基于深度增强学习的无人机赋能雾无线电接入网络的能效优化

梅海波，杨鲲，范新宇
（电子科技大学，四川 成都 610054）

摘　要：雾无线电接入网络适合用于广域范围内的诸如管线管网监测等国家重要行业的物联网应用场景。然而基于地面雾接入节点的网络将受到环境、地形等影响，无法及时有效地提供雾接入服务。利用低空无人机作为雾接入点实现空地的边缘通信和雾计算方面引起了普遍的关注。本文探讨怎样利用深度增强学习来提高无人机雾接入点的能效，延长无人机的任务时间。深度增强学习可以保障无人机雾接入点及时地调整空地通信和计算的配置策略，包括资源优化、动态任务卸载以及缓存，也可以优化无人机在三维空间中的飞行航迹，提高无人机赋能的雾无线电接入网络的总体性能。研究的创新性在于综合论述了深度增强学习用于无人机赋能的雾无线电接入网络要解决的主要优化问题，并且总结了解决相关优化问题的技术细节，最后对深度增强学习应用于无人机赋能的雾无线电接入网络的技术挑战和未来研究方向展开讨论。

关键词：无人机；雾无线电接入网络；深度增强学习；航迹规划；网络配置
中图分类号：TP393
文献标识码：A
doi: 10.11959/j.issn.2096−3750.2021.00234

Deep reinforcement learning to enhance the energy-efficient performance of UAV-enabled F-RAN

MEI Haibo, YANG Kun, FAN Xinyu
University of Electronic Science and Technology of China, Chengdu 610054, China

Abstract: Fog radio access network (F-RAN) is suitable for Internet of things applications of national important industries, such as pipeline network monitoring in wide area. However, the performance of the F-RAN based on the territorial fog access point will be affected greatly by the complicated territorial environment. This causes F-RAN not able to provide fog access service in a timely and effectively manner. To this problem, the research was proposed to utilize low altitude UAV as the fog access point to realize air ground edge communication and fog computing, which has attracted enormous research interests. How to use deep reinforcement learning (DRL) to improve the energy efficiency of UAV fog access point and extend the mission time of UAV were discussed. Deep reinforcement learning can ensure the UAV fog access point to adjust the configuration strategy timely of air ground communication and computing, including resource optimization, dynamic task offloading and caching. DRL can also optimize the UAV trajectory in 3-D space, and improve the overall performance of UAV enabled fog access network. The innovation of the research lies in the comprehensive discussion of the main optimization problems to be solved in the UAV-enabled F-RAN using DRL. The technical details were also summarized to solve the related optimization problems. Finally, the technical challenges and future research directions of the application of DRL in the UAV-enabled F-RAN were discussed.

Key words: unmanned aerial vehicle, fog radio access network, deep reinforcement learning, trajectory design, network configuration

收稿日期：2021−01−11；**修回日期：**2021−04−29
通信作者：梅海波，haibo.mei@uestc.edu.cn
基金项目：国家自然科学基金资助项目（No.61620106011，No.U1705263，No.61871076）
Foundation Items: The National Natural Science Foundation of China (No.61620106011, No.U1705263, No.61871076)

1　引言

雾无线电接入网络（F-RAN, fog radio access network）首次于 2014 年 6 月在下一代移动网络论坛上被提出。在雾无线电接入网络中，着重讨论利用边缘设备，如移动边缘服务器、设备到设备（D2D）通信的用户终端，构架分布式的通信和计算框架。这些边缘设备都可以作为雾接入点（F-AP）为周围本地设备（如物联网终端）提供通信和计算服务。具体来讲，在雾无线电接入网络中，本地设备可以卸载通信、计算、存储以及网络功能到附近的雾接入点来拓展自身的处理能力，并且不依靠中心云，通过协作的方式来处理通信和计算任务。雾无线电接入网络可以节约后传网络的数据传输量，因此比中心网络更能减少任务处理过程中的通信代价。雾无线电接入网络逐渐吸引了各界的普遍关注，其在物联网及低时延要求的垂直行业中的应用越来越多[1]。

针对传统雾无线电接入网络在偏远地区、环境恶劣地区等有挑战性的环境里的覆盖问题，无人机可以及时且方便地部署于网络中充当雾接入点，以适用很多场景应用[2-3]。例如，我国幅员辽阔，地质环境复杂，特别是四川地区是我国地质灾害的多发地区，5·12 汶川地震以来四川西部的地震灾区地质灾害的发生频率呈上升趋势。余震、山体滑坡等自然灾害时刻影响着四川经济生产的各类基础设施尤其是山区的移动通信网络设施的正常工作。针对此情况，无人机赋能的雾无线电接入网络可以满足以上灾害预警、应急通信保障等需求。例如，四川的地质灾害预警机构（如地震局等）所能提供的安全预警机制不能完全满足主要四川工农业及生活设施对灾害预警的切实需求，很多信息和数据存在滞后。因此有必要利用无人机赋能的雾无线电接入网络，建立独立的地质灾害监测预警系统，为各行业的生产、运输和经营管理提供切实可靠的信息保障。无人机赋能的雾无线电接入网络可以大幅度提升地区预警系统的可靠性、精准性和快速反应能力，最大限度地减少自然灾害带来的生命和财产损失。

无人机雾接入点主要在空地视距传输以及三维空间的可控高移动性两方面有明显的优势。一方面，无人机处于空中相对高的位置，可以和地面设备大概率地建立视距无线连接。因而，与地面间通信比较，无人机到地面终端的通信较少受到阴影衰减的影响，地面终端可以通过视距连接很方便地卸载任务数据到无人机雾接入节点上。另一方面，由于无人机三维空间高移动性的支持，无人机雾接入节点可以根据地面终端位置，实时地调整自身的位置和航迹，保持和地面终端良好的视距连接[4]。无人机赋能的雾无线电接入网络如图 1 所示，总的来说，无人机雾接入点可以很好地实现空中雾计算平台，和中心云协作一起形成云雾结合的接入网络[5]。

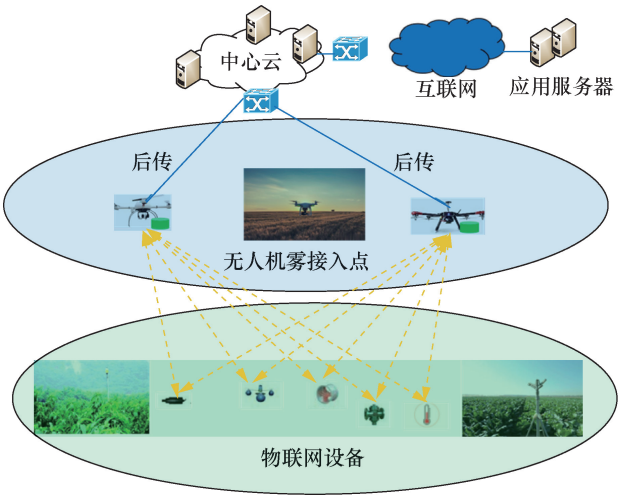
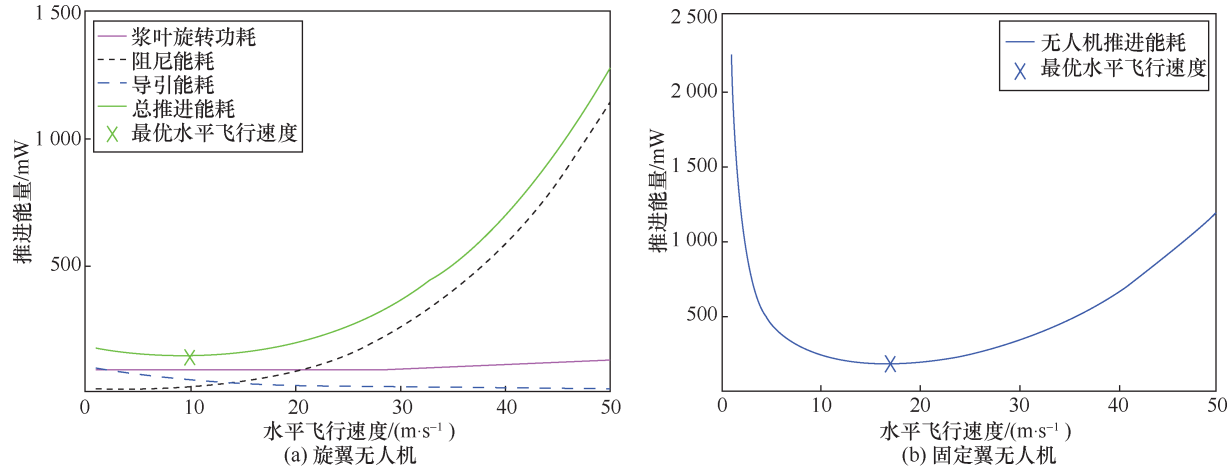


图 1　无人机赋能的雾无线电接入网络

尽管无人机赋能雾无线电接入网络有诸多优势，但无人机能量限制将严重阻碍其应用，保障无人机的高能效成为一个待解决的问题。文献[6-7]提出无人机的大部分能耗是飞行能耗，其难点问题是设计出能效最大化的无人机三维飞行航迹，以保障雾无线电接入网络的通信和计算服务质量并延长无人机的任务时间。另外，为支持雾无线电接入网络中的地面终端以较小的代价卸载任务到无人机上，地面终端倾向于无人机靠近它们并且在其附近尽量长时间悬停。这将给无人机带来更大的任务压力并快速消耗完无人机的能量。此外，固定翼无人机的能耗问题将更严峻，因为固定翼无人机无法悬停，其需要不停地保持滞空前行。无人机的能耗问题在复杂的任务环境里将变得更明显。在复杂任务环境里，无人机需要频繁地进行上升/下降以及避障飞行，同时需要付出更多的努力来调整位置，以便和地面终端实现视距传输[7]。以上情况都会给无人机的能耗带来更大的压力。

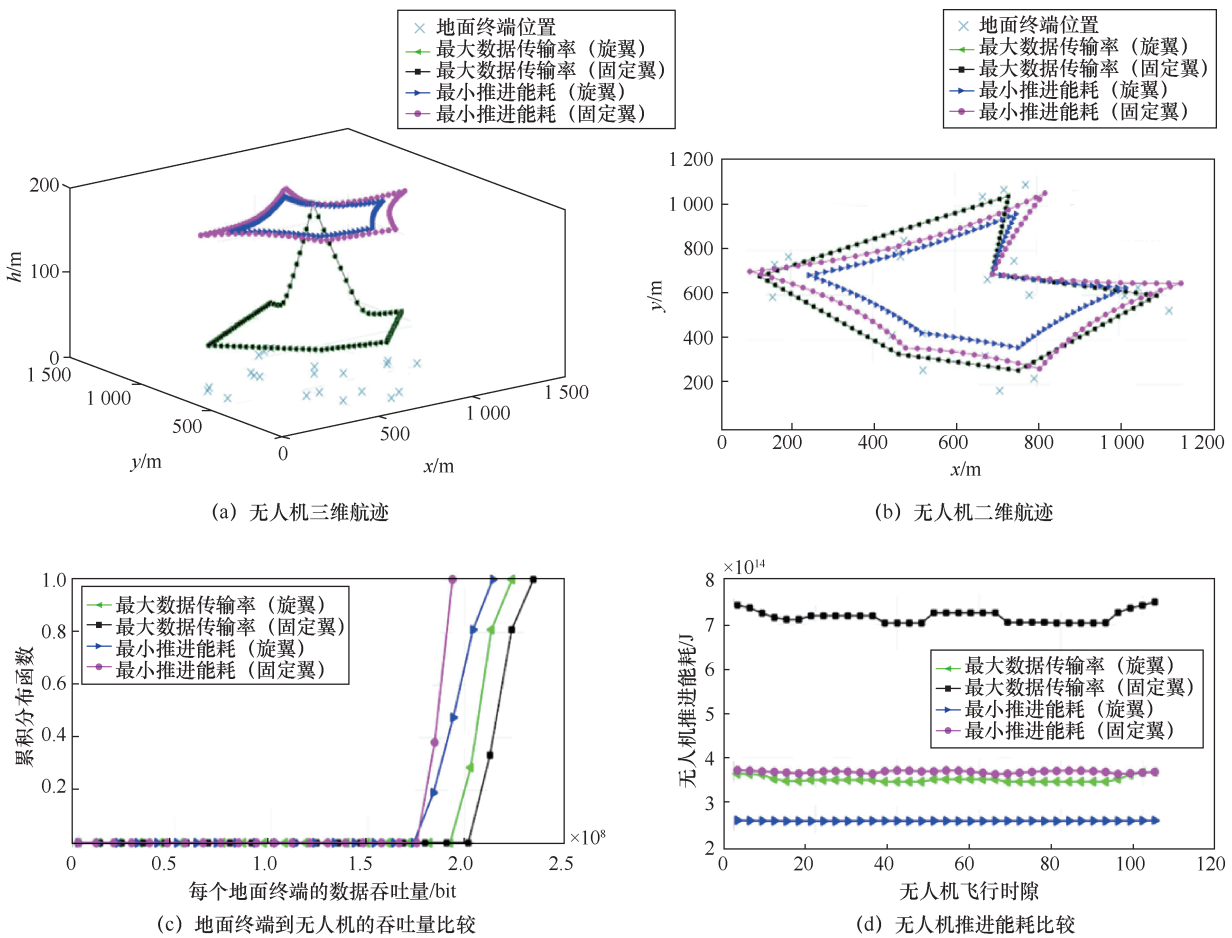
从无人机本身的特点来讲，无人机在水平前进过程中倾向于保持一个最优的水平推进速度，以尽量节约推进能量。无人机推进能耗和水平飞行速度的关系如图 2 所示，无人机最优的水平推进速度决定于无人机的特点。对于旋翼无人机，其水平推进

能量由桨叶旋转功耗、阻尼能耗、导引能耗组成，分别由桨叶的叶尖速度、悬停时的平均旋翼速度、机身阻力比和旋翼固实度决定。与旋翼无人机相比较，固定翼无人机的推进能量组成相对简单，只决定于无人机的重量和机翼面积。另外，旋翼无人机和固定翼无人机都倾向于水平飞行，以减少上升/下降行为造成的垂直维度的能量消耗。图 2 的具体技术细节和讨论参考文献[6-7]。无人机和地面终端连接的数据吞吐量和无人机推进能量消耗之间的关系如图 3 所示，图 3 计算了固定翼无人机和旋翼无人机的飞行推进能耗以及其作为空中接入节点服务地面终端的数据传输率和吞吐量，具体的仿真



(a) 旋翼无人机　(b) 固定翼无人机

图 2　无人机推进能耗和水平飞行速度的关系



（a）无人机三维航迹　（b）无人机二维航迹

（c）地面终端到无人机的吞吐量比较　（d）无人机推进能耗比较

图 3　无人机和地面终端连接的数据吞吐量和无人机推进能量消耗之间的关系

设置参考文献[8]。假设无人机尽量节约其飞行推进能量，图 3 中无人机将处于水平飞行状态，不会故意降低飞行高度以接近地面终端，这样地面终端的数据传输率和吞吐量将比较低。相反，无人机考虑通信的数据传输优先的情况下，固定翼无人机和旋翼无人机将尽量保障和地面终端保持较近的水平距离，同时高度也接近地面终端。这样地面终端的数据传输率和吞吐量将得以保障，代价就是无人机将消耗更多的飞行能耗。如果地面终端想取得较高的数据传输吞吐量，其需要和无人机建立长时间的通信连接，导致很高的无人机推进能量的消耗。相反，如果无人机采取最节能的水平飞行速度，并避免垂直高度变化，这将造成地面终端和其通信的数据吞吐量过低的问题。图 3 中的地面终端的数据吞吐量是以累积分布函数（CDF, cumulative distribution function）值来衡量的。当地面终端的数据吞吐量大时，数据吞吐量累积分布函数的曲线分布在 x 值较大的部分（x 轴偏右）；相反，地面终端的数据吞吐量小时，数据吞吐量积累分布式值的曲线分布在 x 值较小的部分（x 轴偏左）。总体来说，无人机本身的特点将直接影响其作为雾接入节点服务地面终端的效果。因此，怎样平衡无人机雾计算服务质量保证和航迹设计间的关系，这是一个复杂的问题。

针对以上无人机的能耗、空地通信和计算的制约关系，本文旨在论证利用深度增强学习[9]来解决以上难题的可能性。本文利用深度增强学习方法，通过资源优化、动态任务卸载以及缓存设置，提高无人机雾接入点的能效。同时探讨无人机在三维空间的航迹规划的方法，提高无人机的工作能效。最后对以上思路的技术挑战和研究方向展开讨论。

2　深度增强学习应用于无人机赋能的雾无线电接入网络

针对无线通信网络，人工智能算法可以挖掘用户服务特性，预测不同类型用户移动性和需求，自适应地调整网络设置，提高网络的性能[10]。在各种各样的人工智能算法中，人工神经网络已经应用到了无人机通信网络中。人工神经网络实现了无人机通信网络用户关联、资源分配、缓存设置及移动模式预测等优化功能[11-12]。然而，人工神经网络需要依靠大量历史数据训练模型，这阻碍了其在无人机通信网络中的广泛应用。与人工神经网络相比较，深度增强学习通过代理和环境实时交互，采取最优的行动获得最大的奖励值，解决了以上训练问题[9]。对比传统的采用离线训练模式的人工智能算法，深度增强学习不需要外部的经验训练样本。相反，深度增强学习依靠自身的实时经验，从不确定的环境中快速找到最优的行动方案。因此，深度增强学习特别适用于无人机赋能的雾无线电接入网络，其克服了无人机移动性带来的无人机与地面终端的连接高动态变化所引起的一系列问题。本文探索利用深度增强学习实现在线的无人机赋能的雾无线电接入网络的网络配置及无人机三维空间的航迹规划，平衡雾无线电接入网络的能耗和服务的关系，整体提升无人机雾无线电接入网络的工作性能。

2.1　利用深度增强学习实现无人机赋能的雾无线电接入网络的网络配置

本节讨论利用深度增强学习实现资源分配、动态任务卸载及缓存设置等网络配置，提高无人机赋能的雾无线电接入网络的空地通信和计算的性能。

2.1.1　无人机雾接入点的资源分配

无人机雾接入点需要利用有限的频谱资源及计算资源，实现高资源利用率的通信与雾计算服务。针对频谱资源，无人机雾接入节点一般可以支持 4G/5G、IEEE 802.11 或者 IEEE 802.15 等无线通信协议。以上协议利用正交频分技术将带宽划分为固定数量的子信道。子信道资源再根据相应机制分配给无人机和地面终端的通信链路，保证较高的数据传输率。同时，以上通信资源分配还需要控制各子信道的传输能量。在一个时间段内，无人机的雾节点能够提供给子信道的总下行传输能量是恒定的。因此，无人机雾接入节点需要采取能量分配方案，将有限的传输能量合理分配给各子信道，并满足一定的数据传输率要求。另外，传输能量分配还需要考虑减少同频干扰。假设一个无人机到地面终端的链路的子信道对其他链路造成了同频干扰，那么需要降低造成同频干扰的子信道的传输能量。针对计算资源，无人机雾无线电接入网络需要分配其有限的计算资源（计算容量）给地面终端执行通信和雾计算任务。与通信资源类似，无人机雾接入点将计算资源划分为数量有限的计算资源块，并利用一定的策略将计算资源块分配给配对的地面终端。总的来说，通信资源和计算资源共同决定了地面终端任务的时延。这是因为地面终端任务的时延是由无人机和地面终端的数据传输时间和无人机雾接

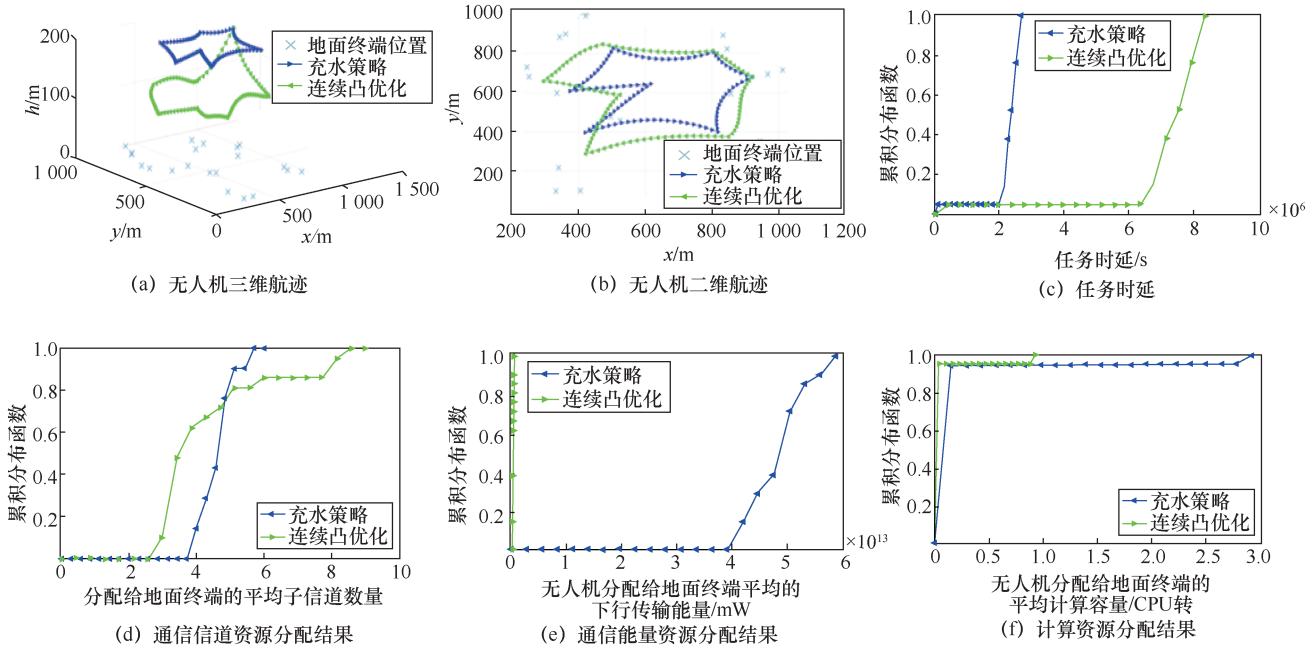
入节点计算任务的时间共同构成的。

为验证以上现象，进行仿真实验。资源分配及无人机航迹导致的地面终端任务的时延结果如图 4 所示。在图 4 中，通信和计算资源通过充水策略（Water-filling）或连续凸优化的方法[13]分配给无人机和地面终端的连接。同时，无人机航迹规划利用旅行推销员问题（TSP, traveling salesman problem）和连续凸优化的方法，使无人机以较短的飞行距离遍历地面终端的所有位置。无人机在三维空间的航迹旨在让无人机尽全力服务于地面终端，以降低地面终端的任务时延。图 4 中的 TSP 和连续凸优化方法的具体实现以及仿真参数设置参考文献[14]。图 4 中的充水策略目的在于给无人机和地面终端的连接尽量分配多的通信和计算资源，不考虑公平性和资源有效性。相反，连续凸优化的方法旨在实现最高的资源有效利用率，因此优于充水策略，能保证更好的无人机赋能的雾无线电接入网络的性能。在图 4 中，即使充分策略结合贪婪的资源分配方法可以实现更低的任务时延，然而从通信、传输能量和计算资源分配的情况来看，更先进的连续凸优化方法能实现更高的资源有效利用率。图 4 中分配给地面终端的平均子信道数量、无人机分配给地面终端的平均下行传输能量、无人机分配给地面终端的平均计算容量类似于图 3 的情况，以累积分布函数值来衡量。如当分配较多的相应资源时，相应的累积分布函数的曲线分布在 x 值较大的部分（x 轴偏右）；相反，当分配较少的资源时，资源分配的相应累积分布函数的曲线分布在 x 值较小的部分（x 轴偏左）。

现实场景中，无人机赋能的雾无线电接入网络的资源分配的优化比传统的地面雾无线电接入网络要复杂得多。这是因为无人机和地面终端的连接因为无人机的高移动性，是动态变化的。无人机的位置随时间变化，网络中的通信和计算资源的分配需要实时更新。这将造成巨大的网络配置消耗，降低了无人机赋能的雾无线电接入网络的适应性。对比图 4 中的充水策略或者连续凸优化的传统方法，深度增强学习可以实时地实现有效的资源分配。接下来将进一步介绍深度增强学习综合考虑无人机和地面终端连接的信道质量（信噪比）、终端任务需求、无人机现有通信及计算资源等数据输入，然后利用深度人工神经网络学习得到资源分配结果。该深度增强学习过程将在线完成，优于传统的资源分配方法。

2.1.2　无人机雾接入点的动态任务卸载和缓存设置

在无人机赋能的雾无线电接入网络中，一个地面终端可以感知其附近的地面雾接入点或者无人机雾接入点，并决定是否全部或部分卸载其任务到某一个雾接入点。该动态卸载旨在保证无人机雾接入节点能够以独立或者和其他接入点合作的方式圆满地完成地面终端的任务。具体来讲，如果无人



(a) 无人机三维航迹　(b) 无人机二维航迹　(c) 任务时延

(d) 通信信道资源分配结果　(e) 通信能量资源分配结果　(f) 计算资源分配结果

图 4　资源分配及无人机航迹导致的地面终端任务的时延结果

机和地面终端之间的射频连接通信速率低或者计算资源贫乏，该无人机雾接入点将不能保证在任务截止时间之前完全完成卸载的任务。这样地面终端将不卸载或只卸载部分任务到该无人机。动态任务卸载和无人机雾接入点的通信、计算资源的分配以及无人机航迹位置密切相关。

另外，地面终端往往会在不同时间请求处理相同的任务。无人机雾接入点可以利用缓存技术保存频繁卸载的任务到自身的存储空间中。这样无人机可以避免和地面终端进行重复的任务数据的传输[11]，任务缓存机制可以节约无人机雾接入节点和地面终端之间的能耗及时延。在实际系统中，为了有效地缓存频繁卸载的任务到无人机的有限的存储空间中，需要预测地面终端的任务。同时，缓存策略需要随时更新无人机存储空间中的缓存内容。无人机缓存机制还和动态任务卸载结果息息相关。具体来讲，一般只有近期会被卸载的任务才会被无人机缓存到存储空间中。从动态任务卸载的角度来讲，如果无人机雾接入点缓存了某一个地面终端频繁卸载的任务，那么该地面终端将更倾向于卸载任务到该无人机雾接入点。动态任务卸载及缓存导致的地面终端任务的时延和无人机飞行能耗结果如图 5 所示，展示了动态任务卸载和缓存机制可以有效减少地面终端的任务时延和卸载的任务量。动态任务卸载和缓存机制将减轻无人机雾接入点的任务压力，有利于无人机的航迹设计，节约无人机飞行推进能量的消耗。以上仿真中的任务动态卸载、缓存、无人机航迹规划方法及具体参数设置参考文献[8]。具体来讲，图 5 中动态卸载+缓存策略通过实际的需求情况选择是否将任务卸载及缓存到无人机，达到较高的任务处理效率。相反，基准方法是基于贪婪的方法将任务卸载并缓存到无人机，这将引起无人机的任务拥堵，导致较低的任务处理效率。图 5 同时也比较了动态任务卸载、缓存策略和不同的无人机航迹规划方法（TSP、连续凸优化）的联合工作情况，在不同的无人机航迹规划方法下，动态卸载和缓存方法都可以带来任务处理性能的提升。图 5 中的任务卸载比例、任务缓存比例、任务时延和无人机飞行推进能耗类似于图 3 和图 4 的情况，以累积分布函数值来衡量。如当无人机飞行推进能耗小时，无人机飞行推进能耗累积分布函数的曲线分布在 x 值较小的部分（x 轴偏左）；类似地，当任务时延小时，任务时延累积分布函数的曲线分布在 x 值较小的部分（x 轴偏左）。

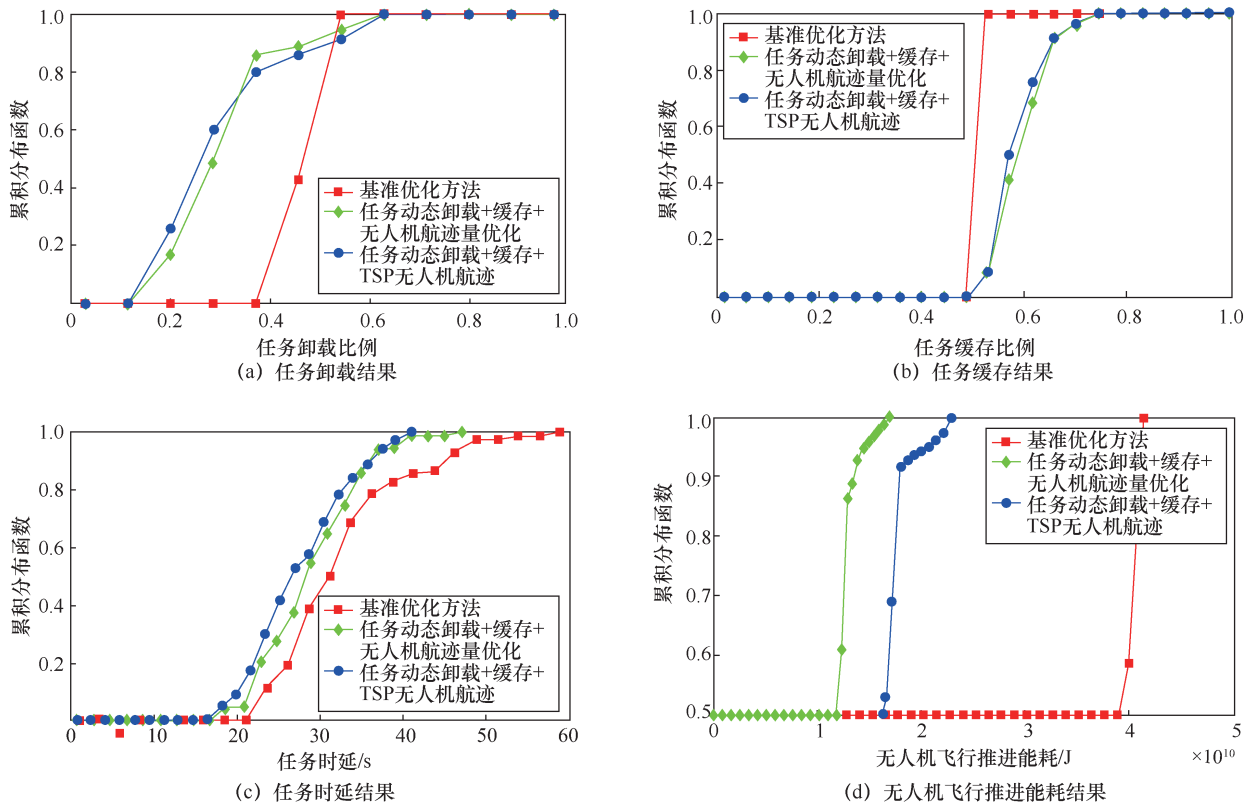


图 5　动态任务卸载及缓存导致的地面终端任务的时延和无人机飞行能耗结果

与无人机雾接入点的资源分配类似，以上无人机雾接入点的动态任务卸载和缓存设置同样是一个高动态变化的优化问题，图 5 的仿真因为利用了多限制条件的目标优化方法，如凸优化的方法，将无法满足任务卸载和缓存的实时性要求。因此，本文同样进一步提出利用深度增强学习的方法动态规划地面终端的任务卸载和缓存策略。由于动态卸载和缓存参数是离散参数，利用深度增强学习的值函数算法，如双 Q 学习网络（double Q-learning network）算法，将很好地实现以上动态任务卸载和缓存的实时完成。

2.1.3　深度增强学习实现无人机赋能雾无线电接入网络的网络配置

通过以上介绍发现，无人机雾无线电接入网络中的资源分配、任务卸载及缓存在网络配置中是高度关联的。需要同时解决以上资源分配、任务卸载及缓存配置的联合问题，该联合问题在无人机雾无线电接入网络环境下将变得异常复杂。与传统优化方法相比，可以利用深度增强学习来实现一个在线的学习框架，快速地解决以上复杂问题。如图 6 所示，深度增强学习框架收集无人机雾接入点、地面雾接入点和地面终端的状态数据，包括无人机和地面终端链路的信噪比、地面终端分配情况及任务需求、无人机的计算容量和缓存容量，将以上信息提炼为综合的系统状体信息。接着，深度增强学习框架将提炼的系统状态发送给代理，然后从代理获得无人机雾无线电接入网络的网络配置输出。深度增强学习优化无人机赋能的雾无线电接入网络的网络配置如图 6 所示，深度学习框架由两个相互迭代的阶段组成，包括动作生成阶段及策略更新阶段，并利用了经验回放技术。动作生成阶段运用深度人工神经网络推导网络配置策略，该深度人工神经网络的特性主要表现在隐藏神经元之间的连接权值。在具体的某一时帧内，深度人工神经网络输入提炼出的系统状态，然后利用 Q 学习的相似过程输出网络配置结果。以上 Q 学习过程根据状态和动作的对应关系的奖励值，选择具体的系统状态对应的最优动作。在无人机雾无线电接入网络中，选择奖励值最高的网络配置为深度人工神经网络输出的最优动作。深度增强学习代理还会把获取的动作、状态及对应的奖励值当作经验存放入记忆库即经验回放缓存中。随后，在同一时帧内，深度学习进入策略更新阶段。在该阶段中，一批次的训练样本将从记忆库中提取出来，用以训练深度人工神经网络。通过训练，深度人工神经网络将更新其神经元之间的连接权值。深度人工神经网络的参数值训练更新后，将更准确地根据下一个时帧的系统状态推导出网络配置结果。以上基于动作生成及策略更新的增强学习过程将在无人机雾无线电接入网络的系统状态变化时实时迭代更新，框架中的深度人工神经网络将持续优化其策略，实现接近最优的雾无线电接入网络的配置。深度增强

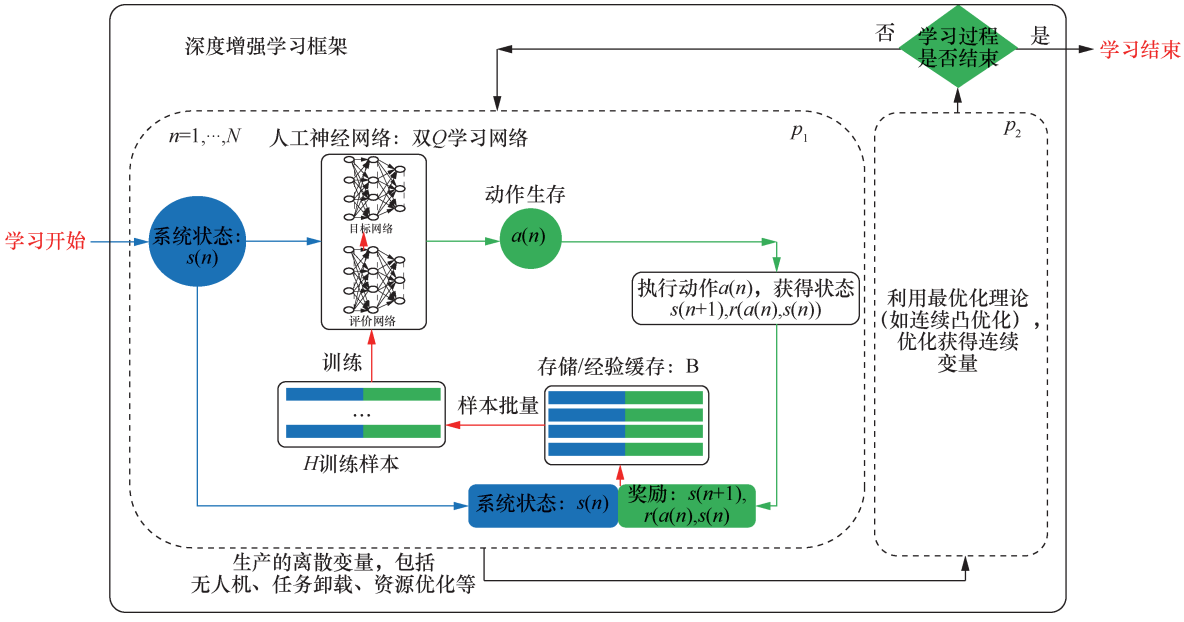


图 6　深度增强学习优化无人机赋能的雾无线电接入网络的网络配置

学习优化无人机赋能的雾无线电接入网络的网络配置如图 6 所示。

在实现层面，深度增强学习框架适合通过双 Q 学习网络算法来实现网络配置。双 Q 学习网络利用目标网络技术[15]可以避免传统深度人工神经网络的不稳定问题。具体来说，双 Q 学习网络构建了两个深度人工神经网络，即评价网络和目标网络。深度增强学习使用经验回放缓存中的批次样本，而不是当前收集的样本，来训练这两个深度人工神经网络。同时目标网络的更新频率要低于主网络，从而实现稳定性。对比利用深度人工神经网络实现的传统 Q 学习，经验回放打破了连续生产样本的关联性，避免了发散和非平滑学习的问题。

2.2　利用深度增强学习实现无人机三维空间的航迹规划

在无人机系统中，利用类似图 6 的深度增强学习过程，无人机的工作航迹设计也可以利用具体的深度增强学习算法实现。通过深度增强学习，无人机可以根据网络中的系统状态来顺序地决定自身在每一个时帧的最优飞行航迹，保障无人机系统进行高能效的工作，利用深度增强学习实现无人机基于无线通信的飞行航迹规划如图 7 所示。类似上一节的网络配置问题，为了在三维空间中规划无人机的飞行航迹，深度增强学习需要根据不同的采集数据，包括用户移动、组网、雾计算以及能耗信息等，构建系统状态，推理合理的动作。推理出的动作主要包括无人机在水平维度、垂直维度以及时间维度的飞行航迹行为。

具体来讲，如图 7 所示的在线学习框架，无人机可以利用双 Q 学习网络的方法选择航迹规划动作。无人机利用上一时帧的网络状态及相应动作的奖励值决定无人机航迹动作。图 7 中，假设无人机飞行的二维水平空间划分为 L 个栅格，每个栅格之间的水平距离固定为 (x_s, y_s)，每个栅格 l 的中心位置为 $L_l^c=[x_l, y_l]^{\mathrm{T}}, l=1,\cdots,L$。同时，将无人机的飞行高度刻画为 $h_{\max}$ 个高度等级，每个等级之间的高度差为 h_s。通过此划分，无人机在时隙 i 的三维位置可以根据其所处的栅格及高度等级表示为：$[L_i^u, h_i^u]$。在每个时隙内，无人机在二维水平空间可采用的飞行动作为：{向东移动一栅格、向西移动一栅格、向南移动一栅格、向北移动一栅格、维持当前栅格}；在垂直空间可采用的飞行动作为：{上升一级、下降一级、维持高度}；以上飞行动作都需要避免通过障碍区域。类似于第 2.1.3 节介绍的学习过程，图 7 中每个时隙 i 内的具体步骤如下。

步骤 1　获取当前状态：根据每个无人机的三维位置、地面终端的位置、通信与计算任务时空分布、地面障碍物的分布情况，得到系统状体 $s(i)$。

步骤 2　制定动作奖励：根据状态 $s(i)$，计算无人机动作 $a(i)$ 的奖励 $r(s(i)、a(i))$。如果无人机的动作更利于空地通信且飞行能耗降低，该动作的奖励值 $r(s(i)、a(i))$ 将较高。相反，如果该无人机动作导致其靠近障碍物、和其他无人机远离、空地通信效果差、飞行能耗较高等任何情况，该动作的奖励值将降低。据此，奖励函数可以定义为

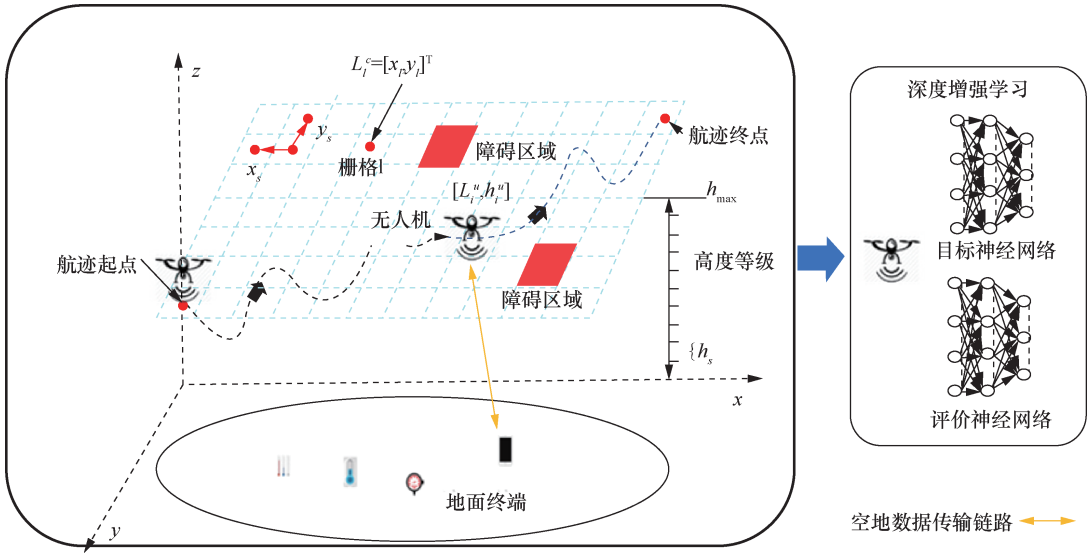


图 7　利用深度增强学习实现无人机基于无线通信的飞行航迹规划

$$r\left(s(i),a(i)\right)=\sum_{a=1}^{A}\left(\sum_{k=1}^{K}\frac{\sum_{j=1}^{i+1}a_{a,k}B\mathrm{lb}\left(1+\frac{P_k h_{a,k,j}}{BN_0}\right)}{\sum_{j=1}^{i+1}e_{a,j}^{\mathrm{uav}}}-p_a^{\mathrm{col}}-p_a^{\mathrm{cnpc}}\right) \tag{1}$$

其中，$e_{a,j}^{\mathrm{uav}}$ 是计算的无人机 α 在时隙 j 的飞行能耗（根据固定翼无人机或者旋翼无人机），p_a^{col} 为无人机 α 无法和地面障碍保持安全距离时的惩罚值，p_a^{cnpc} 为无人机 α 无法和需要的无人机保持有效控制和非载荷通信的惩罚值，式(1)中分数的分子部分是无人机和地面终端卸载数据时的传输吞吐量。

步骤 3 选择动作：根据动作奖励，利用深度神经网络评价状态动作对 $(s(i),a(i))$ 的 Q 值（值函数）如式(2)所示。

$$Q\left(s(i),a(i)\right)=\mathrm{E}\left[\sum_{n'=i}^{N}\gamma r\left(s(n'),a(n')\right)\middle|s(i),a(i)\right] \tag{2}$$

其中，$Q(s(i),a(i))$ 为状态和动作对 $(s(i),a(i))$ 产生的折算积累回报，$\gamma=(0,1]$ 为折扣因子。随后，算法根据策略 π 选择最大的 Q 值，以决策无人机在三维空间中的最佳动作 $a(i)$。策略 π 可以定义为

$$\pi\left(s(i)\right)=\arg\max_{a(n)}Q\left(s(i),a(i)\right) \tag{3}$$

当最佳动作选择后，算法将更新系统状态为 $s(i+1)$；将产生的{动作 $a(i)$、奖励 $r(s(i),a(i))$、更新状态 $s(i+1)$}存储到经验库。

步骤 4 训练深度神经网络：利用经验回放从经验库中抽取 H 个样本来训练评价网络 $Q(\cdot)$ 的权值 θ^Q。评价网络的训练旨在最小化损失函数 $L(\theta^Q)$为

$$L\left(\theta^Q\right)=\mathrm{E}\left[y(i)-Q\left(s(i),a(i)\right)\middle|\theta^Q\right] \tag{4}$$

式(4)中，$y(i)$ 是目标值。

$$y(i)=r\left(s(i),a(i)\right)+\gamma Q'\left(s(i+1),\arg\max_a Q\left(s(i+1)\middle|\theta^Q\right)\middle|\theta^{Q'}\right) \tag{5}$$

其中，权值为 $\theta^{Q'}$ 的目标网络 $Q'(\cdot)$ 被用来计算目标值 $y(i)$。目标网络和评价网络拥有相同的深度神经网络，只是其权值 $\theta^{Q'}$ 相对更新的速度大大慢于评价网络的权值 θ^Q。这样目标网络可以避免深度增强学习过程的过拟合问题。

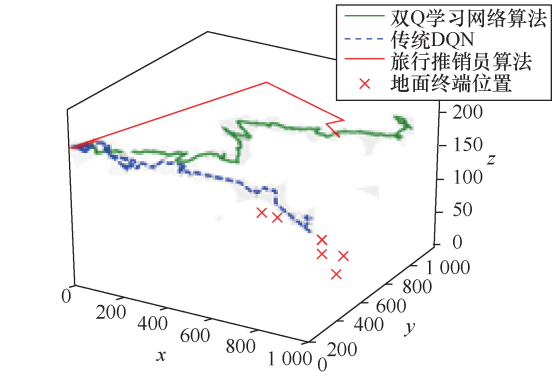
步骤 5 更新目标网络的权值：根据一定的频率（远远小于评价网络的权值更新频率）更新 $Q'(\cdot)$的权值$\theta^{Q'}$。

$$\theta^{Q'}=\sigma\theta^Q+(1-\sigma)\theta^{Q'} \tag{6}$$

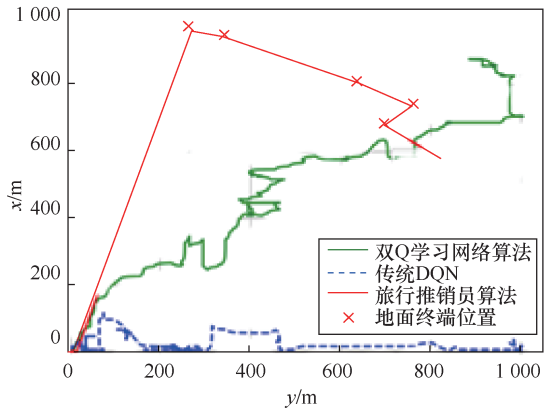
当经历过 N 个片段的以上 5 个步骤的循环迭代后，深度增强学习过程将收敛并获得优化的无人机的飞行航迹规划结果。

进一步，根据如上讨论得知无人机在雾无线电接入网络中的航迹设计和雾无线电接入网络的配置息息相关，这导致无人机航迹规划和网络配置问题一起变成了一个高度复杂的联合优化问题。例如，在文献[16]中，在无人机的二维空间的航迹的基础上，优化了传输能量和用户关联配置。在文献[17]中，一个名为 DRL-EC2 的学习框架被设计用来控制无人机固定高度时的水平飞行航迹，以保障稳定并能效最优地对地无线覆盖。然而文献[16-17]中的研究只利用深度增强学习解决了无人机航迹规划和网络配置的联合问题的局部子问题，没有完全发挥无人机在三维空间中的移动性，也没有考虑资源、动态任务卸载、缓存配置等问题。

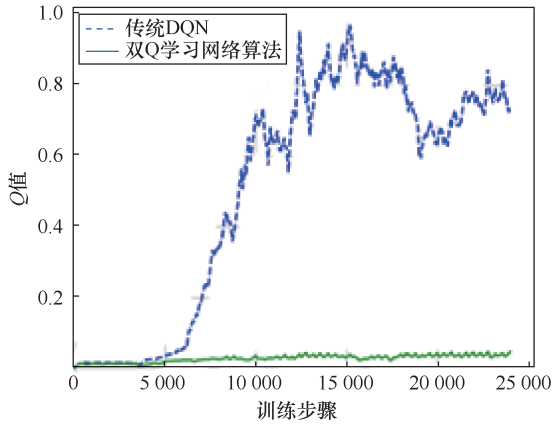
当前研究表明，还不能利用深度增强学习完全解决以上无人机系统的复杂联合优化问题。具体来说，该复杂的联合优化问题将造成深度增强网络很难构建出奖励函数来合理评价无人机针对网络配置和飞行航迹的待选动作，很难在范围广泛的搜索空间中找到最优的动作。另外，利用深度增强学习优化无人机飞行时间等连续的变量，学习的动作空间过大，面临学习精度过低的问题。因此，深度增强学习训练人工神经网络得到的策略将有很大的可能性造成欠拟合的情况。该问题在无人机雾无线电接入网络工作于应急、复杂环境及临时布置、学习框架欠缺先验知识时，变得更明显。利用双 Q 学习网络算法、传统深度 Q 学习（DQN，deep Q-learning network）和旅行推销员算法实现无人机雾无线电接入网络的航迹规划、任务卸载和缓存如图 6 所示，为了避免以上欠拟合问题，深度学习框架将无人机的飞行能效、数据传输效果、任务时延以及无人机飞行速度当作约束条件，利用最优化理论及方法（如凸优化）优化得出连续变量结果。然后在连续变量已知的前提下，深度学习框架将在有限的动作空间中有效地搜索出最优的飞行行为。该最优化理论和深度增强学习结合的方式可以发挥两者的优势，实现高效的无人机飞行航迹规划算法，发挥无人机赋能雾无线电接入网络的优势。图 8 中，



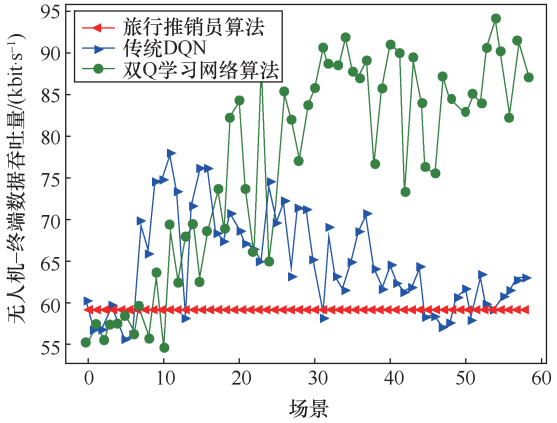
(a) 无人机三维航迹



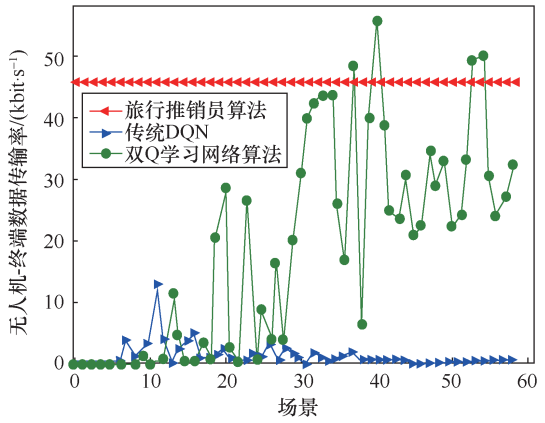
(b) 无人机二维航迹



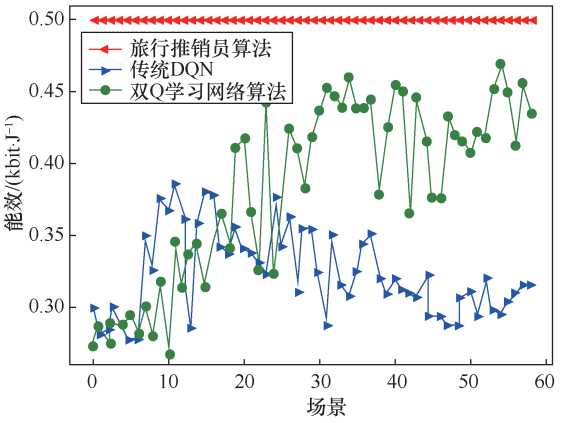
(c) 深度增强学习的Q值比较



(d) 无人机–终端吞吐量比较



(e) 无人机–终端数据率比较



(f) 无人机能效比较

图 8　利用双 Q 学习网络算法、传统 DQN 和旅行推销员算法实现无人机雾无线电接入网络的航迹规划、任务卸载和缓存

对比了双 Q 学习网络算法和传统 DQN 及旅行推销员算法，来联合规划无人机的航迹、任务卸载及缓存。通过对比发现，双 Q 学习网络算法明显优于传统 DQN，并能够获得类似旅行推销员算法的通信和雾计算效果。最重要的是双 Q 学习网络算法获得的无人机能效是 3 个算法中最高的。通过深度增强学习的 Q 值比较可以发现，双 Q 学习网络算法 Q 值上升缓慢且收敛快，避免了传统 DQN 算法的不稳定问题。图 7 的双 Q 学习网络算法、传统 DQN 算法及旅行推销员算法的实现以及相关仿真参数配置详见文献[18]。

另外，以上双 Q 学习网络算法是基于值函数的深度增强学习过程，其通过计算无人机的动作的奖励值来决策。学习过程在确定了奖励函数 $r(s(i),a(i))$ 的基础上，采用如式(3)所示的贪婪策略的方式来选取无人机的动作。当要解决的问题动作

空间很大或者动作为连续集时，式(3)将无法有效求解出最优动作。因此图 7 中针对无人机的飞行时间连续变量是利用优化理论及方法（如凸优化）求解的，这样可以简化基于值函数的深度增强学习的动作搜索空间。此外，奖励值的计算涉及的相关参数较多，也会导致值函数求解 Q 值过于复杂。研究可以进一步考虑利用策略梯度的思想，根据直接策略搜索的方法对策略 π 进行参数化表示，其参数可以表示为 θ^{π}。这样，无人机每一步可以通过参数化的最优行为策略函数 $\pi(\cdot)$ 直接获得确定的动作为

$$a(i)=\pi\left(s(i)\middle|\theta^{\pi}\right) \tag{7}$$

与式(2)的值函数进行参数化表示相比，策略参数化更简单，有更好的收敛性。基于以上理论，研究可以利用基于策略梯度的深度增强学习方法实现无人机飞行航迹的联合优化，更好地实现无人机的飞行动作选择。该方法设计策略网络 π，利用深度增强学习实现策略 π 的参数化(θ^{π})，以支持无人机的动作选择。另外，方法同时构建评价 Q 网络，通过深度 Q 学习来评价策略网络的选择动作的优劣，输出无人机的动作梯度给策略网络，以便后者进一步更好地训练其网络。基于梯度策略的深度增强学习算法实现无人机航迹规划的联合优化在这里将不详细介绍，更多细节可以参考文献[19]。

2.3 技术挑战和未来研究方向

未来无人机雾无线电接入网络将考虑多无人机间的配合以及无人机和地面诸如基站这样的高性能雾接入点的空地协同工作，这将使无人机雾无线电接入网络具备更好的服务性能及更持久的工作时间。然而，多无人机及空地协同的雾无线电接入网络将带来更多烦琐的问题，如防止无人机的碰撞、无人机和其他雾接入节点的通信频谱共享、无人机间协同通信以及更复杂的网络配置等。因此，深度增强学习框架需要扩展，以支持以上复杂的无人机雾无线电接入网络的学习问题。更进一步，当多无人机联合地服务地面终端时，在考虑提供雾计算功能之外，多架无人机在三维空间中的航迹规划需要考虑彼此的能量情况。如各个无人机间的推进能量的消耗需要在航迹设计的时候得以协同均衡，最大化多无人机网络的工作时间。以上新技术将给深度增强学习带来新的挑战。多无人机的深度学习框架需要实现协同学习，利用人工神经网络算法构建系统状态及动作空间，并考虑其他无人机学习得到的新行为的影响。针对以上问题，多智能体强化学习将是一个很好的解决方法。

另外，无人机未来会考虑利用太阳能或者激光束的无线充能技术实时获得能量补充，这样无人机雾无线电接入网络的工作时间可以得到有效提高。然而，这项新技术将给无人机在三维空间中的航迹设计带来新的设计挑战。具体来讲，为了获得有效的能量补充，无人机将改变如图 2～图 3 所示的飞行行为，不再倾向于采取最优速度飞行以及水平飞行。相反，无人机将在垂直维度拔升以获得更好的太阳能或者激光无线充能，然后再下降为地面终端提供通信和计算服务。因此，深度增强学习将更改无人机在垂直维度的飞行航迹的学习策略。

无人机也可以作为能量提供者，通过无线充能的方式为地面终端进行充能。该功能在终端能量短缺且终端没有其他能量供应的物联网系统中，显得尤为有用。无人机因此不仅作为雾接入节点，同时也作为能量提供者。该无人机的航迹规划、资源分配和任务卸载将面临新的限制。例如，为了给地面终端很好的充能，无人机需要下降靠近终端。但这样会造成无人机过高的飞行能量消耗。总的来说，无人机同时为地面终端提供通信雾计算服务以及无线充能服务，其航迹规划及相关网络配置将面临全新的挑战。未来利用深度增强学习需要支持无人机在不同场景中的应用，特别是在物联网系统中，实现能效最优的工作。

3 结束语

本文探索了利用深度增强学习来解决无人机赋能的雾无线电接入网络服务质量和消耗的折中平衡问题。无人机在雾无线电接入网络中作为接入点，为地面终端提供通信及雾计算服务。本文提到深度增强学习可以用来解决无人机雾无线电接入网络中的网络配置涉及的联合的资源、任务及缓存优化问题，以及无人机在三维空间的航迹规划问题。利用双 Q 学习网络算法以及策略梯度算法等，应用经验回放及目标网络技术，实现在线的深度增强学习框架，实时实现网络的优化，保障无人机较高的能效比。本文同时也讨论了无人机赋能的雾无线电接入网络面临的新功能以及挑战，以及利用深度增强学习解决新挑战的技术前景。

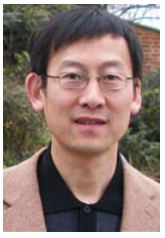
参考文献：

[1] CHIANG M, ZHANG T. Fog and IoT: an overview of research op-

portunities[J]. IEEE Internet of Things Journal, 2016, 3(6): 854-864.
[2] LOKE S W. The Internet of flying-things: opportunities and challenges with airborne fog computing and mobile cloud in the clouds[E]. arXiv preprint arXiv:1507.04492, 2015.
[3] 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.
[4] WU Q Q, LIU L, ZHANG R. Fundamental trade-offs in communication and trajectory design for UAV-enabled wireless network[J]. IEEE Wireless Communications, 2019, 26(1): 36-44.
[5] KU Y J, LIN D Y, LEE C F, et al. 5G radio access network design with the fog paradigm: confluence of communications and computing[J]. IEEE Communications Magazine, 2017, 55(4): 46-52.
[6] ZENG Y, XU J, ZHANG R. Energy minimization for wireless communication with rotary-wing UAV[J]. IEEE Transactions on Wireless Communications, 2019, 18(4): 2329-2345.
[7] ZENG Y, ZHANG R. Energy-efficient UAV communication with trajectory optimization[J]. IEEE Transactions on Wireless Communications, 2017, 16(6): 3747-3760.
[8] MEI H B, WANG K Z, ZHOU D D, et al. Joint trajectory-task-cache optimization in UAV-enabled mobile edge networks for cyber-physical system[J]. IEEE Access, 2019(7): 156476-156488.
[9] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
[10] LI R P, ZHAO Z F, ZHOU X, et al. Intelligent 5G: when cellular networks meet artificial intelligence[J]. IEEE Wireless Communications, 2017, 24(5): 175-183.
[11] CHEN M Z, MOZAFFARI M, SAAD W, et al. Caching in the sky: proactive deployment of cache-enabled unmanned aerial vehicles for optimized quality-of-experience[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(5): 1046-1061.
[12] CHEN M Z, SAAD W, YIN C C. Liquid state machine learning for resource and cache management in LTE-U unmanned aerial vehicle (UAV) networks[J]. IEEE Transactions on Wireless Communications, 2019, 18(3): 1504-1517.
[13] BOYD S, VANDENBERGHE L. Convex optimization[M]. Cambridge: Cambridge University Press, 2004.
[14] MEI H B, YANG K, LIU Q, et al. Joint trajectory-resource optimization in UAV-enabled edge-cloud system with virtualized mobile clone[J]. IEEE Internet of Things Journal, 2020, 7(7): 5906-5921.
[15] HE Y, YU F R, ZHAO N, et al. Software-defined networks with mobile edge computing and caching for smart cities: a big data deep reinforcement learning approach[J]. IEEE Communications Magazine, 2017, 55(12): 31-37.
[16] CHALLITA U, SAAD W, BETTSTETTER C. Deep reinforcement learning for interference-aware path planning of cellular-connected UAVs[C]//2018 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2018: 1-7.
[17] LIU C H, CHEN Z Y, 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.
[18] MEI H B, YANG K, SHEN J, et al. Joint trajectory-task-cache optimization with phase-shift design of RIS-assisted UAV for MEC[EB]. IEEE Wireless Communications Letters, 2021.
[19] PENG H X, SHEN X S. DDPG-based resource management for MEC/UAV-assisted vehicular networks[C]//2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall). Piscataway: IEEE Press, 2020: 1-6.

[作者简介]

梅海波（1983− ），男，博士，电子科技大学讲师，主要研究方向为无线移动通信、移动边缘计算、无人机通信、人工智能。



杨鲲（1969− ），男，电子科技大学教授，主要研究方向为无线通信网络、数能一体化通信网、移动计算。

范新宇（1997− ），男，电子科技大学硕士生，主要研究方向为物联网、数能一体化通信网。