

物联网数据收集中无人机路径智能规划

付澍^{1,2}, 杨祥月¹, 张海君³, 陈晨¹, 喻鹏⁴, 简鑫¹, 刘敏¹

(1. 重庆大学微电子与通信工程学院, 重庆 400030; 2. 重庆大学信息物理社会可信服务计算教育部重点实验室, 重庆 400030;
3. 北京科技大学计算机与通信工程学院, 北京 100083; 4. 北京邮电大学网络与交换技术国家重点实验室, 北京 100876)

摘 要: 为解决无人机在数据收集过程中的路径规划问题, 将其分为全局路径规划和局部路径规划。针对全局路径规划, 将其建模为一个定向问题, 定向问题是背包问题和旅行商问题 2 种经典优化问题的组合。采用指针网络深度学习对该模型进行求解, 并在无人机能量约束下得到其服务节点集合及服务顺序。针对局部路径规划, 基于无人机接收到节点的参考信号强度, 通过深度 Q 网络学习对无人机局部飞行路径进行规划, 使无人机逼近节点位置并服务各节点。仿真结果表明, 所提方案能够在无人机能量约束下有效提升其数据收集的收益。

关键词: 无人机; 数据收集; 路径规划; 指针网络; 深度 Q 网络

中图分类号: TN92

文献标识码: A

DOI: 10.11959/j.issn.1000-436x.2021036

UAV path intelligent planning in IoT data collection

FU Shu^{1,2}, YANG Xiangyue¹, ZHANG Haijun³, CHEN Chen¹, YU Peng⁴, JIAN Xin¹, LIU Min¹

1. School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400030, China

2. Key Laboratory of Dependable Service Computing in Cyber Physical Society of Ministry of Education, Chongqing University, Chongqing 400030, China

3. School of Computer & Communication Engineering, University of Science and Technology Beijing, Beijing 100083, China

4. State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: To solve the problem of path planning of UAV data collection, it was generally be divided into global path planning and local path planning. For global path planning, it was modeled as an orientation problem, which was a combination of two classical optimization problems, the knapsack problem and the traveling salesman problem. The pointer network of deep learning was used to solve the model to obtain the service node set and service order under the energy constraint of the UAV. In terms of the local path planning, the reference signal strength (RSS) of the sensor node received by UAV was employed to learn the local flight path of UAV by deep Q network, which enabled the UAV to approach and serve the nodes. Simulation results show that the proposed scheme can effectively improve the revenue of UAV data collection under the energy constraint of UAV.

Keywords: unmanned aerial vehicle, data collection, path planning, pointer network, deep Q network

1 引言

近年来, 物联网产业的飞速发展, 极大地推动了无线传感器网络 (WSN, wireless sensor network)

技术的应用, 其承载的业务数据量呈几何式增长。WSN 中存在大量的数据需要被收集, 根据收集方式的不同, 可将其分为 2 种类型, 静态数据收集和移动数据收集。静态数据收集是指传感器网络中的

收稿日期: 2020-08-26; 修回日期: 2020-11-20

通信作者: 刘敏, liumin@cqu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61701054); 中央高校基本科研业务费专项资金资助项目 (No.2020CDJQY-A001, No.2020CDJGFWZ014)

Foundation Items: The National Natural Science Foundation of China (No.61701054), The Fundamental Research Funds for the Central Universities (No.2020CDJQY-A001, No.2020CDJGFWZ014)

节点通过自组网,将自身采集的传感器数据经过多跳上传到数据中心^[1];移动数据收集是指在被监测环境中设置一个可移动的数据收集器进行数据收集。针对部署在地表交通困难的大规模无线传感器网络,无人机提供了一种有效的方式来对传感器设备移动式地进行数据辅助收集^[2]。

无人机是我国人工智能产业体系的重点培育产品。与静态数据收集方法相比,基于无人机的移动数据收集可以显著降低数据传输的能耗,减少多跳间数据路由中存在的隐藏终端及其发送冲突问题带来的射频干扰,并有效延长网络的使用寿命。

无人机数据收集克服了地面数据采集的局限性,但仍然有一些关键的问题需要解决。具体而言,无人机数据收集包括网络节点部署、节点定位、锚点搜索、无人机路径规划、网络数据采集5个部分^[3]。无人机最致命的缺点是续航时间短^[4-6],因此其能耗问题是系统稳定性的关键。近几年,关于无人机在数据收集的能耗研究中,无人机路径规划是一个开放性的研究课题,引起学术界的广泛关注。无人机路径规划是一个复杂的网络优化问题^[7],一般可分为全局路径规划和局部路径规划。

一般而言,无人机将在可用能量限制下,根据任务环境信息事先规划一条全局最优或次优路径,得到访问节点的访问顺序,再通过局部路径规划进行实时单个节点的搜索与逼近。近年来,无人机路径规划已经得到了广泛的研究。文献[3]把无人机路径规划问题建模成经典的旅行商问题,并执行快速路径规划(FPPWR, fast path planning with rule)。文献[8]利用马尔可夫链对单个无人机从远处传感器收集数据的移动过程进行建模,并模拟无人机运行过程中的不规则运动。文献[9]利用部分可观察的马尔可夫决策过程(POMDP, partially observable Markov decision process)对无人机路径进行规划。文献[10]基于Q学习算法对无人机的路径和避障等问题进行学习,并采用自适应随机探测的方法实现无人机的导航和避障。文献[11-14]基于深度强化学习(DRL, deep reinforcement learning)实现无模型的无人机路径规划。除了利用机器学习方法外,研究者还提出了很多启发式算法^[15-17]来解决无人机路径规划问题。

定向问题^[18]为节点选择和确定所选节点之间最短哈密顿路径的组合,可以看作背包问题和旅行

商问题2种经典问题的组合。本文将无人机数据收集过程的全局路径规划问题建模为定向问题。本文考虑的全局路径规划是指综合考虑无人机自身的能量约束、节点收益等,在指针网络深度学习架构下进行的路径规划。其中,背包问题的目标是在可用资源限制下,选择一部分节点并使之获得的收益最大化;旅行商问题的目标是试图使无人机服务所选节点的旅行时间或距离最小化。

文献[19]对定向问题最近的变化、解决方案及应用等进行了综述。近几年,关于求解定向问题的启发式方法很多,例如遗传算法^[20]、动态规划法^[21]、迭代局部搜索法^[22]等。

2015年,Vinyals等^[23]在人工智能顶级会议NIPS上提出了一个用于解决变长序列到序列的神经网络模型——指针网络,还验证了该模型可以单独使用训练示例来学习3个几何问题的近似解,即寻找平面散点集的凸包、Delaunay三角剖分算法和平面旅行商问题。指针网络深度学习被提出后,近几年被研究者多次引用,文献[24]将指针网络深度学习结合强化学习解决旅行商问题,并提出该模型也可用于解决背包问题。文献[25]使用指针网络模型结合强化学习技术来优化3D装箱序列以最大化其收益。受这些模型的启发,本文首先将无人机全局路径规划建模为定向问题,接着采用指针网络深度学习对其进行求解。

在局部路径规划方面,无人机将根据节点广播参考信号强度(RSS, received signal strength)的特征^[26]对其局部路径进行规划。文献[27]采用Q学习利用无人机对非法无线电台进行定位和寻找。然而,传统Q学习很难解决具有大量状态空间的模型,这导致其很难适用于大规模节点网络中的无人机路径规划。本文通过深度Q网络(DQN, deep-Q network)学习机制对大规模的Q表进行模拟与近似,从而极大地降低了Q学习的计算复杂度。

综上,本文首先将无人机的全局路径规划建模为定向问题并通过指针网络深度学习求解;然后在局部路径规划方面,利用DQN使无人机逼近目标节点。仿真结果表明,在无人机能耗限制下,所提方案能极大地提升物联网中的数据收集收益。

2 系统模型及问题建立

2.1 全局路径规划

本文综合考虑无人机在数据采集过程中面临

的能量约束问题和路径规划问题。无人机能量消耗不仅与航行时间、航行速度有关，还与所处环境中的风速、障碍物等有关^[28]。文献[29]将无人机的路由算法分类为恒定速度无人机、自适应速度无人机、悬停最大服务时间（HMS, hover with maximum service time）等。本文采用 HMS 的路由方法，即无人机悬停在相应节点上方，并以最大悬停时间 $t_{\max}$ 对用户进行数据传输，且假设无人机以恒定的速度 v 飞行。

在图 1 所示的系统模型中，无人机的起点和终点均为无人机服务站 D_{depot} 。 D_{depot} 可对无人机收集到的数据进行处理，并对无人机充电，需要收集数据的传感器节点随机分布在地图上，可通过聚类算法对随机分布的传感器节点进行分簇，并得到簇的中心坐标（图 1 中黑点）^[30]。关于无人机以什么样的顺序访问这些簇才能在有限的能量约束下取得最大收益的问题，可以被建模为一个定向问题，即选取点和确定最短路径 2 种问题的结合。由于无人机数据收集存在无人机能量限制，因此不是所有的簇都会被服务。令 $S \in \{1, 2, \dots, k\}$ 表示簇的集合，其中 k 表示簇的数目。第 i 个簇的奖励值为 p_i ，簇 i 到 j 的距离为 $\text{dist}_{i,j}$ ， $l_{i,j} = 1$ 表示 i 与 j 之间有路径，那么无人机全局路径规划问题可以表示为

$$\max_N \sum_{i,j \in N} l_{i,j} p_i, \min_N \sum_{i,j \in N} l_{i,j} \text{dist}_{i,j} \quad (1)$$

$$\text{s.t. } N = k\delta, \quad 0 \leq \delta \leq 1 \quad (2)$$

$$\sum_{j=1}^k l_{\text{depot},j} = \sum_{i=1}^k l_{i,\text{depot}} = 1 \quad (3)$$

$$\sum_{i=1}^{k-1} l_{i,s} = \sum_{j=1}^k l_{s,j} \leq 1, \quad \forall s \in 2, \dots, k-1 \quad (4)$$

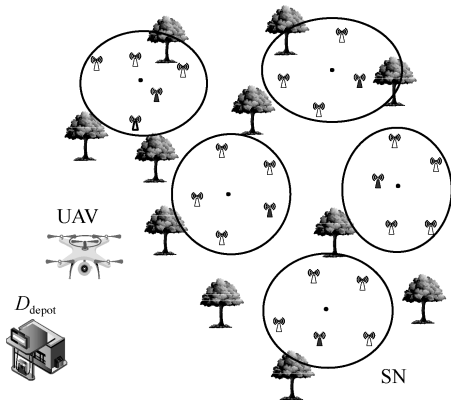


图 1 系统模型

目标方程(1)表示最大化数据收集的奖励值，并且最小化无人机的飞行路径；约束(2)表示无人机可服务簇的份额；约束(3)表示起点和终点均为无人机服务站 D_{depot} ；约束(4)表示每个簇最多被服务一次。

在优化目标式(1)中，关于每个簇奖励值的设定，如果简单设定为簇内所有节点存储数据的总和，则可能对稍远的节点不公平，所以可将每个节点的奖励值设置为

$$p_i = I_i \frac{\text{dist}_{0,i}}{\max_{j=1}^N \text{dist}_{0,j}} \quad (5)$$

其中， I_i 表示簇 i 内所有节点数据量总和的值（无量纲）。因此，奖励值的设定不仅与节点到服务站的距离有关，还与数据的存储量有关，且 p_i 无量纲^[31]。

2.2 局部路径规划

假设无人机对目标传感器的位置是未知的，无人机以固定的高度飞行，只考虑二维平面的运动，目标传感器节点的位置坐标为 (x, y) ，无人机当前状态的坐标为 (x_i, y_i) ，无人机与目标节点之间的距离可以表示为

$$d = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (6)$$

无人机通过配备天线来测量目标节点的 RSS，无人机可移动方向被相等地划分为 8 个方向，具体如图 2 所示。

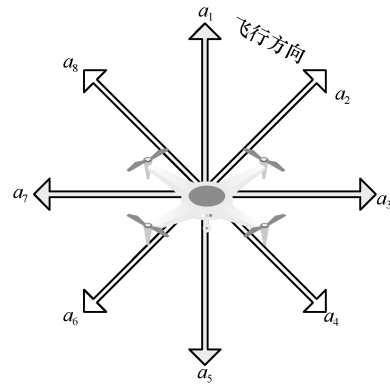


图 2 无人机移动方向

RSS 值 P_R 可以通过以下计算式求得，它与距离 d (单位为 m) 有关，具体为

$$P_R = P_T - \text{PL}(d) - f \quad (7)$$

$$\text{PL}(d) = 128.1 + 37.6 \lg \left(\frac{d}{1000} \right) \quad (8)$$

其中, P_t 为目标节点发射功率; $PL(d)$ 为距离 d 处的路径损耗, 此处的路径损耗模型^[32]采用 3GPP TR 38.814, 本文主要参考天线接收信号强度值, 为简化系统模型, 只考虑了地对地大尺度信道衰落, 后期研究无人机数据传输过程中将进一步同时考虑视距和非视距对传输性能的影响^[33-35]; f 为捕获信道衰落变量。

深度 Q 学习^[36]融合了神经网络和 Q 学习的方法, 属于强化学习的一种, 当然也应该具有强化学习的基本组成部分, 即智能体、环境、动作、奖励、策略、值函数等。强化学习智能体与环境的交互过程如图 3 所示。智能体通过与环境进行交互, 循环迭代产生新的状态并结合环境给出奖励值。

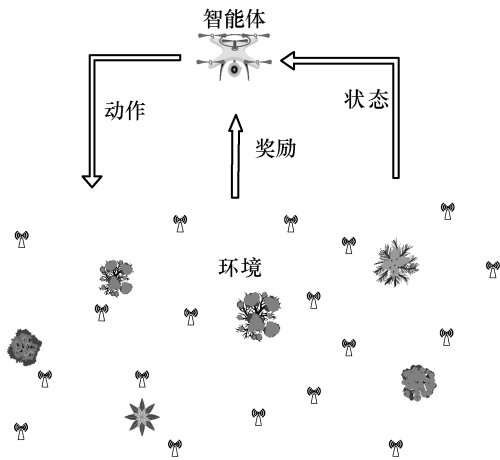


图3 强化学习智能体与环境的交互过程

Q 值函数更新式为

$$Q(s, a) = Q(s, a) + \alpha[r(s, a) + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (9)$$

其中, $\alpha \in [0, 1]$ 是学习率, $\gamma \in [0, 1]$ 是折扣因子。

无人机的状态 s 取决于各个方向测量的平均 RSS 值中最大的 RSS 值, 动作空间对应图 2 中的 8 个方向, 即 $a \in \{a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8\}$ 。奖励值 $r(s, a)$ 的设定为: 如果当前位置各个方向测量的平均 RSS 值中的最大值减去上一位置的各个方向测量的平均 RSS 值中的最大值为正, 则给一个正的奖励值, 该奖励值设为固定值; 如果为负, 则给一个负的奖励值。如果无人机达到终止条件, 则给其较大奖励值, 本文中奖励值设定为

$$r(s, a) = \begin{cases} 2, \text{RSS增加} \\ -2, \text{RSS减小} \\ 100, \text{达到终止条件} \end{cases}$$

目标方程为

$$Q_{\text{target}} = r + \gamma(\max_{a'} Q(s', a') - Q(s, a)) \quad (10)$$

无人机的终止条件为当距离 $d < 7.0$ 时, 目标节点定位成功。深度 Q 学习的原理框架如图 4 所示。

3 算法建立

3.1 指针网络深度学习建立

指针网络深度学习的结构如图 5 所示, 它是由序列到序列模型^[37]和注意力机制^[38]结合改进得到

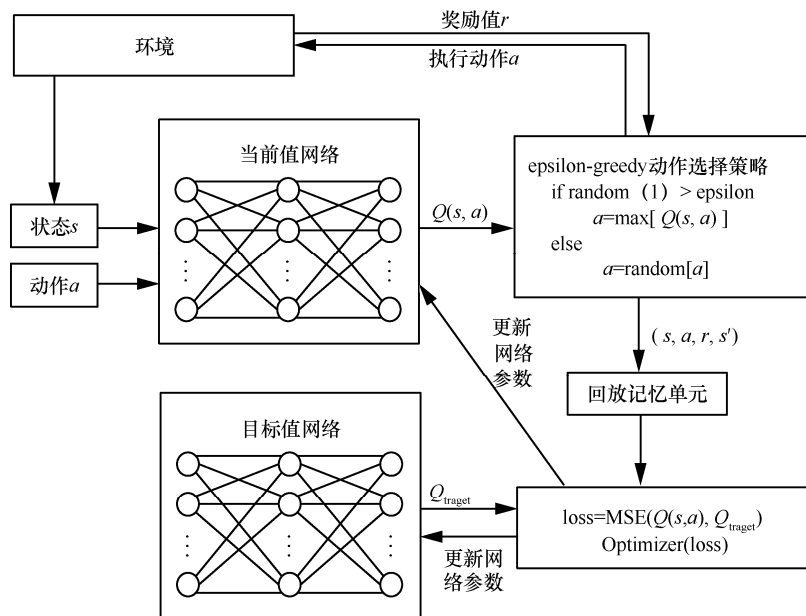


图4 深度Q学习的原理框架

的, 由 Encoder 和 Decoder 这 2 个阶段组成。在 Encoder 阶段, 只考虑输入 x_j 对输出 y_i 的影响; 在 Decoder 阶段, 解码输出注意力概率矩阵, 并通过 softmax 得到序列的输出概率分布。由于长短期记忆网络 (LSTM, long short-term memory)^[39]能够成功学习具有远距离时间依赖性数据的特征, 其被用作网络单元构建指针网络深度学习模型。Encoder 部分使用 LSTM 多层神经网络 (记为 LSTM-e), Decoder 部分使用 LSTM 多层神经网络 (记为 LSTM-d)。

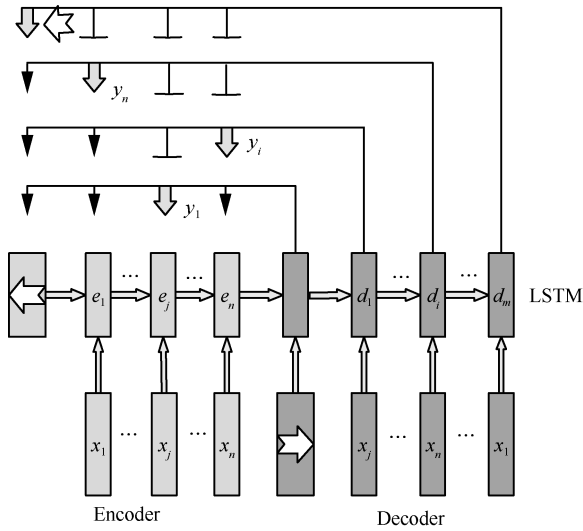


图 5 指针网络深度学习的结构

第 2 节对无人机路径规划问题进行了建模, 分别确定指针网络 (PN, pointer network) 深度学习模型的输入输出如下所示。

1) 输入

$D_{\text{coords}} = \{(x_0, y_0), (x_1, y_1), \dots, (x_k, y_k)\}$ 为无人机服务站 D_{depot} 和每个簇中心坐标 D_{loc} 的并集。假设 D_{depot} 处的奖励值 $p_0 = 0$, $P_{\text{prize}} = \{p_0, p_1, \dots, p_k\}$ 为 p_0 和 p_i 的并集, $D_{\text{inputs}} = \{(x_0, y_0, p_0), (x_1, y_1, p_1), \dots, (x_k, y_k, p_k)\}$ 。

2) 输出

输出序列 $D_{\text{roads}} = \{P_0, P_1, \dots, P_n\}$ 表示无人机数据收集过程中簇的收集顺序, P_n 对应 D_{inputs} 值的索引。指针网络深度学习的编解码过程为: 输入序列 D_{inputs} 经过 $k+1$ 步依次输入 Encoder 模块, 然后通过 Decoder 模块依次输出 D_{roads} 中的元素。

因此指针网络的原理可以表示如下^[40]。

编码过程。将输入序列 D_{inputs} 经过 $k+1$ 步依次

输送给 LSTM-e, 得到每一步输入所对应的 LSTM-e 网络状态 $e_j (j = 0, 1, \dots, k)$, 如式(11)所示。当 D_{inputs} 输入完毕后, 将得到的隐藏层状态 $\text{Enc} = (e_1, \dots, e_i, \dots, e_n)$ 进行编码后输入解码模块。

解码过程。由式(12)计算出 LSTM-d 网络的隐藏层状态 $\text{Dec} = (d_1, \dots, d_i, \dots, d_n)$; 由 LSTM-e 网络的隐藏层状态 $(e_1, \dots, e_i, \dots, e_n)$ 和 LSTM-d 网络的隐藏层状态 $\text{Dec} = (d_1, \dots, d_i, \dots, d_n)$ 分别计算出每个输入对当前输出带来的影响, 如式(13)所示。将其 softmax 归一化后得到注意力矩阵 a_i , 如式(14)~式(16)所示。然后选择矩阵中权重占比最大的指针作为输出, 如式(17)所示。

$$e_j = f((x_j, y_j, p_j), e_{j-1}) \quad (11)$$

$$d_i = f(P_{i-1}, d_{i-1}) \quad (12)$$

$$u_{ij} = \mathbf{v}^T \tanh(W_1 e_j + W_2 d_i) \quad (13)$$

$$p(P_i | P_0, \dots, P_{i-1}, D_{\text{inputs}}) = a_i \quad (14)$$

$$a_i = \{a_{i1}, a_{i2}, \dots, a_{in}\} \quad (15)$$

$$a_{ij} = \frac{\exp(u_{ij})}{\sum_{j=1}^n \exp(u_{ij})} \quad (16)$$

$$P_i = \arg \max p(P_i | P_0, \dots, P_{i-1}, D_{\text{inputs}}) \quad (17)$$

其中, f 为非线性激活函数; $\mathbf{v}$ 、 W_1 和 W_2 为输出模型的可学习参数; a_{ij} 由 u_{ij} 经过 softmax 后得到, 其作用是将 u_{ij} 标准化为输入字典上的输出分布^[23]。

在解码过程中, 还要考虑到定向问题模型中的约束问题。首先, 对于约束(2), 预设一个服务份额值 δ 。对于约束(3), 无人机的起点和终点均为无人机服务站 D_{depot} , 因此, 第一步和最后一步将 P_0 和 P_n 设置为 0。对于约束(4), 根据禁忌搜索的思想, 在每一步添加 D_{roads} 元素时, 将其作为禁忌元素添加到 D_{action} 表中, 每一步输出将根据 D_{action} 表, 在注意力矩阵中选择非 D_{action} 表中权值最大的作为输出。

3.2 指针网络+主动搜索策略

根据文献[25]提出的主动搜索 (AS, active search) 策略, 将 D_{inputs} 中的元素 (除索引 0 的位置) 随机排列组合生成 B 个批次的输入序列, 通过梯度下降法优化目标函数, 最后输出路径。

通过多目标优化中的线性加权法将目标函数(1)改写成

$$\min \sum_{i,j \in N} (\omega_1 l_{i,j} \text{dist}_{i,j} - \omega_2 l_{i,j} p_i^2) \quad (18)$$

其中, $\omega_1 + \omega_2 = 1$ 表示在减少无人机飞行距离与提升无人机服务奖励之间的折中关系, 具体可由工程经验得到。本文设置 $\omega_1 = 0.9$, $\omega_2 = 0.1$, 通过梯度下降法, 式(18)可以同时优化减小距离和增加奖励值。

指针网络+主动搜索策略的算法描述如下。

- 1) 初始化输入序列, 将 D_{inputs} 中的元素 (除索引 0 的位置) 随机排列组合生成 B 个批次的输入序列。
- 2) 将序列输入指针网络, 得到一系列结果。
- 3) 使用梯度下降法优化目标函数(18)。
- 4) 重复执行步骤 1)~步骤 3), 直到达到终止条件。
- 5) 选择最小的目标函数值的路径作为输出路径 D_{roads} 。

3.3 DQN 算法建立

DQN 算法描述如下。

- 1) 初始化经验重放缓存区。
- 2) 预处理环境: 把状态-动作输入 DQN, 返还所有可能动作对应的 Q 值。
- 3) 利用 ε 贪心策略选取一个动作 a , 以概率 ε 随机选择动作, 以概率 $1-\varepsilon$ 选取具有最大 Q 值的动作。
- 4) 选择动作 a 后, 智能体在状态 s 执行所选的动作, 得到新的状态 s' 和奖励 r 。
- 5) 把该组数据存储在经验重放缓冲区中, 并将其记作 s, a, r, s' 。
- 6) 计算目标方程(10), 更新 Q 网络权重。
- 7) 重复执行步骤 3)~步骤 6), 直到达到终止条件。

4 实验结果与分析

4.1 参数设置

为验证指针网络模型对无人机全局路径规划的优化性能, 实验主要对比了指针网络深度学习、基于主动搜索策略的指针网络深度学习方法。为了对比 AS 方法的效果, 本文设计贪婪奖励 (GP, greed prize) 方法与其进行比较。GP 方法受贪婪优化方法的影响, 先贪婪地选择奖励值大小为前

N 簇的坐标, 然后通过 PN+AS 方法求这些簇的最短路径。

实验令无人机服务站 D_{depot} 的坐标为 (0,0), 在 $[0,1] \times [0,1]$ (单位为 km) 的范围内分别随机生成 50 个簇和 100 个簇的中心位置坐标, 分别为 D_{50} 和 D_{100} 。每个簇的奖励值设定按照式(5)得到。

表 1 和表 2 分别给出了 AS 方法和 GP 方法使用的参数及其相应值。

表 1 AS 方法使用的参数及其相应值

参数	含义	值
H_s	PN 的隐藏层数目	256
E_s	PN 的 embedding 层数目	256
B	批次	256
T	迭代次数	100
γ	学习率	0.02
α	衰减系数	0.9
δ	簇总数份额	0.5

表 2 GP 方法使用的参数及其相应值

参数	含义	值
H_s	PN 的隐藏层数目	128
E_s	PN 的 embedding 层数目	128
B	批次	128
T	迭代次数	100
γ	学习率	0.02
α	衰减系数	0.9
δ	簇总数份额	0.5

为验证 DQN 的性能, 本文主要分析 Q 学习和 DQN 这 2 种方法在无人机数据集中单个节点定位的仿真效果。为模拟无人机接收信号强度值, 采用网格法确定当前位置距离目标节点的距离, 通过式(7)和式(8)计算接收信号强度值, 实现 DQN 状态输入。本文主要对 2 种方法迭代次数内的成功率、步数及其最优路径进行比较。2 种方法均使用 ε 贪心策略, 在仿真中将无人机到目标节点的距离小于 7 m 视为成功, 为防止算法无限次迭代, 将无人机步数大于 200 步视为失败。仿真结果表明, DQN 的性能优于 Q 学习, 能够达到一个较高的成功率。DQN 仿真各参数的设置如表 3 所示。

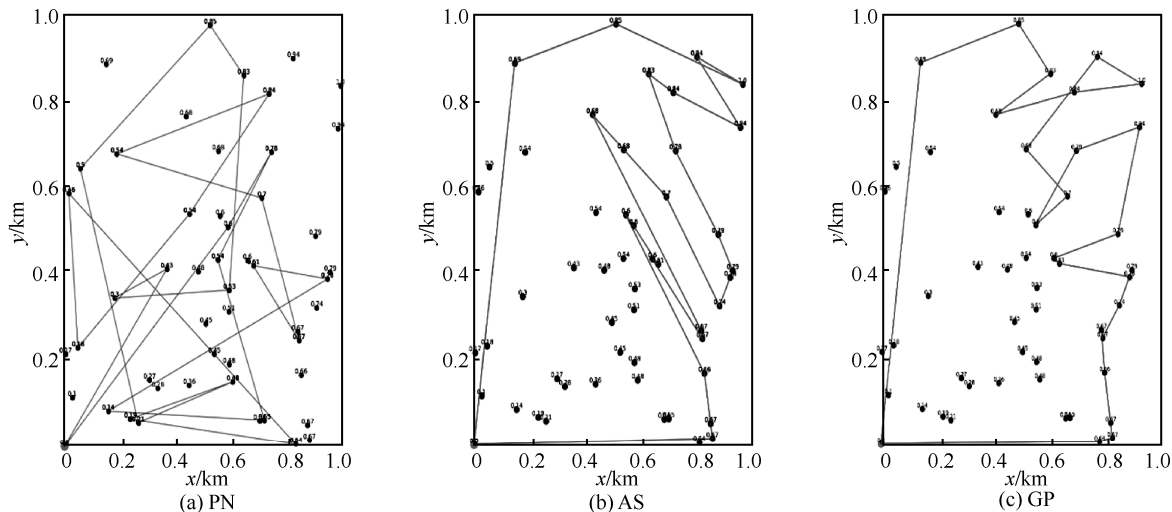
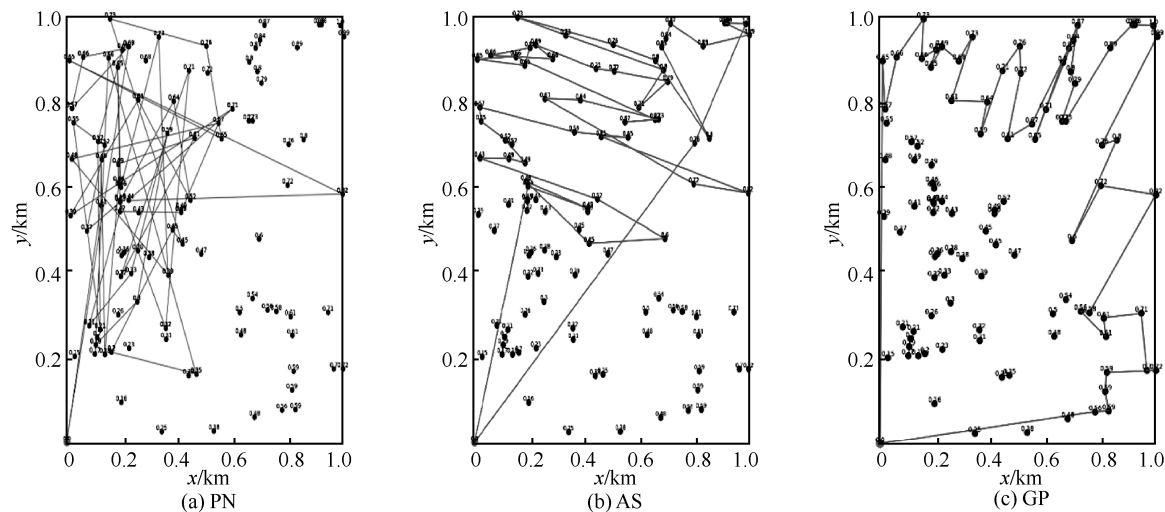
表3 DQN 仿真各参数的设置

仿真参数	取值
无人机初始位置	(20,20)
目标节点位置	(183,265)
步长/m	8
探索速率 ϵ	0.1
传感器发射功率 P_t /dBm	13
迭代次数 T	100
更新周期 t	30
记忆重放批次	400
折扣因子 γ	0.9

4.2 结果分析

图6和图7分别是 D_{50} 和 D_{100} 下使用 PN、AS

和 GP 的路径规划效果,表4是 D_{50} 和 D_{100} 下使用 PN、AS 和 GP 的距离和奖励值,其中,距离的单位为 km。根据式(9)可知,奖励值越大越好,距离越小越好,这样可以使模型的收益能效更高。从图6、图7和表4中可以直观地看出,使用 PN 方法比 AS 方法的路径规划图交叉点多,总路径距离大,总奖励值小;将 GP 方法与 AS 方法进行比较,GP 方法交叉点少,总路径距离小,不过 GP 方法存在贪婪的性质,相当于将全局路径规划问题变为一个简单的旅行商问题来解决,使该算法获得的奖励值更好,但在某些场景下有可能导致更大的飞行距离。虽然 AS 方法不能完全达到 GP 方法的效果,但 AS 方法的效果接近 GP 方法,且 AS 方法最大的特点就是同时优化距离和奖励值,虽然奖励值可

图6 D_{50} 下使用 PN、AS 和 GP 的路径规划效果图7 D_{100} 下使用 PN、AS 和 GP 的路径规划效果

能不如 GP 方法但其距离有可能更小，且更具随机性、更适合动态环境中的无人机路径规划。

表 4 D_{50} 和 D_{100} 下使用 PN、AS 和 GP 的距离和奖励值

任务	PN		AS		GP	
	距离/km	奖励	距离/km	奖励	距离/km	奖励
D_{50}	11.71	12.89	6.45	17.78	5.71	17.82
D_{100}	24.56	19.32	10.87	33.68	7.08	35.07

图 8 是 AS 方法下使用梯度下降法训练 PN 的损失值。从图 8 中可以看出，训练 PN 的损失值随着迭代次数的增加先快速下降，而后趋于稳定，在 0 值上下波动，这表明该深度模型可以在训练后达到收敛，网络性能可靠。

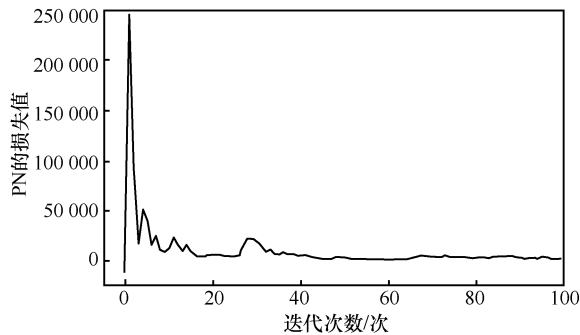


图 8 AS 方法下使用梯度下降法训练 PN 的损失值

图 9 为 DQN 和 Q 学习的成功次数的步数变化波动曲线。从图 9 中可以明显看出，DQN 步数的变化只在一开始波动较大，经过一个更新周期（30 次）后波动趋于平稳，且步数较小，迭代次数为 100 次，成功率接近 100%；Q 学习的成功次数只有 7 次，其余均大于 200 步。

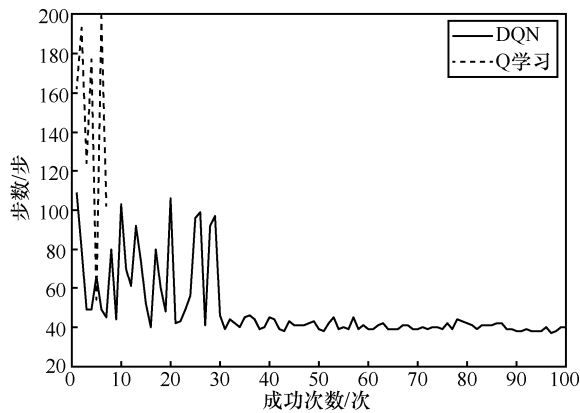


图 9 DQN 和 Q 学习成功次数的步数变化波动曲线

从图 10 可以更清晰地看到，DQN 的最优路径与 Q 学习的最优路径相比更平缓、拐点较少。表 5 为不

同起点和目标位置时 Q 学习和 DQN 的最优步长比较。从表 5 可以看出，针对不同起点和目标位置，除了第三组 $(0,0) \rightarrow (68,78)$ 这 2 种方法效果一样外，其余场景中的 DQN 都优于 Q 学习，可见 DQN 的泛化性能强，可以适应不同的场景。

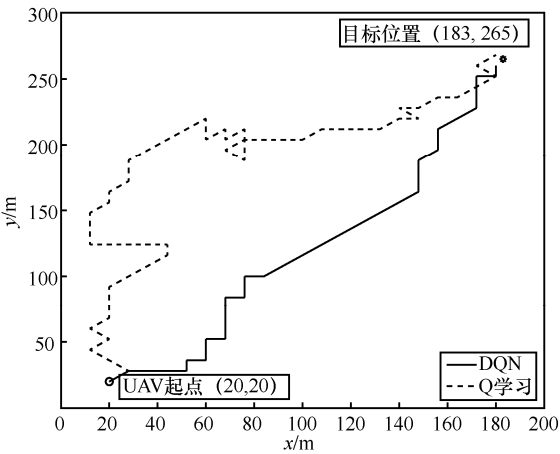


图 10 DQN 和 Q 学习最优路径对比

表 5 不同起点和目标位置时 DQN 和 Q 学习的最优步长比较

UAV 起点→目标节点坐标	Q 学习 最优步长/m	DQN 最优步长/m
$(20,20) \rightarrow (183,265)$	56	39
$(10,10) \rightarrow (388,378)$	179	64
$(0,0) \rightarrow (68,78)$	12	12

5 结束语

本文首先使用指针网络深度学习来解决无人机数据收集过程中的全局路径规划问题，并将该问题建模成定向问题，利用指针网络深度学习得到无人机服务节点集合及服务顺序。然后，根据无人机接收目标节点的 RSS 通过 DQN 来定位目标节点并接近目标节点，经仿真验证，DQN 在时延等方面的性能优于 Q 学习。最后，通过仿真验证了所提学习机制的有效性。

参考文献：

[1] LI C, ZHANG H, HAO B, et al. A survey on routing protocols for large-scale wireless sensor networks[J]. Sensors, 2011, 11(12): 3498-3526.

[2] ZENG Y, ZHANG R, LIM T J. Wireless communications with unmanned aerial vehicles: opportunities and challenges[J]. IEEE Communications Magazine, 2016, 54(5): 36-42.

[3] WANG C, MA F, YAN J, et al. Efficient aerial data collection with UAV in large-scale wireless sensor networks[J]. International Journal

- of Distributed Sensor Networks, 2015, 2015: 1-19.
- [4] UTKARSHA S. P., RAJIV K. G. Clustering and compressive data gathering in wireless sensor network[J]. *Wireless Personal Communications*, 2019, 109(2): 1-21.
 - [5] JIE G., TSUNG-HUI C., CHAO S., et al. Flight time minimization of UAV for data collection over wireless sensor networks[J]. *IEEE Journal on Selected Areas in Communications*, 2018, PP(99): 1.
 - [6] ZAIN A. A., SUHAIB M., MUHAMMAD A. UAV based data gathering in wireless sensor networks[J]. *Wireless Personal Communications*, 2019, 106(4): 1-11.
 - [7] DUAN H., LI P. UAV path planning[M]. Berlin: Springer, 2014.
 - [8] ARVANITAKI A., PAPPAS N. Modeling of a UAV-based data collection system[C]//2017 IEEE 22nd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks. Piscataway: IEEE Press, 2017: 1-6.
 - [9] RAGI S., CHONG E. K. P. UAV path planning in a dynamic environment via partially observable markov decision process[J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2013, 49(4): 2397-2412.
 - [10] ZHAO Y. J., ZHENG Z., ZHANG X. Y., et al. Q learning algorithm based UAV path learning and obstacle avoidance approach[C]//2017 36th Chinese Control Conference. Piscataway: IEEE Press, 2017: 3397-3402.
 - [11] HUANG H., YANG Y., WANG H., et al. Deep reinforcement learning for UAV navigation through massive MIMO technique[J]. *IEEE Transactions on Vehicular Technology*, 2020, 69(1): 1117-1121.
 - [12] WANG C., WANG J., SHEN Y., et al. Autonomous navigation of UAVs in large-scale complex environments: a deep reinforcement learning approach[J]. *IEEE Transactions on Vehicular Technology*, 2019, PP(99): 1.
 - [13] ZHANG B., LIU C. H., TANG J., et al. Learning-based energy-efficient data collection by unmanned vehicles in smart cities[J]. *IEEE Transactions on Industrial Informatics*, 2017, PP(99): 1.
 - [14] LIU C. H., CHEN Z., TANG J., et al. Energy-efficient UAV control for effective and fair communication coverage: a deep reinforcement learning approach[J]. *IEEE Journal on Selected Areas in Communications*, 2018, 36(9): 2059-2070.
 - [15] NA L., SI-MING H., CHANG-QING G. UAV path planning based on adaptive weighted pigeon-inspired optimization algorithm[J]. *Computer Simulation*, 2018, 35(1): 38-42, 125.
 - [16] ARANTES J. D. S., MARCIO D. S. A., TOLEDO C. F. M., et al. Heuristic and genetic algorithm approaches for UAV path planning under critical situation[J]. *International Journal of Artificial Intelligence Tools*, 2017, 26(1): 1760008.
 - [17] 刘丽珏, 罗舒宁, 高琰, 等. 基于回溯蚁群-粒子群混合算法的多点路径规划[J]. *通信学报*, 2019, 40(2): 102-110.
 - [18] LIU L. J., LUO S. N., GAO Y., et al. Multi-point path planning based on the algorithm of colony-particle swarm optimization[J]. *Journal on Communications*, 2019, 40(2): 102-110.
 - [19] VANSTEENWEGEN P., SOUFFRIAU W., OUDHEUSDEN D. V. The orienteering problem: a survey[J]. *European Journal of Operational Research*, 2011, 209(1): 1-10.
 - [20] GUNAWAN A., LAU H. C., VANSTEENWEGEN P. Orienteering problem: a survey of recent variants, solution approaches and applications[J]. *European Journal of Operational Research*, 2016, 255(2): 315-332.
 - [21] ABBASPOUR R. A., SAMADZADEGAN F. Time-dependent personal tour planning and scheduling in metropolises[J]. *Expert Systems with Applications*, 2011, 38(10): 12439-12452.
 - [22] LI J., WU Q. M., LI X. Q., et al. Study on the time-dependent orienteering problem[C]//International Conference on E-product E-service & E-entertainment. Piscataway: IEEE Press, 2010: 1-4.
 - [23] BRAHIM A., HASSAN E. F. Iterated local search algorithm for solving the orienteering problem with soft time windows[J]. *SpringerPlus*, 2016, 5(1): 1-36.
 - [24] VINYALS O., FORTUNATO M., JAITLY N. Pointer networks[C]//International Conference on Neural Information Processing Systems. Massachusetts: MIT Press, 2015: 2692-2700.
 - [25] BELLO I., PHAM H., LE Q. V., et al. Neural combinatorial optimization with reinforcement learning[J]. *arXiv Preprint*, arXiv:1611.09940v1, 2016.
 - [26] HU H., ZHANG X., YAN X., et al. Solving a new 3D bin packing problem with deep reinforcement learning method[J]. *arXiv Preprint*, arXiv:1708.05930, 2017.
 - [27] 王天荆, 李秀琴, 白光伟, 等. 无线传感器网络中基于自适应网格的多目标定位算法[J]. *通信学报*, 2019, 40(7): 197-207.
 - [28] WANG T. J., LI X. Q., BAI G. W., et al. Multi-target localization algorithm based on adaptive grid in wireless sensor network[J]. *Journal on Communications*, 2019, 40(7): 197-207.
 - [29] WU S. Illegal radio station localization with UAV-based Q-learning[J]. *China Communications*, 2018, 15(12): 122-131.
 - [30] FARAJZADEH A., ERCETIN O. UAV data collection over NOMA backscatter networks: UAV altitude and trajectory optimization[J]. *arXiv Preprint*, arXiv:1902.03061, 2019.
 - [31] JAWHAR I., MOHAMED N., AL-JAROUDI J. UAV-based data communication in wireless sensor networks: models and strategies[C]//International Conference on Unmanned Aircraft Systems. Piscataway: IEEE Press, 2015: 687-694.
 - [32] 王继红, 石文孝. 认知无线传感器网络分簇路由协议综述[J]. *通信学报*, 2018, 39(11): 156-169.
 - [33] WANG J. H., SHI W. X. Survey on cluster-based routing protocols for cognitive radio sensor networks[J]. *Journal on Communications*, 2018, 39(11): 156-169.
 - [34] KOOL W., VAN H. H., WELLING M. Attention, learn to solve routing problems[J]. *arXiv Preprint*, arXiv:1803.08475v3, 2018.
 - [35] 3GPP. Further advancements for E-UTRA physical layer aspects[S]. TR 36.814 (V9.0.0), 2010.
 - [36] PAN Y., DA X. Y., HU H., et al. Energy-efficiency optimization of UAV-based cognitive radio system[J]. *IEEE Access*, 2019, 7: 155381-155391.
 - [37] YAO Y., ZHU Z., HUANG S., et al. Energy efficiency characterization in heterogeneous IoT system with UAV swarms based on wireless power transfer[J]. *IEEE Access*, 2020, 8: 967-979.
 - [38] ZHU Z., WANG N., HAO W., et al. Robust beamforming designs in secure MIMO SWIPT IoT networks with a non-linear channel model[J]. *IEEE Internet of Things Journal*, 2020, PP(99): 1.
 - [39] VINCENT F. O., RAPHAEL F., DAMIEN E. Playing atari with deep reinforcement learning[J]. *arXiv Preprint*, arXiv: 1312.5602, 2013.
 - [40] SUTSKEVER I., VINYALS O., LE Q. V. Sequence to Sequence Learning with Neural Networks[J]. *Advances in Neural Information Processing Systems*. Massachusetts: MIT Press, 2014: 3104-3112.
 - [41] VINYALS O., KAISER L., KOO T., et al. Grammar as a foreign language[J]. *arXiv Preprint*, arXiv: 1412.7449, 2014.

[39] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.

[40] 曾广贤. 基于神经网络学习方法的单机调度问题研究[D]. 厦门: 厦门大学, 2018.

ZENG G X. Research on single machine scheduling problem based on neural network learning method[D]. Xiamen: Xiamen University, 2018.

[作者简介]



付澍(1985-), 男, 贵州贵阳人, 博士, 重庆大学副教授、硕士生导师, 主要研究方向为星地通信、NOMA、物联网、网络一体化等。



陈晨(1988-), 男, 江苏淮安人, 博士, 重庆大学研究员、博士生导师, 主要研究方向为面向 6G 和物联网的无线光通信技术。



喻鹏(1986-), 男, 湖北随州人, 博士, 北京邮电大学副教授、博士生导师, 主要研究方向为 5G/6G 网络智能管控。



杨祥月(1995-), 女, 贵州毕节人, 重庆大学硕士生, 主要研究方向为机器学习、无人机数据收集、路径规划等。



简鑫(1987-), 男, 四川自贡人, 博士, 重庆大学副教授、硕士生导师, 主要研究方向为应用统计学、物联网无线接入技术及应用、分布式算法等。



张海君(1986-), 男, 辽宁朝阳人, 博士, 北京科技大学教授、博士生导师, 主要研究方向为 6G 移动通信、人工智能与无线网络等。



刘敏(1975-), 女, 四川大竹人, 重庆大学教授、博士生导师, 主要研究方向为光通信与光网络。