智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于强化学习的超高层建筑非法入侵情景推演方法

胡今鸣, 胡啸峰, 石磊, 石拓, 滕腾

引用本文:
胡今鸣, 胡啸峰, 石磊, 等. 基于强化学习的超高层建筑非法入侵情景推演方法[J]. 智能系统学报, 2025, 20(4): 958-968.
HU Jinming, HU Xiaofeng, SHI Lei, et al. Method of unauthorized intrusion scenario simulation in super high-rise building based on reinforcement learning[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(4): 958-968.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202408002

DOI: 10.11992/tis.202408002
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20250114.1913.004

基于强化学习的超高层建筑非法入侵情景推演方法

胡今鸣[1], 胡啸峰[1,2,3], 石磊[4], 石拓[5], 滕腾[1]
（1. 中国人民公安大学 信息网络安全学院, 北京 100038; 2. 中国人民公安大学 首都社会安全研究基地, 北京 100038; 3. 安全防范技术与风险评估公安部重点实验室, 北京 102623; 4. 中国传媒大学 媒体融合与传播国家重点实验室, 北京 100024; 5. 北京警察学院 公安管理系, 北京 102202）

摘　要: 为计算超高层建筑潜在非法入侵者的“最优”入侵路径, 本文提出了一种基于强化学习的情景推演方法。该方法将建筑公共走廊抽象为拓扑结构, 利用贝叶斯网络计算入侵者通过每个拓扑节点的概率, 结合强化学习算法获得外部人员的最优入侵路径, 为超高层建筑非法入侵的高效防范提供精准依据。为验证方法的有效性, 以北京市 CBD 地区某超高层建筑为例, 将入侵终点设置为顶层, 设计了 3 种不同的入侵情景。情景推演结果表明: 在初始状态下（未进行任何优化措施）, SARSA 模型的训练性能最佳。优化安防系统后发现, 在建筑内的层间节点增加安防系统投入最有效。该优化情景下, 安防系统投入与风险值的非线性拟合结果显示, 随着安防系统投入的增加, 入侵风险显著降低。

关键词: 非法入侵; 情景推演; 超高层建筑; 强化学习; 贝叶斯网络; 安防系统; SARSA 模型; 非线性回归
中图分类号: TP18; X937　　**文献标志码**: A　　**文章编号**: 1673−4785(2025)04−0958−11

中文引用格式: 胡今鸣, 胡啸峰, 石磊, 等. 基于强化学习的超高层建筑非法入侵情景推演方法 [J]. 智能系统学报, 2025, 20(4): 958–968.
英文引用格式: HU Jinming, HU Xiaofeng, SHI Lei, et al. Method of unauthorized intrusion scenario simulation in super high-rise building based on reinforcement learning[J]. CAAI transactions on intelligent systems, 2025, 20(4): 958–968.

Method of unauthorized intrusion scenario simulation in super high-rise building based on reinforcement learning

HU Jinming[1], HU Xiaofeng[1,2,3], SHI Lei[4], SHI Tuo[5], TENG Teng[1]
(1. School of Information and Cyber Security, People's Public Security University of China, Beijing 100038, China; 2. Center for Capital Social Safety, People's Public Security University of China, Beijing 100038, China; 3. Key Laboratory of Security Technology & Risk Assessment, Ministry of Public Security, Beijing 102623, China; 4. State Key Laboratory of Media Integration and Communication, Communication University of China, Beijing 100024, China; 5. Department of Public Security Management, Beijing Police College, Beijing 102202, China)

Abstract: To calculate the “optimal” intrusion path of potential illegal intruders in super high-rise buildings, a scenario simulation method based on reinforcement learning is proposed in the paper. This method provides a precise basis for efficiently preventing illegal access in super high-rise buildings by abstracting the buildings' public corridors into a topological structure, calculating the probability of an intruder passing through each node based on a Bayesian network, and exploring the optimal intrusion path by means of reinforcement learning algorithms. To validate this method, a super high-rise building in the CBD area of Beijing was taken as an example, where the intrusion endpoint was assumed as the top floor and three different intrusion scenarios were designed. Results reveal that the SARSA model has the best training performance in the initial state (without any optimization measures). After optimizing the security system, increasing security system investment at interfloor nodes within the building is the most effective. In this context, a nonlinear fit between security investment and risk values shows that as investment in a security prevent system increases, intrusion risk remarkably decreases.

Keywords: unauthorized intrusion; scenario simulation; super high-rise building; reinforcement learning; Bayesian network; security system; SARSA model; nonlinear regression

超高层建筑因出入口众多且人流量巨大, 成为各类安全风险高度聚集的区域。其中, 人员非法入侵作为一类高频发生的风险事件, 始终是超高层建筑安防较为脆弱的环节, 现有的安防技术

收稿日期: 2024−08−03.　**网络出版日期**: 2025−01−15.
基金项目: 中国人民公安大学拔尖创新人才培养研究生科研创新重点项目 (2024yjsky009); 国家自然科学基金项目 (72174203); 中国人民公安大学安全防范工程双一流专项 (2023SYL08).
通信作者: 胡啸峰. E-mail: huxiaofeng@ppsuc.edu.cn.

水平亟需进一步提升。与其他类型建筑相比，超高层建筑内部结构更为复杂，相应的安防系统通常由多个子系统和大量监控、管理终端组成[1]。各子系统及设备设施之间的联动能力不足、智能化水平欠缺，极易造成系统漏洞，从而为非法入侵提供多种可能路径。随着楼层数量的增加，潜在入侵路径的数量通常呈指数增长。因此，针对超高层建筑的安防系统，识别上述隐患漏洞，计算出相对于非法入侵者的“最优路径”，对于安防系统的优化以及超高层建筑非法入侵风险的整体防控具有显著的现实意义。

目前，国内外研究多集中于视频监控子系统中的人员入侵行为识别技术，以及入侵探测子系统的优化方法[2-4]，主要聚焦于安防子系统的技术细节。针对超高层建筑安防系统这一整体，开展非法入侵风险识别、尤其是关注具体情景下潜在入侵者在建筑内部入侵路径选择的研究极少。从研究方法来看，以推演非法入侵者的最优入侵路径作为目标时，传统的贝叶斯网络方法[5]在应对不确定性和动态交互程度较高的超高层建筑非法入侵场景时，缺乏足够的自适应能力。而现有的多智能体仿真技术[6-7]尽管能够模拟具有一定的不确定性的场景，但其自主学习能力和与真实环境的互动能力，仍然难以满足实际应用需求。

基于此，本文围绕 3 种不同入侵情景，提出一种基于强化学习的超高层建筑非法入侵推演方法，用于精准模拟复杂环境下的路径选择，特别适用于应对超高层建筑的安防需求。以位于北京 CBD 地区的一栋超高层建筑作为研究对象，获取研究所需的建筑结构、安防系统点位等数据。在此基础上，将建筑内的公共过道抽象为拓扑结构中的节点，使用贝叶斯网络计算入侵者成功通过每个节点的概率，并通过 4 种强化学习模型 (Q-learning、SARSA、DQN (deep Q-learning) 和 DDQN (double deep Q network)) 计算入侵者的最优入侵路径。旨在为超高层建筑非法入侵的高效防范提供精准依据，切实增强超高层建筑的整体安全性。

本文的主要贡献有：1）将超高层建筑内部的公共过道抽象为拓扑结构中的节点，结合贝叶斯网络和强化学习模型，提出了一种识别潜在入侵者最优入侵路径的方法；2）通过评估不同情景下的风险补偿效果，本文提供了一套有效的安防系统优化方案，为提升超高层建筑的整体安全性提供了实用的策略支持。

1　相关工作

本章从超高层建筑人员非法入侵领域、情景推演领域以及强化学习算法 3 个方面介绍相关工作。

1.1　超高层建筑人员非法入侵领域的研究现状

在建筑物的非法入侵研究中，以往的工作主要集中在识别人员入侵行为领域。例如，Huang 等[2]提出了一种基于计算机视觉的方法，用于评估建筑工地的人员入侵情况，并证明该方法能有效提高安全性。同样，Li 等[3]提出了一种结合基于空间位置的技术与行为安全原则的方法，并通过香港建筑工地的实际应用，验证了其在降低入侵风险和提高安全性方面的效果。此外，Arslan 等[4]提出了一种使用低功耗蓝牙（BLE）信标和建筑信息建模（BIM）技术来追踪人员的移动，从而阻止建筑工地上的入侵行为。这些研究为防止建筑物入侵事件提供了全面的见解。然而，关于超高层建筑的室内入侵研究却很少被关注，目前尚未有关于超高层建筑室内入侵最优路径识别的相关研究。

1.2　情景推演领域研究现状

在应急情景推演方面，许多研究采用贝叶斯网络方法来预测和评估突发安全事件[8]。由于仅依靠专家判断获得的结果往往过于主观，这些研究通常依赖于现实世界的数据来支撑其分析。然而，大多数情景推演研究往往难以获取真实数据[9]。这可能归因于以下几个因素：数据收集成本过高，难以复制真实效果（例如地震等自然灾害），情景本身具有高度的不确定性。这些因素使得单一的贝叶斯网络方法不足以解决超高层建筑中复杂的非法入侵情景推演问题，潜在入侵者在选择入侵路径时的不确定性给相关研究带来了巨大的挑战。

在难以获取真实数据的情景中，以往的研究通常采用多智能体仿真技术对未来可能发生的情景进行预测。然而，这种方法存在一定的局限性，包括缺乏足够的自主学习能力以及无法与真实环境交互的问题。而本研究中提出的入侵情景推演问题需要自主学习能力以确定入侵者最可能的入侵路径。此外，还需要基于确定的最优路径来优化安防系统，并与现实情景相互验证。因此，多智能体仿真技术不足以胜任这一特定任务。

1.3　强化学习算法

强化学习作为一种自主学习算法[10-12]，通过对真实环境的响应动态地采用策略。它常应用于机器人研究、策略游戏以及自主系统等领域。通过基于真实情景建立规则，它能够有效地解决多智能体仿真的局限性。此外，强化学习通常分为两大类：基于值的强化学习和基于策略的强化学

习。基于值的方法[13]专注于评估智能体在每个状态下相对于目标的价值贡献，从而为实现目标提供明确的方向。具体而言，该方法旨在识别回报最大的路径，这与本研究中提到的搜索“最优入侵路径”高度契合。随后，管理者沿着确定的最优路径优化安防系统，通过多种强化学习算法的训练来观察最优路径和惩罚值的变化。通过这一过程，可以得出潜在的最优入侵路径。实际上，Hu 等[14]结合贝叶斯网络和数字孪生技术，从安防系统脆弱性分析的角度为安防系统提供针对入侵事件的优化建议。相比之下，本文从入侵者入侵路径的角度为安防系统提供优化建议，显著提高了超高层建筑整体安全性，很好地弥补了以往研究中对入侵路径分析的不足。

2　本文方法

基于强化学习的超高层建筑非法入侵情景推演研究的技术路线如图 1 所示，这一过程包含 3 个模块，分别为环境构建、训练与优化以及情景推演。环境构建部分包括构建状态空间、制定行动规则和奖励函数，用于指导入侵者（智能体）的决策。训练与优化部分利用从环境构建部分获得的 Q 值形成经验回放池，然后比较 4 种强化学习算法以获得最优的入侵路径。总而言之，环境部分提供了情景推演的模拟框架，而训练与优化部分为情景推演阶段提供了策略指导。情景推演部分包括 3 个情景，评估指标包括收敛所需步数、奖励总得分、安防系统成本和风险值。

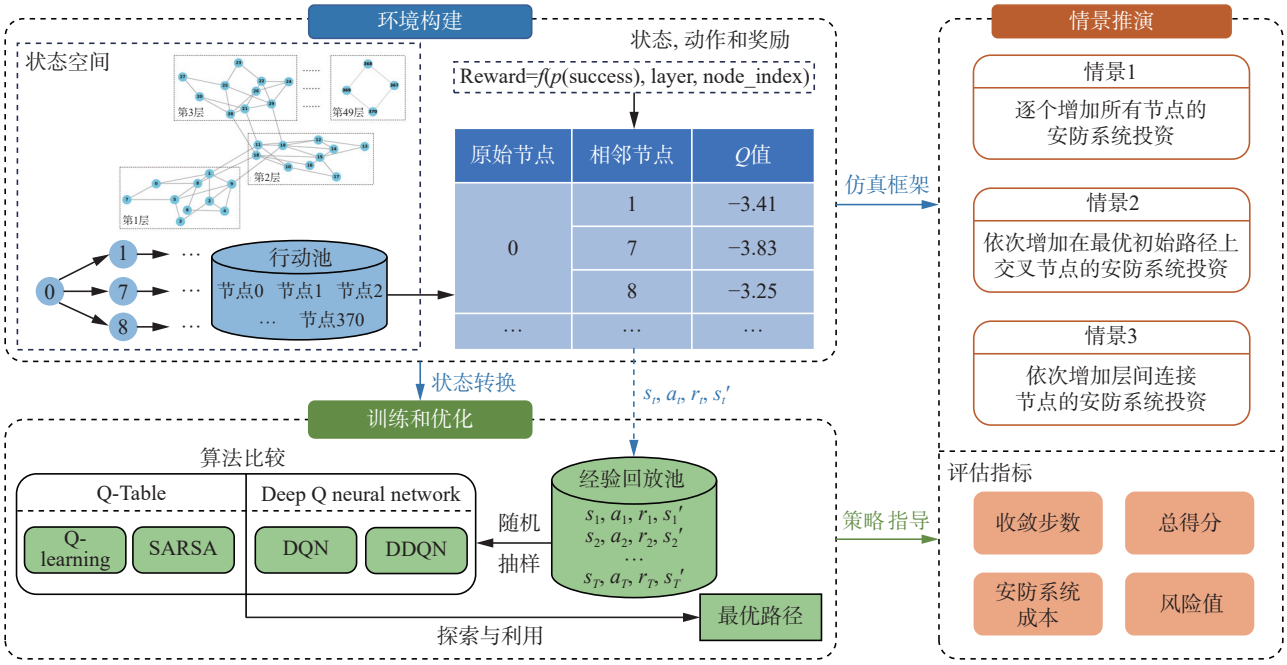

图 1　基于强化学习的超高层建筑入侵情景推演技术路线

Fig. 1　Process for simulating personnel intrusion scenarios in super high-rise building based on reinforcement learning.

2.1　环境构建

如图 1 所示，环境中的状态空间由超高层建筑内部真实结构抽象成的拓扑结构构成，拓扑结构中的每个节点对应建筑内的每一条公共过道。入侵者的移动仅限于相邻节点（相邻过道）。在本文算法中，奖励惩罚机制在引导其选择最优路径时起到关键作用。这种奖惩机制通过函数 $R_e(i)$ 来表示，该函数能够动态评估基于当前状态和所选行动的即时后果，具体定义为

$$R_e(i)=\begin{cases}1\,000, & a=d\\ -0.1(\max(l(a))-l(a)+1)p_a-10s_{c,i}, & \text{其他}\end{cases}$$

式中：i 代表步数，决定了所选择的行动；a 是当前节点所能选择的行动，指向决策空间中的相邻节点；d 表示目标节点；$l(a)$ 表示当前所在节点的楼层数；$\max(l(a))$ 表示该建筑的最大楼层数；p_a 是选择该行动的概率；$s_{c,i}$ 是对入侵者重复之前走过的节点所施加的惩罚。如果入侵者走过该节点，则 $s_{c,i}$ 为 1；如果未走过该节点，则 $s_{c,i}$ 为 0。当入侵者到达目标节点后，能够获得 1 000 分的奖励。

该奖励函数采用贝叶斯网络对各节点的入侵风险进行动态评估，其中入侵成功率 p_a 的计算基于 5 位专家对各安防系统防御效果的评分，从而使评估结果更加科学严谨。通过动态惩罚机制，该函数有效避免入侵者反复沿用相同路径，并设定不同楼层的风险权重，使模型更符合实际场景。到达目标节点的高额奖励则反映了入侵成功

的严重后果，确保模型能够在复杂环境中优化出更合理的路径选择方案。

2.2　训练与优化

如图 1 所示，为计算超高层建筑潜在非法入侵者的“最优入侵路径”，本文选择了 4 种基于值的强化学习模型：Q-learning[15-16]、SARSA[17]、DQN[18] 和 DDQN[19]。这些模型的作用是指导智能体（入侵者）从初始节点开始到目标节点的路径决策。Q-learning 模型的公式表示为

$$Q(N_s,N_a)\leftarrow Q(N_s,N_a)+\alpha[P(N_s,N_a)+\gamma\min_{N_{a'}}Q(N_{s'},N_{a'})-Q(N_s,N_a)]$$

式中：N_s 表示当前所在节点，N_a 代表从节点 N_s 采取行动以移动到下一个节点，$N_{s'}$ 是采取行动 N_a 后到达的后续节点，$N_{a'}$ 是根据下一个状态 $N_{s'}$ 进行当前选择的动作。$P(N_s, N_a)$ 量化了从节点 N_s 过渡到节点 $N_{s'}$ 的即时惩罚。$Q(N_s, N_a)$ 估计在节点 N_s 采取动作 N_a 的预期总惩罚值。α 是学习率，用于控制新获得的信息如何迅速取代旧信息。最后，γ 是折扣因子，用来平衡即时惩罚与未来惩罚的重要性。

类似于 Q-learning，SARSA 模型公式表示为

$$Q(N_s,N_a)\leftarrow Q(N_s,N_a)+\alpha[P(N_s,N_a)+\gamma Q(N_{s'},N_{a'})-Q(N_s,N_a)]$$

DQN 模型通过最小化损失函数来训练网络：

$$L(\theta)=E[(P(N_s,N_a)+\gamma\min_{N_{a'}}Q(N_{s'},N_{a'};\theta^{-})-Q(N_s,N_a;\theta))^2]$$

式中：$Q(N_{s'},N_{a'};\theta)$ 表示 DQN 模型中深度神经网络 Q 函数的近似值，该函数预测在节点 N_s 采取行动 N_a 的预期总惩罚值；其中 θ 代表网络参数；θ^{-} 表示目标网络参数。这些参数从当前网络参数周期性更新，以稳定训练过程。此外，折扣因子 γ 用于调整未来的惩罚值。

DDQN 模型的损失函数表示为

$$L(\theta)=E[(P(N_s,N_a)+\gamma Q(N_{s'},\arg\max_{N_{a'}}Q(N_{s'},N_{a'};\theta);\theta^{-})-Q(N_s,N_a;\theta))^2]$$

式中：$\arg\max_{N_{a'}}Q(N_{s'},N_{a'};\theta)$ 表示入侵者的行动选择，$N_{a'}$ 用于优化预期的惩罚值。

在探索拓扑结构类型的最优路径时，采用贪婪算法[20] 具有较高效率。这种方法基于当前状态和可用信息做出决策，通过在每一步选择惩罚值最低的行动，高效地到达目标节点。在节点惩罚值固定的拓扑结构中，进行路径规划的贪婪策略运算机制为

$$a'=\begin{cases}\text{随机选择 } a\in A(s), & \xi<\varepsilon,\\ \arg\max\limits_{a\in A(s)}Q(s,a), & \text{其他}\end{cases}$$

式中：a'是选择的行动，s 对应一个节点，$A(s)$ 表示入侵者所能采取的行动空间集合。在文本中，入侵者的每一次行动仅能够选择移动至与其相邻的节点。ξ是从 $U(0,1)$ 中取得的均匀随机变量，ε 是探索率，在本文中取值为 0.9，有助于在探索新节点和利用已知路径之间的拓扑结构中实现平衡。argmax 函数识别具有最高预估得分的相邻节点的行动，从而根据学习到的 Q 值优化入侵路径。

2.3　情景推演

情景推演方案包含 3 个不同的情景，如图 2 所示。

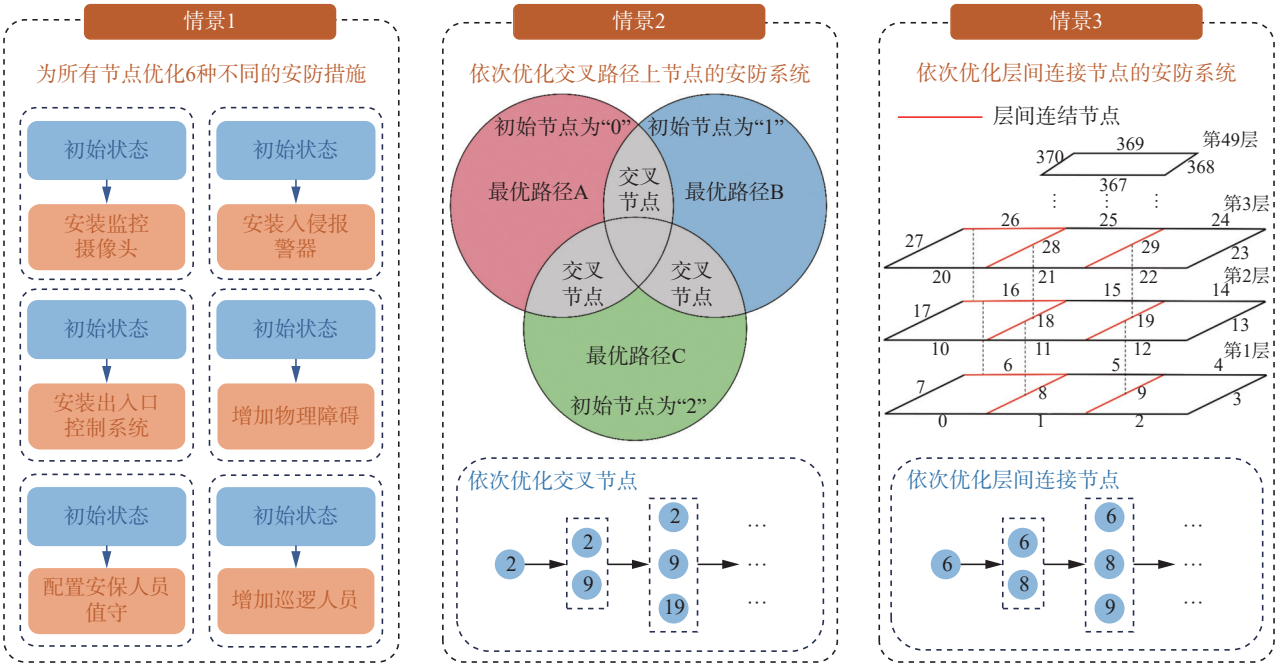

图 2　情景推演方案

Fig. 2　Specific details of three scenario deduction

第 1 个情景在初始状态下为超高层建筑内的全部节点分别进行不同的安防措施优化，以评估不同安防措施对非法入侵的影响大小。具体包括 6 种安防措施：安装监控摄像头、安装出入口控制系统、安装入侵报警器、增加巡逻人员、配置安保人员值守和增加物理障碍（带磁锁的门）。

在初始状态下，由于 3 个起始节点开始的最优入侵路径相互交叉。情景 2 通过在这些交叉节点优化安防系统（为每个节点配备一套全面的六大安防措施）。该情景旨在评估这种策略是否可以显著增强建筑对入侵的防御作用。

第 3 个情景探讨在初始状态下，依次逐层优化层间连接节点的安防系统。这类节点是入侵者在楼层之间移动的关键，极可能成为超高层建筑中的潜在入侵点。

情景 2 和情景 3 对比证明哪些关键节点能够有效降低总体入侵风险。此外，这两个情景对比了二者之间的风险补偿效应（即随着安防系统投资的增加，建筑物风险的变化）[21]。为更好比较两种情景策略的风险补偿效应，本文建立建筑物风险估值 R_i：

$$R_i = 0.5 \times \left(1 - \frac{L - L_{\min}}{L_{\max} - L_{\min}}\right) + 0.5 \times \left(\frac{R_e - R_{e_{\min}}}{R_{e_{\max}} - R_{e_{\min}}}\right)$$

式中：L 表示最优路径的长度（即该路径包含的节点数量），R_e 表示入侵者通过最优路径进入时获得的总奖励值。

3　情景推演实例分析

3.1　研究区域

本文针对位于北京市 CBD 的一个超高层建筑进行了入侵情景推演研究。该建筑高 230.9 m，共 49 层。其中，顶层停机坪以及计算机机房、电力设备室、监控室等点位被认定为建筑内的重点保护区域。每层的公共区域有多条走廊，数量从 4 到 10 条不等。这些公共走廊被抽象为拓扑结构网络节点[22]。建筑内部抽象结构与其相应的拓扑结构如图 3 所示。其中，左侧面板是建筑内公共过道的抽象布局，实线代表公共区域的走廊，虚线表示楼层之间的连接（包括电梯、观察楼梯和楼梯）。右侧面板为对应左侧结构的室内多层级拓扑结构图，每一个节点对应为真实场景中的一个过道。

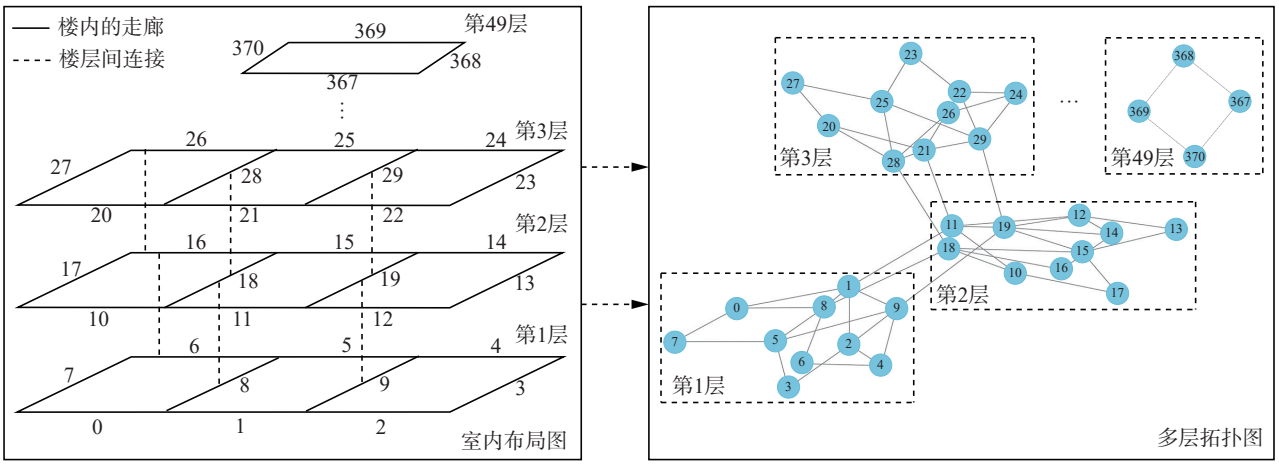

图 3　建筑内部抽象结构与对应的拓扑结构

Fig. 3　Actual interior structure of the building and the corresponding topological structure diagram

在实际场景下，非法入侵的入口选择往往多样且复杂。本文仅考虑 3 个起始入口、1 个入侵终点的简单情景，寻找入侵者入侵该建筑的最优路径，验证所提出方法的可行性。由于该超高层建筑的一层有 3 个访客入口，因此将首层的“0”、“1”和“2”节点定义为初始节点。由于电力设备室和监控室位于地下一层，并由专人把守，因此，将顶层节点“370”定义为主要目标节点。同时，还选择了位于 38 层的计算机机房节点“319”作为另一个目标节点作为对比实验。

3.2　数据集与数据处理

为确定入侵者成功入侵每个节点的概率，并将其作为奖励惩罚函数的重要组成部分，本文根据中国国家标准 GB 55029—2022[1] 对超高层建筑的安防系统进行了针对性调研。如图 4 所示，本文依据调研结果建立了一个 3 层贝叶斯网络，重点关注 3 个方面 (人力防护措施、实体防护系统以及电子防护系统)，共 11 个节点。其中包括 7 个根节点、3 个中间节点和 1 个代表入侵者成功入侵概率的叶节点。表 1 提供了每个节点详细信息。

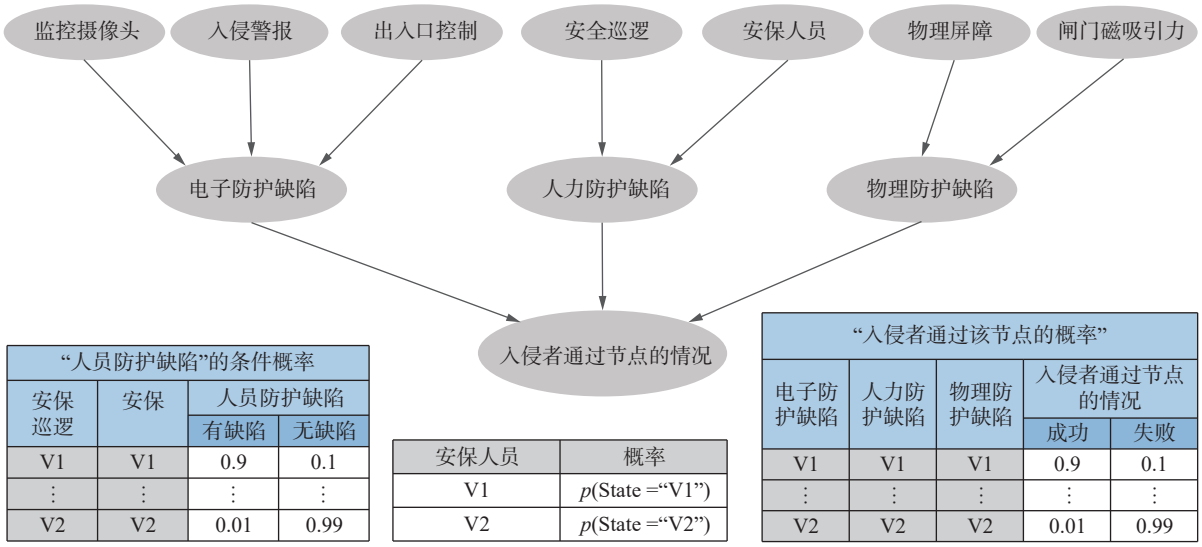

图 4　贝叶斯网络结构

Fig. 4　Bayesian network's structure

表 1　贝叶斯网络节点详细信息

Table 1　Nodes description of the Bayesian network

节点名称	数据类型	节点详细描述	成本/元
有无视频监控系统摄像机	Bool	是否有一个视频监控系统摄像头全覆盖照射该过道	500
有无入侵探测系统探测器	Bool	该过道是否有一个能够检测人员非法入侵的探测器	300
有无出入口控制系统闸机	Bool	是否有一个有效限制人员进出的出入口控制系统	800
有无安保人员巡逻	Bool	是否有安保人员在该过道巡逻	200
有无安保人员值守	Bool	是否有安保人员在该过道值守	200
有无实体防护门	Bool	是否在关键的过道有实体防护门	500
门磁吸力是否大于250kg	Bool	过道中实体防护门的磁力吸引强度是否大于250kg	500
电子防护系统是否存在缺陷	Float	电子防护系统是否有漏洞, 可能被入侵者利用	—
实体防护系统是否存在缺陷	Float	实体防护系统是否有漏洞, 可能被入侵者利用	—
人力防范措施是否存在缺陷	Float	人力防范措施是否有漏洞, 可能被入侵者利用	—
入侵者的入侵情况	Float	入侵者成功通过此节点的可能性	—

在构建出贝叶斯网络的节点和结构后, 为每个节点分配两种状态, 分别为 V1 和 V2。在父节点中, V1 表示该节点拥有该安防措施, 而 V2 表示该安防措施缺失。对于中间节点, V1 表示存在缺陷, 而 V2 表示无缺陷。对于叶节点, V1 代表入侵者可以成功入侵, 而 V2 表示无法通过该节点。

考虑到本文构建的贝叶斯网络中节点的显著不确定性, 本文采用德普斯特–沙弗(D-S)证据理论[23]来收集专家经验。该方法用于确定所有节点在两种状态下的条件概率分布表(CPTs)。作为传统概率理论的扩展, 这一理论便于表示和综合来自不同来源的证据。

为了增强获得的 CPTs 的可信度, 本文邀请 5 位专家(3 位技术专家和 2 位业务专家)在互不交流的情况下进行了 5 轮匿名概率赋值[24]。最终, 专家间的意见通过了一致性检验[25], 得到了叶节点入侵者成功通过每个节点的概率 p_a, 每一个中间节点存在缺陷的概率 p_b(缺陷指数)。

相关模型代码存储在 Github(https://github.com/lhys-lhyn/SUS-RL)以供参考。

3.3　初始状态下的最优入侵路径

为评估不同强化学习模型的有效性, 对比了 4 种模型针对两个不同终点的训练结果, 如图 5 与图 6 所示, 以确定建筑物的最优入侵路径。每次训练从 3 个不同初始节点“0”、“1”和“2”开始, 并在进行了 1 000 次迭代后达到收敛。训练的对比结果如表 2 所示, 当初始节点为“0”时, 对于两个不同终点, SARSA 模型均表现最佳, 入侵路径分别为经过 54 和 44 个节点, 奖励值分别为 936 和 952。当初始节点为“1”时, SARSA 模型同样呈现出最优的训练结果。最优路径分别包含 56 个节点和 50 个节点, 奖励值分别为 933 和 949。最后, 当初始节点为“2”时, 针对到达顶层的情况, SARSA 和 DQN 模型训练结果一致, 经过了 54 个节点入侵者到达目标节点, 总奖励值为 936。而针对到达计算机机房的情况, SARSA 模型的效果最佳, 入

侵者经过 46 个节点到达终点，获得奖励值为 548。总的来说，SARSA 模型在 4 种模型中拥有显著优势。

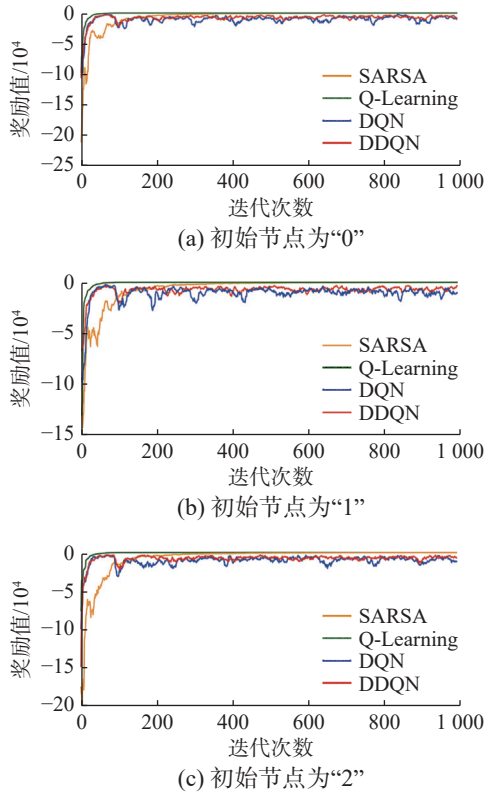

图 5　在不同初始节点出发并到达顶层终点的训练过程

Fig. 5　Training process starting from different initial nodes and reaching the top-level endpoint

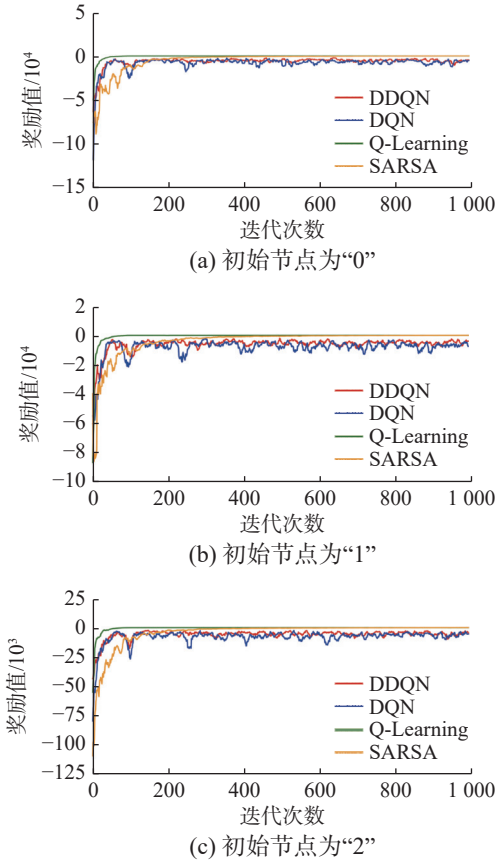

图 6　在不同初始节点出发到达计算机机房的训练过程

Fig. 6　Training process starting from different initial nodes and reaching the computer room

表 2　4 个模型的对比实验结果

Table 2　Training results for comparative experiments among four models

初始节点	模型	到达顶层所通过的节点数	到达顶层所获得的奖励值	到达计算机机房所通过的节点数	到达计算机机房所获得的奖励值
0	Q-learning	65	916	70	923
	SARSA	**54**	**936**	**44**	**952**
	DQN	57	932	47	950
	DDQN	54	925	62	807
1	Q-learning	66	921	63	935
	SARSA	**56**	**933**	**50**	**949**
	DQN	57	917	55	825
	DDQN	83	674	53	945
2	Q-learning	85	889	77	919
	SARSA	**54**	**936**	**46**	**948**
	DQN	**54**	**936**	51	901
	DDQN	53	928	60	850

注：粗体表示该模型在4个模型中具有最佳的训练性能。

在获得从 3 个不同初始节点出发的最优入侵路径后，本文重点针对到达顶层“370”号节点的 3 条不同初始节点的最优入侵路径的交叉节点进行进一步分析与情景推演。如图 7 所示为所有

交叉节点的安防系统缺陷指数 p_b, 蓝色表示电子防护系统的缺陷概率, 橙色代表实体防护系统的缺陷概率, 而绿色表示人力防护措施的缺陷概率。在所有交叉节点中, 节点“369”表现出最高的总缺陷指数 2.94, 表明急需增强安防系统防护能力。

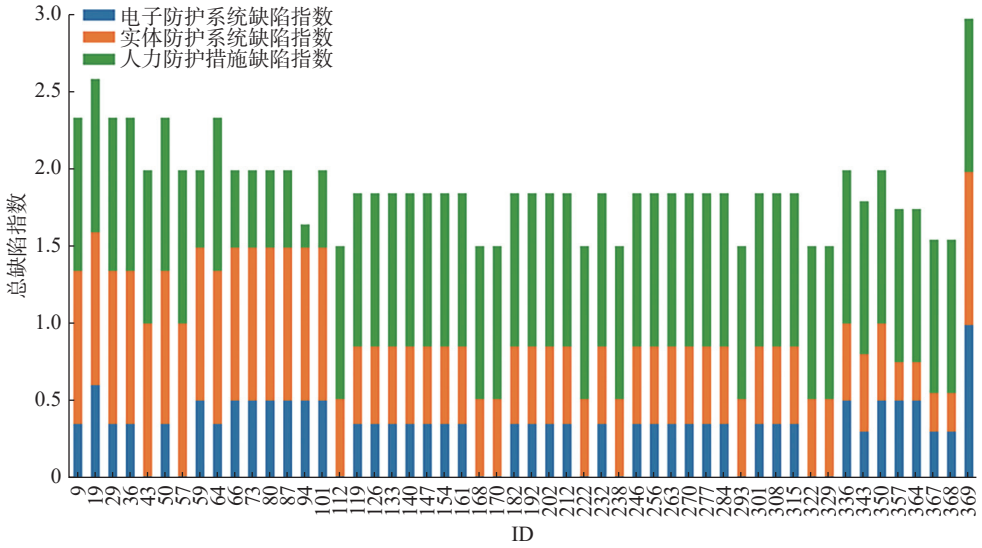

图 7　初始状态下 3 条最优路径上交叉节点的安防系统缺陷指数

Fig. 7　Security system defect index of intersecting nodes in optimal paths from three different initial points

3.4　情景 1 推演结果

情景 1 探讨了 6 种不同安防措施对入侵者入侵路径的影响。如图 8 与图 9 所示, 通过在所有节点上分别优化 6 种不同的安防措施, 入侵者所能获得的最大奖励值均低于 930, 这比初始状态下 3 个不同初始节点所对应 3 条最优路径可获得的最大奖励值都要低。在所有节点上安装入侵探测器相对来说是 6 种措施中最有效的。它确保了无论入侵者从哪一个初始节点出发, 其能获得的最大奖励值均保持在 910 以下。此外, 入侵者到达目的地所需的步数也比初始状态下任一路径所需的步数都要多, 表明了 6 种安防措施均产生了显著的效果。

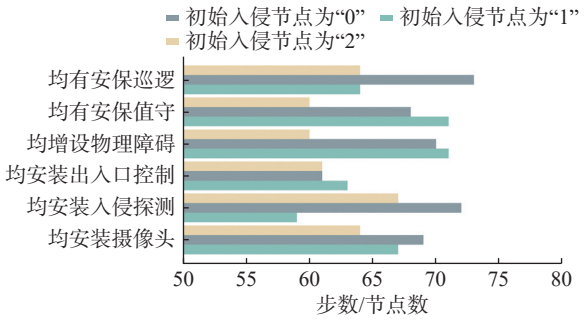

图 8　优化 6 种不同的安防措施的最优路径步数对比

Fig. 8　Optimal path's total rewards for implementing six different security measures

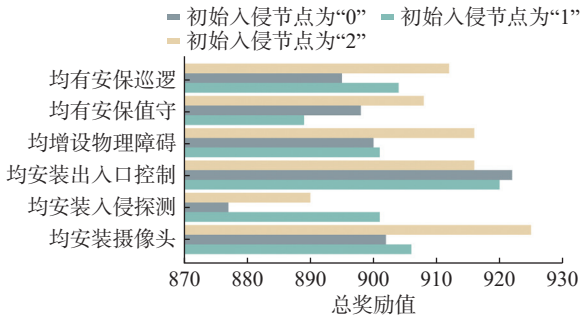

图 9　优化 6 种不同安防措施的最优路径总奖励值对比

Fig. 9　Optimal path's total steps for implementing six different security measures

3.5　情景 2 推演结果

如 2.3 节所述, 情景 2 探究了在初始状态下针对 3 个不同最优路径的交叉节点增加投入优化安防系统的效果, 对每个节点的优化操作为完整配置 6 种安防措施。图 10 给出了风险值随安防系统成本变化的二次拟合曲线, 二次拟合曲线指标及其对应的显式公式如表 3所示。初始节点为“0”所对应的散点图拟合的二次曲线 R^2 值为 0.340, p 值为 0.539, 表明其相关性的显著性并不明显。此外, 初始节点为“1”和“2”所对应的散点图拟合的二次曲线 R^2 值分别为 0.248 和 0.206, p 值分别为 0.111 和 0.641, 表明其相关性较弱且不显著。总而言之, 随着对交叉节点安防系统投资的增加, 建筑的整体风险出现显著波动, 难以保证此类投资方案能够稳定提高建筑的安全性。

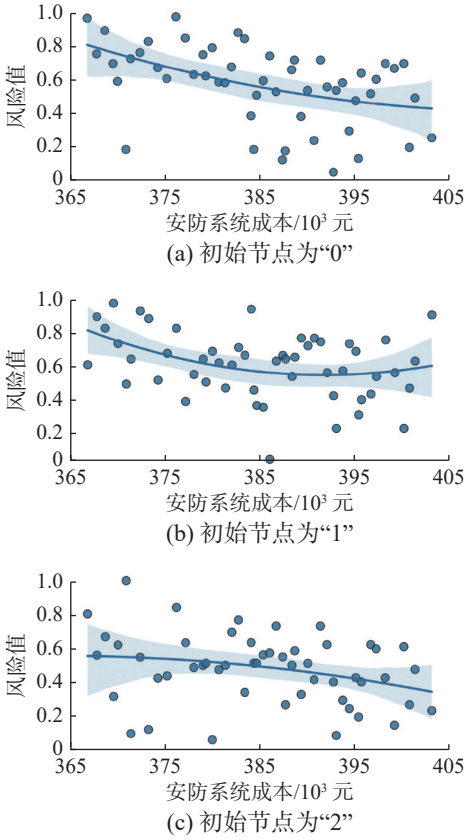

图 10　情景 2 中风险值随安防系统成本变化的二次拟合曲线

Fig. 10　Quadratic regression analysis plot of risk variation with cost in Scenario Two

表 3　情景 2 的二次拟合曲线指标及显式公式

Table 3　Quadratic fitting curve indicators and explicit formula for Scenario Two

初始节点	R^2	p值	显式公式
0	0.340	0.539	$f(x)=1.790\times10^{-10}x^2-1.480\times10^{-4}x+31.020$
1	0.248	0.111	$f(x)=4.273\times10^{-10}x^2+3.347\times10^{-4}x+66.072$
2	0.206	0.641	$f(x)=-1.287\times10^{-10}x^2+9.328\times10^{-5}x-16.350$

3.6　情景 3 推演结果

情景 3 探究了在层间连接节点增加投入优化安防系统配置的效果，对每个节点的优化操作与情景 2 相同。图 11 同样给出了风险值随安防系统成本变化的二次拟合曲线。表 4 中的拟合结果显示，初始节点为“0”与初始节点“1”所对应的散点图拟合的二次曲线 R^2分别为 0.643 与 0.647，p 值分别为 0.022 与 0.016，表明安防系统投入与建筑风险值之间具有强相关性和统计显著性。而对于初始节点“2”，R^2为 0.684，p 值为 0.067，表明有强关系，但并不显著。总体而言，对于 3 个初始节点，在层间节点上增加安防系统投入会使入侵风险降低约 50%。相对来说，情景 3 的推演结果明显优于情景 2 中的推演结果，二次曲线与数据之间更具统计意义。尽管如此，随着安防系统投入的增加，建筑的风险仍然存在波动，这表明在增加投资时仍需谨慎。

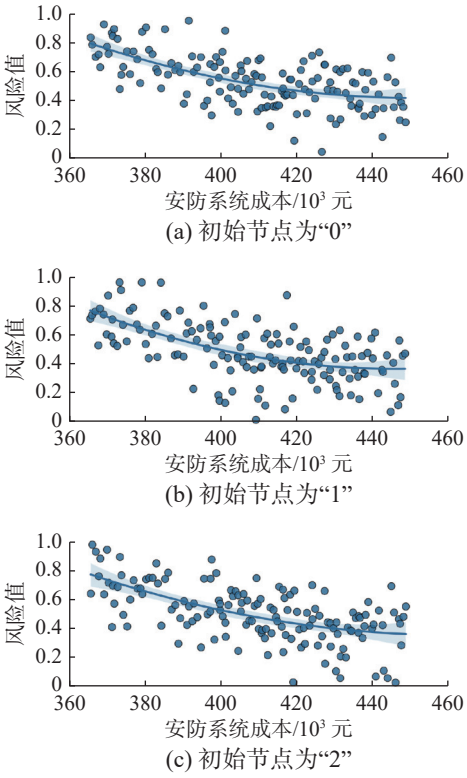

图 11　情景 3 中风险值随安防系统成本变化的二次拟合曲线

Fig. 11　Quadratic regression analysis plot of risk variation with cost in Scenario Three

表 4　情景 3 的二次拟合曲线指标及显式公式

Table 4　Quadratic fitting curve indicators and explicit formula for Scenario Three

初始节点	R^2	p值	显式公式
0	0.643	0.022	$f(x)=5.202\times10^{-11}x^2-4.693\times10^{-5}x+10.999$
1	0.647	0.016	$f(x)=6.236\times10^{-11}x^2-5.563\times10^{-5}x+12.764$
2	0.684	0.067	$f(x)=4.560\times10^{-11}x^2-4.209\times10^{-5}x+10.066$

4　结束语

本文比较了 4 种基于值的强化学习模型，发现 SARSA 模型在识别入侵者最优路径上效果最佳。情景 1 推演结果表明，入侵探测器对防止入侵者入侵的效果最为显著。情景 2 与情景 3 的对比结果表明，在楼层连接节点增加安防系统投入比在初始交叉节点更为有效，突显增加层间节点安全性在防入侵方面的关键作用。情景 3 推演结

果显示, 对层间节点增加安防系统投入可将入侵风险减少约 50%, 安防系统成本与建筑风险的拟合曲线 R^2值大于 0.6。

本文的主要创新点在于将建筑内的公共过道抽象为拓扑结构中的节点, 作为强化学习的状态空间。并基于贝叶斯网络计算入侵者入侵每个走廊的入侵成功率, 作为奖励函数的关键组成部分。此外, 3 种情景推演结果为建筑管理者提供了增强建筑安全性的重要见解。然而, 研究的局限性在于情景推演的多样性不足, 仅探讨了 3 种不同的情景推演方案。推演结果表明, 随着安防系统投资的增加, 建筑风险表现出波动, 因此管理者在增加投入时需谨慎。未来的研究应探索更多样化的投资方案, 为管理者提供更多参考, 从而更全面高效地增加建筑安防系统投入。

参考文献:

[1] 中华人民共和国住房和城乡建设部. 安全防范工程通用规范: GB 55029—2022[S]. 北京: 中国计划出版社, 2022.
Ministry of Housing and Urban-Rural Development of the People's Republic of China. General code of security engineering: GB 55029—2022[S]. Beijing: China Planning Press, 2022.

[2] HUANG He, HU Hao, XU Feng, et al. Skeleton-based automatic assessment and prediction of intrusion risk in construction hazardous areas[J]. Safety science, 2023, 164: 106150.

[3] LI Heng, DONG Shuang, SKITMORE M, et al. Intrusion warning and assessment method for site safety enhancement[J]. Safety science, 2016, 84: 97−107.

[4] ARSLAN M, CRUZ C, GINHAC D. Visualizing intrusions in dynamic building environments for worker safety[J]. Safety science, 2019, 120: 428−446.

[5] 王润芳, 陈增强, 刘忠信. 融合朴素贝叶斯方法的复杂网络链路预测[J]. 智能系统学报, 2019, 14(1): 99−107.
WANG Runfang, CHEN Zengqiang, LIU Zhongxin. Link prediction in complex networks with syncretic naive Bayes methods[J]. CAAI transactions on intelligent systems, 2019, 14(1): 99−107.

[6] GUO Kai, ZHANG Limao, WU Maozhi. Simulation-based multi-objective optimization towards proactive evacuation planning at metro stations[J]. Engineering applications of artificial intelligence, 2023, 120: 105858.

[7] 李冰, 杨薪玉, 王延锋. 轨道交通车站乘客集散系统 Anylogic 仿真优化[J]. 智能系统学报, 2020, 15(6): 1049−1057.
LI Bing, YANG Xinyu, WANG Yanfeng. Simulation and optimization of the passenger distribution system Anylogic in rail transit stations[J]. CAAI transactions on intelligent systems, 2020, 15(6): 1049−1057.

[8] HOSSEINI S, IVANOV D. Bayesian networks for supply chain risk, resilience and ripple effect analysis: a literature review[J]. Expert systems with applications, 2020, 161: 113649.

[9] ZHU Rongchen, HU Xiaofeng, BAI Yiping, et al. Risk analysis of terrorist attacks on LNG storage tanks at ports[J]. Safety science, 2021, 137: 105192.

[10] BIAN Tao, JIANG Zhongping. Reinforcement learning and adaptive optimal control for continuous-time nonlinear systems: a value iteration approach[J]. IEEE transactions on neural networks and learning systems, 2022, 33(7): 2781−2790.

[11] 于泽, 宁念文, 郑燕柳, 等. 深度强化学习驱动的智能交通信号控制策略综述[J]. 计算机科学, 2023, 50(4): 159−171.
YU Ze, NING Nianwen, ZHENG Yanliu, et al. Review of intelligent traffic signal control strategies driven by deep reinforcement learning[J]. Computer science, 2023, 50(4): 159−171.

[12] WU Yan, LUO Shixian, DENG Feiqi. Reinforcement learning for optimal control of linear impulsive systems with periodic impulses[J]. Neurocomputing, 2024, 585: 127569.

[13] 高玉钊, 聂一鸣. 基于值函数分解的多智能体深度强化学习方法研究综述[J]. 计算机科学, 2024, 51(S1): 34−42.
GAO Yuzhao, NIE Yiming. Review of multi-agent deep reinforcement learning method based on value function decomposition[J]. Computer science, 2024, 51(S1): 34−42.

[14] HU Jinming, HU Xiaofeng, KONG Feng, et al. Vulnerability analysis of super high-rise building security system based on Bayesian network and digital twin technology [J]. Process safety and environmental protection, 2024, 187: 1047−1061.

[15] ASGHARNIA A, SCHWARTZ H, ATIA M. Multi-objective fuzzy Q-learning to solve continuous state-action problems[J]. Neurocomputing, 2023, 516: 115−132.

[16] KIUMARSI B, ALQAUDI B, MODARES H, et al. Optimal control using adaptive resonance theory and Q-learning[J]. Neurocomputing, 2019, 361: 119−125.

[17] GARÍ Y, PACINI E, ROBINO L, et al. Online RL-based cloud autoscaling for scientific workflows: Evaluation of Q-Learning and SARSA[J]. Future generation computer systems, 2024, 157: 573−586.

[18] YANG Xu, LIU Pei, LIU Fang, et al. A DOD-SOH bal-

ancing control method for dynamic reconfigurable battery systems based on DQN algorithm[J]. Frontiers in energy research, 2023, 11: 1333147.

[19] WU Peiliang, ZHANG Yan, LI Yao, et al. A robot pick and place skill learning method based on maximum entropy and DDQN algorithm[J]. Journal of physics: conference series, 2022, 2203(1): 012063.

[20] SENTHIL KUMAR S, ALZABEN N, SRIDEVI A, et al. Improving quality of service (QoS) in wireless multimedia sensor networks using epsilon greedy strategy[J]. Measurement science review, 2024, 24(3): 113−117.

[21] MILI K, BENGANA I, OUASSAF S, et al. Testing the co-integration relationship between auto insurance premiums and risk compensation amount[J]. Computers in human behavior reports, 2024, 13: 100377.

[22] HOU Miaomiao, HU Xiaofeng, CAI Jitao, et al. An integrated graph model for spatial-temporal urban crime prediction based on attention mechanism[J]. ISPRS international journal of geo-information, 2022, 11(5): 294.

[23] WANG Lina, XU Mengjie, ZHANG Ying. An intelligent decision algorithm for a greenhouse system based on a rough set and D-S evidence theory[J]. IAENG international journal of applied mathematics, 2024, 54(6): 1240−1250.

[24] 秦荣水, 石晨晨, 陈超, 等. 基于模糊贝叶斯网络的城市商业综合体火灾风险分析[J]. 中国安全科学学报, 2023, 33(12): 176−182.
QIN Rongshui, SHI Chenchen, CHEN Chao, et al. Risk analysis on fire accident of urban commercial complex based on fuzzy Bayesian network[J]. China safety science journal, 2023, 33(12): 176−182.

[25] COPPA E, IZZILLO A. Testing concolic execution through consistency checks[J]. Journal of systems and software, 2024, 211: 112001.

作者简介：

胡今鸣，硕士研究生，主要研究方向为强化学习、社会公共安全风险评估。发表学术论文 4 篇。E-mail：hujinming2024@163.com。

胡啸峰，副教授，博士，主要研究方向为人工智能、社会公共安全风险评估。主持国家自然科学基金项目 2 项，发表学术论文 60 余篇。E-mail：huxiaofeng@ppsuc.edu.cn。

石磊，助理研究员，博士，中国人工智能学会智能服务专委会委员，主要研究方向为智能信息处理、大数据分析与挖掘、社交网络搜索及人工智能。发表学术论文 20 余篇。E-mail：leiky_shi@cuc.edu.cn。