具有精英策略的深度强化学习无人机集群通信网络拓扑设计

董方昊[1]，冯有前[1]✉，尹忠海[1]，梁晓龙[2]，周　诚[1]，李明杰[1]
(1. 空军工程大学基础部，西安，710051；2. 空军工程大学空管领航学院，西安，710051)

摘要　针对集群无人机背景下定向天线网络拓扑设计的 NP-hard 特点，基于网络高抗毁、低功耗、高稳定性等要求，以抗毁性(3-连通)、链路量、链路功耗和稳定性为奖励，提出了一种具有精英策略的深度强化学习通信网络拓扑生成算法，验证了精英经验池加速训练效果。与传统 DQN 相比，引入精英经验池能够有效加速模型收敛，训练时间减少 3 倍以上。与遗传算法相比，算法分离了训练与使用过程，当网络训练完成后，能够根据场景需要实时计算通信网络拓扑。实验阶段设计了随机给定空间位置的 6 节点、10 节点、24 节点和 36 节点的 3-连通通信网络拓扑。实验结果表明：所提算法具有强的实时性和适用性，对于不大于 36 节点的网络，可在 183 ms 内实现网络拓扑的更新计算，达到了实际应用的实时性要求。

关键词　深度强化学习；精英经验池；通信网络连通度；通信网络拓扑
DOI　10.3969/j.issn.1009-3516.2019.04.008
中图分类号　TP393　　**文献标志码**　A　　**文章编号**　1009-3516(2019)04-0052-07

Topology Design of Network Based on Deep Reinforcement Learning with Strategy of Elite

DONG Fanghao[1], FENG Youqian[1]✉, YIN Zhonghai[1], LIANG Xiaolong[2], ZHOU Cheng[1], LI Mingjie[1]
(1. Department of Basic Science, Air Force Engineering University, Xi′an 710051, China;
2. Air Traffic Control and Navigation College, Air Force Engineering University, Xi′an 710051, China)

Abstract: Aiming at the NP-hard characteristics of directional antenna network topology design under cluster UAV background, an elite strategy for deep reinforcement learning communication network topology generation algorithm is introduced with the requirements of high survivability, low power consumption and high stability of the network, which has the rewarding of invulnerability (3-connectivity), link quantity, link power consumption and stability. Compared with traditional DQN, elite experience pool verifies the acceleration training effect by effectively accelerating the convergence of the model and reducing the training time by more than three times. Rather than genetic algorithm, this algorithm separates the processes of use and training . When the network training is completed, the communication network topology can be calculated in real time with the needs of scene. In experimental stage, a 3-connected communication network topology with randomly given spatial location is designed which includes 6 nodes, 10 nodes,

收稿日期：2019-03-22
基金项目：国家自然科学基金(61472443)
作者简介：董方昊(1996—)，男，山西运城人，硕士生，主要从事无人机通信研究。E-mail：dongfanghao1996@163.com
通信作者：冯有前(1960—)，男，陕西富平人，博士生导师，教授，主要从事计算机视觉、云计算研究。E-mail：417461659@qq.com

引用格式：董方昊，冯有前，尹忠海，等. 具有精英策略的深度强化学习无人机集群通信网络拓扑设计[J]. 空军工程大学学报(自然科学版)，2019，20(4)：52-58. DONG Fanghao, FENG Youqian, YIN Zhonghai, et al. Topology Design of Network Based on Deep Reinforcement Learning with Strategy of Elite[J]. Journal of Air Force Engineering University (Natural Science Edition), 2019, 20(4): 52-58.

24 nodes and 36 nodes . The experimental results has shown that this proposed algorithm has strong real-time and applicability, it can help network topology which has less than 36 nodes update in 183ms so that meeting the real-time requirements of practical application.

Key words: deep reinforcement learning; elite experience pool; connectivity; communication network topology

随着智能技术快速发展，集群无人机大幅提升了无人机的作战效能。自组织通信网络不依赖于卫星或地面站等基础设备，而是以无人机作为独立节点来感知态势、传递消息，为集群无人机作战提供可能[1]。基于定向天线的集群无人机通常形成特定的通信拓扑结构来提高抗毁性，目前，国内外对于编队控制已有了广泛研究。文献[2]证明无人机集群在模型控制下可以稳定的达到预设编队，文献[3～6]提出多种不同编队算法并加以验证。然而对于编队形成的集群无人机，文献[1～6]均未涉及编队模型应采取何种通信网络拓扑结构。对于单机损毁失效或能量耗尽等情况，设计 1-连通以上的通信拓扑网络十分必要，而过冗余的通信网络将造成巨大的能量损耗。集群无人机通信网络拓扑设计的主要目的是在保证连通性的情况下，尽可能降低节点通信功耗，提高通信质量，加强网络拓扑的稳定性。

本文针对平面结构，根据降低节点通信功耗和网络连通度的要求，利用深度强化学习建立了 3-连通定向天线通信网络拓扑设计模型。以构建无人机定向天线网络为背景，设计 3-连通通信拓扑为具体要求，采用带有精英策略的深度强化学习算法(ES-DQN)设计功耗低、抗毁性强的自组网拓扑结构。

1　基本原理

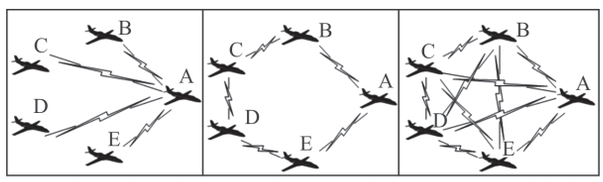
无人机集群基于合作策略与协调机制，通过传感器、通信、自主决策规划等方式达成合作、完成任务[7]。运用组网技术，集群无人机通过直通链路或多跳链路，无人机之间互联、信息共享、多路由信息传输，当单机的某条信息获取链路受到干扰时，可以选择其他未干扰的路由链路获取所需信息，从而加强无人机单机对抗各种干扰的能力。

1.1　网络的 k-连通和节点度

评价无人机集群通信网络性能的核心指标之一是抗毁性[8]，度量网络抗毁性的指标是网络的连通度，本文从网络的 k-连通性出发，讨论无人机集群通信网络拓扑设计。

定义 1　(k-连通)一个具有 N 个点的图 G 中，在去掉任意 $k-1$ 个顶点后($1\leqslant k\leqslant N$)，所得的子图仍然连通，去掉某 k 个顶点后不连通，则称 G 是 k 连通图[9]，k 称作图 G 的连通度，记作 $k(G)$。

网络连通度 k 决定了通信网络保持连通性的条件下，可允许失去节点的最大数目[10]。图 1(a)表示 1-连通通信网络中，网络抗毁性较差。图 1(b)所示 2-连通图损失任意 1 个节点，通信网络在多跳链路下扔保持连通。图 1(c)所示的 3-连通图可保证在损失任意 2 个节点情况下仍然保持整个通信网络具有连通性，具有更好的抗毁性。因此，k 值很好地衡量了网络的抗毁性和可靠性，k 越大，网络稳定性越强。



(a) 1-连通通信拓扑　(b) 2-连通通信拓扑　(c) 3-连通通信拓扑

图 1　不同连通度控制下的通信拓扑示意图

定义 2　(节点度)节点度是指和该节点相关联的边的条数，又称关联度，表示为 $\deg(u)$。

节点度衡量了网络中某个节点的重要程度。为保持图 G 连通度为 k，图中任意节点度均大于等于 k，如图 1(c)所示通信拓扑节点度均为 4。

1.2　集群通信网络的 3-连通特征分析

本文假设算法重构网络的时间间隔为 ΔT，无人机在 ΔT 时间内损毁失效的概率为 q。现假设集群内共有 N 架无人机，设每架无人机损毁记为事件为 A，则 N 次伯努利实验中事件 A 发生的次数 X 为随机变量，服从二项分布，即 $X\sim b(N,q)$[11]。在 ΔT 时间内，N 架无人机损毁超过 k 架的概率为

$$P\{X>k\}=1-\sum_{l=0}^{k}C_N^l q^l(1-q)^{N-l} \qquad (1)$$

假设集群有 20 架无人机，单架无人机在 ΔT 时间内损毁失效的概率为 0.000 1，(ΔT 可取拓扑重建时间，如本文所提算法的 183 ms)。表 1 表示连通度为 1～4 时网络失效概率。可以发现选定通信网络连通度为 3，即可保证 ΔT 网络失效概率小于 10^{-8}，低于一般的计算机网络误码率要求，足以保证战场环境下集群无人机网络的连通性要求，因此 3-连通通信组网在实际应用中具有更广泛的现实意义。

表 1　不同连通度下的网络失效概率

连通度	1	2	3	4
失效概率	1.998×10^{-3}	1.898×10^{-6}	1.139×10^{-9}	4.837×10^{-13}

2 相关工作

2.1 定向天线网络连通性

定向天线即通过天线技术,在某个方向或某几个方向上提高发射和接收电磁波能力,而在其他方向上发射及接收电磁波能力几乎为零的一种天线[12]。相比于全向天线,采用定向天线进行通讯可以提高辐射功率的利用率,有效提高通信保密性,增强信号抗干扰能力。

如图 2 所示,全向天线将信息以电磁波形式向空间全向辐射,而定向天线将辐射角控制在 θ 内,有效保持功率定向发射。定向天线的波束宽度越窄,天线增益越高,通信隐蔽性更好。

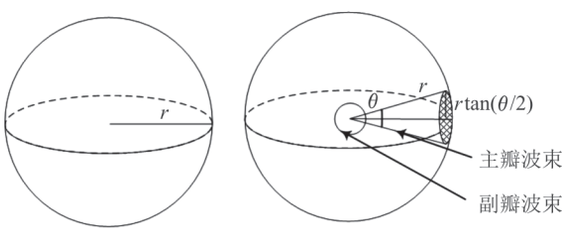


图 2 全向天线与定向天线信号辐射图

自 2008 年以来,众多学者对定向天线通信拓扑问题展开了研究,主要工作集中在连通拓扑存在的充要条件、产生算法、干扰性、频谱可用性等方面,理论性的工作较多,实时性算法成果较少。文献[13]提出采用定向天线建立构建通信网络,并证明给定 1~5 个方向的定向天线能够产生强连通的通信拓扑。文献[14]提出任意两节点之间通过多跳链路连通的概率 P,指出多向天线可以提高网络连通性性能:

$$P=\frac{\sum_{i=1}^{v}\frac{1}{2}n_i(n_i-1)}{\frac{1}{2}n(n-1)} \tag{2}$$

式中:n 表示网络中节点总数;v 表示连接的节点个数。

式(2)表明:网络中任意 2 个节点通过单跳或多跳互相连接的概率与参与连接的节点数目有关。当连接点数越多,网络中任意 2 点基于多跳连接的概率越大,仅从理论上探讨了网络节点的连通性。文献[15]推导出全向天线和定向天线的连通性闭合表达式,指出定向天线由于干扰更低,频谱可用性更高,可以形成比全向天线连接性更高的拓扑网络。文献[16]建立了有效的网络模型,采用遗传算法探索 k-连通条件下的最优网络拓扑,但是响应时间较长,难以满足战场环境下灵活多变的拓扑变换需求。

与编队控制不同,通信拓扑控制不关心个体的拓扑位置关系、鲁棒控制、滑膜变结构控制等技术,而是在现有编队的基础上,基于节点功耗、通信质量及网络稳定性等因素,通过建立适宜的通信链路,保持集群信息无障碍交互,为编队控制、任务规划铺设信息网络。

衡量通信网络优劣性的指标除抗毁性外,还包括网络的功耗和可靠性,采用合理的拓扑网络可以有效降低通信组网的总耗能。文献[17]指出:与全向天线相比,采用功率控制算法对于纯定向天线 MAC 协议进行控制时,系统耗能降低 30%到 80%。其中影响网络节点功耗的主要原因之一是通信链路总长度。在接收灵敏度一致的情况下,无线发射功率 P 与传输距离 R 的关系为:$P\propto R^2\sim R^5$,为降低通信功耗,应尽可能使通信链路总长度最短。同时,文献[18]提出功率控制方案表明:全体节点功耗水平一致,有利于保证通信质量稳定,因此拓扑中尽可能降低最长链路与最短链路之比,按照就近原则建立通信网络。

基于上述关于定向天线形成通信网络的论证工作,本文聚焦于建立更适用于战场环境的集群无人机通信拓扑网络。对于该类 NP-hard 问题[13],本文结合深度强化学习能够有效感知环境并提取特征的优势,提出采用深度强化学习来建立通信网络拓扑。

2.2 深度强化学习和经验回放机制

深度强化学习[19]是人工智能发展历史上的里程碑,实现了智能体从感知到动作的端对端学习。在深度 Q 网络(Deep Q Network,DQN)中,智能体将探索过程的当前状态、当前动作、当前奖励和下一状态以记忆单元的形式存入经验池中用于网络训练。为打破数据关联性,Nature DQN 采用随机取样法选取经验池部分数据对神经网络进行训练,大幅提高了算法稳定性。随机取样意味着经验池中数据地位平等,致使探索过程中缺稀的优质经验利用率不高,网络收敛速度较慢。因此 Schaul 提出 Prioritizedreplay 方法[20],对于经验池内容按照中重要性进行排序,有效提升了收敛速度。但是更新记忆单元优先级过程提高了算法复杂性,文献[21]提出基于 TD-error 重抽样优选机制的改进,通过自适应采样方法保证了低优先级的记忆单元仍可更新经验池并参与训练。

3 模型框架

3.1 模型设定

对于二维平面(三维空间)内集群无人机,各节点可由定向天线与任意其他节点建立至多一条双向

通信链路，相当于图中的一条边，因此给定编队控制下各个节点相对坐标，以图 G 表示集群无人机通信网络拓扑结构。其中顶点集 $V=\{v_1,v_2,\cdots,v_n\}$，每架无人机由一个顶点 v 表示。设邻接矩阵为：

$$\boldsymbol{E}=\begin{bmatrix}0 & e_{12} & \cdots & e_{1n}\\ 0 & 0 & \cdots & e_{2n}\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 0\end{bmatrix} \tag{3}$$

与通信组网的中通信链路相对应，$e_{i,j}=0$ 表示 v_i 与 v_j 之间不存在通信链路；$e_{i,j}=1$ 表示 v_i 与 v_j 之间存在双向通信链路。

这里定义动作 a 由一系列点动作 $a_1\sim a_6$ 组成，点动作 a_i 来自于点动作集 A_i。由于设计双向链路通信组网，为避免重复连接，设定 A_i 表示顶点 v_i 连接至多 3 个顶点 v_l、v_m、v_k（$i<l,m,k$）。例如 6 节点自组织网络中，动作集 A_3 如图 3 所示，表示 v_3 共有 7 个可选动作。

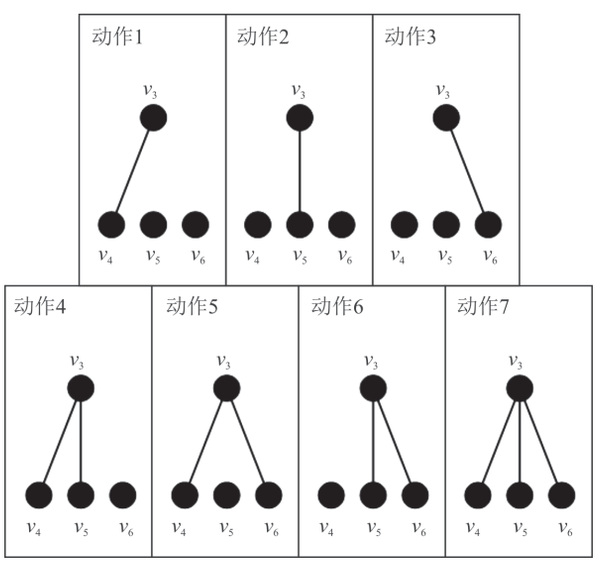


图 3　6 节点动作集

执行动作 a_i 之后，即由状态 s_0 转到状态 s，同时将本次迭代过程的当前状态 s_0、动作 a、下一状态 s、奖励 r 以记忆单元形式存入经验池中。记忆单元中，奖励 r 的设定应基于网络抗毁性、网络功耗、网络稳定性等多方面考虑，接下来介绍奖励设定。

3.2　奖励设定

DQN 算法中，奖励设定不仅关乎算法的收敛速度，也直接影响最终通信拓扑结构的优劣。利用深度强化学习设计网络拓扑的效果直接体现在奖励值的大小上，因此如何设置合理的奖励十分关键。由于节点所携带多向天线数目有限，因此限制网络中节点度最多为 A。基于网络综合分析，奖励应包含 4 个部分：抗毁性奖励 E_1，链路量奖励 E_2，低功耗奖励 E_3 和稳定性奖励 E_4。

为设计最优网络拓扑，建立最优化奖励模型见式(4)：

$$\begin{cases}\max\ c_1E_1+c_2E_2+c_3E_3+c_4E_4\\ \text{s.t.}\quad \lambda(G)=3\\ \max\limits_{u\in G}\deg(u)\leqslant 4\end{cases} \tag{4}$$

式中：c_1，c_2，c_3，c_4 分别表示各个奖励的权重，其大小根据具体情况设定。

本模型针对定向天线双向通信链路组网，因此设网络节点个数为 n，连通度为 $\lambda(G)$，组网拓扑邻接矩阵为式(4)，其距离矩阵为：

$$\boldsymbol{P}=\begin{bmatrix}0 & d_{12} & \cdots & d_{1n}\\ 0 & 0 & \cdots & d_{2n}\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 0\end{bmatrix} \tag{5}$$

模型旨在设计抗毁性能优异的无人机拓扑网络，以连通度 k 衡量网络抗毁性，经论证以 3-连通通信组网抗毁性最佳，因此网络连通度奖励为：

$$E_1=\begin{cases}1,\lambda(G)=3\\ 0,\lambda(G)\neq 3\end{cases} \tag{6}$$

对于一个 n 节点的 3-连通网络，每个节点连通度至少为 3，因此网络边数至少为 $3\times n/2$。网络边数越多意味着冗余链路更多，而较少的通信链路意味着每架无人机更低的通信功耗。由于 n 节点全连接网络边数为 $n^2/2$，因此链路边数奖励为：

$$E_2=\frac{n}{n-1}\left(-\frac{2}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}e_{i,j}+n-1\right) \tag{7}$$

在自由空间传播时，功耗与通信距离相关，传输距离越短节点功耗越小，对于链路个数相等的 3-连通网络，通信链路总长度越小意味着组网系统内总功耗越低。令全连通网络链路总长度为 L，链路长度总功耗奖励为：

$$E_3=\frac{L-\sum_{i=1}^{n}\sum_{j=1}^{n}e_{i,j}d_{i,j}}{L-L_{\boldsymbol{P}}},L=\sum_{i=1}^{n}\sum_{j=1}^{n},$$
$$L_{\boldsymbol{P}}=\min\nolimits_{*}^{\left\lceil\frac{3n}{2}\right\rceil}\boldsymbol{P} \tag{8}$$

式中：$L_{\boldsymbol{P}}$ 表示距离矩阵 $\boldsymbol{P}$ 中除 0 元素外距离最短的 $3\times n/2$ 条边的总长。

由于通信链路的长短不仅影响通信质量，也影响自组网络的稳定性。基于此，在上述情况均一致的前提下，通信链路应尽可能短，设 $\max'=\max D_{i,j}$，$\min'=\min D_{i,j}$，因此网络稳定性奖励为：

$$E_4=\frac{1}{\max'-\min'}\left(\max'\frac{\min(V_{i,j}\times D_{i,j})}{\max(V_{i,j}\times D_{i,j})}-\min'\right) \tag{9}$$

根据以上要求，联立式(4～7)，则模型最终奖励为：

$$E=c_1E_1+c_2E_2+c_3E_3+c_4E_4 \tag{10}$$

3.3 精英算法

模型中由于最终奖励对于前期动作的反馈相对稀疏,因此经验累积过程选择放弃单步动作奖励,选取最终奖励作为累积经验存入经验池。为使模型加速向预期方向收敛,简化算法流程,这里提出 ESDQN 模型。图 4 为模型流程示意图。

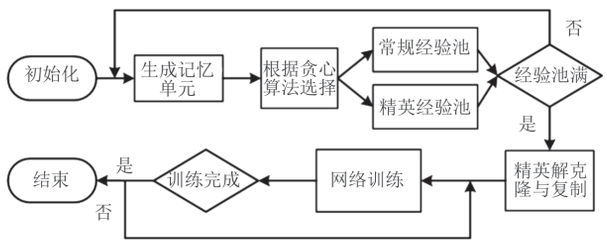


图 4 ESDQN 模型流程示意图

模型收敛速度缓慢关键在于,网络在训练时应选取大量优质的记忆单元作为训练集。但事实上,训练集中的优质记忆单元利用率较低,易致使模型偏移;训练集缺乏多样性又易致使网络陷入局部极值。因此本文结合贪心算法思想,构建 2 个不同的经验池:常规经验池和精英经验池。引入系数 ε,依概率选取记忆单元填充精英经验池或常规经验池。待累积到一定程度之后,对于精英经验池中奖励值最高的记忆单元进行复制,并根据贪心算法依概率选取精英经验池或常规经验池对预测网络进行训练。ESDQN 算法具体过程为:

1)初始化常规经验池 N 和精英经验池 E,网络 H,参数 θ,贪心概率 ε,最小贪心概率 $\min\varepsilon$ 以及最大迭代次数 lp;

2)初始状态 s,动作 a;

3)计算 $p(\varepsilon)=\max(\varepsilon^{\lg lp},\min\varepsilon)$;

4)以概率 $p(\varepsilon)$ 选取随机动作 a,计算下一状态 s_{i+1} 和奖励 r_i,并将记忆单元$<s_i,a_i,r_i,s_{i+1}>$存入精英经验池 E;

5)否则选取随机动作 a_i,计算下状态 s_{i+1} 和奖励 r_i,并已记忆单元形式存入常规经验池 N;

6)若迭代到最大次数并且精英经验池 E 积满,选取 E 中最大奖励值的记忆单元复制并完全覆盖 E;

7)分别在精英经验池 E 和常规经验池 N 随机采样 M 个记忆单元到 minibatch$E$、minibatch$N$;

8)以概率 $p(\varepsilon)$ 选择 minibatch$E$ 训练网络;

9)否则选择 minibatch$N$ 训练网络;

10)结束。

4 实验及分析

本实验使用环境:Windows 系统,建模环境为 matlab2017,CPU 为酷睿 i5。实验内容包括对于 6 节点、10 节点、12 节点、24 节点和 36 节点和的网络拓扑设计,ESDQN 与 NatureDQN 迭代次数对比以及 ESDQN 算法和遗传算法对于解算拓扑网络响应时间的对比。

4.1 网络拓扑设计分析

为验证深度强化学习的拓扑设计效果,以 6 节点、10 节点和 12 节点的集群无人机为例,设计该奖励设定下的最优拓扑,并分析 ESDQN 算法相比于传统深度强化学习算法的加速效果。

在某种编队控制下,6 节点定向天线无人机自组网顶点编号及相对坐标如图 5(a)所示。以邻接矩阵 $\boldsymbol{E}_0=[0]_{6\times6}$ 作为初始状态 s_0,学习率 $\alpha=0.9$,$A=4$,部分参数选取 $\varepsilon=0.8$,$\min\varepsilon=0.3$,$c_1=10$,$c_2=5$,$c_3=2$,$c_4=1$,$lp=10\ 000$。

图 5(b)表示在该奖励设定下,6 节点无人机通信组网的最优拓扑。该拓扑具有特性,保证该网络有较强的抗毁性,并且在此前提下,系统内通信总功耗最低,各个节点尽可能就近选择其他建立链路,具有一定的稳定性。图 5(c)、(d)分别展示 10 节点和 12 节点的最优网络拓扑示意图。

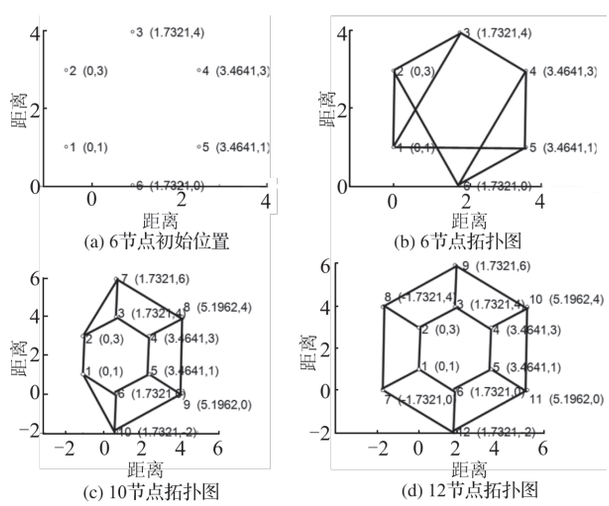


图 5 网络拓扑示意图

图 6 分别展示了该算法控制下随机位置下 24 节点、36 节点的最佳网络拓扑结构。

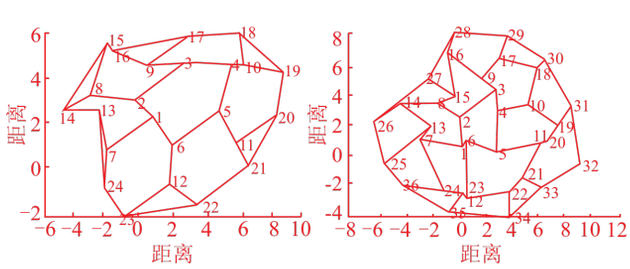
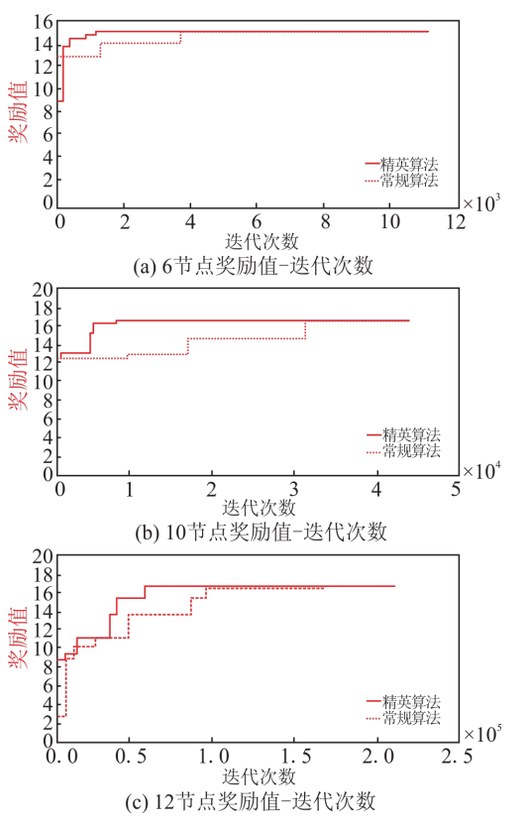


图 6 随机位置多节点网络拓扑示意图

对于节点数目不同的通信网络,训练网络所需

时间也不相同。引入精英策略后，训练过程的迭代次数明显减少。

图 7(a)记录了 6 节点网络 ESDQN 算法与常规算法的奖励值随训练次数的变化对比情况。由图可知，奖励值最终收敛，表示网络根据输入状态准确预测价值量，形成状态到动作的映射。其中，实线表示精英算法奖励值经过约 1 175 次迭代收敛于 14.312 7，虚线表示常规算法经 3 710 次训练后奖励值收敛。图 7(b)、(c)分别表示 10 节点和 12 节点的算法收敛情况对比。对比可知，随着网络节点个数的增加，各节点的动作集 A_i 大小和网络状态 s_i 的个数也随之增加，导致模型收敛所需的迭代次数和时间都大幅增加。



(a) 6节点奖励值-迭代次数

(b) 10节点奖励值-迭代次数

(c) 12节点奖励值-迭代次数

图 7　奖励变化趋势

结果表明：引入精英经验池大幅减少了训练次数，提升效果在 315.74% 以上，对于加速模型收敛有很好的效果。

4.2　响应时间对比

实际情况下，网络拓扑构建情况分为 3 种模式：①组网模式：N 架无人机按照预设编队形成 6 节点自组织网络；②并网模式：按照需求对于 M 组集群无人机进行通信网络合并；③重组模式：任务过程中某 L 个节点脱离组网，形成子集群，其余部分根据当任务需要形成 3-连通网络。表 2 对比了 3 种模式下基于 ESDQN 算法和基于遗传算法[16] 2 种网络拓扑设计方法的指标。

表 2　ESDQN 与遗传算法网络响应时间对比

算法	组网模式下用时/s	并网模式下用时/s	重组模式下用时/s
DQN 算法训练	11 959.486 2	71 335.632 8	54 392.123 7
DQN 算法线上	0.183 1	0.164 2	0.142 1
遗传算法	35.193 4	39.590 2	48.925 7

可以看出，基于遗传算法的网络拓扑设计方法响应时间较长，网络拓扑形成时间远远不能满足战场环境下的动态组网需求。基于 ESDQN 算法设计方法训练与测试分离，尽管训练用时较长，但是新网络拓扑响应时间远小于遗传算法，达到了实时水平，实现线下预先学习，线上实时使用，保证无论节点损毁或是不同任务需要情况下，短时间内形成新拓扑并保持网络连通性、抗毁性良好。同时，线上使用 ESDQN 算法设计网络拓扑占用计算资源远小于遗传算法，更适用于计算能力有限的无人机飞行控制器。

5　结语

本文提出一种基于深度强化学习的网络拓扑设计算法，通过设置奖励函数满足网络图拓扑的抗毁性、功耗、稳定性等需求，提出 ESDQN 加速模型收敛，相比于其他算法大大缩短了响应时间，实现了实时解算最优拓扑。

参考文献(References)：

[1] 陈瑶，梁加红，邹顺，等. 无人机 Ad Hoc 网络拓扑控制算法研究[J]. 计算机仿真，2010，27(7)：33-37.
CHEN Y，LIANG J H，ZOU S，et al. Research on Network Topology Control Algorithm of UAV Ad Hoc [J]. Computer Simulation，2010，27 (7)：33-37.(in Chinese)

[2] 景晓年，梁晓龙，张佳强，等. 航空集群作战编队优化控制研究[J]. 计算机仿真，2017，34(4)：90-94.
JING X N，LIANG X L，ZHANG J Q，et al. Study on Optimal Control of Air Cluster Combat Formation [J]. Computer Simulation，2017，34 (4)：90-94.(in Chinese)

[3] ZHOU C，Y FENG Y Q，YIN Z H，et al. Formation Control of Multi-Agents Based on Method of Matrices [J]. MATEC Web of Conferences，2018，173 (5)：03091.

[4] 陈侠，鹿振宇. 风场干扰下基于一致性卡尔曼滤波的 UAV 编队控制算法[J]. 兵工自动化，2013，32(10)：28-32.

CHEN X, LU Z Y. UAV Formation Control Algorithm Based on Consistent Kalman Filter under Wind Disturbance [J]. Military Automation, 2013, 32 (10): 28-32. (in Chinese)

[5] 贺坤. 无人机编队重构算法研究[D]. 沈阳:沈阳航空航天大学,2016.
HE K. Research on UAV Formation Reconstruction Algorithm [D]. Shengyang: Shenyang University of Aeronautics and Astronautics, 2016. (in Chinese)

[6] 丁贤达. 多机器人编队控制及连通性保持研究[D]. 南昌:华东交通大学,2014.
DING X D. Study on Formation Control and Connectivity Maintenance of Multi-Robots [D]. Nanchang: East China Jiaotong University, 2014. (in Chinese)

[7] 牛轶峰, 肖湘江, 柯冠岩. 无人机集群作战概念及关键技术分析[J]. 国防科技, 2013, 34(5):37-43.
NIU Y F, XIAO X J, KE G Y. Analysis of the Concept and Key Technology of UAV Cluster Operations [J]. National Defense Science and Technology, 2013, 34 (5): 37-43. (in Chinese)

[8] 吴俊, 谭跃进. 复杂网络抗毁性测度研究[J]. 系统工程学报, 2005, 20(2):128-131.
WU J, TAN Y J. Research on the Survivability Measurement of Complex Networks [J]. Journal of Systems Engineering, 2005, 20 (2): 128-131. (in Chinese)

[9] 王班, 马润年, 王刚. 基于自然连通度的复杂网络抗毁性研究[J]. 计算机仿真, 2015, 32(8):315-318.
WANG B, MA R N, WANG G. Research on the Invulnerability of Complex Networks Based on Natural Connectivity [J]. Computer Simulation, 2015, 32 (8): 315-318. (in Chinese)

[10] 袁培燕, 李腊元. Ad Hoc 网络连通度的研究[J]. 计算机工程与应用, 2006, 44(2):177-178.
YUAN P Y, LI L Y. Research on Connectivity of Ad Hoc Network [J]. Computer Engineering and Application, 2006, 44 (2): 177-178. (in Chinese)

[11] 张艳娥, 刘国义, 孙建平,等. 二项分布及其应用[J]. 数理医药学杂志, 2004, 17(5):390-391.
ZHANG Y E, LIU G Y, SUN J P, et al. Binomial Distribution and Its Application [J]. Journal of Mathematical Medicine, 2004, 17 (5): 390-391. (in Chinese)

[12] 李彦平. 移动自组网的定向天线组网技术研究[D]. 西安:西安电子科技大学, 2003.
LI Y P. Research on Directional Antenna Networking Technology for Mobile Ad-Hoc Networks [D]. Xi′an: Xidian University, 2003. (in Chinese)

[13] DOBREV S, KRANAKIS E, KRIZANC D, et al. Strong Connectivity in Sensor Networks with Given Number of Direction Antennae of Bounded Angle [J]. Discrete Mathematics, Algorithms and Applications, 2012, 4(3):1250038.

[14] XU H, DAI H N, ZHAO Q. On the Connectivity of Wireless Networks with Multiple Directional Antennas [C]// IEEE International Conference on Networks. Singapore:IEEE, 2013.

[15] CARAGIANNIS I, KRIZANC D, KAKLAMANIS C, et al. Communicationin is Wireless Networks with Directional Antennas[C]//SPAA′08. New York, NY, USA:ACM, 2008:334-357.

[16] 王亚利,冯有前,刘志国,等. 基于遗传算法的定向天线网络拓扑控制[J]. 空军工程大学学报(自然科学版), 2018,19(2):51-55.
WANG Y L, FENG Y Q, LIU Z G, et al. Directional Antenna Network Topology Control Based on Genetic Algorithm [J]. Journal of Air Force Engineering University (Natural Science Edition), 2018, 19 (2): 51-55. (in Chinese)

[17] HUANG Z, ZHANG Z, RYU B. Power Control for Directional Antenna-Based Mobile Ad-Hoc Networks [C]// International Conference on Wireless Communications & Mobile Computing. New York, NY, USA: ACM, 2006:917-922.

[18] CHATTERJEE S, ROY S, SOMPRAKASH B, et al. A Power Aware Routing Strategy for Ad-Hoc Networks with Directional Antenna Optimizing Control Traffic and Power Consumption[C]// Proceeding of the 7th International Conference on Distributed Computing. Kharagpur, India: Springer-Verlag, 2005: 275-280.

[19] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-Level Control through Deep Reinforcement Learning[J]. Nature, 2015, 518(7540):529-533.

[20] SCHAUL T, QUAN J, ANTONOGLOU I, et al. Prioritized Experience Replay [R]. ArXiv: 1511.05952v4, 2016.

[21] 陈希亮,曹雷,李晨溪,等. 基于重抽样优选缓存经验回放机制的深度强化学习方法[J]. 控制与决策,2018,33(4):600-606.
CHEN X L, CAO L, LI C X, et al. Deep Reinforcement Learning via Good Choice Resampling Experience Replay Memory[J]. Control and Decision, 2018, 33 (4): 600-606. (in Chinese)

(编辑:徐楠楠)