基于深度强化学习的无人机自主部署及能效优化策略

周毅[1,2]，马晓勇[1,2]，郜富晓[1,2]，李伟[1,2]，承楠[3]，路宁[4]

（1. 河南大学计算机与信息工程学院，河南 开封 475004；2. 河南省车联网协同技术国际联合实验室，河南 开封 475004；
3. 滑铁卢大学电气与计算机工程学院，安大略 滑铁卢 N2L 3G1；
4. 汤普森河大学计算机科学系，不列颠哥伦比亚 坎路普斯 V2C 0C8）

摘 要：利用无人机组建空中移动基站，可为地面终端用户提供更灵活、高效的接入服务。受无人机覆盖范围和有限能量的约束，研究如何建立快速、高效、节能的空地协同网络至关重要，无人机需要根据复杂动态场景进行最优覆盖部署，同时要减少部署过程中的路径损耗和能量消耗。基于深度强化学习提出了无人机自主部署和能效优化策略，建立无人机覆盖状态集合，以能效作为奖励函数，利用深度神经网络和 *Q*-learning 引导无人机自主决策，部署最佳位置。仿真结果表明，该方法的部署时间能够有效减少 60%，能耗可降低 10%~20%。

关键词：空地协作组网；无人机；自主部署；能效优化；深度强化学习

中图分类号：TP393

文献标识码：A

doi: 10.11959/j.issn.2096−3750.2019.00106

Autonomous deployment and energy efficiency optimization strategy of UAV based on deep reinforcement learning

ZHOU Yi[1,2], MA Xiaoyong[1,2], GAO Fuxiao[1,2], LI Wei[1,2], CHENG Nan[3], LU Ning[4]

1. School of Computer and Information Engineering, Henan University, Kaifeng 475004, China
2. International Joint Research Laboratory for Cooperative Vehicular Networks of Henan, Kaifeng 475004, China
3. Department of Electrical and Computer Engineering, University of Waterloo, Waterloo N2L 3G1, Canada
4. Department of Computing Science, Thompson Rivers University, Kamloops V2C 0C8, Canada

Abstract: Utilizing a UAV to build aerial mobile small cell can provide more flexible and efficient access services for ground terminal users. Constrained by the coverage and limited energy of the UAV, it is necessary to study how to build a fast, efficient and energy-saving air-ground collaborative network. To deal with complex dynamic scenarios, the UAV needs to deploy an optimal coverage position, and meanwhile reduce both path loss and energy consumption in the deployment process. Based on the deep reinforcement learning, a strategy of autonomous UAV deployment and efficiency optimization was proposed. The coverage state set of UAV was established, and the energy efficiency was used as a reward function. Depth neural network and *Q*-learning were used to guide UAV to make autonomous decision and deploy the optimal position. The simulation results show that the deployment time of the proposed method can be effectively reduced by 60%, while the energy consumption can be reduced by 10%~20%.

Key words: aerial-ground cooperative networking, unmanned aerial vehicle, autonomous deployment, efficiency optimization, deep reinforcement learning

收稿日期：2019−04−05；**修回日期：**2019−05−10

基金项目：国家自然科学基金资助项目（No.61701170）；河南省国际科技合作项目（No.182102410050）；河南省青年骨干教师计划（No.2016GGJS-018）

Foundation Items: The National Natural Science Foundation of China (No.61701170), Henan International Science & Technology Cooperation Program (No.182102410050), Henan Young Scholar Promotion Program (No.2016GGJS-018)

1 引言

由于无人机（UAV，unmanned aerial vehicle）的灵活性和机动性能，它在城市勘测、交通诱导、应急通信、军事作战、工业物联网等领域都有潜在的应用价值。无人机与地面基础设施的协作，既可以业务分载，也可以提供增强服务。在满足快速、高效的覆盖前提下，无人机能够更好地进行 3D 时空态势感知和协作动态组网，从而提高空地协同的响应能力和无缝接入能力。无人机可以作为空中移动基站的有效部署，为用户提供更好的网络覆盖率和更高的数据传输率[1-4]。利用无人机组网具有以下优点：1) 在高空水平位置适当部署无人机基站，可为地面用户提供更好的视距（LoS，line-of-sight）传播路径，从而提高通信质量；2) 无人机能够向通信覆盖中断的区域提供快速、可靠、高效的网络接入服务[5]；3) 无人机能够通过协作组网快速形成智能体集群，从而对目标场景进行全方位的时空态势感知。由于无人机组网的高动态性及其面临的复杂环境，还存在诸多技术挑战。首先，地面基站是固定的，因此，路径损耗仅取决于用户的位置，而空对地（A2G，air-to-ground）信道模型是由用户位置和无人机位置共同决定的。其次，不同于地面基站部署的二维问题，无人机的位置部署是一个非线性约束的三维问题。另外，目前的小微型无人机通常采用电池供电，执行任务时间受限[6]。因此，无人机组网的关键挑战是如何以最快速、最有效、最低能耗的方式来部署无人机基站，从而实现高效、节能的空地协作组网。空地协作组网总体架构如图 1 所示，在该空地协作网络中，无人机作为空中移动基站，可利用 5G 等移动组网技术为地面用户提供高效、可靠的通信链路，同时能与地面网络进行协同交互，提供智能优化及辅助决策服务。

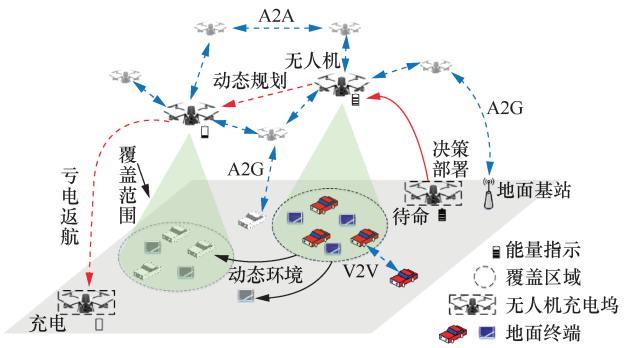


图 1 空地协作组网总体架构

空地协同网络有较高的应用价值，已成为近年的研究热点，特别是在组网架构、资源优化、覆盖部署和能效优化等方面已有诸多研究。在组网架构方面，作者已于 2015 年提出了一种空地协同的车联网体系架构，该架构通过空中自组织网络与地面车辆自组织网络的协同，提升特殊应用场景下的空地协作能力，并进行了实测验证[3]。Cheng 等[4]在此基础上引进边缘计算思想，提出空地协作的移动边缘组网架构，并利用软件定义网络（SDN，software defined network）有效部署和调度异构网络资源，充分利用边缘网络促进空地一体化网络的实时交互和高效协作。随着新一代移动通信技术的发展，空地一体化网络在持续演进中，网络架构的优化仍是研究焦点。

在资源优化方面，Lu 等[1]基于马尔可夫近似方法设计了无人机基站动态部署机制，通过无人机基站间的紧密协作，为热点区域数据分流提供服务，可有效卸载蜂窝宏基站的负荷。基于参考文献[7]提出的 A2G 路径损失模型，参考文献[8-10]利用该模型研究了无人机的部署问题，目的是使无人机的数量最小化。其中，Lyu 等[8]提出了一个具有连续移动基站部署的多项式时间算法，解决了不同用户密度下的无人机基站数量最小化问题。Al-Hourani 等[9]采用三维泊松点过程（PPP）方法，在传统蜂窝网络的频谱共享环境下，研究了部署无人机基站的最优密度问题。Kalantari 等[10]提出了一种基于粒子群优化的启发式算法，次优地找到了服务于特定区域的无人机基站最小数量及其位置。经典的粒子群算法通过搜索可以找到次优解，但是搜索速度相对较慢。

在覆盖部署方面，Bor-Yaliniz 等[2]假设无人机基站以全功率传输，并将无人机基站部署问题表示为一个平方约束的混合整数非线性问题，使无人机覆盖的用户数量最大化。Alzenad 等[11]提出了一个评估无人机基站的三维位置框架，使用最小的传输功率使覆盖用户的数量最大化。Ruan 等[12]利用博弈论的方法，解决了多无人机覆盖范围最大化的问题。Chen 等[13]提出了一种改进的多种群遗传算法，可以最大限度地满足不同服务质量要求的覆盖用户数量。Lai 等[14]提出了一种基于密度感知的部署算法，在每个用户所需数据速率最小的约束下，最大限度地增加被覆盖用户的数量。以上部署方法均可将无人机基站部署在覆盖范围最大的位置，但是

在实际情况中，还要充分考虑能耗优化问题，需要在部署和能耗之间寻求平衡。

在能效优化方面，Wang 等[15]通过在水平和垂直方向解耦部署问题来解决无人机基站满足覆盖所需的最小发射功率问题。Shakhatreh 等[16]提出了一种粒子群优化算法进行无人机三维部署，考虑室内用户的位置均匀分布在每一层，使覆盖用户所需的总发射功率最小化。以上能效优化算法可以较好地减少通信能耗或水平移动能耗，但在实际中，无人机从起始位置移动到最优部署位置所消耗的能量不容忽视。

为了进一步提升无人机部署效率，本文提出了基于深度强化学习（DRL，deep reinforcement learning）[17]的无人机自主部署及能效优化策略。DRL 能够很好地处理复杂的状态空间和时变环境，在无人驾驶、移动网络、边缘计算等领域应用广泛。DRL 利用深度神经网络（DNN，deep neural network）来指导决策，体现了优越的性能。由于空地协作组网涉及覆盖率、路径损耗、能耗等多个目标，具有不同的目标、约束、状态和动作空间，因此，利用 DRL 来实现复杂网络环境下的决策控制具有很大挑战。本文尝试在空地协同网络环境下解决基于 DRL 的无人机最优部署问题，主要贡献如下。

1) 提出了基于深度强化学习的无人机自主部署及能效优化策略，将无人机的初始位置和地面终端的分布作为状态集合，将无人机的能效作为奖励函数，根据 DNN 和 *Q*-learning 指导无人机的动作，以找到满足覆盖率、路径损耗和最小能耗的位置。

2) 解决了无人机从起始位置到最优位置的最小移动能耗问题，将无人机的初始位置和移动能耗作为状态集合的一部分，无人机根据不同的初始位置可以选择相应的最优位置。

3) 利用 Python 搭建了一个 DRL 仿真环境，对所提算法进行了仿真，大量仿真结果表明，该算法的部署时间能够有效减少 60%，能耗可降低 10%~20%。

2　系统模型和问题规划

2.1　系统模型

空地协作组网总体架构如图 1 所示，无人机为目标区域的地面用户提供通信覆盖。目标区域被划分为 $I\times J$ 个单元，无人机部署在单元格中心上方，并且覆盖整个单元格。

给定地面终端的位置，用 $B=\{b_{11},b_{12},\cdots,b_{IJ}\}$ 表示地面终端分布指示变量的集合，单元 ij ($i\in\{1,2,\cdots,I\}$，$j\in\{1,2,\cdots,J\}$)中有地面终端时，指示变量 $b_{ij}=1$，否则为 0。$C=\{c_{11},c_{12},\cdots,c_{ij}\}$ 表示地面无人机覆盖指示变量的集合，单元 ij 被覆盖时，指示变量 $c_{ij}=1$，否则为 0。由集合 B 和 C 可以得到地面终端覆盖状态的集合 $W=\{w_{11},w_{12},\cdots,w_{ij}\}$，同理，若单元 ij 有地面终端 $b_{ij}=1$ 且被覆盖 $c_{ij}=1$ 时，$w_{ij}=1$，否则为 0。假设无人机的初始位置为 $(u_{x_0},u_{y_0},0)$，在 t 时刻无人机的位置为 (u_x^t,u_y^t,u_z^t)。假设无人机的覆盖区域是半径为 R 的圆形区域，其覆盖半径随无人机部署高度的增加而增加。

由于无人机能量有限，无法长时间飞行。因此，需要找到一个最优位置，使无人机在满足覆盖率和路径损耗的前提下，从起始位置到最优位置的能耗最小。该能耗包括水平移动的能耗和垂直移动的能耗，其中，随飞行距离的增加，能量消耗分别在水平方向和垂直方向呈线性增加。

2.2　问题陈述

为了找到一个最优控制策略，在满足覆盖率和路径损耗的前提下，无人机以最少的能耗从初始位置移动到目标位置。覆盖率定义为被覆盖的地面终端数量与总地面终端数量之比，在 t 时刻，覆盖率用 α^t 表示，则

$$\alpha^t=\frac{\sum_{i=1}^{I}\sum_{j=1}^{J}w_{ij}^t}{\sum_{i=1}^{I}\sum_{j=1}^{J}b_{ij}^t} \tag{1}$$

显然，$\alpha^t\in[0,1]$，且 α^t 越大，对整个系统越有利。

采用参考文献[9]中提出的 A2G 信道模型，地面单元 ij 和无人机基站之间 LoS 链路的概率定义如下

$$P_{\mathrm{LoS}}(r_{ij},h)=\frac{1}{1+a\exp\left\{-b\left(\arctan\left(\frac{h}{r_{ij}}\right)-a\right)\right\}} \tag{2}$$

其中，a 和 b 是与环境类型相关的常数[7]，h 表示无人机基站的高度，r_{ij} 为无人机基站与单元 ij 之间的水平距离。

$$r_{ij}=\sqrt{(i-u_x)^2-(j-u_y)^2} \tag{3}$$

此外，非视距（NLoS，non line-of-sight）链路的概率为

$$P_{\mathrm{NLoS}}=1-P_{\mathrm{LoS}} \tag{4}$$

无线传播信号除了自由空间传播损耗外，还会因城市环境的遮挡和散射而遭受损失，本文处理的是平均路径损失，而不是随机行为。通常基站部署时主要考虑信道的长期变化，而不是小尺度的变化。因此，无人机基站中 LoS 和 NLoS 链路的路径损失模型[9]分别为

$$L_{\mathrm{LoS}}=20\log\left(\frac{4\pi f_{\mathrm{c}}d_{ij}}{c}\right)+\eta_{\mathrm{LoS}} \tag{5}$$

$$L_{\mathrm{NLoS}}=20\log\left(\frac{4\pi f_{\mathrm{c}}d_{ij}}{c}\right)+\eta_{\mathrm{NLoS}} \tag{6}$$

其中，f_{c} 为载波频率，d_{ij} 为无人机基站与单元 ij 之间的距离

$$d_{ij}=\sqrt{h^2+r_{ij}^2} \tag{7}$$

在 LoS 和 NLoS 模型下，A2G 链路平均路径损失为

$$L(h,r_{ij})=L_{\mathrm{LoS}}\times P_{\mathrm{LoS}}+L_{\mathrm{NLoS}}\times P_{\mathrm{NLoS}} \tag{8}$$

对于给定的发射功率 P_{t}，单元 ij 接收的功率取决于其通信链路所经历的路径损耗，可以写成

$$P_{\mathrm{r}}=P_{\mathrm{t}}-L(h,r_{ij}) \tag{9}$$

为了保证服务质量，单元的接收功率 P_{r} 必须超过一定的阈值 $P_{\min}$，也就是说，如果单元 ij 的链路经历的路径损耗小于或等于某个阈值 κ，则单元 ij 将被覆盖，即

$$L(h,r_{ij})\leqslant\kappa \tag{10}$$

此外，无人机在空中飞行会导致能源消耗，为了延长网络寿命，需要将能耗降至最低。由于无人机水平移动和垂直移动相等距离所消耗的能量不同，假设无人机水平移动一个单位的能耗为 e_1，垂直移动一个单位的能耗为 ξe_1，其中，ξ 由无人机特性决定。那么，无人机从初始位置到部署位置的移动能耗为

$$e=e_1\sqrt{(u_x-u_{x_0})^2+(u_y-u_{y_0})^2}+\xi e_1 u_z \tag{11}$$

3 DRL 自主部署算法

无人机位置部署的目标是设计一个控制策略，满足 3 个约束条件：最大化覆盖率、路径损耗最低和最小化能源消耗。因此，无人机应部署在合适的高度，使覆盖率最大化，且路径损耗和能源消耗相对较低。由于该问题是非凸的，无法得到解析解，因此，寻找无人机的最优位置颇具挑战。权衡以上因素，考虑利用 DRL 来寻找有效解。

3.1 算法框架

DRL 将深度学习的感知能力和强化学习的决策能力相结合，不断以试错的方式与环境进行交互，通过最大化累积奖赏的机制来获得最优策略[17]。

本文采用深度 Q 网络（DQN，deep Q-network）来求解无人机位置部署问题，核心思想是将 Q 函数值网络作为评判模块，基于值网络遍历当前观测状态下的各种动作，与环境进行实时交互。将状态、动作和奖惩值存储在回放记忆单元中，采用 Q-learning 算法来反复训练值网络，最后选择能获得最大价值的动作来部署无人机。基于 DRL 的位置部署算法框架如图 2 所示。其中，DQN 模型表示为集合 $\{s,a,R\}$，如下所述。

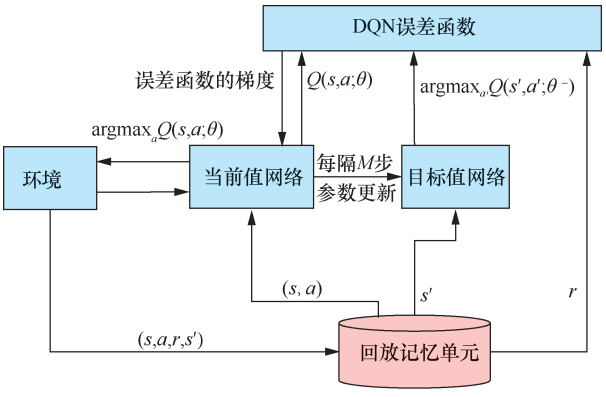


图 2 基于 DRL 的位置部署算法框架

1) 状态 s_t（t 时刻）

s_t 包含 3 个部分：$b_{ij}^t\in[0,1]$ 表示第 ij 单元是否有地面终端；$\{u_x^t,u_y^t,u_z^t,u_{x_0},u_{y_0}\}$ 表示无人机的初始位置以及 t 时刻的位置；β 表示期望的覆盖率。

所以 $s_t=\{b_{11}^t,b_{12}^t,\cdots,b_{ij}^t,u_x^t,u_y^t,u_z^t,u_{x_0},u_{y_0},\beta\}$，元素总量是 $(i\times j+6)$，通过这种方式定义状态，DRL 智能体可以根据当前地面终端的位置、无人机的位置以及期望覆盖率做出决策。

2) 动作 a_t（t 时刻）

将无人机的动作离散化，定义 11 个动作，UAV 水平运动方向示意图如图 3 所示。水平面上假定 8 个方向的动作，另外包括垂直方向上下两个动作以及悬停动作[18]。其中，悬停动作是让无人机找到最优位置后可以保持当前状态。假设无人机执行每个动作都移动固定的距离，根据环境的反馈信息，如果执行该动作后无人机没有到达最优位置，无人机将会继续采取相应动作，直到处于最优位置，完成自主部署。总体动作空间为 $\{0,1,2,3,4,5,6,7,8,9,10\}$，选择哪个动作，该动作值就为 1，其他动作值则为 0。

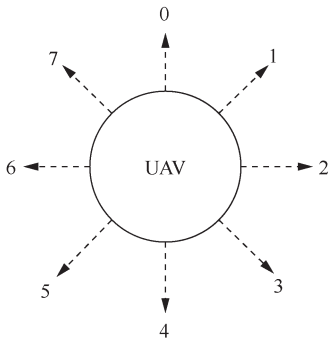


图 3　UAV 水平运动方向示意图

3) 奖励函数 r_t（t 时刻）

奖励函数 r_t 定义为

$$r_t = \frac{\Delta\alpha^t}{\Delta e^t} \tag{12}$$

其中，$\Delta\alpha^t = \alpha^t - \alpha^{t-1}$ 为覆盖率增量；$\Delta e^t = e^t - e^{t-1}$ 为能耗增量。同时，为了保证满足覆盖率的要求，如果 α 小于要求的覆盖率 β 或路径损耗 $L(h,r_{ij})$ 大于 κ 时

$$r_t = r_t - p \tag{13}$$

相当于给一个惩罚。这里考虑的奖励需要恰当地处理 3 个目标：覆盖率、路径损耗和能量消耗。将 $\Delta\alpha^t$ 看作覆盖增量，因此，奖励函数的分子相当于收益，分母相当于成本（以能量消耗为单位）。因此，奖励可以看作能量效率（单位能耗带来的收益）。那么，累积回报的最大化就等于平均能量效率的最大化。

DQN 模型主要包括卷积神经网络（CNN，convolutional neural network）和基于 Q-learning 的决策模型。随着训练时间的增加，系统可以利用 CNN 提高 Q-learning 模块的学习速度。与 Q-learning 类似，DQN 为每个状态—动作对更新 Q 函数，这是状态 s 和动作 a 在 n 个时间段的长期折扣奖励的期望。Q 函数定义如下[17]

$$Q(s,a) = E_{s'}[R_s + \gamma \max Q(s',a) \mid s,a] \tag{14}$$

其中，R_s 是奖励函数，折扣因子 γ 衡量奖励值在 Q 函数中作用的大小。Q 函数通过使用具有可调权值参数的 CNN 来近似，它是每个动作的非线性逼近器。然而，由于网络的动态性，CNN 模型需要进行再训练以适应无人机的位置部署。DQN 使用来自经验回放缓冲区的小批次数据，包括在学习期间收集的状态样本（而不是立即收集的样本）。与传统 Q-learning 中直接采样相比，经验回放打破了顺序生成样本之间的相关性，从而避免了发散，加快了收敛速度。本文的 DQN 模型通过最小化损失函数 $L(\theta^Q)$ 来训练

$$L\left(\theta^Q\right) = E\left[y_t - Q\left(s_t, a_t \mid \theta^Q\right)\right] \tag{15}$$

其中，θ^Q 是 DQN 的权重向量，y^t 是目标值[19]

$$y_t = r(s_t,a_t) + \gamma Q(s_{t+1}, \pi(s_{t+1} \mid \theta^\pi) \mid \theta^Q) \tag{16}$$

同时，DQN 使用了另外一个目标网络来估计 DNN 训练的目标值 y_t。目标网络具有与原始 DNN 相同的结构，但其权值每隔几个时刻才会随着原始 DNN 权值的变化而更新一次，是为了克服 DNN 的不稳定性。

3.2　算法流程

DRL 智能体定期收集无人机和地面终端的状态，使用所提算法找到无人机的最佳动作，基于 DRL 的无人机自主部署算法如算法 1 所示。

算法 1　基于 DRL 的无人机自主部署

1) 初始化记忆回放单元 D，其容量为 N
2) 初始化 Q 网络，随机生成权重 θ
3) 初始化目标 Q 网络，权重为 $\theta^- = \theta$
4) for episode =1, 2,⋯, M do
5)　初始化状态 s_1
6)　for step =1, 2,⋯, T do
7)　以 ε 概率选择一个随机的 action
8)　以 $1-\varepsilon$ 概率选择 $a_t = \max_a Q(s_t,a;\theta)$
9)　执行动作 a_t，得到奖励 r_t 及新的状态 s_{t+1}
10)　for i=1, 2,⋯, I do
11)　for j=1, 2,⋯, J do

12) if $L(h,r_{ij})>k$ then
13) $r_t=r_t-p$
14) end if
15) end for
16) end for
17) if $\alpha<\beta$ then
18) $r_t=r_t-p$
19) end if
20) 将数据样本(s_t,a_t,r_t,s_{t+1})存入 D 中
21) 从 D 中随机抽取一个小样本$(s_{t'},a_{t'},r_{t'},s_{t'+1})$
22) 如果$t'+1$步是终点，令 $y_{t'}=r_{t'}$，否则，
23) 令 $y_{t'}=r_{t'}+\gamma\max_{a'}Q(s_{t'+1},a';\theta^-)$
24) 对$(y_{t'}-Q(s_{t'},a_{t'};\omega))^2$关于 ω 使用梯度下降法进行更新
25) 每隔 M 步更新目标 Q 网络，$\theta^-=\theta$
26) end for
27) end for

3.2.1 *初始化*（第 1 ~ 5 行）

先初始化记忆回放单元 D，然后该算法随机初始化了动作值函数 Q 网络的权重 θ，如上所述，使用目标网络 Q^ 提高学习稳定性。目标网络与原始 Q 网络具有相同结构，采用与原始网络（第 2 行）相同的方式初始化权重 θ^-，但为了提高稳定性，经过一段时间后再更新（第 26 行）。仿真中，使用的更新速率值为 0.001。

3.2.2 *自主探索*（第 6 ~ 19 行）

在探索过程中，算法以 ε 的概率随机地选择动作，以$1-\varepsilon$ 的概率选择 Q 网络的动作，ε 初始化为 1，并以一定的速率随时间的延长而衰减。同时，加入了一个惩罚过程（第 10～19 行）。仿真中，惩罚被设置为一个很大的值，是相应奖励的 100 倍。

3.2.3 *更新神经网络*（第 20 ~ 25 行）

使用了一个记忆回放缓冲区更新 Q 网络，算法刚开始时已经被初始化（第 3 行）。具体而言，先将收集的样本放到记忆回放缓冲区（第 20 行）中，再用随机抽取的缓冲区样本更新 Q 网络（第 24 行）。如上所述，Q 网络的参数 θ 通过最小化损失函数 L 式(15)更新。在仿真中，设置 minibatch 的长度 H 为 1 024、折扣系数 γ 为 0.9。经过一段时间后再更新目标网络。仿真中，使用了一个两层全连接的前馈神经网络作为 Q 网络，其中，第一层和第二层分别包含400个和300个神经元，并利用ReLU函数进行激活[20]。在最后的输出层中，使用 Tanh 函数作为激活函数，这些过程均在 TensorFlow 1.4 上实现。

3.3 算法扩展

上述算法对单无人机进行了自主部署，如果定义新的动作和状态空间，也可以将该算法应用于多无人机的联合部署中。假设有 K 个无人机进行联合部署，按照集中式的部署原则，将动作空间设为 $A_t=\{a_t^1,a_t^2,\cdots,a_t^K\}$，其中，$a_t^k$ 表示 t 时刻第 k 个无人机采取的动作。对于状态空间来说，需要获取每个无人机的位置，无人机群的位置坐标集合定义为 $U=\{u_x^1,u_x^2,\cdots,u_x^K;u_y^1,u_y^2,\cdots,u_y^K;u_z^1,u_z^2,\cdots,u_z^K\}$，其中，$u_x^k$ 表示第 k 个无人机的 X 轴坐标，u_y^k 表示第 k 个无人机的 Y 轴坐标，u_z^k 表示第 k 个无人机的 Z 轴坐标，定义 $S_t=\{b_{11}^t,b_{12}^t,\cdots,b_{ij}^t,U,\beta\}$ 为新的状态空间。

根据所定义的动作和状态空间，结合本文设计的奖励函数，可以完成对多无人机的联合部署。本文的奖励函数是根据所考虑的覆盖率、能耗和链路质量等指标所设计的，如果考虑其他指标，只需要设计新的奖励函数，同样可以完成多无人机的联合部署。

4 仿真与分析

4.1 仿真设置

利用 Python 搭建了 DRL 仿真环境，DRL 仿真运行场景如图 4 所示，将目标区域设置为 11×11 个单元的正方形。其中，每个单元对应 100 m。把这个区域分成 121 个单元，每个单元都有一个单位大小，假设无人机的覆盖半径与高度比为 1。使用 Tensor Flow 1.4 和 Python 3.5 在 Ubuntu 16.04.3 服务器上运行仿真。将提出的基于 DRL 的方法训练 2 000 个周期，每个周期有 2 000 个时刻。经过训练，对其进行了 $T=2\,000$ 个周期的测试，使用以下指标进行性能评估。

寻找时间：从无人机初始位置出发到找到最优点的时间。在这方面对比了 DRL 和粒子群优化算法。

粒子群算法是目前寻找最优位置普遍采用的方法，该算法初始化为一群随机的粒子，然后通过多次迭代找到最优解。在每次迭代过程中，粒子通过本身所找到的最优解（个体极值）和整个种群

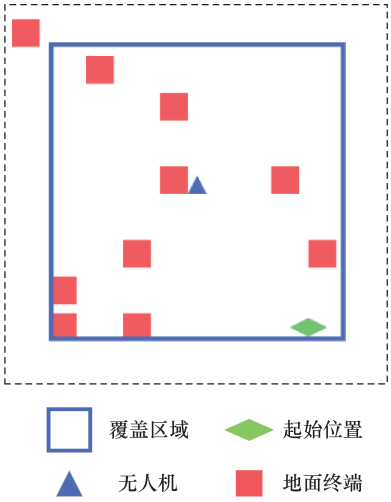


图 4　DRL 仿真运行场景

目前找到的最优解（全局极值）来更新自己，也可以使用粒子本身的邻居极值（局部极值）来更新自己[16]。

平均能耗：在同一个地面终端覆盖密度、同一个期望覆盖率下，不同初始位置所对应的能耗平均值。在这方面对比了 DRL 和中心布局方法。

中心布局是常用的无人机位置部署方法，将无人机放置在整个目标区域的中心，根据目标区域的范围，得到无人机的覆盖半径，再由无人机的覆盖半径得到无人机的高度。

4.2　结果与分析

累计奖励值随时间的变化如图 5 所示，累计奖励值随着时间的增加而增加。当累计奖励值达到 600 个训练周期后停止增长，趋于稳定。这是因为在部署任务开始时，许多地面终端还未被覆盖，有很大的空间来提高奖励值。当覆盖率满足需求时，奖励值提高的空间很小，因此会达到稳定值。

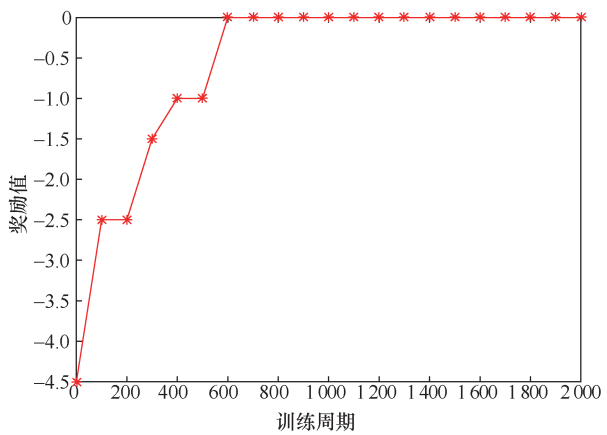
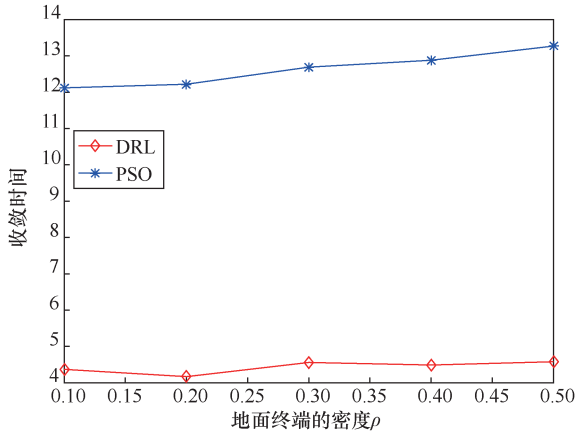
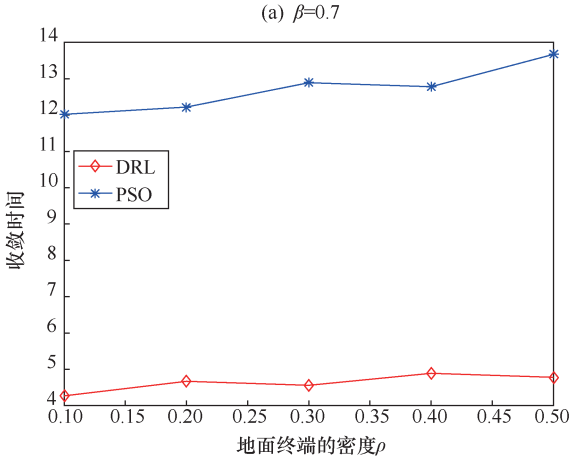


图 5　累计奖励值随时间的变化

不同地面终端密度下的部署时间如图 6 所示，图 6 描述了无人机寻找最优位置的时间随着地面终端密度变化的情况，其中，图 6(a)和图 6(b)分别是不同期望覆盖率下的收敛时间，可见，不同的期望覆盖率对收敛时间的影响不大。结果表明，无人机使用 DRL 算法找到最优位置的时间明显小于使用粒子群算法，这是因为 DRL 具有强大的“记忆功能”，能够根据记忆快速找到最优值，而粒子群每次寻找最优值都需要进行全局搜索。与粒子群算法相比，DRL 算法受地面终端密度 ρ 的影响小，这同样得益于 DRL 的“记忆功能”，而粒子群需要计算与所有地面终端之间的路径损耗，地面终端密度越大，计算的时间越长。



(a) β=0.7



(b) β =0.9

图 6　不同地面终端密度下的部署时间

不同覆盖率下的能耗分析如图 7 所示，为了避免偶然性，能耗设为同一个覆盖率、同一个密度下的不同初始位置对应的平均能耗，同时，比较了不同的地面终端密度与常规算法的能耗差异，常规算法选为按照中心布局的覆盖范围最大的点。结果表明，基于 DRL 的无人机部署算法能耗随着覆盖率

的增加而增加，中心布局的能耗不变，因为常规算法只是把无人机部署到覆盖范围最大的点，而 DRL 算法可以根据地面终端的不同位置部署到最佳的地方。

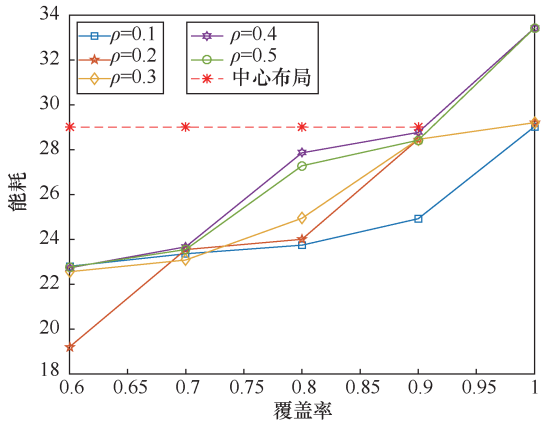


图 7　不同覆盖率下的能耗分析

另外，常规算法的覆盖率只能达到 90%，而 DRL 算法的覆盖率可以达到 100%，并且常规算法的能耗始终高于 DRL 算法，是因为 DRL 算法的自主性和智能性使其可以根据地面终端的分布，用最小的能耗满足覆盖率和路径损耗的要求，而常规算法并没有考虑地面终端的分布情况。当地面终端的密度较低时，DRL 算法的能耗比常规算法的能耗显著降低。因此，在地面终端密度较低时，更能体现 DRL 算法的优越性。

5　结束语

受无人机覆盖范围及有限能量的约束，本文提出了一种基于 DRL 的无人机自主部署及能效优化策略。无人机根据自身初始位置、地面通信终端位置以及期望的覆盖率等因素进行 DRL 自主决策部署，以找到满足覆盖率、路径损耗和最小能耗的最优位置。通过对该算法的大量仿真分析和数值验证，结果表明，所提出的 DRL 自主部署及优化策略相对于粒子群算法有较显著的速度提升，相对于其他传统方法也体现了能耗上的优势。空地协作组网是复杂的网络优化问题，下一步将对异构网络的协同优化及多无人机动态演进机制等进行更深入的研究。

参考文献：

[1] LU N, ZHOU Y, SHI C, et al. Planning while flying: a measurement-aided dynamic planning of drone small cells[J]. IEEE Internet of Things Journal, 2019, 6(2): 2693-2705.

[2] BOR-YALINIZ R, EL-KEYI A, YANIKOMEROGLU H, et al. Efficient 3D placement of an aerial base station in next generation cellular networks[C]//2016 IEEE International Conference on Communications(ICC). IEEE, 2016: 1-5.

[3] ZHOU Y, CHENG N, LU N, et al. Multi-UAV aided networks: aerial-ground cooperative vehicular networking architecture[J]. IEEE Vehicular Technology Magazine, 2015, 10(4): 36-44.

[4] CHENG N, XU W, SHI W, et al. Air-ground integrated mobile edge networks: architecture, challenges and opportunities[J]. IEEE Communications Magazine, 2018, 56(8): 26-32.

[5] MOZAFFARI M, SAAD W, BENNIS M, et al. Unmanned aerial vehicle with underlaid device-to-device communications: performance and tradeoffs[J]. IEEE Transactions on Wireless Communications, 2016, 15(6): 3949-3963.

[6] ZENG Y, ZHANG R. Energy-efficient UAV communication with trajectory optimization[J]. IEEE Transactions on Wireless Communications, 2017, 16(6): 3747-3760.

[7] AL-HOURANI A, KANDEEPAN S, JAMALIPOUR A. Modeling air-to-ground path loss for low altitude platforms in urban environments[C]//2014 IEEE Global Communications Conference. IEEE, 2014: 2898-2904.

[8] LYU J, ZENG Y, ZHANG R, et al. Placement optimization of UAV-mounted mobile base stations[C]//IEEE Communications Letters. 2017, 21(3): 604-607.

[9] AL-HOURANI A, KANDEEPAN S, LARDNER S. Optimal LAP altitude for maximum coverage[J]. IEEE Wireless Communications, 2014, 3: 569-572.

[10] KALANTARI E, YANIKOMEROGLU H, YONGACOGLU A. On the number and 3D placement of drone base stations in wireless cellular networks[C]//2016 IEEE 84th Vehicular Technology Conference (VTC-Fall). IEEE, 2016: 1-6.

[11] ALZENAD M, EL-KEYI A, LAGUM F, et al. 3D placement of an unmanned aerial vehicle base station (UAV-BS) for energy-efficient maximal coverage[C]//IEEE Wireless Communications Letters, 2017, 6(4): 434-437.

[12] RUAN L, WANG J, CHEN J, et al. Energy-efficient multi-UAV coverage deployment in UAV networks: a game-theoretic framework[J]. China Communications, 2018, 15(10): 194-209.

[13] CHEN Y, LI N, WANG C, et al. A 3D placement of unmanned aerial vehicle base station based on multi-population genetic algorithm for maximizing users with different QoS requirements[C]//2018 IEEE 18th International Conference on Communication Technology (ICCT). IEEE, 2018: 967-972.

[14] LAI C, CHEN C, WANG L. On-demand density-aware UAV base station 3D placement for arbitrarily distributed users with guaranteed data rates[C]//IEEE Wireless Communications Letters. IEEE, 2019: 1.

[15] WANG L, HU B, CHEN S. Energy efficient placement of a drone base station for minimum required transmit power[C]//IEEE Wireless Communications Letters. IEEE, 2018.

[16] SHAKHATREH H, KHREISHAH A, ALSARHAN A, et al. Efficient 3D placement of a UAV using particle swarm optimization[C]//2017

8th International Conference on Information and Communication Systems (ICICS). IEEE, 2017: 258-263.

[17] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

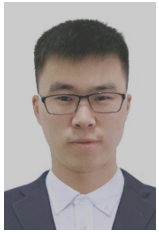
[18] XU Y, LIU Z, WANG X. Monocular vision based autonomous landing of quadrotor through deep reinforcement learning[C]//2018 37th Chinese Control Conference (CCC). IEEE, 2018: 10014-10019.

[19] ZHANG Y, SUN P, YIN Y, et al. Human-like autonomous vehicle speed control by deep reinforcement learning with double *Q*-learning[C]//2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018: 1251-1256.

[20] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning[M]. Cambridge: MIT Press, 2016.

[作者简介]

周毅（1981− ），男，河南信阳人，河南大学副教授、博士生导师，主要研究方向为车联网、空地协同组网、平行增强学习、协作机器人等。



马晓勇（1993− ），男，河南洛阳人，河南大学硕士生，主要研究方向为无人机组网、边缘计算等。

郜富晓（1992− ），女，河南洛阳人，河南大学硕士生，主要研究方向为空地协同组网、深度学习等。

李伟（1979− ），女，河南济源人，河南大学副教授，主要研究方向为车联网优化控制、协作通信等。

承楠（1987− ），男，辽宁锦州人，滑铁卢大学在站博士后，主要研究方向为车联网、人工智能、空地协同等。

路宁（1984− ），男，山西长治人，汤普森河大学助理教授，主要研究方向为车联网、移动边缘计算等。