# Multicrew Scheduling and Routing in Road Network Restoration Based on Deep Q-learning

**SU Zhaopin[1,2], DUAN Lianqi[2], ZHANG Guofu[1,2,3,4], CHANG Jiayuan[2] , SHEN Yufeng[2],**

Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Ministry of Education

School of Computer Science and Information Engineering, Hefei University of Technology

Intelligent Interconnected Systems Laboratory of Anhui Province (Hefei University of Technology)

Anhui Province Key Laboratory of Industry Safety and Emergency Technology (Hefei University of Technology)

szp@hfut.edu.cn[1], dlq@163.com[2], zgf@hfut.edu.cn[3], cjy@163.com[4], syf@163.com[5]

## Abstract

Road network restoration is an important issue in the post-disaster disposal and rescue, especially when extraordinarily serious natural disasters (e.g., floods and earthquakes) occur. Central to this endeavor is the problem of determining how to reasonably schedule and route the repair crew to quickly restore the damaged road network and establish reliable supply lines from supply nodes to demand nodes. However, most existing work focuses on the activities of the single repair crew, and rarely considers the problem of continuously damaged road sections, especially for the road network with enormous demand nodes and serious damage. Consequently, this work is concentrated on multicrew scheduling and routing for the damaged road network with enormous demand nodes. Specifically, a model of multicrew scheduling and routing is first presented. Next, a road network partition strategy is first proposed to make different repair crews responsible for different subnets. Then, a deep Q-learning based multicrew scheduling and routing algorithm is proposed for the damaged road network with enormous demand nodes, which utilizes the learning experience of multiple repair crews to achieve hybrid learning. Finally, experimental results demonstrate that the proposed method can make repair crews adjust their scheduling and routing strategies according to the damaged road network and provides a useful attempt to restore the damaged road network in complex emergency scenarios of post-disaster.

## 1 Introduction

Over the past decade, natural disasters, such as floods, volcanic eruptions, earthquakes, and tsunamis, happened frequently around the world and have caused enormous losses of life and property to human society, which poses a common challenge to countries all over the world. Particularly, China is among those suffer most. Along with the global climate change and the economic takeoff and progress in urbanization, China is suffering from increasing pressure on resources, environment, and ecology. Consequently, the situation in the response to natural disasters has become more severe and complicated, which puts higher demands on the functionality, flexibility, and interactivity of the decision support systems used in natural disaster response and recovery.

Today, there can be no doubt that in the immediate hours after a disaster, effectively and timely restoring the damaged road network to establish supply lines and enable rescue teams and emergency resources in the supply nodes to be delivered to different demand nodes in time, has become even more important for rapid transfer and resettlement of injured persons and minimizing disaster losses. Accordingly, a natural question is how to reasonably schedule and route the repair crew to determine which damaged road sections will be repaired and in what order these selected sections should be repaired. Such a problem was formally defined as the crew scheduling and routing problem (CSRP) (Duque 2016).

For CSRP to be solved effectively, much effort has been devoted to this topic in the way of proposing mathematical models of the CSRP and utilizing effective problem-solving techniques. Duque et al. (Duque 2016) presented a dynamic programming (DP) model of the CSRP and developed an exact DP algorithm and an iterated greedy-randomized constructive procedure to optimize accessibility. Moreno et al. (Moreno 2019) proposed a branch-and-Benders-cut (BBC) algorithm to decompose CSRP into scheduling decisions and subproblems with routing decisions. In their serial work (Moreno 2020), they combined BBC with genetic algorithm and simulated annealing to minimize the total time that demand nodes remain inaccessible from the depot. Akbari et al. (Akbari 2021) developed a polynomial time online algorithm for minimizing the summation of the total traveled distance and the total unblocking time. Shin et al. (Shin 2019) proposed a mathematical model based on Mixed Integer Linear Programming (MILP) and an Ant Colony Optimization (ACO) algorithm providing optimal scheduling by taking into account both reconstruction and delivery. Su et al. (Su 2020) introduced the Markov decision process to simulate the crew activities and used the Q-learning algorithm to minimize the total traveled distance and the total repair time. Zhang et al. (Zhang 2021) developed an algorithm for solving CSRP on the severely damaged road network based on Q-learning and action set reduction.

However, the above research focuses on a single repair crew and is only suitable for simple, small-scale emergency scenarios of post-disaster. When some extraordinarily serious natural disasters occur, such as floods and earthquakes, the scale of the damaged road network often increases sharply. It is clear that only a single repair crew cannot meet the timeliness requirements of the post-disaster disposal and rescue.

In this case, different repair crews may start from different supply nodes to repair the damaged road network simultaneously, so as to establish supply lines for all the demand nodes as soon as possible and significantly improve the emergency efficiency and effect.

Against this background, this work studies multicrew scheduling and routing in road network restoration. Specifically, we first propose a model of multicrew scheduling and routing. Next, a road network partition strategy is first proposed to make different repair crews responsible for different sub-nets, and a decision model of repair crew is present. Then, a deep Q-learning based multicrew scheduling and routing algorithm is proposed for the damaged road network with enormous demand nodes. Finally, the effectiveness of the proposed method is verified by comprehensive experiments.

## 2 Problem Formulation

To describe the damaged road network, as shown in Fig.1, we use an undirected graph $G=\{V,E\}$ to construct the damaged road network.

$V=\{1,2,...,n\}$ represents the set of all the nodes in the road network, including a set of supply nodes $V_s$, and a set of demand nodes $V_d$. Namely, $V=V_s\cup V_d$. For $i\in V_d$, there is a weight $I_i\in(0,1]$, which indicates the disaster severity of demand node $i$. The larger $I_i$ is, the more seriously node $i$ is damaged.

$E$ denotes the set of all the road sections, including a set of passable road sections $\bar{E}=\{\bar{e}_1,\bar{e}_2,...,\bar{e}_p\}$ and a set of damaged road sections $\tilde{E}=\{\tilde{e}_1,\tilde{e}_2,...,\tilde{e}_m\}$, that is, $E=\tilde{E}\cup\bar{E}$. $\forall\tilde{e}_j\in\tilde{E}$ and $\forall\bar{e}_k\in\bar{E}$ has a road length $a_j$ and $b_k$, respectively. Additionally, each $\forall\tilde{e}_j\in\tilde{E}$ has a repair-time overhead $t_j$.

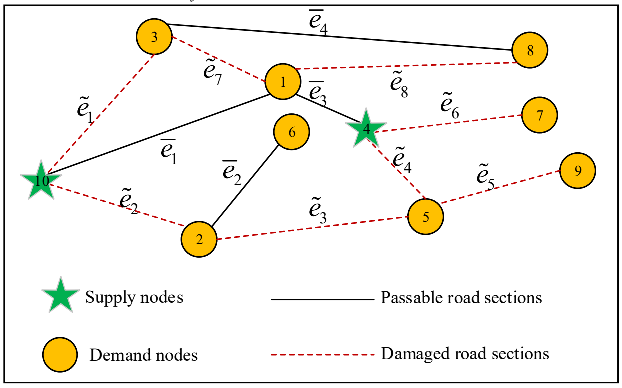



Fig 1. The Damaged Road Network with Many Supply Nodes.

Each supply node $\delta\in V_s$ is assigned only a repair crew $c_\delta$ with a travel speed $v_\delta$. It is clear that there are at total $|V_s|$ repair crews in the damaged road network.

Note that for a damaged road section, at most a repair crew is assigned to repair it.

Each $c_\delta$ has a repair strategy $H_\delta\subseteq E$, which denotes an ordered set of road sections, including its travelling passable sections and damaged sections.

It is expected that each demand node should be connected as early as possible. That is, the connection time of each $H_\delta$ should be as small as possible. The connection time $f_1(H_\delta)$ is defined as:

$$f_1(H_\delta)=\sum_{\tilde{i}\in V_{d_\delta}} T_{\tilde{i}}^{\delta}\cdot I_{\tilde{i}} \tag{1}$$

$$T_{\tilde{i}}^{\delta}=\sum_{\tilde{j}\in\tilde{E}_\delta}\left(t_{\tilde{j}}+\frac{a_{\tilde{j}}}{v_\delta}\right)+\sum_{\bar{k}\in\bar{E}_\delta}\frac{b_{\bar{k}}}{v_\delta} \tag{2}$$

Where $V_{d_\delta}$ is the set of demand nodes repaired by the repair crew $c_\delta$. $T_{\tilde{i}}^{\delta}$ is the connection time of demand node $\tilde{i}$ repaired by the repair crew $c_\delta$.

On the other hand, it is expected that after road network restoration, the transportation time between each demand node and the corresponding supply node should be as little as possible. That is, the transportation time $f_2(H_\delta)$

$$f_2(H_\delta)=\sum_{\tilde{i}\in V_{d_\delta}} D_{\tilde{i}}\cdot I_{\tilde{i}} \tag{3}$$

Where $D_{\tilde{i}}$ is the shortest path length between demand node $\tilde{i}$ and its corresponding supply node $\delta$.

Based on the above notations, the multicrew scheduling and routing (MCSR) in road network restoration can be described as the following constrained optimization problem:

$$\min f(H_\delta)=w_1\cdot f_1(H_\delta)+w_2\cdot f_2(H_\delta) \qquad \delta\in V_s \tag{4}$$

where $w_1, w_2 \in (0,1)$ is a weight of the trade-off between the connection time and the transportation time.

From the proposed model, the number of repair crews is consistent with the number of supply nodes, and repair crews cooperate with each other to restore the damaged road network. For demand nodes, it is hoped that the connection time and the transportation time should be as little as possible. For damaged road sections, it can be restored by any repair crew. However, to obtain less connection time and transportation time, they should be repaired by a nearby repair crew. Therefore, the policy of divide and conquer can be utilized. That is, the damaged road network is first partitioned to make different repair crews responsible for different damaged road sections, and repair crews share their repair experience to cooperate with each other.

## 3 Road Network Partition

The damaged road network is partitioned into $|V_s|$ intersected sub-networks according to the distance between the supply nodes and damaged road sections.

For each damaged road section $j \in \tilde{E}$, the travelling time $\tau_{j\delta}$ with every supply node $\delta \in V_s$ is computed according to (5).

$$\tau_{j\delta} = \frac{D_{j\delta}}{v_\delta} \tag{5}$$

Where $D_{j\delta}$ is the shortest path between $\delta$ and two nodes of $j$ computed by Dijkstra algorithm.

If there is a minimum $\tau_{j\delta^*}$, and $\tau_{j\delta^*} << \tau_{j\delta}$ ($\delta^* \neq \delta$), the damaged road section $j$ is classified into the set $\tilde{E}_{\delta^*}$. And the two nodes of $j$ are put in to the set $V_{\delta^*}$.

If there are several $\tau_{j\delta_1}, \tau_{j\delta_2}, \cdots$, and these values have little difference, the damaged road section $j$ is simultaneously classified into $\tilde{E}_{\delta_1}, \tilde{E}_{\delta_2}, \cdots$. And the two nodes of $j$ are put in to the sets $V_{\delta_1}, V_{\delta_2}, \cdots$.

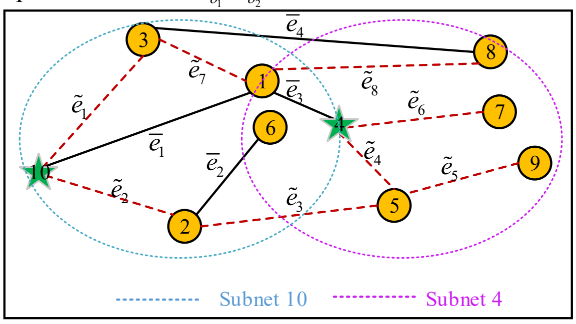



Fig 2. An Example of Road Network Partition.

Fig 2 gives an example of road network partition. As there are 2 supply nodes, the damaged road network is partitioned into 2 intersected sub-networks. After road network partition, $V_4 = \{4,1,6,5,7,8,9\}$, $V_{10} = \{10,1,2,3,5,6\}$, $\tilde{E}_4 = \{\tilde{e}_3, \tilde{e}_4, \tilde{e}_5, \tilde{e}_6, \tilde{e}_8\}$ and $\tilde{E}_{10} = \{\tilde{e}_1, \tilde{e}_2, \tilde{e}_3, \tilde{e}_7\}$.

## 4 Decision Model of Repair Crew

It is clear that MCSR is a sequential decision problem which can be described as Markov decision process (MDP) including three basic elements: state space $S$, action space $A$, and reward function $r$. In this work we regard the repair crew $c_\delta$ as an agent, $S_\delta \leftarrow V_\delta$, and $A_\delta \leftarrow \tilde{E}_\delta$.

The goal of agent exploration is to make the adjacent and seriously damaged demand nodes connected as soon as possible. Assumed that the state of the agent $c_\delta$ changes by repairing a damaged road section $a$, the immediate reward value $r$ of the action $a$ of $c_\delta$ is defined as:

$$r \leftarrow \begin{cases} -I_{i^*} - \frac{1}{t_{j^*}}, N = 0 \\ \frac{\vec{I}}{N} + \frac{1}{t_j}, N \geq 1 \end{cases} \tag{6}$$

$$i^* = \arg\max_{i \in V_\delta} I_i \tag{7}$$

$$j^* = \arg\min_{j \in \tilde{E}_\delta} t_j \tag{8}$$

Where $N$ is the number of connected demand nodes by repairing a damaged road section $a$. $\vec{I}$ is the sum of the disaster severity of demand nodes which are newly connected.

When $N = 0$, it means no demand node can be connected, and repairing $a$ cannot bring significant effect to the repair of the whole road network. Therefore, the emergency repair team will get a certain benefit punishment (i.e. negative feedback).

When $N \geq 1$, it means repairing $a$ can bring great performance to road network restoration. At this time, the agent should be rewarded (i.e. positive feedback), and the reward is related to the number of newly connected demand nodes, the disaster severity of demand nodes, and the repair time.

## 4 Deep Q-learning based Scheduling and Routing Algorithm

In this section, we proposed a deep Q-learning to solve the proposed MCSR problem, which combines deep Q-network (DQN) (Volodymyr, 2015), dueling reinforcement learning (Wang, 2016), and prioritized experience replay (Schaul, 2016). Each agent achieves the damaged road sections repairing in its own subnet through training and learning, and cooperates with other agents for the overlapped parts through exchanging information in the shared experience buffer.

The basic flow of the proposed deep Q-Learning for solving MCSR (hence called DQL-MCSR) is shown in Fig.3.

The specific steps are as follows:

**(1)** Initialize the corresponding parameters of the road network model, the decision model of repair crew, and the deep Q-learning.

**(2)** Partition the road network into $|V_s|$ parts.

**(3)** Initialize shared experience buffer of a fixed size $L$.

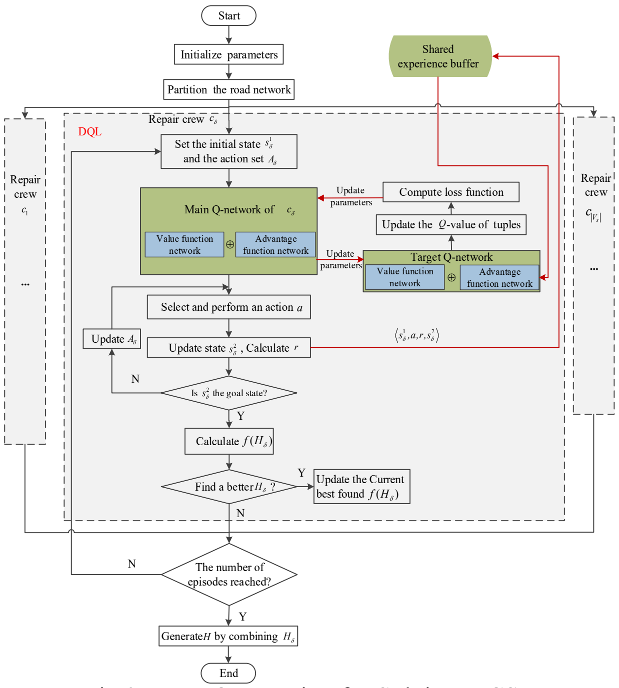


Fig 3. Deep Q-Learning for Solving MCSR.

**(4)** For each repair crew $c_\delta$, create a main Q-network and a target Q-network, and execute steps as follows:

**(4.1)** Set the initial state $s_\delta^1 = V_{d_\delta}$ and the action set $A_\delta = \tilde{E}_\delta$.

**(4.2)** Select and perform an action $a \in A_\delta$ according to the $\varepsilon$-greedy ($\varepsilon \in (0,1)$) strategy. Specifically, $a$ is chosen randomly from the optimal action set in the $Q$-value table with a high probability $1-\varepsilon$, or $a$ is randomly selected from $A_\delta = \tilde{E}_\delta$ with a low probability $\varepsilon$.

**(4.3)** Update the state $s_\delta^2$ according to the current status of $c_\delta$, and calculate the reward $r$.

**(4.4)** Add the experience tuples $\langle s_\delta^1, a, r, s_\delta^2 \rangle$ to the end of the shared experience buffer.

**(4.5)** Take $x$ experience tuples from the shared experience buffer, and compute the $Q$-value of each tuples according to (9).

$$Q(s_\delta^1, a, r, s_\delta^2; \theta, \phi, \varphi) \leftarrow Value(s_\delta^1, a, r, s_\delta^2; \theta, \phi) + \left( Adv(s_\delta^1, a, r, s_\delta^2; \theta, \varphi) - \frac{1}{|A_\delta|} \sum_{a \in A_\delta} Adv(s_\delta^1, a, r, s_\delta^2; \theta, \varphi) \right) \tag{9}$$

Where $\phi$ and $\varphi$ are the parameters of Value Function and Advantage Function, respectively.

**(4.6)** Compute the loss function, and update the gradient of main Q-network.

**(4.7)** Check whether target network need to be updated. If yes, update the target network with the parameters of the main Q-network.

**(4.8)** Check whether $s_\delta^2$ is the goal state (i.e., $s_\delta^1 = \varnothing$ and all the demand nodes are connected). If no, update the current $A_\delta$ and go to (4.2). If yes, calculate the objective function value of $H_\delta$ based on (4).

**(4.9)** If a better $H_\delta$ is found, update the current best found $f(H_\delta)$.

**(5)** If the number of episode is reached, generate $H$ by combining each repair crew's repair strategy $H_\delta$. Otherwise, go to (4.1).

## 5 Experimental Results

To comprehensively evaluate the performance of the proposed DQL-MCSR, in this section we compared DQL-MCSR with the state-of-the-art BQL (Su 2020), IQL (Zhang 2021) and ACO (Shin 2019).

### 5.1 Parameter Settings

We randomly generate 4 road networks with different sizes and the proportion of damaged road sections as shown in Table 1.

Table 1. 4 Road Networks.

| Instance | $\lvert V_s \rvert$ | $\lvert V_d \rvert$ | $\lvert \tilde{E} \rvert / \lvert E \rvert$ |
|---|---|---|---|
| 1 | 2 | 99 | 30% |
| 2 | 2 | 99 | 50% |
| 3 | 5 | 196 | 30% |
| 4 | 5 | 196 | 50% |

For each road network, the weight $I_i$ of each demand node is randomly generated according to the normal distribution in $(0,1]$. $a_j$, $b_k$ and $t_j$ are randomly generated according to the normal distribution in $[1,10]$. Every $v_\delta = 1$.

For BQL and IQL, the learning rate is 0.4. For ACO, the number of ants is 80, the iteration number is 300. For MRC-DOL, $L = 3000$, $x = 32$, the number of episode is 6000.

Each algorithm was run on 4 instances, each of which was tested repeatedly for 30 independent runs with different random seeds.

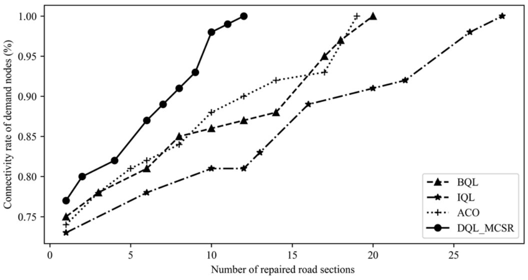
### 5.2 Performance Evaluation

In this experiment, we analyze the performance of BQL, IQL, ACO and DQL-MCSR based on 4 different instances.

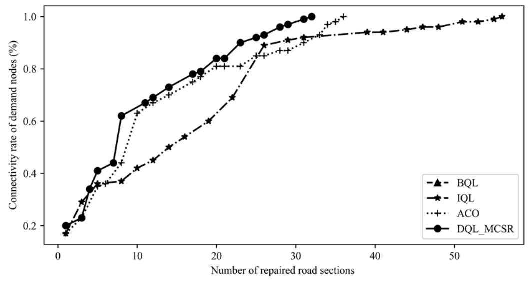
Table 2 depicts the objective function value (mean and standard deviation) obtained by different algorithms under 4 different instances, where "\" means the algorithm does not find an effective solution on the instance. It can be observed that under all the 4 damaged road networks, the solutions obtained by the DQL-MCSR are significantly better than that achieved by the BQL, IQL and ACO.

Table 2. Mean Value of Objective Function.

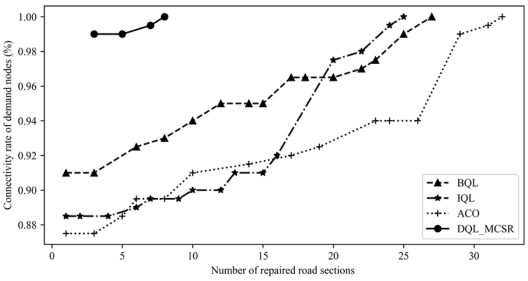
| Instance | BQL | IQL | ACO | DQL-MCSR |
|---|---|---|---|---|
| 1 | 305.10 | 308.22 | 259.58 | **130.16** |
| 2 | | 617.91 | 588.85 | **335.84** |
| 3 | 430.70 | 433.59 | 431.02 | **34.35** |
| 4 | | 1926.64 | 2046.38 | **523.15** |



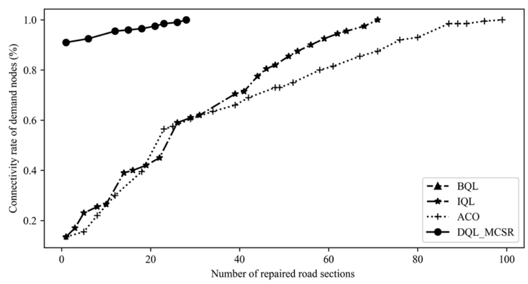
(a) Instance 1



(b) Instance 2



(c) Instance 3



(d) Instance 4

Fig 4. Connectivity Changes of Demand Nodes by Repairing Road Sections.

In particular, the larger the road network size, the more the damaged road sections, the better results of the proposed DQL-MCSR.

Fig 4 illustrates the connectivity changes of demand nodes by repairing road sections. Under all the 4 damaged road networks, DQL-MCSR repairs less damaged road sections to connect all the demand nodes. Particularly, with the increase of the road network size, the number of repaired road sections by DQL-MCSR is much less than that by BQL, IQL and ACO. When the proportion of damaged road sections is 50%, BQL cannot find any effective solution. Take Instance 4 for example, to achieve the full connectivity of demand nodes, AQL and ACO needs to repair 71 and 99 damaged road sections, respectively, while DQL-MCSR only needs to repair 28.

The above observations indicate that DQL-MCSR is able to well adapt to the serious damaged road network with enormous demand nodes. An explanation is that DQL-MCSR utilizes the policy of divide and conquer, and learn with each other by exchanging information, which is beneficial for DQL-MCSR to learn good repair strategies in the course of the remaining episodes.

## 6 Conclusions

To address the complex emergency scenario in extraordinarily serious natural disasters, this work is concentrated on multicrew scheduling and routing in road network restoration. More specifically, a model of multicrew scheduling and routing is first proposed. Next, a decision model of the repair crew is presented based on the Markov decision process. Then, a deep Q-learning algorithm is developed to find the best strategy. Experimental results demonstrate that the proposed DQL-MCSR can make repair crews adjust their scheduling and routing strategies according to the damaged road network through learning and cooperation. In the future, we will focus on dynamic emergency scenario.

## References

Akbari, V.; Shiri, D.; and Salman, F. S. 2021. An online optimization approach to post-disaster road restoration. *Transportation Research Part B-Methodological* 150:1-25.

Duque, P. A. M.; Dolinskaya, I. S.; and Sorensen, K. 2016. Network repair crew scheduling and routing for emergency relief distribution problem. *European Journal of Operational Research* 248(1): 272-285.

Hasselt, H. V.; Guez, A.; and Silver, D. 2015. Deep reinforcement learning with double q-learning. arXiv:1509.06461.

Moreno, A.; Munari, P.; and Alem D. 2019. A branch-and-Benders-cut algorithm for the crew scheduling and routing problem in road restoration. *European Journal of Operational Research* 275(1): 16-34.

Moreno, A.; Munari, P.; and Alem D. 2020. Decomposition-based algorithms for the crew scheduling and routing problem in road restoration. *Computers & Operations Research* 119, article 104935. doi.org/10.1016/j.cor.2020.104935.

Schaul, T.; Quan, J. ; Antonoglou, I.; and Silver, D. 2016. Prioritized experience replay. Paper presented at the 2016 ICLR conference, San Juan, Puerto Rico, May 2-4.

Shin, Y.; Kim, S.; and Moon, I. 2019. Integrated optimal scheduling of repair crew and relief vehicle after disaster. *Computers & Operations Research* 105: 237-247.

Su, Z.; Li, M.; Zhang, G.; and Yue, F. 2020. Modeling and solving the repair crew scheduling for the damaged road networks based on Q-learning. *Acta Automatica Sinica* 46(7):1467-1478.

Volodymyr, M.; Koray, K.; David, S. ; Rusu, A. A. ; Joel, V. , & Bellemare, M. G. , et al. 2015. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529-533.

Wang, Z.; Schaul, T.; Hessel, M.; Hasselt, H.; Lanctot, M.; and F, N. 2016. Dueling network architectures for deep reinforcement learning. In Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York: JMLR.

Zhang, G.; Tu, B.; Su, Z.; and Liu, Y. 2021. An algorithm for repair crew scheduling on severely damaged road network. *Control and Decision* 36(7):1663-1671.

## Acknowledgments

This work is supported by the Anhui Provincial Key Research and Development Program under Grant 202004d07020011 and Grant 202104d07020001; the MOE (Ministry of Education in China) Project of Humanities and Social Sciences under Grant 19YJC870021; the Guangdong Provincial Key Laboratory of Brain-inspired Intelligent Computation under Grant GBL202117; and the Fundamental Research Funds for the Central Universities under Grant PA2020GDKC0015, Grant PA2021GDSK0073, and Grant PA2021GDSK0074.