IDW	174.07	10.22	13.19	7634	225.69	10.84	15.02	89.67
	RFR	57.54	4.55	7.58	31.27	217.05	11.91	14.76	101.04
<i>Cr</i>	SAIDW	97.41	7.11	9.87	28.19	144.41	8.20	12.02	35.25
	IDW	119.63	8.40	10.93	33.22	158.79	8.49	12.60	36.46
	RFR	48.17	3.89	6.94	15.24	168.83	8.91	12.99	38.12
<i>Ni</i>	SAIDW	53.57	5.17	7.31	27.41	69.52	5.52	8.34	27.48
	IDW	76.06	6.59	8.72	33.60	73.63	5.77	8.58	28.04
	RFR	31.61	3.04	5.62	14.40	76.71	6.23	8.76	32.12
<i>Pb</i>	SAIDW	1037.1	18.91	18.91	27.89	690.88	20.63	26.28	34.03
	IDW	1279.24	22.52	35.76	33.87	712.62	21.02	26.69	34.44
	RFR	456.46	10.34	21.36	14.89	886.51	22.03	29.77	39.06

To quantitatively analyze the performance of each model, the evaluation indicators of each model are compared, and the results are shown in Table 2. As can be seen, in training data set, the indicator values of RFR model are lower than those of proposed model and IDW model. However, in

the test set, the indicator values of RFR model are greater than those of proposed model, which indicates that RFR overfits on the training set. To avoid such a situation, the parameters of RFR model need to be adjusted, which will consume a lot of time cost. Compared with IDW and RFR

models, the MSE, RMSE, MAPE, and MAE values of proposed model on the test set are lower. The reason is that the adaptive deep Q network can adaptively allocate the corresponding hyperparameters of each prediction point, which makes model more suitable for the interpolation spatial characteristics of prediction points. Here the prediction results are more accurate and consistent with the above conclusion, which shows that proposed model has the best performance in all indicators. Compared with RFR model and IDW model, the prediction accuracy of proposed model increases by 13.03% and 7.47%, respectively, and the prediction performance of soil heavy metal content is the best.

5. Conclusion

In conclusion, the proposed prediction method of soil heavy metal content based on deep reinforcement learning uses deep Q network as basic model, and it utilizes state value reuse to promote agent to learn the training samples quickly; thus the convergence rate of model is improved. At the same time, adaptive fuzzy membership factor is introduced to change the sensitivity of agent to environmental feedback value in different training periods, which improves the stability of model after convergence. Moreover, adaptive inverse distance interpolation method is adopted to predict the observed values of interpolation points; thus the prediction accuracy of model is improved. Compared with RFR model and IDW model, the proposed model performs better in MSE, RMSE, MAPE, and MAE. The prediction accuracy of soil heavy metal content is higher, which increases by 13.03% and 7.47%, respectively. Although certain research results have been achieved, there are still some shortcomings. Due to the high interpolation accuracy, the proposed prediction model takes a lot of time to complete a training, which has certain disadvantages for the actual prediction of soil heavy metal content. Therefore, a more suitable method needs to be found to train the model, so as to shorten the training time.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

Acknowledgments

This work was sponsored in part by science and education joint project of Hunan Natural Science Foundation (2020JJ7058).

References

- [1] X. Liu, S. Gu, S. Yang, J. Deng, and J. Xu, "Heavy metals in soil-vegetable system around E-waste site and the health risk assessment," *The Science of the Total Environment*, vol. 779, pp. 146438–146448, 2021.
- [2] J. Lan, S. Zhang, Y. Dong et al., "Stabilization and passivation of multiple heavy metals in soil facilitating by pinecone-based biochar: mechanisms and microbial community evolution," *Journal of Hazardous Materials*, vol. 420, pp. 126588–126598, 2021.
- [3] I. Khan, S. A. Awan, M. Rizwan et al., "Effects of silicon on heavy metal uptake at the soil-plant interphase: a review," *Ecotoxicology and Environmental Safety*, vol. 222, pp. 112510–112516, 2021.
- [4] D. Yang, S. Yang, H. Yuan et al., "Co-benefits of biochar-supported nanoscale zero-valent iron in simultaneously stabilizing soil heavy metals and reducing their bioaccessibility," *Journal of Hazardous Materials*, vol. 418, pp. 126292–126296, 2021.
- [5] X. Guo, Y. Wu, N. Li, Y. Tian, Y. Peng, and D. Yuan, "Effects on the complexation of heavy metals onto biochar-derived WEOM extracted from low-temperature pyrolysis," *Ecotoxicology and Environmental Safety*, vol. 221, pp. 112456–112460, 2021.
- [6] A. Rana, M. Sindhu, A. Kumar et al., "Restoration of heavy metal-contaminated soil and water through biosorbents: a review of current understanding and future challenges," *Physiologia Plantarum*, vol. 173, no. 1, pp. 394–417, 2021.
- [7] J. Zhang, G. Mu, Z. Zhang, X. Huang, and H. Fang, "Speciation variation and bio-activation of soil heavy metals (Cd and Cr) in rice-rape rotation lands in karst regions," *International Journal of Environmental Research and Public Health*, vol. 18, no. 3, pp. 1364–1368, 2021.
- [8] Y. P. López, F. A. da Fonseca Breda, E. S. A. Lima, C. da Costa Barros de Souza, J. M. F. González, and N. M. B. do Amaral Sobrinho, "Variability factors of heavy metals in soils and transfer to pasture plants of Mayabeque in Cuba," *Environmental Monitoring and Assessment*, vol. 193, no. 5, pp. 245–247, 2021.
- [9] S. Shao, B. Hu, Y. Tao, Q. You, and Z. Shi, "Comprehensive source identification and apportionment analysis of five heavy metals in soils in Wenzhou City, China," *Environmental Geochemistry and Health*, vol. 38, pp. 1–24, 2021.
- [10] S. Mandal, S. Pu, S. Adhikari et al., "Progress and future prospects in biochar composites: application and reflection in the soil environment," *Critical Reviews in Environmental Science and Technology*, vol. 51, no. 3, pp. 219–271, 2021.
- [11] R. Ramlan, M. B. Cyio, M. Napitupulu et al., "Pollution and contamination level of Cu, Cd, and Hg heavy metals in soil and food crop," *International journal of Environmental Science and Technology*, vol. 25, pp. 1–12, 2021.
- [12] M. Xiang, Y. Li, J. Yang et al., "Heavy metal contamination risk assessment and correlation analysis of heavy metal contents in soil and crops," *Environmental Pollution*, vol. 278, pp. 116911–116917, 2021.
- [13] J. Moon, M. Yang, and J. Jeong, "A novel approach to the job shop scheduling problem based on the deep Q-network in a cooperative multi-access edge computing ecosystem," *Sensors*, vol. 21, no. 13, p. 4553, 2021.
- [14] B. Peng, Q. Sun, S. E. Li, D. Kum, and T. Gu, "End-to-End autonomous driving through dueling double deep Q-network," *Automotive Innovation*, vol. 45, pp. 1–12, 2021.
- [15] T. H. Ho, J. W. Kim, S. H. Son, H. Kim, K. Lee, and J. M. Lee, "Automatic control of simulated moving bed process with deep Q-network," *Journal of Chromatography A*, vol. 1647, pp. 462073–462078, 2021.

- [16] S.-H. Kim, Y.-R. Yoon, J.-W. Kim, and H.-J. Moon, "Novel integrated and optimal control of indoor environmental devices for thermal comfort using double deep Q-network," *Atmosphere*, vol. 12, no. 5, pp. 629–632, 2021.
- [17] Y. Wang, X. Li, P. Wan, L. Chang, and X. Deng, "Dueling deep Q-networks for social awareness-aided spectrum sharing," *Complex & Intelligent Systems*, vol. 23, pp. 1–12, 2021.
- [18] Y. Dai, K. Lee, and S. G. Lee, "A real-time HIL control system on rotary inverted pendulum hardware platform based on double deep Q-network," *Measurement and Control*, vol. 54, no. 3-4, pp. 417–428, 2021.
- [19] D. Kim, S. Y. Kwon, J. Woo, W. M. Kang, and J. H. Lee, "On-chip trainable hardware-based deep Q-networks approximating a backpropagation algorithm," *Neural Computing & Applications*, vol. 46, pp. 1–12, 2021.
- [20] L. Chen, Y. Zhao, H. Zhao, and B. Zheng, "Non-communication decentralized multi-robot collision avoidance in grid map workspace with double deep Q-network," *Sensors*, vol. 21, no. 3, pp. 841–845, 2021.
- [21] Y. Xue, R. Wu, J. Liu, and X. Tang, "Crowd evacuation guidance based on combined action reinforcement learning," *Algorithms*, vol. 14, no. 1, p. 26, 2021.
- [22] A. Skrynnik, A. Staroverov, E. Aitygulov, K. Aksenov, and A. Panov, "Hierarchical deep Q-network from imperfect demonstrations in minecraft," *Cognitive Systems Research*, vol. 65, no. prepublish, pp. 74–78, 2021.
- [23] Z. He, K. P. Tran, S. Thomassey, X. Y. Zeng, J. Xu, and C. H. Yi, "A deep reinforcement learning based multi-criteria decision support system for optimizing textile chemical process," *Computers in Industry*, vol. 56, pp. 103373–103376, 2020.
- [24] S. Shivanshu, B. Chen, C. Chen, H. Wang, and M. Dai, "Deep Q-network learning based downlink resource allocation for hybrid RF/VLC systems," *IEEE ACCESS*, vol. 8, pp. 149412–149434, 2020.