

# 结合 Double DQN 算法的园区应急避难规划研究

王晨鹭, 李佳琦, 马菲娅, 陆一平

(北京交通大学 机械与电子控制工程学院, 北京 100044)

**【摘要】**人员密集型组织机构的防灾避难功能研究具有重要的理论和现实意义。园区应急避难规划主要包括应急避难场所选址、路径规划以及后续医疗服务配合,是园区应急管理在设计初期就需要予以重视的问题。选择以对应急避难需求度较高的校园作为研究对象,统筹考虑地震、火灾以及军事打击三种灾难形式下园区人员在工作、就餐以及就寝三种时段的分布情况,对医疗救治方案进行规划,运用深度强化学习作为研究方法,编写出计算机程序并训练得到一个良好结果,建立以得到最短疏散时间、最佳医疗服务方案和最小规划成本为目标的园区应急避难规划模型。为了验证规划模型的有效性,以北京市某大学校园对象进行分析,结果表明规划模型效果良好,能够有效地对人员密集园区的应急避难方案进行合理规划。

**关键词:** 应急避难规划; 多灾难情况; 深度学习; 强化学习; 优化模型; Double DQN

## 0 引言

近年来,城市作为现代社会人类主要的聚居点,在现代化进程中有着至关重要的作用,城市人口密度越来越高,同时城市需要应对频繁的自然灾害,如地震、洪水等。工业革命以来,由于物质资料不断丰富,人口增长、城市化水平也越来越高,社会经济不断发展、灾难的种类以及造成的破坏性后果都有了更为复杂的特点。例如,我国是多地震的国家,也是不时遭受严重地震震害的国家之一。中国地震局在 2022 年发布的文章表明[1],我国占全球 7 级以上的地震约 1/3,我国有 40% 以上地区属于 7 级地震烈度区,且有 70% 的百万以上人口大城市处于地震区。虽然我国城市发展进程加快与高精度技术迅猛发展,但是城市的整体防灾减灾功能一直远远滞后于城市发展。因此,如何在灾难发生时做好应急疏散避难工作、将损失降到最低;如何在灾难发生时紧急疏散人群,如何进行安置这些问题引起了更多人的思考和关注,也是当今社会需要重视的一大难题。

应急避难场所是为了人们能在灾难发生后一段时间内,躲避由灾害带来的直接或间接伤害,并能保障基本生活而事先划分的带有一定功能设施的场地,现代城市中建筑密度大,可用的土地资源有限,而城市公园、绿地、广场、学校操场、大型露天停车场等因空间大,也因此作为应急避难场所的首选,将城市应急避难场所纳入到城市规划体系中,能为城市居民提供安全避难、基本生活场地及灾后救援的场所,具有“直接启用”的特点。为应对突发事件,经规划建设具有应急避难生活服务的设施,可供居民应急疏散、临时生活的安全场所,加强应急避难场所的规划建设,是提高城市综合防灾能力、减轻灾害影响、增强政府应急管理工作能力的重要举措。在 21 世纪的今天[2],除自然灾害外,公共卫生事件、恐怖事件、战争等人为灾害也在不断地威胁着人类的生命和财产安全。[3]我国又是一个自然灾害频发的国家,几乎每年都会发生洪水、干旱等自然灾害,灾害一旦发生,造成的损失将非常严重。这时,应急避难规划研究显得极为重要,我国

的城市应急避难场所建设正是基于先进的应急管理理念和为城市居民提供更加安全的城市基础环境, 顺应我国经济社会新发展。

此前已有许多学者在单一灾害场景中的疏散路径以及选址优化问题上展开研究, 李梦雅等[4]考虑需求控制、容量限制、交通延误、公平分配和资源节约等约束条件。对经典 Dijkstra 算法进行改进, 并采用混合拆分疏散方法, 构建洪灾避难应急疏散路径规划模型。运用 C# 语言编写算法, 求解最佳路径。王国英[5]基于自主开发的应急疏散分析工具 MiniGIS。对规划路径进行动态模拟。王泽等人[6]选址-路径双层规划模型, 从疏散的快速性、全面性两个角度。为保证疏散方案的高效性提供重要的模型。决策考虑如何制定疏散公共车辆的最佳疏散路线, 使受灾人群尽可能快速的到达安全区域、给定若干疏散公交集结点的位置和疏散需求。通过求解车辆路径问题 (Vehicle Routing Problems, VRP)、达到疏散总时间、也包括等待时间。黄梦溪等人[7]2BM 技术在火灾疏散中的应用研究 BM 的作用是创建一个包含了所有建筑信息的虚拟建筑模型, 以多视角的、参数化的、智能的建筑构件为基础, 建立一个完全的 3D 环境, 以此 BM 作为构建基础的项目管理系统可以持续的、及时的提供各项实时数据。张佳瑜等人[8]通常使用平面两点间的坐标来获取欧氏距离作为避震距离 16, 这与微观角度如在校园道路网中的实际避震距离并不能很好拟合。网络路径长度能反映实际路网对空间可达性的影响, 可以更精确地表示出避震服务范围。采用真实道路网中的路径长度作为避震距离度量。迪杰斯特拉 (Dijkstra) 算法是网络中求最短路径问题的经典算法。

根据国内外研究现状, 学者主要关注单一灾害形式下的应急避难场所规划, 致力于解决单一灾害形式下应急避难场所的选址以及人员疏散的路线问题。其中, 以单一的地震灾害场景中应急避难规划为主, 洪水和火灾等灾害场景的研究相对较少, 虽然有研究考虑次生灾害的影响, 但也是同类灾害, 比如地震和余震, 因此, 对于多灾害形式下优化模型的研究较少, 缺多统筹多种灾害的应急避难规划方案, 并且少有研究关注灾害在人员密集型组织机构中的应用, 而深度学习 (DeepLearning, DL) 作为机器学习领域一个重要的研究热点, 已经在图像分析、语音识别、自然语言处理、视频分类等领域取得了令人瞩目的成功。DL 的基本思想[9]是通过多层的网络结构和非线性变换, 组合低层特征, 形成抽象的、易于区分的高层表示, 以发现数据的分布式特征表示。因此 DL 方法侧重于对事物的感知和表达。强化学习 (Reinforcement Learning, RL) 作为机器学习领域另一个研究热点, 已经广泛应用于工业制造、仿真模拟、机器人控制、优化与调度、游戏博弈等领域。RL 的基本思想[10]是通过最大化智能体 (agent) 从环境中获得的累计奖赏值, 以学习到完成目标的最优策略。因此 RL 方法更加侧重于学习解决问题的策略。随着人类社会的飞速发展, 在越来越多复杂的现实场景任务中, 需要利用 DL 来自动学习大规模输入数据的抽象表征, 并以此表征为依据进行自我激励的 RL, 优化解决问题策略。Mnih 等人[11]在《Nature》上发表论文, 提出深度学习与强化学习相互结合的深度 Q 网络 (DeepQNet-work, DQN) 模型, 经过训练后在 Atari2006 游戏中的表现超过人类水平。DeepMind 团队[12]更进一步将深度强化学习应用到动作空间更大、策略更复杂的围棋游戏上, 开发的 AlphaGo 程序基于深度神经网络, 采用蒙特卡洛树搜索算法 (MonteCarloTreeSearch, MCTS), 同时融合监督学习和强化学习的训练方式, 学习到超出人类水平的围棋策略, 于 2016 年击败顶级人类棋手李世石, 并于 2017 年击败围棋世界冠军柯洁。此外, 面对环境更复杂的在线战术竞技游戏, 深度强化学习也有不俗的表现。OpenAI 团队[13]面向 DOTA2 游戏开发的机器人程序,

在 1 对 1 对战中击败顶级人类玩家。除了游戏中的机器人程序，近年来研究者们基于深度强化学习提出许多新的概念和方法，并应用于工程项目中。DeepMind 团队[14]用模仿学习实现机器人的类人运动控制，而这一过程只需要少量的人类专家样本。Finn 等人[15]结合深度强化学习与机器人抓取动作的预测，在训练机器人抓取策略的同时实现图像预测算法的自监督训练。深度强化学习在自然语言处理、自动驾驶、推荐搜索系统等领域也有应用。

Double DQN 算法是一种改进的深度强化学习算法，最早由 Hado 等人[16]于 2015 年提出，旨在解决传统 DQN 算法中存在的过高估计问题。DQN 算法是一种基于神经网络的 Q 学习算法，用于处理强化学习中的离散状态、离散动作空间问题。Double DQN 算法通过引入两个神经网络来解决这个问题：一个是行动值函数网络，用于选择下一个动作；另一个是目标值函数网络，用于计算下一个状态的目标值。在实际应用中，Double DQN 算法在处理连续动作空间问题时，通常会将 DQN 算法与其他算法进行结合，如深度确定性策略梯度（DDPG）算法，以获得更好的性能。

根据国内外研究现状，学者主要关注单一灾害形式下的应急避难场所规划，致力于解决单一灾害形式下应急避难场所的选址以及人员疏散的路线问题，少有关注多灾难的情况，且基本采用的方法均为传统的运筹学方法[18]来解决问题，而采用深度强化学习的方法来对应急避难规划的研究此前没有，故本文采用较为新颖的深度强化学习的方法来对园区应急避难规划进行研究。

## 1 问题描述与假设

### 1.1 问题描述

在人员密集型园区面临灾害时，在给定的园区网络中，考虑园区内人员不同时段的分布情况以及潜在灾害的类型，依据目前园区内地理情况及设施条件，在最短撤离路径及最小规划成本下为人员提供避难点和必要的医疗保障服务。

### 1.2 问题假设

为了进行下一步的研究，我们对园区应急避难情况做出了如下假设：

- ①园区可能面临火灾、地震、军事打击等灾难；
- ②园区建筑视为平面点，不考虑内部结构及建筑面积；
- ③所有应急避难点都可用于任一需求点；
- ④所有应急避难点的容量已知；
- ⑤所有需求点都能被园区内应急避难点所覆盖；
- ⑥备选应急避难点与需求点位置都已知且离散；
- ⑦医疗服务能覆盖所有应急避难点；
- ⑧灾难发生地理位置及严重程度已知。

## 2 地图环境建模

通过实际收集整理数据，将各地址按顺序编号，建立邻接矩阵，若两个位置之间存在直接路径则将权重填入邻接矩阵，不存在则将值设为正无穷（计算机中设为一个大的整形数值）。



### 3 算法介绍

#### 3.1 Double DQN

Double DQN (Double Deep Q-Networks) 是一种深度强化学习算法, 是对原始的 DQN 算法进行了改进。DQN 算法的目标是学习一个将状态映射到动作的 Q 函数，并使用神经网络来估计给定状态下每个动作的 Q 值。但是，DQN 算法存在 Q 值高估的问题，可能导致次优的策略。Double DQN 通过引入第二个网络来估计 Q 值并调整学习过程，以解决这个问题。

在 Double DQN 中，有两个神经网络：在线网络和目标网络。在线网络用于根据当前状态选择动作，而目标网络用于评估所选动作的值。在线网络在学习过程中的每个步骤都会更新，而目标网络定期通过复制在线网络的权重来更新。Double DQN 的主要思想是使用在线网络选择具有最高 Q 值的动作，然后使用目标网络评估其价值。这是通过使用目标网络计算所选动作的 Q 值来完成的，这有效地减少了标准 DQN 中可能发生的 Q 值高估。

更具体地说，Double DQN 算法可以概括为：

1. 使用随机权重初始化在线网络和目标网络。
2. 重复以下步骤，直到完成一定数量的学习周期：
  - a. 观察当前状态。
  - b. 使用在线网络使用 epsilon-greedy 策略选择动作。
  - c. 执行动作并观察奖励和下一个状态。
  - d. 计算 TD 目标：

- i. 使用在线网络选择具有最高  $Q$  值的动作。
- ii. 使用目标网络计算所选动作的  $Q$  值。
- iii. 计算 TD 目标为  $r + \gamma * Q(s', \arg\max(Q(s', a; \text{在线网络}); \text{目标网络}))$ 。
- e. 使用 TD 误差更新在线网络的权重。
- f. 定期通过复制在线网络的权重来更新目标网络的权重。

### 3.2 基于 Double DQN 的路径规划

首先将邻接矩阵的数据以字典的格式读入计算机，并将每条路径上的人数约束加入字典，其中每个键表示节点，每个值表示与该节点相邻的节点及其对应的边的权重，每条边的人数限制，以及将需求点设置为起点，避难点设置为终点，以及设置当前各个需求点的人数，分别将之后设置 DDQN 所需的超参数，如折扣因子、学习率、记忆缓冲区大小等。DDQN 类是一个简单的全连接神经网络，用于从状态空间到动作空间的函数近似。之后定义了一些辅助函数，如选择动作、获取奖励等。训练过程分为多个阶段。在每个阶段，从起点开始，不断选择动作并转移到下一个状态，直到达到终点，在每次迭代中，将经历的状态、动作、奖励、下一个状态以及是否达到终点的信息添加到记忆缓冲区中，其中之前文章中的约束条件会被写入到选择动作函数中以确保过滤掉不符合约束条件的动作，在每个阶段结束时，对缓冲区中的经验进行采样，并利用这些经验更新神经网络参数。

具体的步骤如下：

1. 从记忆缓冲区中随机采样一批经验。
2. 计算当前状态的  $Q$  值。
3. 计算下一个状态的  $Q$  值，但用于计算目标  $Q$  值的是当前状态的最大  $Q$  值。
4. 计算目标  $Q$  值，即折扣后的最大  $Q$  值加上当前状态的即时奖励。
5. 使用均方误差损失函数计算损失。
6. 对神经网络进行反向传播以更新参数。

在训练过程中，会逐渐减小探索参数  $\epsilon$ ，使得模型在训练后期更多地利用当前知识进行决策。

在完成所有阶段的训练后，将  $\epsilon$  设置为 0，从起点开始，根据神经网络做出决策，生成从起点到终点的最短路径，以及人数，直到最终需求点的人都到达避难点。

### 3.3 结果分析

在以（运动场，思源草坪）为应急避难场所位置的基础上，使用与路径规划以及相应约束限制相结合的 Double DQN 算法，对避难需求的疏散路径进行优化决策，为管理部门提供决策参考，通过调查以及估算，得出以大学校园为例的 3 种灾难形式下的 3 种时间段内需求区域的人数，相应数据收集的部分由于篇幅过大，此处略去。

由于九种场景过多，我们以火灾场景为例展开结果分析。我们将介绍三种紧急情况下的避难路径规划结果，分别是火灾-上课-教室、火灾-就寝-宿舍和火灾-就餐-餐厅。其中，每种情况下都列出了需求点、交叉口、避难点的编号，以及



根据 Double DQN 算法模型计算得出的最优疏散路径方案。详细结果如下所示：

(1) 火灾-上课-教室避难路径规划：

根据 Double DQN 算法模型的计算结果，如果思源楼、思源西楼和思源东楼发生火灾，所有师生需要在最短时间内到达网球场和思源草坪这两个应急避难场所。在路径规划图中，数字 1 至 3 代表需求点，即思源楼、思源西楼和思源东楼；数字 4 至 22 代表路径中转节点，即校园道路交叉口；数字 23 和 24 分别代表避难点，即网球场和思源草坪。在区域 1 中，共有 1089 名师生需要通过节点 6 前往避难点 24，而 2829 名师生需要通过节点 8 和 14 前往避难点 23；在区域 2 中，共有 4125 名师生需要通过节点 6 前往避难点 24；在区域 3 中，共有 1580 名师生需要通过节点 10 前往避难点 24。火灾-上课-教室的疏散路径方案如图 2 所示。



(2) 火灾-就寝-宿舍避难路径规划：

根据 Double DQN 算法模型的计算结果，如果宿舍楼 16 号楼和 18 号楼发生火灾，所有师生需要在最短时间内到达东体育场这个应急避难场所。在路径规划图中，数字 1 和 2 代表需求点，即宿舍楼 16 号楼和 18 号楼；数字 4 至 9 代表路径中转节点，即校园道路交叉口；数字 10 代表避难点，即东体育场。在区域 1 中，共有 1621 名师生需要通过节点 5 前往避难点 10；在区域 2 中，共有 811 名师生需要依次通过节点 3、4 和 5 前往避难点 10。火灾-就寝-宿舍的疏散路径方案如图 3 所示。



### (3) 火灾-就餐-餐厅避难路径规划:

根据 Double DQN 算法模型的计算结果,如果学活餐厅、22 号楼、15 号楼和西餐厅发生火灾,所有师生需要在最短时间内到达篮球场、网球场和篮/排球场这三个应急避难场所。在路径规划图中,数字 1 至 4 代表需求点,即学活餐厅、22 号楼、15 号楼和西餐厅;数字 5 至 24 代表路径中转节点,即校园道路交叉口;数字 25 至 27 代表避难点,即篮球场、网球场和篮/排球场。在区域 1 中,共有 1508 名师生需要通过节点 6、13 和 16 前往避难点 25,其余 654 名师生依次通过节点 7、10、24、5、12 前往应急避难场所 27;在区域 2 中,共有 369 名师生需要依次通过节点 24、5 和 12 前往避难点 27;在区域 3 中,共有 201 名师生需要依次通过节点 14 和 18 前往避难点 25。火灾-就餐-餐厅的疏散路径方案如图 4 所示。



根据 Double DQN 算法避难路径优化模型的计算结果，我们得到了 9 种人员疏散路径方案，覆盖了 9 种不同的灾害场景。虽然这 9 种场景只考虑了单一的灾害形式和时段，但是在面对多种灾害形式时，可以将不同场景的方案进行组合，从而实现更全面的综合预防。例如，在地震灾害场景下，可以将火灾场景下的部分避难场所作为避难点，而不必再搭建新的避难场所。此外，如果地震和火灾同时发生，也可以根据各个场景下的疏散方案，将灾害区域内的人员疏散至相关的避难点，实现更好的综合预防。

## 4 结论以及未来展望

在人员密集型组织机构中，应急避难规划的研究至关重要。为了应对这种特殊场景，我们通过使用深度强化学习算法 Double DQN 算法进行疏散路径优化。该模型考虑了地震、火灾和军事打击等 3 种不同的灾难形式，以及工作、就餐和就寝等 3 种不同的人员分布时段的影响，旨在实现应急避难中规划成本最小和疏散时间最短的目标。此外，我们还通过实地测量不同时段和地点的人数，针对高校校园等特殊的人员密集型组织机构进行了测试和优化。实验结果表明，在 9 种灾害场景中，我们提出的优化模型能够整体上筛选出 8 个露天应急避难场所和 4 个地下避难场所，并且在最短时间内将灾害发生地的人员疏散到各个应急避难场所。通过使用新兴的深度强化学习算法 Double DQN 算法与在一定约束条件下的路径规划问题相结合，最终求得结果，通过与传统的运筹学方法进行对比可以发现，结果基本一致。

相对于单一灾害场景中的避难规划问题，我们的模型可以统筹考虑 3 种灾害类型的应急预案，在减少二次搭建避难场所时间的同时，将灾害区域内的人员安全疏散至各个避难点或者其他灾害场景中的避难点。因此，我们的模型不仅能够提供准确的规划方案，还可以适应多灾害场景下的应急避难规划需求，为应急管



理部门提供科学合理的应急规划依据。

我们采用的方法的优势在于不依赖于商用的收费软件，较为新颖，可以迅速部署到各个高校，模型收敛得也较快，在更大型的问题上可能有更快更好的表现。我们由于是第一个将深度强行学习应用于此种路径规划问题，或许还存在较多的改进的地方，未来考虑将其他的深度强化学习运用于重点研究建筑物楼层、出口数量和避难场所失效约束等方面的问题。

- [1]中国地质局.(2022).地震基础知识.取自 <http://www.cga.gov.cn/cga/xwzx/zfxgzdt/2016ztjj/2022ztjj/20220722/index.html>
- [2]Wannous C, Velasquez G. United nations office for disaster risk reduction (unisdr)—unisdr's contribution to science and technology for disaster risk reduction and the role of the international consortium on landslides (icl)[C]//Advancing Culture of Living with Landslides: Volume 1 ISDR-ICL Sendai Partnerships 2015-2025. Springer International Publishing, 2017: 109-115.
- [3]Zhang X, Yi L, Zhao D. Community-based disaster management: a review of progress in China[J]. Natural hazards, 2013, 65: 2215-2239.
- [4]李梦雅,王军,沈航.洪灾应急疏散路径规划算法优化[J].地球信息科学学报,2016,18(03):362-368.
- [5]王国英. 基于 QTE 的嵌入式 GPS/GIS 车载导航系统设计与实现[D].浙江大学,2007.
- [6]王泽. 台风灾害下区域疏散公交集结点选址和车辆路径规划[D].哈尔滨工业大学,2010.
- [7]黄梦溪,王哲.火灾应急疏散路径优化研究综述[J].科学技术创新,2020(02):104-105.
- [8]张佳瑜,白林波,杨文伟.基于 GIS 的社区地震应急避难场所配置模型构建——以银川市育林巷社区为例[J].地震工程学报,2019,41(06):1650-1658.
- [9]LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436-444.
- [10]Kaelbling L P, Littman M L, Moore A W. Reinforcement learning: A survey[J]. Journal of artificial intelligence research, 1996, 4: 237-285.
- [11]Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. nature, 2015, 518(7540): 529-533.
- [12]Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. nature, 2016, 529(7587): 484-489.
- [13]Berner C, Brockman G, Chan B, et al. Dota 2 with large scale deep reinforcement learning[J]. arXiv preprint arXiv:1912.06680, 2019.
- [14]Heess N, TB D, Sriram S, et al. Emergence of locomotion behaviours in rich environments[J]. arXiv preprint arXiv:1707.02286, 2017.
- [15]Finn C, Levine S. Deep visual foresight for planning robot motion[C]//2017 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2017: 2786-2793.
- [16]Van Hasselt H, Guez A, Silver D. Deep reinforcement learning with double q-learning[C]//Proceedings of the AAAI conference on artificial intelligence. 2016, 30(1).
- [17]焦思童,张淑菁,秦圣奇,李登飞,陆一平.三种灾难形式下的园区应急避难规划研究[J].数学建模及其应用,2021,10(04):55-63.DOI:10.19943/j.2095-3070.jmmia.2021.04.07.