控制与决策
Control and Decision

基于双层交互*Q*学习的路网抢修和物资配送联合调度

张国富, 朱前顺, 苏兆品, 岳峰

引用本文:
张国富, 朱前顺, 苏兆品, 等. 基于双层交互*Q*学习的路网抢修和物资配送联合调度[J]. 控制与决策, 2024, 39(12): 4109-4117.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2023.1222

基于双层交互Q学习的路网抢修和物资配送联合调度

张国富[1,2,3†], 朱前顺[1], 苏兆品[1,2,3], 岳 峰[1,2]

(1. 合肥工业大学 计算机与信息学院，合肥 230601；
2. 工业安全与应急技术安徽省重点实验室(合肥工业大学)，合肥 230601；
3. 安全关键工业测控技术教育部工程研究中心，合肥 230601)

摘 要: 受损路网修复和物资配送是灾后应急响应初期的两个重要环节,已有研究大都将路网修复和物资配送割裂开来考虑,难以满足实际救援需求. 为此,在构建抢修队与运输队联合调度的路网模型的基础上,引入马尔科夫决策过程来模拟抢修队的修复活动和运输队的救援活动,分别设计相应的状态、动作集和即时奖励函数,并提出一种基于双层交互Q学习的路网抢修和物资配送联合调度算法. 对比实验表明,所提方法能有效提高路网抢修和物资配送的效率,可为应急响应初期的救援与处置提供及时可靠的物资保障.

关键词: 灾后应急响应；路网修复；物资配送；马尔科夫决策；双层交互Q学习；联合调度优化

中图分类号: TP181 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2023.1222

引用格式: 张国富, 朱前顺, 苏兆品, 等. 基于双层交互Q学习的路网抢修和物资配送联合调度 [J]. 控制与决策, 2024, 39(12): 4109-4117.

Joint scheduling of road network restoration and emergency relief supplies delivery based on double-layer interactive Q-learning

ZHANG Guo-fu[1,2,3†], *ZHU Qian-shun*[1], *SU Zhao-pin*[1,2,3], *YUE Feng*[1,2]

(1. School of Computer Science and Information Engineering，Hefei University of Technology，Hefei 230601，China；
2. Anhui Province Key Laboratory of Industry Safety and Emergency Technology (Hefei University of Technology)，Hefei 230601，China；3. Engineering Research Center of Safety Critical Industrial Measurement and Control Technology of Ministry of Education，Hefei 230601，China)

Abstract: Road network restoration and emergency relief supplies delivery are two important aspects in the early stages of post-disaster emergency response. Most existing studies have considered road network restoration and emergency relief supplies delivery separately, making it difficult to meet the actual needs of emergency rescue and disposal. Therefore, a road network model for joint scheduling of repair crews and transportation teams is first constructed. Next, the Markov decision process is adopted to simulate the activities of repair crews and transportation teams, in which the corresponding state spaces, action spaces, and reward functions are designed, respectively. Then, a joint scheduling algorithm for road network restoration and emergency relief supplies delivery is developed based on customised double-layer interactive Q-learning. Finally, comparative experiments demonstrate that the proposed algorithm can improve the efficiency and effectiveness of road network restoration and emergency relief supplies delivery, and provide timely and reliable emergency relief supplies support for the rescue and disposal in the early stages of post-disaster emergency response.

Keywords: post-disaster emergency response；road network rehabilitation；material distribution；Markov decision making；two-layer interactive Q-learning；joint scheduling optimisation.

0 引 言

随着全球气候变化、经济发展以及城市化进程的推进,中国在资源、环境和生态等方面也在接受着愈发严峻的挑战. 根据应急管理部于2023年1月13日发布的《2022年全国自然灾害基本情况》统计,2022年全年我国因自然灾害而导致的受灾人次达到1.12亿,其中554人失踪或丧生,242.8万人次被紧急转移安置;同时,共有79.6万间房屋承受了不同程度

收稿日期: 2023-08-28；**录用日期:** 2024-03-11.

基金项目: 安徽省重点研究与开发计划项目(202104d07020001)；安徽省自然科学基金面上项目(2208085MF166)；中央高校基本科研业务费专项资金项目(PA2023IISL0097, PA2023GDSK0049).

†通讯作者. E-mail: zgf@hfut.edu.cn.

的损坏,其中倒塌的房屋数目为4.7万间;农作物也遭受了严重影响,共12 071.6千公顷的农作物受灾;造成了2 386.5亿元的经济损失[1]. 这些数据表明,我国自然灾害形势依旧复杂严峻和复杂,洪涝灾害点多面广,地震活动强度有所提高,区域内极端性暴雨与强震频发. 国家在《“十四五”国家应急体系规划》提出,充分利用云计算、大数据、物联网等新一代信息技术完善综合风险预警制度,增强风险早期识别能力,系统推进“智慧应急”建设,建立符合大数据发展规律的应急数据治理体系[2],这对进一步建设功能性更强、灵活性更高、交互性更好的应急决策系统有了越来越急切的需求.

当地震、洪水等重特大自然灾害发生后,尽快修复受损道路、恢复运输系统正常工作,是开展应急救援的前提与关键. 同时,应当快速地对需求点提供应急物资并及时地进行救援工作,最大限度地减少伤亡,保证人员和财产安全. 因此,各国研究人员和学者开始注重对灾后路网修复与物资配送问题的研究.

1 相关工作

在物资配送方面,Rabiei等[3]引入了一个模型来处理车辆路径问题与需求点物资分配问题,并将模糊推理系统嵌入到NSGA-II与NRGA中,最后分别在大规模和小规模路网上进行了测试;Chang等[4]建立了一个网络流动模型,该模型不仅连接了物资分配中心与救济中心,还允许物资在救济中心之间运输,从而在灾后反应阶段高效地分配救援物资;Huang等[5]认为使用无人机可大大减少救援时间和成本,设计一种遗传算法来解决基于无人机物流网络的供应配送中心规划问题;孙笑等[6]为了帮助决策者根据实际需求设定目标函数权值以得出最佳调度方案,建立了一个多目标多约束优化模型;张国富等[7]设计一种针对单一事故点、多种应急物资和多个储备站的应急物资多目标分配模型,结合非支配排序遗传算法和启发式策略,旨在为化工园区设计一种高效的应急物资多目标分配算法;刘扬等[8]构建了救援物资多阶段分配和调度模型,设计了基于蚁群优化和NSGA-II的多目标求解算法,并设计相应的编码调整策略以解决可能出现的物资分配冲突问题;宋英华等[9]设计了一种综合考虑应急车辆在应急配送中心等待情况的多级配送、多种物资的应急物资调配方案优化模型,并提出采用基于实数编码的遗传算法进行求解,最后对所提出的模型进行有效性和可行性验证;张国富等[10]构建了一种面向多储备点、多发放点、多应急物资并发分配与调度的多目标优化模型,并提出一种混合智能搜索算法进行求解. 由于突发事件的紧急性,往往不计物资成本,所以本文的物资配送部分从路径优化和配送调度出发,探讨如何快速高效地将救援物资发放到灾区.

在路径修复方面,Ajam等[11]为了最小化路网修复的延迟,提出了一种基于贪婪随机自适应搜索程序与可变邻域搜索组合的元启发式算法;李兆隆等[12]提出了一种基于弹复性的交通网络应急恢复阶段策略优化模型,并设计一种交互式双层算法,证明了其算法的有效性;Maya等[13]开发了一种基于启发式算法的背包问题和可变邻域搜索来提升路网连通率,并通过测试验证了其有效性;苏兆品等[14]认为灾后路网修复是典型的时序决策模型,而强化学习又是一种通过交互的目标导向学习方法,非常契合灾后应急响应的并发性、动态性和连续性,因此提出了一种基于Q学习算法求解抢修队的最优调度策略;在此基础上,张国富等[15]考虑了严重受损的路网模型,简化了决策模型,并对动作集进行了裁剪;随后,张国富等[16]针对大量需求点的路网模型,设计了一种双反馈奖励函数,并基于深度Q学习解决受损路网修复问题;Su等[17]针对灾后出现连续受损路段和大量需求点的问题,提出了一种基于深度Q学习的多抢修队调度算法,利用各个抢修队的学习经验来实现混合学习,实验结果证明该算法能使抢修队根据受损路网状态及时调整抢修策略.

为了提高灾后应急效率,应在路网修复的基础上考虑需求点的物资配送,若将二者单独分开考虑将会造成一定的延时. Souza等[18]对用于道路修复的短期和长期网络模型的特征进行了定性分析,例如网络类型、不确定性和相互依赖性,认为未来的研究应侧重于将道路维修和恢复与其他紧急活动(如救济物资分发)相结合;Farzaneh等[19]开发了一个综合决策支持框架用于在线协调灾害响应阶段的3个相互依存的应急救援行动:损失评估、道路恢复和救济分配;陈钢铁等[20]设计了一种启发式算法来解决应急道路修复和物资配送优化调度问题,并对比分析了3个不同应急配送路网特征;张梦玲等[21]基于手机定位数据,同时结合物资配送和路网修复的相互关系,建立了线性规划模型,并设计了一种启发式算法进行求解;Ransikarbum等[22]针对灾后应急相应阶段和恢复阶段的受损路网修复和应急物资配送问题,提出了一个多目标综合网络优化模型;Shin等[23]建立了基于混合整数线性规划的灾后抢修人员和救援车辆综合

优化调度数学模型,并提出了一种改进的蚁群优化算法(ant colony optimization, ACO),将其用于优化路网修复后总的运输时间. 上述工作中的传统算法主要面对小规模、受损程度较低的路网,一定程度上可以提升应急救援的效率. 然而,对于受损路网规模大、应急需求点多的地震、洪涝等应急场景而言,现有方法的求解效率和求解质量难以满足实际需求.

基于上述背景,本文在总结和分析已有工作的基础上,针对受损路网修复和应急物资配送联合优化问题,分别构建联合调度路网模型和基于马尔科夫决策过程的抢修队和运输队的决策模型,并提出一种基于双层交互Q学习和改进的最优动作集更新策略的联合调度算法,最后通过对比实验验证算法的有效性.

2　问题描述

灾后受损的路网可以通过无向图$G=(V,E)$表示. 在本文将路网模型图抽象为图的形式,用顶点来表示对象,用边表示对象之间的关系.

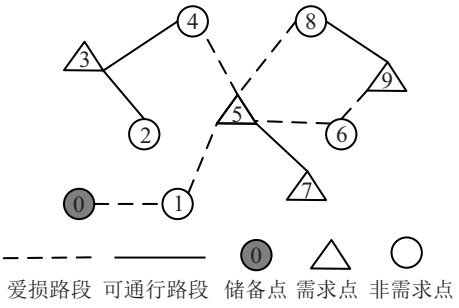


图 1　受损路网示意图

对于路网中任意一个节点$i\in V$,赋予一个权重$I_i\in\mathbb{R}_0^+$,I_i表示节点i的受灾程度. 当节点$i\in V_d$时,$I_i>0$,且I_i越大,代表节点i的受灾程度越严重,救援紧迫度越高;当节点$i\in V_z$时,$I_i=0$.

对于需求点$i\in V_d$,$\mu_i\in\{0,1\}$,表示需求点的应急状态. 若$\mu_i=0$,则表示需求点处于待救援的状态;否则,表示需求点已经被救援.

对于路网中任意一个路段$e_{ij}\in E$,通过参数$\varepsilon_{ij}\in\{0,1\}$表示路段$e_{ij}$的通行状态. 当$\varepsilon_{ij}=1$时,$e_{ij}\in E_z$,说明路段$e_{ij}$处于连通状态,可通行;否则,表示路段$e_{ij}$处于受损状态,不可通行.

运输队对每个需求节点$i\in V_d$进行救援的时间开销为$d_i\in\mathbb{R}_0^+$,救援时间越长,说明需求点对物资的需求越大.

通过参数$\theta_i\in\{0,1\}$来表示i的可达状态. 当$\theta_i=1$时,表示节点i可达;否则,表示节点i不可达. 抢修队从受损路段集中选取一条受损路段修复,使所有需求点可达的受损路段称为一个修复方案,用有序集合H表示:

$$\begin{aligned}&H=\langle e_{i_1j_1},\dots,e_{i_kj_k}\rangle,\ \forall k\in\mathbb{Z}^+;\\&\text{s.t. } D_{i0}\leqslant D_i,\ \forall i\in V_d.\end{aligned}\tag{1}$$

其中$e_{i_1j_1},\dots,e_{i_kj_k}$表示抢修队的修复次序,完整的修复方案$H$满足所有的需求点可达. 在未达到完整的修复方案之前,用$\tilde{H}$记录已修复的受损路段. 本文希望每个需求点i要尽可能快地与储备点连通,即

$$\begin{aligned}&\min f_L(H)=\sum_{i\in V_d}(C_{i0}^r\cdot I_i);\\&\text{s.t. } D_{i0}\leqslant D_i,\ \forall i\in V_d.\end{aligned}\tag{2}$$

其中$C_{i0}^r\in\mathbb{R}_0^+$表示抢修队从储备点“0”到节点$i$打通时的累积时间开销. 例如,在图1中,$H=\langle e_{01},e_{15},e_{54},e_{45},e_{58}\rangle$为抢修队的一个修复方案,抢修队的累积时间开销为

$$C_{i0}^r=\frac{l_{01}}{v}+t_{01}+\frac{l_{15}}{v}+t_{15}+\frac{l_{45}}{v}+t_{45}+\frac{l_{45}}{v}+\frac{l_{58}}{v}+t_{58}.$$

抢修队只需要修复使所有需求点可达的受损路段,即满足$\theta_i=1,i\in V_d$;而运输队则需要抵达各个需求点并进行救援工作,即满足$\mu_i=1,i\in V_d$. 运输队从需求点集合中选取一个需求点进行救援,所有被救援的需求点称为一个救援方案,用有序集合P表示:

$$\begin{aligned}&P=\langle i_1,\dots,i_k\rangle|H,\ \forall k\in\mathbb{Z}^+;\\&\text{s.t. } \mu_{i_k}=1,\ \forall i_k\in V_d.\end{aligned}\tag{3}$$

其中i_1到i_k表示运输队的救援次序. 救援方案P必须满足所有的需求点被救援,运输队的救援方案P在很大程度上取决于抢修队的修复策略H. 如果节点j不可达,即$\theta_j=0$,运输队则无法通行,只能选择其他路段或者原地等待路段e_{ij}被修复. 当没有其他需求点需要救援时,救援队在节点i的等待时间为$w_i\in\mathbb{R}_0^+$. 记$C_{i0}^d\in\mathbb{R}_0^+$表示运输队从储备点“0”到完成节点$i$救援时的累积时间开销. 在图1中,运输队在修复策略$H=\langle e_{01},e_{15},e_{54},e_{45},e_{58}\rangle$下的救援方案$P=\langle 5,7,3,9\rangle$,此时

$$\begin{aligned}C_{i0}^d=&w_0+\frac{l_{01}}{v}+w_1+\frac{l_{15}}{v}+d_5+\frac{l_{57}}{v}+\\&d_7+\frac{l_{57}}{v}+w_5+\frac{l_{45}}{v}+\frac{l_{34}}{v}+d_3+\\&\frac{l_{34}}{v}+\frac{l_{45}}{v}+w_5+\frac{l_{58}}{v}+\frac{l_{89}}{v}+d_9.\end{aligned}$$

完整的应急救灾活动由抢修队和运输队协同完成. 抢修队负责路段修复,保证需求点可达,运输队负

责将救援补给送至需求点. 抢修队间接决定应急救灾活动的效率,运输队直接决定应急救灾活动的效率. 根据上述描述,路网抢修和物资配送联合调度可以描述为如下的双层规划问题:

$$\min f_U(P)=\sum_{i\in V_d}(C_{i0}^d\cdot I_i);$$
$$\text{s.t. } H\neq\varnothing. \tag{4}$$

其中,在优化上层f_U之前,需要先优化下层的f_L并反馈一个H给上层. 本文考虑了需求点的救援效率:对于受损程度比较严重的需求点,它们的时间紧迫度较高,需要在尽可能短的时间内与储备点“0”连通,打通生命线路并得到及时的救援.

3 基于双层交互Q学习的联合调度算法

在本文研究中,受损路网是部分可观测的,尤其是对于运输队而言,网络中有哪些路段是可通行的,以及每个应急点之间的距离相对关系均是未知的,因为可通行路段会随着抢修队的修复动态变化. 对于这些未知信息,抢修队和运输队只能在修复和运输的过程中通过探索逐步获取. 这就意味着,抢修队和运输队在决策时有必要综合考虑以前的观测和当前的状态信息. 而Q学习中的智能体不需要知道全局环境,仅需知道当前状态下可以选择哪些动作. 基于上述考虑,本文设计一种双层交互Q学习来求解抢修队和运输队的联合调度问题. 基于双层交互Q学习的联合调度算法流程见图2.

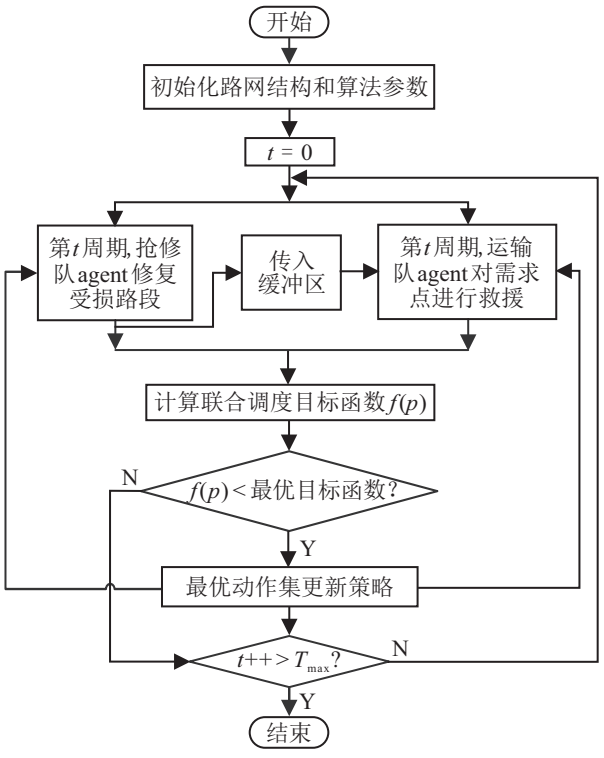


图 2 求解联合调度算法流程

3.1 决策模型构建

马尔科夫决策过程是智能体从当前的状态根据当前的环境做出动作,从而改变环境的状态并获得回报的过程,它与联合调度决策过程十分契合. 马尔科夫决策过程通常包括状态、动作和回报函数3个基本要素,根据这3个基本要素来描述联合调度决策过程. 需要指出的是,在Q学习中,通常只需根据目标赋予agent合理的奖励刺激,agent就会按照奖励的趋势自主学习,并朝着目标积极探索,而不必完全按照目标函数(4)来设计即时奖励.

3.1.1 抢修队agent的建模

设抢修队agent的状态s^r由三元组$\langle\mathcal{V}^r,\mathcal{D}^r,\mathcal{E}^r\rangle$构成,即

$$\begin{cases}\mathcal{V}^r=\{i|\theta_i=1,i\in V_d\},\\ \mathcal{D}^r=\{\langle i,D_{i0}\rangle|\theta_i=1,\forall i\in V_d\},\\ \mathcal{E}^r=\{e_{ij}|\xi_{ij}=1,e_{ij}\in E_d\}.\end{cases} \tag{5}$$

其中:$\mathcal{V}^r$为可达需求点的列表;$\mathcal{D}^r$为可达需求点i到储备点“0”的最短路径长度列表;$\mathcal{E}^r$为已修复路段列表. 如图1所示,若抢修队agent从储备点“0”出发,先后修复受损路段e_{01}和e_{15},则抢修队agent在节点5的状态为$\langle\{5,7\},\{\langle5,D_{50}\rangle,\langle7,D_{70}\rangle\},\{e_{01},e_{15}\}\rangle$.

本文约定抢修队的动作集$\mathcal{A}^r$为抢修队agent当前位置可达的受损路段集合,即

$$\mathcal{A}^r=\{e_{ij}|\varepsilon_{ij}=0\bigwedge(\theta_i=1\bigvee\theta_j=1)\}. \tag{6}$$

在决策时,通常选择能使较多需求点连通的受损路段进行修复,从而达到尽快完成全部需求点连通的目的. 当抢修队agent在状态$s_1^r=\langle\mathcal{V}_1^r,\mathcal{D}_1^r,\mathcal{E}_1^r\rangle$从动作集$\mathcal{A}^r$中选择受损路段$e_{ij}$进行抢修后,达到下一个状态$s_2^r=\langle\mathcal{V}_2^r,\mathcal{D}_2^r,\mathcal{E}_2^r\rangle$. 通常会有以下两种情况:

1) 没有新的需求点打通,即$\mathcal{V}_1^r=\mathcal{V}_2^r$. 抢修队agent执行动作$e_{ij}$后的即时奖励设为

$$R^r=\frac{1}{\dfrac{l_{ij}}{v}+t_{ij}}. \tag{7}$$

2) 有新的需求点打通,即$\mathcal{V}_1^r\subset\mathcal{V}_2^r$. 抢修队agent执行动作$e_{ij}$后的即时奖励设为

$$R^r=\sum_{i\in\mathcal{V}_2^r-\mathcal{V}_1^r}\left[(1-\lambda)\cdot\frac{I_i}{D_{i0}}+\lambda\cdot\frac{I_i}{C_{i0}^r}\right], \tag{8}$$

其中$\lambda\in(0,1)$为加权权重,表示新打通需求节点在最短路径长度和累计时间开销上的综合奖励.

根据上述设计,抢修队agent新打通需求节点的最短路径长度越短、累计时间开销越少、受灾程度越

严重,则抢修队的即时奖励越大. 抢修队间接影响目标函数(4),抢修队的修复效率越高,路段越能尽早打通,从而一定程度上减少运输队的通行时间,可以使得目标函数(4)更优.

3.1.2 运输队agent的建模

设运输队agent的状态s^d由二元组$\langle \mathcal{V}^d, \mathcal{D}^d\rangle$构成,即

$$\begin{cases} \mathcal{V}^d = \{i|\mu_i = 1, i \in V_d\}, \\ \mathcal{D}^d = \{\langle i, D_{i0}\rangle|\mu_i = 1, \forall i \in V_d\}. \end{cases} \tag{9}$$

其中:$\mathcal{V}^d$为已救援需求点的列表,$\mathcal{D}^d$为已救援需求点i到储备点"0"的最短路径长度列表. 如图1所示,运输队agent从储备点"0"出发,救援需求节点5以后,运输队agent在节点5的状态为$\langle\{5\}, \{\langle 5, D_{50}\rangle\}\rangle$.

本文将运输队agent的动作集$\mathcal{A}^d$设为其当前位置可达的待救援需求点,即

$$\mathcal{A}^d = \{i|\theta_i = 1 \bigwedge \mu_i = 0, i \in V_d\}. \tag{10}$$

设运输队agent在节点j的状态为$s_1^d = \langle \mathcal{V}_1^d, \mathcal{D}_1^d\rangle$,然后其从动作集$\mathcal{A}^d$中选择需求点$i$进行响应后,达到下一个状态$s_2^d = \langle \mathcal{V}_2^d, \mathcal{D}_2^d\rangle$. 此时,运输队agent执行动作$i$后的即时奖励设为

$$R^d = \frac{1}{D_{ji}}, \tag{11}$$

其中D_{ji}为从节点j到需求点i的最短距离.

3.2 双层规划Q学习的迭代公式

基于下层Q学习优化受损路网抢修策略,Q^r值的迭代公式为

$$\begin{aligned} Q^r \leftarrow &(1-\alpha)\cdot Q^r(s_1^r, e_{ij}) + \\ &\alpha\cdot[R^r + \gamma\cdot \max Q^r(s_2^r, e_{ij})]. \end{aligned} \tag{12}$$

其中:$\alpha \in (0,1)$为学习速率,用于调控前期训练产生的奖励对当前Q值更新的影响,α越大,保留之前训练的效果越少;$e_{ij} \in \mathcal{A}^r$为抢修队agent从动作集中选择的动作;R^r为抢修队agent选择动作e_{ij}可获得的即时奖励,可根据式(7)或(8)计算得到.

基于上层Q学习优化物资配送策略,Q^d值的迭代公式为

$$\begin{aligned} Q^d \leftarrow &(1-\alpha)\cdot Q^d(s_1^d, i) + \\ &\alpha\cdot[R^d + \gamma\cdot \max Q^d(s_2^d, i)]. \end{aligned} \tag{13}$$

其中:$i \in \mathcal{A}^d$是运输队agent从动作集中选择的动作;R^d是运输队agent选择动作i可获得的即时奖励,可根据式(11)计算得到;$\max Q^d(s_2^d, i)$是在状态s_1^d下选择动作i进入下一个状态s_2^d时能得到的最大Q值.

3.3 联合调度算法

结合上述思想,联合调度算法如算法1所示.

算法1 基于Q学习的受损路网抢修调度/物资配送调度

输入:路网模型,训练周期T,探索率ε,折扣因子γ,学习速率α;

输出:目标函数$f_U(P)$,运行时间,路网修复率.

1) 对路网模型、抢修队决策模型、运输队决策模型和Q学习的相关参数进行初始化.

2) for t := 1 to T

3) 将抢修队agent放置在储备站"0",并设置初始状态$s_1^r = \langle \mathcal{V}_1^r, \mathcal{D}_1^r, \mathcal{E}_1^r\rangle$和初始动作集$\mathcal{A}^r$.

4) 抢修队agent根据$\varepsilon \in (0,1)$贪心策略从$\mathcal{A}^r/\mathcal{A}^d$中选取并执行动作$e_{ij}/i$,抢修队agent从状态$s_1^r$过渡到状态$s_2^r = \langle \mathcal{V}_2^r, \mathcal{D}_2^r, \mathcal{E}_2^r\rangle$,根据式(12)计算对应的$Q^r$值,并更新$Q^r$值表. 往缓冲表$W$中输入已修复的受损路段集和累计时间开销$\langle \tilde{H}, C_{i0}^r\rangle$.

5) 若当前状态s_2^r已达到最终状态,即$\mathcal{V}_2^r = V_d$,则终止本轮抢修队的学习;若s_2^r不是最终状态,即$\mathcal{V}_2^r \neq V_d$,则转1)继续学习.

6) 将运输队agent放置在储备站"0",并设置初始状态$s_1^d = \langle \mathcal{V}_1^d, \mathcal{D}_1^d\rangle$和初始动作集$\mathcal{A}^d$.

7) 从缓冲表W中读取抢修队的累计时间开销C_{i0}^r,更新当前的路网状态.

8) 运输队agent根据$\varepsilon \in (0,1)$贪心策略从$\mathcal{A}^d$中选取并执行动作i,从状态s_1^d过渡到状态$s_2^d = \langle \mathcal{V}_2^d, \mathcal{D}_2^d\rangle$,根据式(13)计算对应的$Q^d$值并更新$Q^d$值表.

9) 如果当前状态s_2^d已达到最终状态(即$\mathcal{V}_2^d = V_d$),则终止本轮运输队的学习,根据式(4)计算配送策略P的目标函数值,更新当前最优策略P^*和最佳目标函数值,更新抢修队与运输队的目标函数值;如果s_2^d不是最终状态(即$\mathcal{V}_2^d \neq V_d$),则转4)继续学习.

10) 如果已达最大训练周期数,则终止训练并输出当前最优策略P^*和最佳目标函数值;否则转2)继续训练.

11) end for

12) 已达最大训练周期数,终止训练并输出当前最优策略H^*、P^*和最佳目标函数值.

4 实验结果与分析

为了验证本文算法的有效性,首先给出实验环境和参数设置,然后对比分析本文所提出的基于双层交互Q学习的路网抢修和物资配送联合调度算法的有效性(以下简称为IQLJS),最后将IQLJS算法与已有的ACO算法[23]、DP (dynamic programming)算法[13]和IQL (improved Q-learning)算法[15]进行深入的对比分析.

4.1 实验环境与参数设计

根据以往的实验研究及结果,本文设计3种不同规模的受损路网:$|V| = 20, |E| = 30$;$|V| = 40, |E| = 60$;$|V| = 60, |E| = 90$. 另外设置3种不同的受损情况$\left(\frac{|V_i|}{|V|}, \frac{|E_d|}{|E|}\right)$:(0.4, 0.6), (0.6, 0.6), (0.6, 0.8). 在本文中,测试数据是根据图结构生成的路网模型,测试实例的各项参数部分都可通过自定义设置,部分实验参数需控制在一定的取值范围内,以保证测试

数据的多样性和符合实际应用场景的特点,如表1所示. 本文测试实例数据的取值参考以往的研究成果[14-16,23]. 基于上述设置,共模拟9种不同的受损路网,每种路网设置10个不同的测试实例. 需要指出的是,由于对比的ACO算法在大规模路网下的测试实例非常耗时,并不能在可接受时间内得出结果. 因此,为了对比的公平性,收集整理3种算法在可接受时间内的所有测试结果,即在9种不同路网下的72个不同测试实例上的测试数据,以深入对比分析IQLJS、ACO、IQL和DP四种算法的性能.

表1 受损路网模型中符号含义及取值范围

符号	含义	取值范围
I_i	需求点i的受灾程度	(1～10)
l_{ij}	路段e_{ij}的通行长度	(1～10)
t_{ij}	受损路段e_{ij}的修复时间	(1～10)
d_i	节点i的救援时间	(1～10)

IQLJS算法的参数如下:探索率$\varepsilon = 0.1$,折扣因子$\gamma = 0.9$,学习速率$\alpha = 0.4$. 不同路网规模下的参数设置见表2,ACO参数与参考文献设置相同:蚂蚁数为80,总迭代次数为300,信息素挥发率为0.5,信息素调控因子为0.003,路程调控因子为0.02.

表2 不同路网规模下的参数设置

路网规模	IQL和IQLJS 最大训练周期数	ACO和DP 总迭代次数
$\|V\| = 20,\ \|E\| = 30$	1 200	300
$\|V\| = 40,\ \|E\| = 60$	1 500	400
$\|V\| = 60,\ \|E\| = 90$	1 800	500

每个测试实例均在Intel Xeon CPU 2.20 GHz、32 GB内存、Windows Server 2012操作系统的个人计算机上独立运行30次,并根据30次不同结果进行统计分析.

4.2 最优动作集更新策略的影响

有无最优动作集下测试的平均目标函数值如表3所示. 通过对比实验结果发现,在引入最优动作集更新策略后得到的目标函数值明显提升. 目标函数值降幅约在15%～30%之间. 尤其值得注意的是,在相同规模的路网下,受损率较高的测试实例所得到的目标函数值降幅更大. 例如,在路网规模为$(V, E) = (40, 60)$的情况下,路网参数$\left(\frac{D_i}{D_{i0}}, \frac{|V_i|}{|V|}, \frac{|E_d|}{|E|}\right)$为$(1.05, 0.6, 0.8)$时,目标函数的降幅为25%,而在参数为$(1.05, 0.4, 0.6)$和$(1.05, 0.6, 0.6)$的情况下,目标函数的降幅分别为19%和15%.

表3 有无最优动作集下测试的平均目标函数值

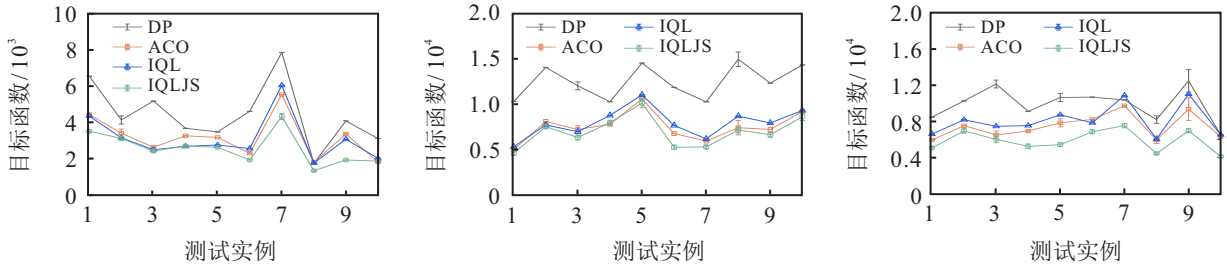
$\left(\frac{D_i}{D_{i0}}, \frac{\|V_i\|}{\|V\|}, \frac{\|E_d\|}{\|E\|}\right)$	$\|V\| = 20, \|E\| = 30$		$\|V\| = 40, \|E\| = 60$		$\|V\| = 60, \|E\| = 90$	
	最优动作集	无	最优动作集	无	最优动作集	无
(1.05, 0.4, 0.6)	1 927.56	2 626.61	14 077.81	18 248.21	45 810.56	57 132.95
(1.05, 0.6, 0.6)	5 283.09	7 192.04	36 044.64	37 995.15	77 214.30	91 784.33
(1.05, 0.6, 0.8)	5 467.36	7 773.40	36 329.54	47 625.65	91 782.32	123 529.67
(1.25, 0.4, 0.6)	1 865.74	2 691.96	13 455.27	16 865.25	37 698.09	45 312.37
(1.25, 0.6, 0.6)	2 761.67	3 816.02	21 666.78	28 796.08	74 556.49	86 900.80
(1.25, 0.6, 0.8)	4 805.57	6 998.72	34 614.66	45 098.03	77 906.45	97 653.48
(1.50, 0.4, 0.6)	2 861.79	4 497.05	11 366.20	13 707.36	41 908.85	53 259.17
(1.50, 0.6, 0.6)	3 105.31	3 409.34	36 723.06	48 296.45	68 168.52	77 359.76
(1.50, 0.6, 0.8)	5 212.45	7 395.21	26 687.31	35 269.15	68 359.52	95 174.63

上述实验结果表明,在相同路网规模和路网受损程度下,采用最优动作集更新策略可以明显提高Q学习的探索效率,并且能让抢修队和运输队选择对联合调度目标函数有利的动作,从而提高联合调度决策的效率.

4.3 不同算法的对比

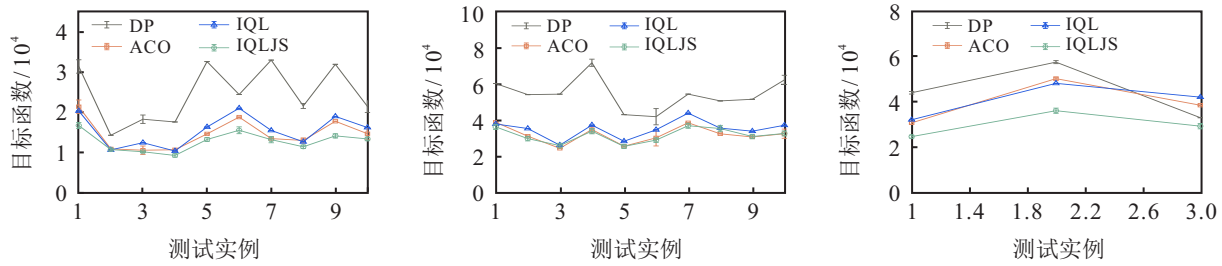
本文对4种算法在不同路网规模及参数下进行对比,结果如图3～图5所示. 对结果进行分析,可以得出以下结论:在大部分测试实例中,本文提出的IQLJS算法在目标函数的均值上优于ACO算法、IQL算法和DP算法. 尤其是在路段受损率和需求点占比率较高的情况下,如$\left(\frac{|V_i|}{|V|}, \frac{|E_d|}{|E|}\right) = (0.6, 0.8)$,IQLJS算法的优势更加明显. 需要注意的是,在少数测试实例中,ACO算法的目标函数值与IQLJS算法相当,甚至更好.

此外,观察到在路网规模和$\frac{D_i}{D_{i0}}$不变的情况下,当路网参数$\left(\frac{|V_i|}{|V|}, \frac{|E_d|}{|E|}\right)$从$(0.4, 0.6)$到$(0.6, 0.6)$,即需求点占比增加0.2时,IQLJS算法的目标函数值显著增加,平均增加了100%以上;而当参数从$(0.6, 0.6)$到$(0.6, 0.8)$,即路段受损率增加0.2时,目标函数值的增加并不明显,平均增加了约8%左右.



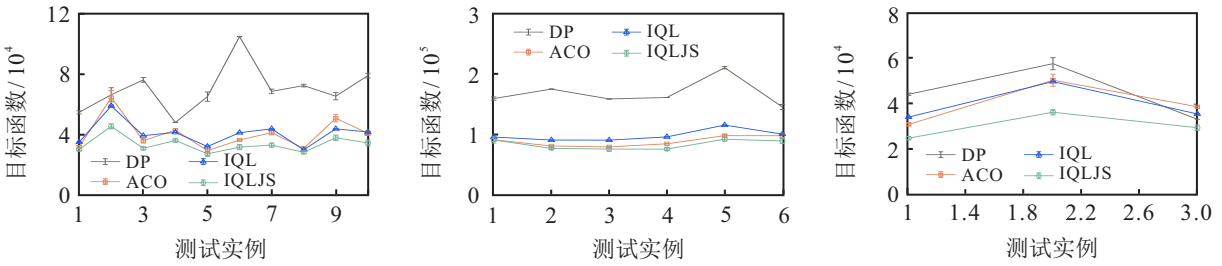
(a) $\left(\frac{D_i}{D_{i0}},\frac{|V_i|}{|V|},\frac{|E_d|}{|E|}\right)$: (1.05,0.4,0.6)　(b) $\left(\frac{D_i}{D_{i0}},\frac{|V_i|}{|V|},\frac{|E_d|}{|E|}\right)$: (1.05,0.6,0.6)　(c) $\left(\frac{D_i}{D_{i0}},\frac{|V_i|}{|V|},\frac{|E_d|}{|E|}\right)$: (1.05,0.6,0.8)

图 3　4种算法在$|V|=20,\ |E|=30$下的目标函数值(均值和标准差)



(a) $\left(\frac{D_i}{D_{i0}},\frac{|V_i|}{|V|},\frac{|E_d|}{|E|}\right)$: (1.05,0.4,0.6)　(b) $\left(\frac{D_i}{D_{i0}},\frac{|V_i|}{|V|},\frac{|E_d|}{|E|}\right)$: (1.05,0.6,0.6)　(c) $\left(\frac{D_i}{D_{i0}},\frac{|V_i|}{|V|},\frac{|E_d|}{|E|}\right)$: (1.05,0.6,0.8)

图 4　4种算法在$|V|=40,\ |E|=60$下的目标函数值(均值和标准差)

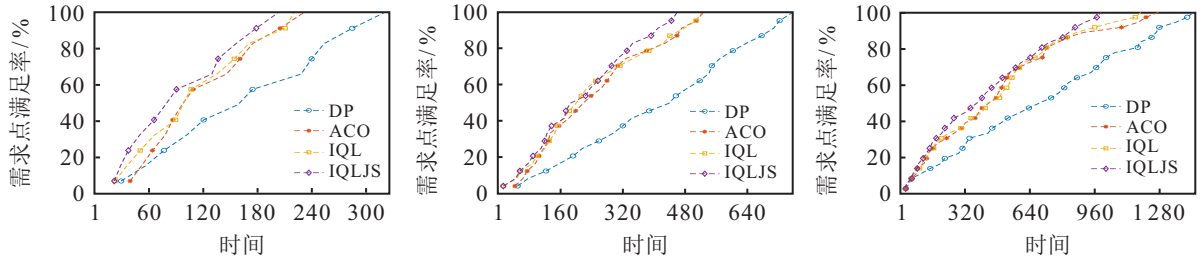


(a) $\left(\frac{D_i}{D_{i0}},\frac{|V_i|}{|V|},\frac{|E_d|}{|E|}\right)$: (1.05,0.4,0.6)　(b) $\left(\frac{D_i}{D_{i0}},\frac{|V_i|}{|V|},\frac{|E_d|}{|E|}\right)$: (1.05,0.6,0.6)　(c) $\left(\frac{D_i}{D_{i0}},\frac{|V_i|}{|V|},\frac{|E_d|}{|E|}\right)$: (1.05,0.6,0.8)

图 5　4种算法在$|V|=60,\ |E|=90$下的目标函数值(均值和标准差)

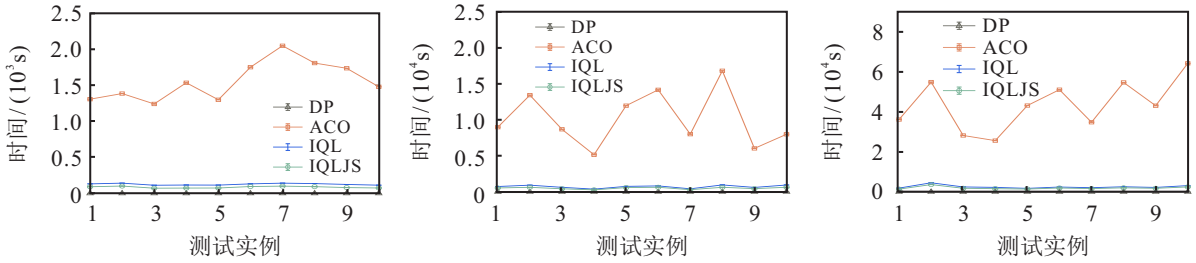
总体而言,IQLJS算法在大多数测试实例中表现优于对比的3种算法,而DP算法由于方法的局限性,效果明显不如IQLJS算法和ACO算法,只在少数测试实例中优于ACO算法.

本文研究了3种不同路网规模下4种算法的需求点满足率随时间的变化情况,如图6所示. 可以观察到,在3种路网规模下,4种算法都能够达到100%的需求点满足率. 具体而言,本文提出的IQLJS算法能够以最短的时间满足所有需求点的救援需求,在更大规模的路网下,尤其是当路网规模为$|V|=60,|E|=90$时,IQLJS算法的需求点满足率达到100%;而ACO算法、IQL算法和DP算法的需求点满足率分别仅为88.8%、90.2%和69.4%. 这显示出IQLJS算法在满足需求点方面具有明显的优势.



(a) $|V|=20,|E|=30$　(b) $|V|=40,|E|=60$　(c) $|V|=60,|E|=90$

图 6　不同路网规模下的需求点满足率



(a) $|V|=20,|E|=30$ (b) $|V|=40,|E|=60$ (c) $|V|=60,|E|=90$

图 7 相同受损程度、不同路网规模下 4 种算法的运行时间

4.4 算法运行时间

图 7 展示了 4 种算法在相同受损程度、不同路网规模下的平均运行时间. 观察结果表明, ACO 算法的运行时间明显远高于其他算法, 而 IQLJS 算法的运行时间略高于 DP 算法. 随着路网规模的增加, 4 种算法的运行时间逐渐增加, 但 ACO 算法的增幅也明显大于其他算法. IQL 算法运行时间略高于 IQLJS 算法, 因此, 相比其他算法, IQLJS 算法能够在合理的时间范围内获得较优的目标函数.

5 结 论

本文针对同时兼顾受损路网修复与灾后物资配送的应急场景构建了受损路网修复和物资配送联合调度模型, 并基于缓冲表、最优动作集更新策略和 Q 学习求解抢修队与运输队联合调度策略, 设计了一种基于双层交互 Q 学习的路网抢修和物资配送联合调度算法. 实验结果表明, 在联合调度的应急场景下, 本文所提方法与现有算法相比能合理地规划抢修队与运输队的应急策略, 花费较短的时间就能让所有需求点得到及时的救援, 并且应急效率要比现有算法更高. 但本文只是在自然灾害应急响应下对受损路网修复与物资配送联合调度问题的一个初步探索, 未来可以考虑多个抢修队和运输队从不同的储备点出发进行抢救活动, 如何协调不同抢修队和不同运输队之间的联系, 及时有效地完成救援工作值得进一步研究.

参考文献(References)

[1] 应急管理部. 应急管理部发布2022年全国自然灾害基本情况[Z]. 防灾博览, 2023(1): 26-27.

[2] 中国应急管理. 国务院印发《“十四五”国家应急体系规划》[J]. 中国应急管理, 2022(2): 4.

[3] Rabiei P, Arias-Aranda D. Introducing a novel multi-objective optimization model for vehicle routing and relief supply distribution in post-disaster phase: Combining fuzzy inference systems with NSGA-II and NRGA[C]. 2021 6th International Conference on Transportation Information and Safety. Wuhan, 2021: 1226-1243.

[4] Chang K H, Hsiung T Y, Chang T Y. Multi-Commodity distribution under uncertainty in disaster response phase: Model, solution method, and an empirical study[J]. European Journal of Operational Research, 2022, 303(2): 857-876.

[5] Huang Y, Han H, Zhang B, et al. Supply distribution center planning in UAV-based logistics networks for post-disaster supply delivery[C]. 2020 IEEE International Conference on E-health Networking, Application & Services. Shenzhen, 2021: 1-6.

[6] 孙笑, 宋卫星, 班利明, 等. 复杂人力资源约束下的抢占式维修工序调度[J]. 控制与决策, 2022, 37(2): 393-400.
(Sun X, Song W X, Ban L M, et al. Preemptive maintenance process scheduling under complex human resource constraints[J]. Control and Decision, 2022, 37(2): 393-400.)

[7] 张国富, 陆淑君, 苏兆品, 等. 化工园区应急物资多目标分配问题建模与求解[J]. 控制与决策, 2022, 37(4): 962-972.
(Zhang G F, Lu S J, Su Z P, et al. Modelling and solving the multi-objective allocation problem of emergency supplies in chemical parks[J]. Control and Decision, 2022, 37(4): 962-972.)

[8] 刘扬, 张国富, 苏兆品, 等. 救灾物资多阶段分配与调度问题建模与求解[J]. 控制与决策, 2019, 34(9): 2015-2022.
(Liu Y, Zhang G F, Su Z P, et al. Modeling and solving multi-phase allocation and scheduling of emergency relief supplies[J]. Control and Decision, 2019, 34(9): 2015-2022.)

[9] 宋英华, 葛艳, 杜丽敬, 等. 考虑车辆等待的应急物资调配方案优化研究[J]. 控制与决策, 2019, 34(10): 2229-2236.
(Song Y H, Ge Y, Du L J, et al. Optimization of emergency materials allocation plan considering vehicle waiting[J]. Control and Decision, 2019, 34(10): 2229-2236.)

[10] 张国富, 王永奇, 苏兆品, 等. 应急救援物资多目标分配与调度问题建模与求解[J]. 控制与决策, 2017, 32(1): 86-92.
(Zhang G F, Wang Y Q, Su Z P, et al. Modeling and solving multi-objective allocation-scheduling of emergency relief supplies[J]. Control and Decision, 2017,

32(1): 86-92.)

[11] Ajam M, Akbari V, Salman F S. Minimizing latency in post-disaster road clearance operations[J]. European Journal of Operational Research, 2019, 277(3): 1098-1112.

[12] 李兆隆, 金淳, 胡畔, 等. 基于弹复性的交通网络应急恢复阶段策略优化[J]. 系统工程理论与实践, 2019, 39(11): 2828-2841.
(Li Z L, Jin C, Hu P, et al. Optimization of emergency recovery phase strategies for transportation networks based on resilience[J]. Systems Engineering — Theory & Practice, 2019, 39(11): 2828-2841.)

[13] Maya Duque P A, Dolinskaya I S, Sörensen K. Network repair crew scheduling and routing for emergency relief distribution problem[J]. European Journal of Operational Research, 2016, 248(1): 272-285.

[14] 苏兆品, 李沫晗, 张国富, 等. 基于Q学习的受灾路网抢修队调度问题建模与求解[J]. 自动化学报, 2020, 46(7): 1467-1478.
(Su Z P, Li M H, Zhang G F, et al. Modeling and solving the repair crew scheduling for the damaged road networks based on Q-learning[J]. Acta Automatica Sinica, 2020, 46(7): 1467-1478.)

[15] 张国富, 涂冰花, 苏兆品, 等. 一种面向严重受损路网的抢修队调度算法[J]. 控制与决策, 2021, 36(7): 1663-1671.
(Zhang G F, Tu B H, Su Z P, et al. An algorithm for repair crew scheduling on severely damaged road network[J]. Control and Decision, 2021, 36(7): 1663-1671.)

[16] 张国富, 常加远, 苏兆品, 等. 大量需求点下基于深度Q学习的受损路网抢修队调度[J]. 控制与决策, 2022, 37(12): 3267-3277.
(Zhang G F, Chang J Y, Su Z P, et al. Repair crew scheduling for damaged road network with enormous demand points using deep Q-learning[J]. Control and Decision, 2022, 37(12): 3267-3277.)

[17] Su Z P, Duan L Q, Zhang G F, et al. Multicrew scheduling and routing in road network restoration based on deep Q-learning[C]. AAAI-22 Workshop on Machine Learning for Operations Research (ML4OR). Vancouver, 2022.

[18] Souza Almeida L, Goerlandt F, Pelot R. Trends and gaps in the literature of road network repair and restoration in the context of disaster response operations[J]. Socio-Economic Planning Sciences, 2022, 84: 101398.

[19] Farzaneh M A, Rezapour S, Baghaian A, et al. An integrative framework for coordination of damage assessment, road restoration, and relief distribution in disasters[J]. Omega, 2023, 115: 102748.

[20] 陈钢铁, 帅斌. 震后道路抢修和应急物资配送优化调度研究[J]. 中国安全科学学报, 2012, 22(9): 166-171.
(Chen G T, Shuai B. Optimizing emergency road repair and distribution of relief supplies after earthquake[J]. China Safety Science Journal, 2012, 22(9): 166-171.)

[21] 张梦玲, 王晶, 黄钧, 等. 基于手机定位数据的突发事件下道路修复和物资配送集成优化研究[J]. 中国管理科学, 2021, 29(3): 133-142.
(Zhang M L, Wang J, Huang J, et al. Research on the integrated optimization of road repair and relief distribution based on mobile phone location data[J]. Chinese Journal of Management Science, 2021, 29(3): 133-142.)

[22] Ransikarbum K, Mason S J. Multiple-objective analysis of integrated relief supply and network restoration in humanitarian logistics operations[J]. International Journal of Production Research, 2016, 54(1): 49-68.

[23] Shin Y, Kim S, Moon I. Integrated optimal scheduling of repair crew and relief vehicle after disaster[J]. Computers and Operations Research, 2019, 105(C): 237-247.

作者简介

张国富 (1979–), 男, 教授, 博士, 主要研究方向为智慧应急、进化算法、软件工程, E-mail: zgf@hfut.edu.cn;

朱前顺 (1998–), 男, 硕士生, 主要研究方向为智慧应急, E-mail: 2021171124@hfut.edu.cn;

苏兆品 (1983–), 女, 副教授, 博士, 主要研究方向为智慧应急、多媒体安全, E-mail: szp@hfut.edu.cn;

岳峰 (1981–), 男, 副研究员, 博士, 主要研究方向为软件工程、多媒体安全, E-mail: yuefeng@hfut.edu.cn.