

DOI: 10.13382/j.jemi.B2205353

基于强化学习的应急无人机通信中继选择策略*

彭 艺^{1,2} 唐 剑¹ 杨青青^{1,2} 李 睿¹

(1. 昆明理工大学信息工程与自动化学院 昆明 650031; 2. 昆明理工大学云南省
计算机技术应用重点实验室 昆明 650500)

摘 要:在基站无法提供通信支撑区域,可引入无人机群作为中继节点来搭建无人机应急通信网络。针对无人机如何高效选择最优中继节点及保证系统吞吐量最大化问题,提出一种基于 SA-SARSA 强化学习算法的无人机中继选择策略。在中继节点通过解码重传 (DF) 之后进行转发,得到用户端最大比合并后的平均吞吐量的表达式,通过设定 SARSA 算法的状态、动作、奖励函数,选择回报值最大的中继节点。同时,引入退火算法来使源节点探索到更多的中继节点,从而使无人机群通信网络性能达到最优状态。仿真结果表明,与改进前的 SARSA 中继选择策略相比,所提的 SA-SARSA 中继选择策略占理想算法比例提升 10%。同时,在相同总功率条件下,所提策略所选的中继节点的吞吐量比 Q-learning 中继选择策略、SARSA 中继选择策略分别提升了 8%、13%。

关键词: 应急通信; 无人机; 中继选择; SA-SARSA; 吞吐量

中图分类号: TN925; V279.3 **文献标识码:** A **国家标准学科分类代码:** 510

Relay selection strategy for emergency UAV communication based on reinforcement learning

Peng Yi^{1,2} Tang Jian¹ Yang Qingqing^{1,2} Li Rui¹

(1. School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650031, China;
2. Yunnan Key Laboratory of Computer Technologies Application, Kunming University
of Science and Technology, Kunming 650500, China)

Abstract: A UAV group can be introduced as a relay node in the base station that cannot provide a communication support area to build a UAV emergency communication network. To solve the problem of how to efficiently select the optimal relay node and maximize the system throughput, we propose a relay selection strategy for UAVs based on the SA-SARSA reinforcement learning algorithm. After all the relay nodes are forwarded after decoding retransmission (DF), the expression of the maximum average throughput of the user is obtained. The relay node with the maximum return value is selected by setting the SARSA algorithm's state, action, and reward function. At the same time, the annealing algorithm is introduced to make the source node explore more relay nodes so that the performance of the UAV swarm communication network can reach the optimal state. The simulation results show that compared with the previous SARSA relay selection strategy, the proposed SA-SARSA relay selection strategy increases the proportion of the ideal algorithm by 10%. At the same time, under the same total power condition, the throughput of relay nodes selected by the proposed strategy is 8% and 13% higher than that of the Q-learning relay selection strategy and SARSA relay selection strategy, respectively.

Keywords: UAV; emergency communication; relay selection; SA-SARSA; simulated annealing arithmetic

0 引言

随着地球环境的变化,近年来自然灾害频发,应急救援能力面临着巨大的挑战。当某地区出现地震等恶劣情况下,基站通信设备遭到破坏,在其通信覆盖范围的灾民无法使用终端发送求救信号。针对这种情况,可以通过引入中继通信设备来解决问题。由于无人机(unmanned aerial vehicle, UAV)的机动性较好^[1],应用在应急通信方面具有很大的优势,因此考虑将无人机作为中继通信设备来进行应急通信。中继选择策略通过“探索”一定范围的中继节点来选出最优的中继节点,从而使应急通信网络的吞吐量达到最大。

近年来,许多学者研究了最优中继节点的选择策略。文献[2]对应急通信中无法进行通信的节点间构建有效链路,考虑多无人机的数量和构建中继链路的时间,提出了一种多项式时间中继节点布置算法 PTRPA。文献[3]研究了无人机在多种情况下的中继通信系统的最优中继布设算法。文献[4]在 DF 协议下提出了联合天线选择和中继选择的算法。文献[5]对不同服务质量需求的地面用户在三维空间进行部署中继节点,来实现用户的全覆盖。文献[6]提出了一种同时考虑无人机编队成员的邻近域及编队成员间的社会性依存关系来优化选择中继成员的中继策略。文献[7]提出一种不确定偏好匹配博弈的分布式中继无人机选择机制。文献[8]建立了一种无人机多跳中继信道模型。文献[9]将基于最大最小中继选择、谐波均值中继选择、基于阈值的最大最小和基于阈值的谐波均值的机会中继技术与随机中继方案进行了比较,得出机会中继在相同误码率情况下优于随机中继。文献[10]提出了一种改进的 Ballman-Ford 算法来寻找中继节点的最优选择。文献[11]针对多用户与单用户的不同情况,分析了无人机辅助移动中继通信系统在海上通信中的性能差异。文献[12]在低轨卫星通信系统中提出一种遍历容量的中继选择算法,动态调整协作卫星的优先级,提高卫星的服务时间。

上述文献考虑使用传统算法来选择最优的中继节点,但是由于传统算法的弊端,可能没有“探索”区域内的全部节点,选出的中继节点可能是局部最优的节点。同时,传统算法的迭代时间更长,算法复杂度更大。因此,如何快速准确的选择中继节点是当前值得研究的问题。

强化学习^[13]是一种新兴的学习策略,主要用来解决智能体(Agent)与环境交互过程中通过不断的学习策略,来达到回报值最大化的问题。近年来,强化学习在 D2D 中继选择中被学者广泛应用。文献[14]使用深度优先搜索算法和人工蜂群算法实现了节点间的可行链路的快

速搜索以及最少中继节点部署方案。文献[15]提出一种基于强化学习中的 Q-learning 算法的多中继选择机制。文献[16]提出了一种基于信噪比和 Q 学习的中继选择机制。文献[17]提出一种分布式 Q 学习的联合中继选择和访问控制方案。总体而言,强化学习在中继选择中具有较好的效果。因此,本文将强化学习应用到应急无人机的中继选择策略之中。

SARSA^[18](state-action-reward-state-action)算法是一种强化学习算法,与 Q-learning 类似,都是采用 Q-table 来保存动作值函数,同时在决策部分采用的是 ϵ -greedy 策略。与 Q-learning 的不同在于 SARSA 采用的更新方式是 on-policy, SARSA 算法的过程不需要先验知识。

综上所述,本文将 SARSA 算法引入到无人机中继通信网络,同时结合模拟退火算法(simulated annealing algorithm, SA)来扩大“搜索”范围,使中继网络寻找到更多可利用的中继节点,从而使应急通信网络“利用”这些最优节点来获得更大的吞吐量。在无人机中继通信网络中将 UAV 源节点视为学习者,通过感知中继通信网络状态,与通信环境进行交互,采取动作来使 Q 值达到最大。在该状态下采取的动作就是无人机中继通信网络中的最优中继选择策略,该策略有效的选择了中继节点,增大了应急通信网络的吞吐量。

1 系统模型

本文考虑的是山区无人机应急通信场景。本文的系统模型是建立在蜂窝网络基础上,由 2 个基站(base station, BS)、K 个用户终端(user equipment, UE)和多个无人机(unmanned aerial vehicle, UAV)组成,如图 1 所示。假设基站 BS1 受灾情影响,导致其覆盖范围内通信中断,用户无法进行正常通信或通信信号较弱,需要无人机对该区域内用户与外部基站 BS2 之间进行中继通信,基站 BS2 与用户群存在通信链路,但是信号较差。多个中继节点构成一个中继集 $\{r_1, r_2, \dots, r_l\} \in L$ 。

本文基于 DF 协议进行分析,在中继协作系统中信号的传输分为两个阶段,在第 1 阶段中,基站 BS2 以广播的形式将信号发送到用户端和所有中继节点,因此在用户端和中继节点 r_i 接收信号分别表示为:

$$y_{sd} = \sqrt{p_s} a_{sd} x_s + n_{sd} \quad (1)$$

$$y_{si} = \sqrt{p_s} a_{si} x_s + n_{si}, \quad i = 1, 2, \dots, K \quad (2)$$

其中, p_s 为基站的节点发射功率; a_{sd}, a_{si} 分别为 S-D 和 S- r_i 信道的衰减系数; x_s 为基站的发射信号; n_{sd}, n_{si} 均表示为功率谱为 N_0 的加性高斯白噪声,由于本文考虑场景为山区,用户数量较小且分布分散,因此本文假设不存在用户之间的干扰。

在第 2 阶段中,有 i 个无人机作为中继节点。如果采

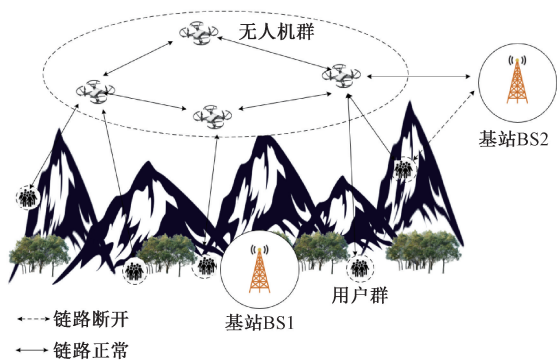


图1 无人机中继通信网络模型

Fig. 1 Relay Communication network model of UAV

用 DF 协作方式, 中继节点将解码源节点发射的信号再编码发送给目的节点。

目的节点收到中继 r_i 的信号为:

$$y_{id} = \sqrt{p_i} a_{id} \bar{x} + n_{id}, \gamma_{si} \geq \gamma_{th} \quad (3)$$

其中, p_i 是中继节点的发射功率, a_{id} 表示中继节点 $r_i - D$ 信道链路的信道参数, $\bar{x}$ 是中继成功解码后的信号, γ_{si} 是中继接收信号后的信噪比 (SNR) 且 $\gamma_{si} = \frac{p_s |h_{si}|^2}{\delta_{si}^2}$ 。 n_{id} 是用户端独立同分布的加性高斯白噪声其功率为 δ_{id}^2 。

DF^[19] 协作模式下, 单中继 r_i 被选中情况下, 在用户端采用最大比合并 (maximal ratio combining, MRC^[20]) 来合成信号, 因此用户端的平均吞吐量表示为:

$$C_{ODF}(r_i) = \begin{cases} \frac{1}{2} \log_2(1 + \gamma_{sd}), \gamma_{si} \leq \gamma_{th} \\ \frac{1}{2} \log_2(1 + \gamma_{sd} + \gamma_{id}), \gamma_{si} \geq \gamma_{th} \end{cases} \quad (4)$$

其中, $\gamma_{sd} = \frac{p_s |h_{sd}|^2}{\delta_{sd}^2}$, $\gamma_{id} = \frac{p_i |h_{id}|^2}{\delta_{id}^2}$ 分别为 $S - D$ 、 $r_i - D$ 链路的信噪比。

当多个中继节点被选中情况下, 用户端的信噪比为:

$$\gamma_{MDF} = \begin{cases} \gamma_{sd}, L = \emptyset \\ \gamma_{sd} + \sum_{i \in L} \gamma_{id}, L \neq \emptyset \end{cases} \quad (5)$$

平均吞吐量为:

$$C_{MDF}(L) = \begin{cases} \frac{1}{l} \log_2(1 + \gamma_{sd}), L = \emptyset \\ \frac{1}{l} \log_2(1 + \gamma_{sd} + \sum_{i \in L} \gamma_{id}), L \neq \emptyset \end{cases} \quad (6)$$

其中, l 为被选中的中继节点数量。

2 基于 SARSA 算法的无人机中继选择策略

2.1 SARSA 学习模型

对于本文提出的基于 SARSA 的无人机中继选择策略如图 2 所示。在 SARSA 算法中, 基本三要素为学习者、状态、动作, 本文将其应用到无人机中继协作网络中。该算法指示源节点选择最佳中继, 选择策略以广播形式在第 1 个时隙中传输到每个中继节点。通过源节点与目的地之间的反馈信道获得的奖励值, 用于更新 Q 矩阵并指导之后的策略选择。每个中继节点被认为是系统中 SARSA 的状态。

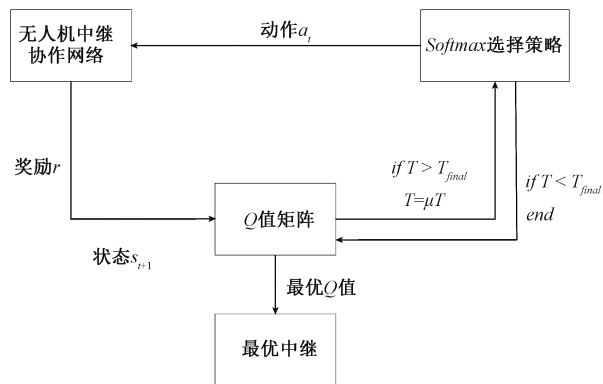


图2 基于 SA-SARSA 的无人机中继策略

Fig. 2 Relay strategy of UAV based on SA-SARSA

2.2 最佳中继选择

最佳中继节点的选择可以看作是一个有限的马尔科夫决策过程, Q 值更新函数、状态空间 S 和动作 A 设置为如下:

1) SARSA 算法的更新 Q 值函数的表达式为:

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)) \quad (7)$$

式中: $\alpha = 1/(1 + \text{Visit}(s, a))$, $\alpha \in (0, 1]$ 为学习速率, $\text{Visit}(s, a)$ 是 SARSA 算法访问状态-动作对 (s, a) 的总次数, γ ($0 < \gamma < 1$) 为折扣因子, r_t 为回报值。随着访问次数 $t \rightarrow \infty$, 状态-动作对就进行了无穷次的更新, 同时 $\alpha \rightarrow 0$, $\lim_{t \rightarrow \infty} Q_t(s, a) = Q^*(s, a)$, 再根据式 (8) 得到最优策略。

$$\pi^*(s) = \arg\max_a [Q^*(s, a)] \quad (8)$$

2) 状态空间 (S)。每一个中继节点代表一个状态, 则状态空间 $S = \{s_1, s_2, \dots, s_m\}$ 为所有中继节点的集合。

3) 动作空间 (A)。由 m 个动作 $\{a_1, a_2, \dots, a_m\}$ 组成, 当动作 a_i 被执行时, 系统将会进入下一个状态 s_i 。

4) 状态转移函数 $S \times A \rightarrow S$ 。当系统在状态 s_j 下, 动作被执行, 状态转移函数定义为:

$$f(s_j, a_k) = s_k \quad (9)$$

5) 奖励函数(r)。无人机应急通信系统的目的就是为了让节点的吞吐量。因此,将目标节点获得的 SNR 作为回报值,SNR 越大,其吞吐量也越大。在 DF 转发协议下,奖励函数设置为:

$$r_{ODF} = \begin{cases} \gamma_{sd}, \gamma_{si} < \gamma_{th} \\ \gamma_{sd} + \gamma_{id}, \gamma_{si} \geq \gamma_{th} \end{cases} \quad (10)$$

6) Q 值矩阵 $Q(s_i, a_i)$ 。 Q 矩阵由每个状态-动作对的 $Q(s, a)$ 值组成。在训练开始前, Q 矩阵的每一项都为 0。在 SARSA 算法中,无人机选择一个动作 $a_i \in A$, 然后根据状态转移函数式(12)转移到对应的状态 $s_{i+1} \in S$ 来更新 Q 表。 $m \times m$ 矩阵包含了所以状态和动作对所对应的 $Q(s, a)$ 值,表达式为:

$$Q = \begin{bmatrix} Q(1,1) & \cdots & Q(1,m) \\ \vdots & \ddots & \vdots \\ Q(m,1) & \cdots & Q(m,m) \end{bmatrix} \quad (11)$$

式中: $Q(s, a) \in \mathbb{R}^{m \times m}$ 。系统通过不断迭代更新 $Q(s, a)$ 值并存储在 Q 矩阵中。

为了防止无人机在选择 $Q(s, a)$ 中陷入过早的收敛,从而取到局部的最大化 T 的动作,引入 *Softmax* 选择策略来使无人机能够遍历所有动作。*Softmax* 选择策略的概率分配满足 *Boltzmann* 分布,当 $Q(s, a)$ 值较大,则动作 a 被赋予较高的概率;当 $Q(s, a)$ 值较小,则动作 a 被赋予较低的概率。表达式为:

$$P(a_i/s) = \frac{\exp(\hat{Q}(s, a_i^*)/T)}{\sum_j \exp(\hat{Q}(s, a_j)/T)} \quad (12)$$

其中, $P(a_i/s)$ 是指在状态 s 下选择下一动作 a_i 的概率。 $T > 0$ 为退火温度, T 的值大小影响选取 $Q(s, a)$ 的优先度。若 T 值较大,源节点可以充分地“探索”当前 Q 值不高的中继节点;若 T 值较小,源节点将会选择 Q 值较高的动作。因此,为了保证 SARSA 算法在训练过程中,既能“探索”更多的中继节点,又能最大化的“利用”和选择回报值最大的中继节点,前期给 T 一个较大的初始值,后期随着训练次数的增加将不断减少到最终值 T_{final} ,实现 SARSA 在时间上从“探索”平滑地过渡到“利用”。

$$T = \mu T \quad (13)$$

式中: μ 为退火因子。

2.3 基于 SA-SARSA 算法的中继选择步骤

本文提出的基于 SA-SARSA 算法中继选择策略以最大化吞吐量为准则,通过 SARSA 算法选择最优中继,同时采用 *Softmax* 选择策略使算法不会过早的结束“探索”,而是以一定的概率进行“探索”从而获得最大的回报值。基于 SARSA 算法的中继选择过程如下:

1) 初始化,将 Q 矩阵初始化为 0 矩阵,参数 η, μ, T, T_{final} 初始化,同时随机选择初始状态 $s \in S$;

2) 动作的选择和执行,根据式(7)计算当前 $Q(s, a)$ 值,同时利用 *Softmax* 选择策略从动作空间 A 中以概率 $P(a/s)$ 选择选择动作 a 并执行;

3) 奖励值获取,当中继节点被激活,源节点选择动作后,地面终端通过式(10)获取奖励值,然后通过反馈信道将其发送到源节点来更新 Q 矩阵;

4) 更新退火温度,每次训练结束后根据式(13)更新退火温度;

5) 重复步骤 2)~4), 直到训练结束。

结合 *Softmax* 选择策略, SARSA 中继选择流程如图 2 所示。基于 SARSA 算法的中继选择算法是经过多次的强化学习结合 *Softmax* 选择策略控制退火温度逐渐减少,最终选择 $Q(s, a)$ 值较大的动作,实现系统的最大吞吐量。SARSA 中继选择算法流程的伪代码如下:

算法 1: 基于 SA-SARSA 的最佳中继选择策略

输入: $Q(s, a) = 0, \text{Visit}(s_i, a_i) = 0, \forall s_i \in S, \forall a_i \in A, \alpha = 1,$
 $T = 10^{50}, T_{final} = 0.1, \mu = 0.9$
 输出: $\max[Q(s, a)]$

1. 随机选取 $s_i \in S$ 作为当前状态
2. Repeat
3. 根据式(12)从状态 s_i 选择 a_i
4. 更新退火温度 T :
5. if $T > T_{final}$
6. $T = \mu T$
7. else
8. $T = T_{final}$
9. 执行动作 a_i , 选取中继节点
10. 根据式(10)获取 DF 转发协议的奖励值 $r(s_i, a_i)$
11. 根据式(9)转移到下一状态 s_{i+1}
12. 根据式(7)更新 $Q(s_i, a_i)$
13. 根据 $\text{Visit}(s_i, a_i) = 1 + \text{Visit}(s_i, a_i)$ 和 $\alpha = 1/(1 + \text{Visit}(s, a))$ 更新学习速率 α
14. $s_i = s_{i+1}$
15. Unit(直到训练结束)
16. 根据式(8)得到最优策略 π^* (即最优中继节点)

3 仿真结果分析

仿真将文中所提的 SA-SARSA 算法的中继选择算法与基于 SARSA 算法、Q-learning 算法的中继选择和随机中继选择进行比较。将用户放置在基站 BS1 覆盖范围半径 2 km 的圆形区域内,基站 BS2 与基站 BS1 的距离 4 km,无人机群在基站 BS1 与 BS2 之间进行中继通信。由于本文主要研究应急通信环境下的最优中继节点的选择,因此在仿真中基站与用户的位置均为固定。基站与用户之间的信道服从瑞利分布 $h_{i,j} \sim CN(0, 1/d_{i,j}^\nu), d_{i,j}$

为两节点之间的距离, V 为信道的路径损耗。为了使 SARSA 中继选择算法在训练过程尽可能的探索和利用最优中继节点, 仿真主要参数设置如表 1 所示。

表 1 仿真参数设置

Table 1 Simulation parameter setting

仿真参数	仿真值
载波频率 f	2.4 GHz
飞行高度 h	300 m
飞行速度 v	15~25 m/s
发射功率 p_s	40 dBm
信道带宽 W	10 MHz
噪声功率 σ_{id}^2	(-103, -108) dBm
初始温度 T	10^{50} °C
退火因子 μ	0.9
最终温度 T_{final}	0.1 °C

图 3 所示为在门限 $\gamma_{th} = -3$ dBw, 中继节点数量 $i = 10$ 情况下, 不同迭代次数下基于 SA-SARSA 中继选择算法与随机中继选择算法、Q-learning 中继选择算法和 SARSA 算法占理想中继选择算法比例的性能比较。随机中继选择就是源节点以相同的概率选择区域内的中继节点进行数据传输。随着迭代次数的增加, SA-SARSA 中继选择算法通过不断与环境交互学习, 不断更新选择策略 π , 从而选择最佳中继节点。基于 SA-SARSA 的中继选择算法的吞吐量占理想吞吐量的比例将近 85%, 优于其他 3 种中继选择算法的性能。因此, 基于 SA-SARSA 算法的应急无人机应用在地震灾区环境中, 可以使得应急网络获得更大的吞吐量。吞吐量越大, 灾区用户持续发送求救信息的时间也就越长, 获救的概率也就越大。

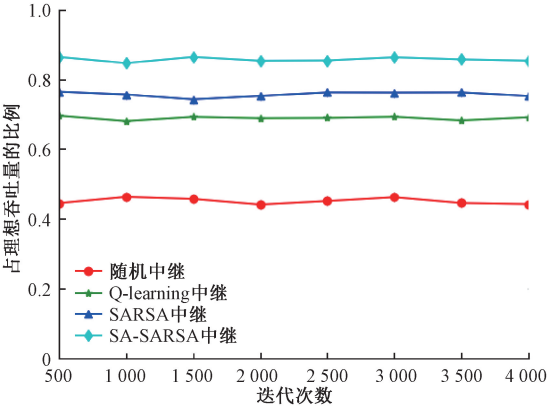


图 3 不同算法占理想吞吐量的比例

Fig. 3 The proportion of different algorithms in the ideal throughput

图 4 所示为基于 SA-SARSA 算法的应急网络吞吐量随着迭代次数变化的曲线。图 4 中 N 值代表应急网络中的中继无人机与用户的数目对。由于中继无人机数目的增多, 更多的中继节点可以被利用, 中继通信的成功率得

到增大, 应急网络的吞吐量也逐渐增大。随着迭代次数的增加, 系统吞吐量趋于平衡。

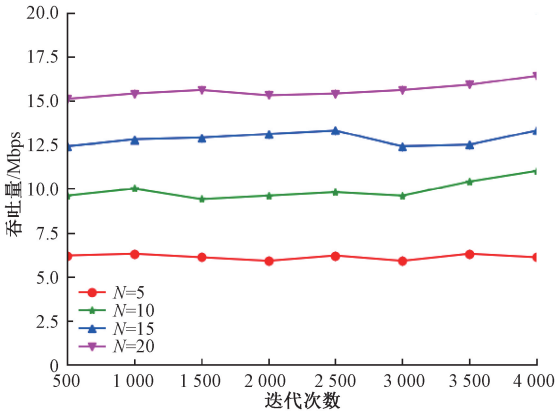


图 4 不同迭代次数下的吞吐量变化

Fig. 4 Throughput variation diagram under different iterations

图 5 所示为不同 UAV-UE 数目对下的基于不同算法的应急网络吞吐量的比较。随着中继无人机的数量增加, UAV-UE 数目对也得到增加, 用户获得中继通信的概率也得到增大, 从而增大了系统吞吐量。与随机中继随机选择信道相比较, SA-SARSA 算法从中继无人机与用户的角度综合考虑合理的选择信道, 所以基于 SA-SARSA 算法的中继选择策略能够获得更大的吞吐量。SARSA 算法和 Q-learning 虽然是通过不断的通过学习来更新选择策略, 但是它们容易受到环境的影响陷入局部收敛。改进后的 SA-SARSA 算法, 通过控制收敛速度, 在前期“探索”到了更多的中继节点, 从而在后期对最优中继节点有了更多的选择, 使得应急通信网络更好的利用其得到更大的系统吞吐量。

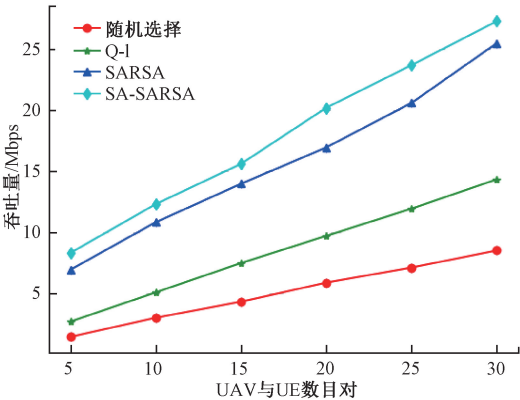


图 5 不同 UAV-UE 数目对下的吞吐量变化

Fig. 5 Throughput variation diagram for different UAV-UE numbers

图 6 仿真了在不同 UAV-UE 数目对下, SA-SARSA 算法与 Q-learning 算法、SARSA 算法随着总功率增加, 吞

吐量的变化曲线。由图 6 所示,随着总功率增加,SA-SARSA 算法的性能优于 Q-learning 算法和 SARSA 算法。总功率越大,中继节点能够更好的进行解码转发,同时越靠近中继节点的用户能够得到更好的信道增益。从图中可以看出中继节点数量越多,基于 SA-SARSA 的中继选择算法可以选择更好的中继节点,从而获得更大的吞吐量。根据式(7)可知,SA-SARSA 算法的 Q 值更新函数式与中继节点的数目无关,因此中继节点数目的增大,并不会增加过多的运算量。并且,在未知瞬时信道状态时,SA-SARSA 算法也能够自适应的选择出最佳的中继节点。因此,可以认为 SA-SARSA 具有更大的实用价值。

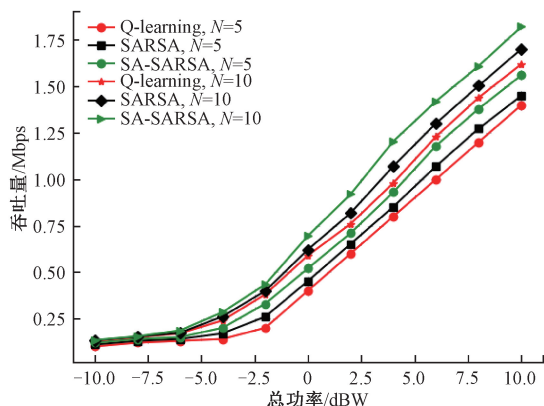


图 6 不同 UAV-UE 数目对下的吞吐量与总功率之比

Fig. 6 Throughput to total power ratio under different UAV-UE number pairs

4 结 论

本文在应急通信网络中提出基于 SA-SARSA 算法的无人机中继选择策略。首先,基站将信号传输到中继无人机,中继无人机采用 DF 协议时将信息进行解码转发系统。然后,通过设置 SARSA 算法的参数,将模拟退火算法的 *Softmax* 策略融合,随后进行中继节点的选取。最后,通过仿真验证了所提中继选择策略的优越性,SA-SARSA 算法中继选择策略优于随机中继、Q-learning 中继选择策略和 SARSA 中继选择策略。在应急救援上,SA-SARSA 中继选择策略能够为应急通信网络带来更大的增益。

参考文献

[1] 张宏宏,甘旭升,李双峰,等.复杂低空环境下考虑区域风险评估的无人机航路规划[J].仪器仪表学报,2021,42(1):257-266.
ZHANG H H, GAN X SH, LI SH F, et al. Route planning of UAV considering regional risk assessment in complex low-altitude environment [J]. Chinese Journal

of Scientific Instrument, 2021, 42 (1): 257-266.
[2] 吴高峰,高晓光,符小卫.一种基于多无人机的中继节点布置问题建模与优化方法[J].航空学报,2017,38(11):241-253.
WU G F, GAO X G, FU X W. Modeling and optimization of relay node arrangement problem based on multi-UAV [J]. Journal of Aviation, 2017, 38(11): 241-253.
[3] 韦薇.中继通信系统的最优中继位置布设算法[D].南京:南京邮电大学,2018.
WEI W. The optimal relay location layout algorithm of relay communication system [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2018.
[4] 袁静珍.基于 DF 中继策略的联合天线选择和中继选择优化算法[J].电子测量与仪器学报,2019,33(6):184-189.
YUAN J ZH. Joint antenna selection and relay selection optimization algorithm based on DF relay strategy [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(6): 184-189.
[5] 朱钟恺.基于无人机的中继部署问题研究[D].合肥:中国科学技术大学,2019.
ZHU ZH K. Research on relay deployment based on UAV [D]. Hefei: University of Science and Technology of China, 2019.
[6] 吴钟博,易建强.无人机编队支撑网络的协同通信中继策略[J].航空学报,2020,41(S2):187-194.
WU ZH B, YI J Q. Cooperative communication relay strategy for UAV formation support network [J]. Journal of Aviation, 2020, 41(S2): 187-194.
[7] LIU D, YANG Y, XU Y, et al. Uncertain preference matching-based relay selection and position adjustment in dynamic UAV systems[C]. 2020 International Conference on Wireless Communications and Signal Processing (WCSP). IEEE, 2020: 1170-1175.
[8] CHEN X, HU X, ZHU Q, et al. Channel modeling and performance analysis for UAV relay systems [J]. China Communications, 2018, 15(12): 89-97.
[9] JAIN N, DONGARIYA A, VERMA A. Comparative study of different types of relay selection scheme for cooperative wireless communication [C]. 2017 International Conference on Information, Communication, Instrumentation and Control (ICICIC). IEEE, 2017: 1-4.
[10] HO D T, GROTLI E I, SHIMAMOTO S, et al. Optimal relay path selection and cooperative communication protocol for a swarm of UAVs[C]. 2012 IEEE Globecom Workshops. IEEE, 2012: 1585-1590.

- [11] ZHANG J, LIANG F, LI B, et al. Placement optimization of caching UAV-assisted mobile relay maritime communication [J]. China Communications, 2020, 17(8): 209-219.
- [12] 穆文静,李勇朝,阮玉晗,等. 基于遍历容量的低轨卫星协作通信中继选择算法[J]. 信号处理, 2017, 33(10): 1317-1323.
MU W J, LI Y CH, RUAN Y H, et al. Ergodic capacity based relay selection algorithm in LEO satellite cooperative communication[J]. Signal Processing, 2017, 33(10): 1317-1323.
- [13] XIAO L, JIANG D, LIU S. Reinforcement Learning Based Communication Security for Unmanned Aerial Vehicles[M]. Cyber Security Meets Machine Learning. Springer, Singapore, 2021: 57-83.
- [14] 张小孟,杨森,宋晓,等. 一种中继无人机快速部署策略[J]. 北京航空航天大学学报, 2021, 47(8): 1705-1711.
ZHANG X M, YAN S, SONG X, et al. A fast deployment strategy for relay UAV [J]. Journal of Beijing University of Aeronautics and Astronautics, 2021, 47(8): 1705-1711.
- [15] 彭艺,张申,朱豪,等. 联合机器学习的 D2D 通信多中继选择机制[J]. 电子测量与仪器学报, 2020, 34(3): 149-154.
PENG Y, ZHANG SH, ZHU H, et al. Multiple relay selection mechanism for D2D communication based on joint machine learning [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34(3): 149-154.
- [16] PAEK M J, NA Y J, LEE W S, et al. A novel relay selection scheme based on Q-learning in multi-hop wireless networks [J]. Applied Sciences, 2020, 10(15): 5252.
- [17] ZHAO B, REN G, DONG X, et al. Distributed Q-learning based joint relay selection and access control scheme for IoT-works [J]. IEEE Communications Letters, 2021, 25(6): 1901-1905.
- [18] MOHAN P, SHARMA L, NARAYAN P. Optimal path finding using iterative SARSA[C]. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2021: 811-817.
- [19] JADOON M A, KIM S. Relay selection algorithm for wireless cooperative networks: A learning-based approach [J]. Iet Communications, 2017, 11(7): 1061-1066.
- [20] JUNG H, KIM K, KIM J, et al. A relay selection scheme using Q-learning algorithm in cooperative wireless communications [C]. Communications. IEEE, 2012.

作者简介



彭艺, 1996 年于电子科技大学获得学士学位, 2001 年于电子科技大学获得硕士学位, 2005 年于电子科技大学获得博士学位, 现为昆明理工大学信息工程与自动化学院副教授, 主要研究方向为无线通信。

E-mail: 527037928@qq.com

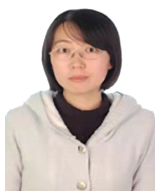
Peng Yi received her B. Sc. degree in 1996, M. Sc. degree in 2001 and Ph. D. degree in 2005 from University of Electronic Science and Technology of China. She is now an associate professor at the School of Information Engineering and Automation, Kunming University of Science and Technology. Her research interest includes wireless communication.



唐剑 (通信作者), 2021 年于湖南工业大学获得学士学位, 现为昆明理工大学信息工程与自动化学院在读硕士研究生, 主要研究方向为应急通信和无人机。

E-mail: jiantang@stu.kust.edu.cn

Tang Jian (Corresponding author) received his B. Sc. degree from Hunan University of Technology in 2021. He is now a M. Sc. candidate in the School of Information Engineering and Automation at Kunming University of Science and Technology. His main research interests include emergency communication and UAV.



杨青青, 2004 年、2007 年获哈尔滨工程大学学士、硕士学位, 2020 年获昆明理工大学博士学位。现为昆明理工大学信息工程与自动化学院讲师、硕士生导师。主要研究方向为信息处理和应急通信。

E-mail: 1016188826@qq.com

Yang Qingqing received her B. Sc. degree and M. Sc. degree from Harbin Engineering University in 2004 and 2007, and her Ph. D. degree from Kunming University of Science and Technology in 2020. She is now a lecturer and master's tutor at the School of Information Engineering and Automation, Kunming University of Science and Technology. Her main research interests include information processing and emergency communication.



李睿, 2021 年获得昆明理工大学学士学位。目前在昆明理工大学信息工程与自动化学院攻读硕士学位, 主要研究方向为无人机应急通信。

E-mail: 1814813449@qq.com

Li Rui received his B. Sc. degree from Kunming University of Science and Technology in 2021. He is now a M. Sc. candidate in School of Information Engineering and Automation, Kunming University of Science and Technology. His main research interest includes UAV emergency communication.