

基于图学习的密集空间网络传输资源调度方法

刘润滋¹, 吴伟华², 张文柱¹, 周 笛², 张 琰²

(1. 西安建筑科技大学信息与控制工程学院, 陕西西安 710055;
2. 西安电子科技大学综合业务网理论及关键技术国家重点实验室, 陕西西安 710071)

摘 要: 面对密集空间网络传输资源调度问题中的复杂度以及有效性挑战, 本文以图论模型为纽带, 将传统数学模型与机器学习方法相结合, 提出一种基于图学习的密集空间网络传输资源调度方法. 该方法基于图论模型对问题结构的认知将密集空间网络资源调度问题分解, 由数学模型与基于图结构的强化学习交替配合完成整个求解过程. 实验结果表明, 与传统的基于数学模型的资源调度方法相比, 所提方法能将资源调度收益提升 25.1%, 且其训练结果对网络场景变化具有较好的适应性.

关键词: 密集空间网络; 资源调度; 机器学习; 深度强化学习; 图论

中图分类号: TN919; TN927

文献标识码: A

文章编号: 0372-2112(2021)11-2133-05

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20201217

Graph Learning Based Transmission Resources Scheduling in Dense Space Networks

LIU Run-zi¹, WU Wei-hua², ZHANG Wen-zhu¹, ZHOU Di², ZHANG Yan²

(1. School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China;

2. The State Key Laboratory of ISN, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract: Facing the challenges of the transmission resource schedule of dense space networks, we combine the mathematical models and machine learning methods, and propose a graph learning based approach for the scheduling of transmission resources in dense space networks. In the proposed method, transmission resource scheduling problem is decomposed based on the knowledge of the problem structure brought by graph theory. On this basis, mathematical model and reinforcement learning alternately complete the whole solution process. Simulation results show that, compared with the traditional mathematical model-based methods, the proposed method improves the scheduling profits by 25.1%, and its training results have better generality.

Key words: dense space network; resource scheduling; machine learning; deep reinforcement learning; graph theory

1 引言

近年来,随着 OneWeb、SpaceX^[1]、鸿雁星座^[2]等大型星座的密集部署,其与位于不同轨道的卫星协同组网,逐渐形成多层、立体的密集空间网络.与传统空间网络相比,密集空间网络庞大的网络规模和复杂的时空结构为网络资源的管理带来了巨大的挑战.作为影响空间网络服务能力的关键问题之一,传输资源调度问题在传统非密集空间网络场景中已经得到了广泛的研究^[3-5].然而,由于空间网络传输资源调度为 NP 难问

题^[6],上述工作仅能通过启发式方法获得次优解,而且所得结果与最优解之间的差距随着网络规模的增加而迅速增大.因此,对于比传统空间网络规模更大、结构更加复杂的密集空间网络,现有的基于“建立-求解数学模型”的资源调度方法不再适用,亟需针对密集空间网络场景特征探索传输资源调度问题的新研究思路.

近年来,机器学习已被应用于以较低的计算成本解决干扰管理、能量分配等地面无线网络典型的 NP 难资源管理问题^[7-9].密集空间网络中,由于卫星节点周

收稿日期:2020-10-30;修回日期:2021-07-08;责任编辑:李勇锋

基金项目:国家自然科学基金(No.61701365, No.61801365, No.62001347);中国博士后科学基金(No.2018M643581, No.2019TQ0241, No.2020M673344);陕西省重点研发计划(No.2021GY-066);陕西省自然科学基金(No.2020JQ-686);陕西省高校科协青年人才托举计划基金(No.20200112);陕西省博士后科研基金(No.2018BSHEDZ47)

期性的绕地飞行,网络拓扑结构演进也具有周期性,从而为使用机器学习方法学习往期经验优化资源调度方案提供了便利条件.目前机器学习在空间网络资源管理中的应用尚在起步阶段^[10],大多采用端到端的学习方式^[10-12],但是这些工作直接使用全网的任务需求、资源状态以及调度结果等高维矢量作为神经网络的输入/输出进行训练,一方面大规模神经网络训练速度慢且需要大量的样本,使得上述方法难以推广到规模较大的密集空间网络;另一方面直接将任务、资源以及调度结果作为神经网络输入/输出进行端到端训练会导致所训练的神经网络难以适应调度周期或者任务需求变化,从而制约了训练结果的一般性.

针对端到端学习方法的弊端,本文以密集空间网络的图论模型为纽带,将机器学习与传统数学模型方法结合,提出一种基于图学习的密集空间网络传输资源调度方法.基于图论模型对问题结构的认知将密集空间网络资源调度问题分解,由数学模型与基于图结构的强化学习交替配合完成整个求解过程,通过仿真分析验证了该方法的有效性.

2 系统模型

2.1 场景与问题描述

考虑图1所示的密集空间网络,由卫星和地面站构成.其中,网络中的卫星包括低轨、中轨以及对地同步轨道卫星,其集合表示为 $S=\{1, 2, \dots, s, \dots\}$,符号 s 表示第 s 颗卫星.地面站集合表示为 $G=\{1, 2, \dots, g, \dots\}$,符号 g 表示第 g 个地面站.网络的规划周期 τ 被划分为 T 个等长的时隙, $\tau=\{1, 2, \dots, t, \dots, T\}$.

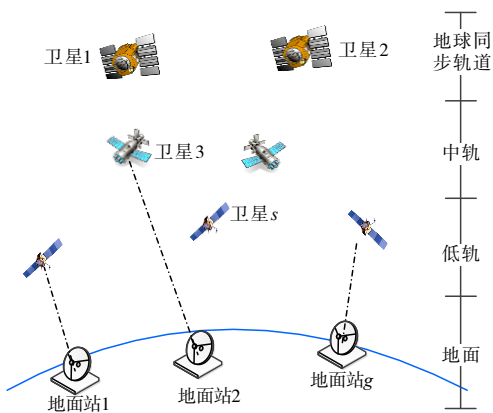


图1 密集空间网络场景

网络中有若干数据传输任务,其集合表示为 $M=\{1, 2, \dots, m, \dots\}$,每个任务用一个六维元组表示,以第 m 个任务为例: $m=[s(m), G(m), p_m, st_m, et_m, w_m]$,其中 $s(m)$ 表示请求第 m 个任务的卫星,即任务 m 的数据由卫星 $s(m)$ 下载到地面, $G(m)$ 表示满足任务 m 的接

收条件的地面站集合. p_m 表示任务 m 所需的传输时间, $[st_m, et_m]$ 为任务 m 的可执行窗口,其中 st_m 、 et_m 分别表示任务 m 的最早开始时间和最晚结束时间, w_m 表示任务 m 成功执行后所获得的收益.

为了完成任务 m ,需要在 $G(m)$ 中选择一个地面站,并调度该地面站在 $[st_m, et_m]$ 内与卫星 $s(m)$ 可见且长度为 p_m 的一个时间窗口接收任务数据.当且仅当该传输窗口不与同一卫星或同一地面站上的其它传输窗口发生重叠时,任务 m 方可被视作成功调度.因此,密集空间网络传输资源调度问题即为在不超出网络资源能力的前提下,为所有任务分配合适的地面站和满足要求的接收时间窗口来最大化网络收益的问题.

2.2 问题模型

空间网络传输资源调度问题已被证明为NP难问题^[6],本文采用图论与机器学习结合的方法实现该问题的高效求解,首先将密集空间网络传输资源调度问题转化为图论中的独立集问题.

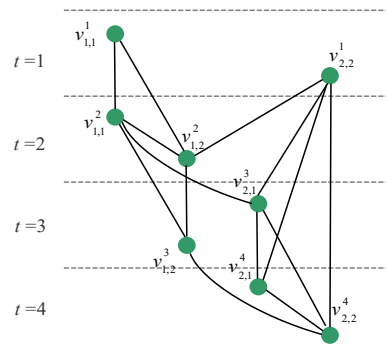


图2 冲突图

建立如图2所示的冲突图模型 $CG(V, E)$,图中顶点对应于空间网络中的资源分配方案,具体而言,顶点 v_{mg}^t 表示将地面站 g 的 $[t, t+p_m-1]$ 时段分配给任务 m ,即地面站 g 从 t 时隙开始接受任务 m 的数据直至 $t+p_m-1$ 时隙为止. CG 中的边表示不同的可行资源分配方案之间的冲突关系:对于任意 $v_{mg}^t, v_{nh}^\tau \in V$ (不失一般性,假设 $t \leq \tau$),如果满足以下条件之一:

- (1) $m = n$;
- (2) $g = h$ 且 $\tau \in Z(m, g, t)$;
- (3) $s(m) = s(n)$ 且 $\tau \in Z(m, g, t)$.

则有 $(v_{mg}^t, v_{nh}^\tau) \in E$.上式中, $Z(m, g, t)$ 表示地面站 g 从 t 时隙开始为任务 m 提供服务后无法开始接收其他任务的时隙集合.上述三个条件分别表示了同一任务不同的资源分配方案、同一地面站上相互重叠的资源分配方案以及同一卫星上相互重叠的资源分配方案之间具有冲突关系.因此,优化密集空间网络传输资源调度问题等价于基于冲突图中的最大权重独立集问题.

3 基于图学习的资源调度问题求解

3.1 基于图结构嵌入的图信息表征

为实现图论分析与机器学习之间的有机结合,需要设计有效的冲突图参数化方法,既保留图中有效信息又便于对其进行挖掘利用. 本文提出一种基于图结构嵌入的图信息参数化方法,其通过邻域卷积的方式将冲突图中顶点关键属性沿拓扑结构聚合到每个顶点,形成嵌入向量.

具体而言,使用 q 维向量 $\mu_v^{(k)}$ 表示递归过程中第 k 次迭代后获得的顶点 v 的嵌入向量. 将顶点 v 嵌入向量的初始值 $\mu_v^{(0)}$ 设置为 0, 在每次迭代中异步地更新冲突图中所有顶点的嵌入向量,即

$$\mu_v^{(k+1)} = \text{relu} \left(\theta_1 y_v + \theta_2 \sum_{u \in N(v)} \mu_u^{(k)} \right) \quad (1)$$

其中, $y_v = (w_v, d_v)$ 表示冲突图中顶点 v 的属性, w_v 为顶点 v 对应的资源分配方案完成后可获得的收益, d_v 表示当前冲突图中顶点 v 的度. $N(v)$ 表示当前冲突图中与顶点 v 相邻的顶点集合. $\theta_1 \in \mathbb{R}^{2 \times q}$ 、 $\theta_2 \in \mathbb{R}^{q \times q}$ 为 Q 函数中待训练参量, $\text{relu}(\cdot)$ 为线性整流函数. 上述嵌入过程在迭代 K 次后终止, 每个顶点获得嵌入向量 $\mu_v^{(K)}$.

根据式(1)可以看出图结构嵌入更新过程是基于图中顶点的邻接关系进行的. 通过图结构嵌入过程, 顶点属性沿图的拓扑结构传播到其它顶点. 因此, 所得嵌入向量既能够反映图中拓扑信息又包含图中顶点的关键属性, 同时, 嵌入向量的维度 q 不受图规模的影响, 从而保障了机器学习结果对网络任务与资源变化的适应性.

3.2 评价函数 Q 的参数化设计

K 步图结构嵌入完成后, 使用嵌入向量定义评价函数 $Q(\text{IS}, v; \Theta)$, 用来衡量最大权重独立集求解过程中可行解的价值. 具体而言, 用嵌入向量 μ_v 表示顶点 v , 池化向量 $\sum_{u \in V} \mu_u$ 表示图 CG, 二者分别代替评价函数中的参量 v 与 IS , $Q(\text{IS}, v; \Theta)$ 定义为:

$$Q(\text{IS}, v; \Theta) = \theta_3^T \text{relu} \left[\theta_4 \sum_{u \in V} \mu_u + \theta_5 \mu_v \right] \quad (2)$$

其中, $\theta_3 \in \mathbb{R}^{2q}$, $\theta_4, \theta_5 \in \mathbb{R}^{q \times q}$, $[\cdot, \cdot]$ 表示向量拼接运算. 由于参量 μ_u, μ_v 由式(1)计算获得, 其受图结构嵌入过程中的参数 θ_1, θ_2 的影响, 因此评价函数 $Q(\text{IS}, v; \Theta)$ 由 5 个参数 $\Theta = \{\theta_i\}_{i=1}^5$ 决定.

3.3 基于深度 Q 学习的评价函数训练

本小节将冲突图中的最大权重独立集求解问题建模为深度强化学习过程, 并完成评价函数 $Q(\text{IS}, v; \Theta)$ 的训练. 首先, 给出强化学习模型中的主要参数:

状态: 强化学习过程中的状态为最加权大独立集的局部解 IS , 由于局部解 IS 对后续独立集迭代构造过

程的影响等同于与其对应的剩余冲突图 CG, 因此, 使用 CG 的嵌入向量, 即 q 维向量 $\sum_{u \in V} \mu_u$ 表示当前状态.

动作: 动作 v 是一个不属于当前局部解的顶点, 用 q 维嵌入向量 μ_v 表示.

奖励: 奖励函数 $r(\text{IS}, v)$ 被定义为在状态 IS 下采取动作 v 转移到状态 $\text{IS}' = \text{IS} \cup \{v\}$ 后收益函数的变化值:

$$r(\text{IS}, v) = c(\text{IS}') - c(\text{IS}) \quad (3)$$

$c(\text{IS}) = \sum_{u \in \text{IS}} w_u$ 表示局部解 IS 对应的资源分配方案获得的收益. 奖励累积函数 R 在终止状态 $\widehat{\text{IS}}$ 的值等于 $\widehat{\text{IS}}$ 的收益, 即 $R(\widehat{\text{IS}}) = \sum_{i=1}^{|\widehat{\text{IS}}|} r(\text{IS}_i, v_i) = c(\widehat{\text{IS}})$.

基于深度强化学习的评价函数训练如算法 1 所示, 其主要思路为, 在每个回合中基于评价函数 Q 贪婪地构造最大权重独立集, 与此同时使用深度神经网络实现对评价函数 $Q(\text{IS}, v; \Theta)$ 中参数 Θ 的优化. 具体而言, 从冲突图中选择一个使得评价函数 Q 值最大的顶点 v 添加到当前局部解 IS 中, 并将该顶点及其邻居顶点从冲突图中删除, 同时, 建立以参数 Θ 为权重的深度神经网络, 通过梯度下降法最小化平方损失 $L(\Theta)$ 来更新参数 Θ :

$$L(\Theta) = (h - Q(\text{IS}, v; \Theta))^2 \quad (4)$$

重复上述过程, 直到冲突图中无剩余顶点为止, 当前局部解即为最大独立集.

考虑到独立集构造过程中部分奖励反馈延迟以及连续的动作之间有较强的相关性等问题, 算法 1 使用更擅长处理收益反馈较慢的延迟奖励问题的 n -步 Q 学习和经验回放技术加快神经网络学习收敛速度并提升训练结果的鲁棒性. 同时, 采用 ϵ -贪婪策略在基于已获知的知识最大化奖励和采取新的动作以获得新知识之间做出有效折衷.

4 仿真分析

构建一个由 54 颗低轨卫星、12 颗中轨卫星、3 颗地球同步轨道卫星, 12 个地面站构成的密集空间网络仿真场景. 仿真周期为 2020-3-10 4:00:00 至 2020-3-11 4:00:00, 网络中任务数为 200~1000 个, 任务收益在 $[1, 5]$ 内均匀分布, 每次实验随机生成 100 组同样数目的任务集合, 其中 50 组作为训练样本, 50 组作为测试样本. 深度 Q 学习训练过程采用 6 层深度神经网络, ϵ 值设为 0.9.

为了分析不同的学习速率对深度 Q 学习训练收敛性的影响, 图 3 展示了当网络中任务数目为 1000 时损失函数 $L(\Theta)$ 随训练步数的变化曲线. 可以看出, 学习速率越大, 训练过程收敛速度越快, 当学习速率为 0.0001 时训练速度过于缓慢, 而当学习速率值为 0.01 时, 虽然收敛速率优于 0.001, 但是容易陷入局部最优.

算法 1 基于深度 Q 学习的评价函数训练

输入: 密集空间网络场景, ε

输出: 训练结果 Θ

```

1: 初始化经验回放存储空间  $\Pi$ 
2: FOR 回合  $e=1$  TO  $LDO$ 
3:   抽取一个任务集合, 构造对应的冲突图  $CG(V, E)$ ;
4:   初始化  $\tau = 1, IS_1 = \emptyset$ ;
5:   WHILE  $V \neq \emptyset$ 
6:      $v_\tau = \begin{cases} \text{random node } v \in V, w.p. \varepsilon \\ \arg \max_{v \in V} Q(IS_\tau, V; \Theta), w.p. 1 - \varepsilon \end{cases}$ 
7:     将顶点  $v_\tau$  加入局部解:  $IS_{\tau+1} = IS_\tau \cup \{v_\tau\}$ 
8:     将顶点  $v_\tau$  及其邻居顶点从  $CG(V, E)$  中删除;
9:     IF  $\tau \geq n$  THEN
10:      将元组  $(IS_{\tau-n}, v_{\tau-n}, R_{\tau-n, \tau}, IS_\tau)$  加入  $\Pi$ 
11:      随机批抽样  $B \sim \Pi$ 
12:      基于梯度下降法使用  $B$  中的样例来训练更新  $\Theta$  值
13:    END IF
14:     $\tau = \tau + 1$ ;
15:  END WHILE
16: END FOR

```

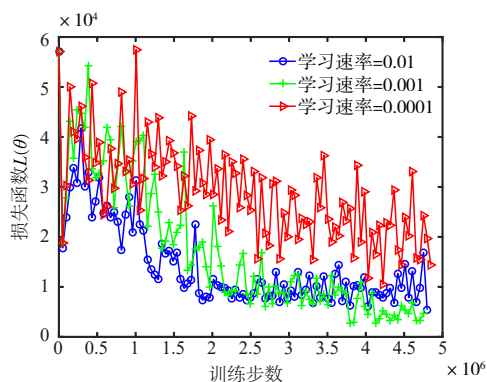


图3 损失函数 $L(\Theta)$ 随训练步数变化

图4对比了所提算法与基于多机调度模型的贪婪随机自适应搜索算法 (GRASP)^[3] 以及混合遗传算法 (HGA)^[13] 的归一化资源调度收益. 归一化资源调度收益值等于成功调度任务收益值之和与请求调度的任务收益值之和的比值. 如图4所示, 三种方法在任务数目较低的时候性能比较接近, 而当任务数目较多时, 所提算法具有明显优势: 当任务数达到1000时, 所提算法收益值比 GRASP 与 HGA 算法分别提升 41.6%、25.1%. 这是因为随着任务数目的增长, 问题规模也越来越大, 启发式算法与最优解之间的性能差距也在逐渐变大, 由于本文所提算法将原问题进行分解后, 仅采用强化学习处理关键步骤, 其性能受问题规模影响较小.

为了验证所提算法训练结果的一般性, 将网络的仿真周期向后推移5日, 即2020-3-15 4:00:00至2020-3-16 4:00:00, 并重新生成100组任务. 图5展示了算法

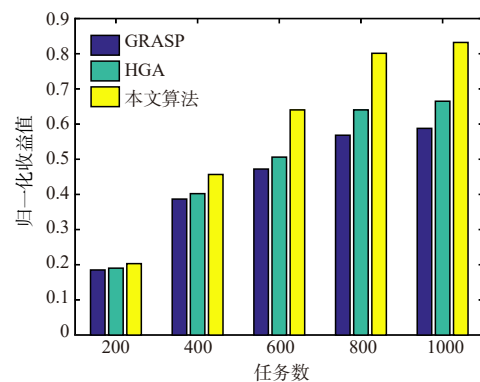


图4 归一化资源调度收益随任务数变化

1应用基于原场景训练得到的 Q 函数与重新训练得到 Q 函数以及对比算法 GRASP、HGA 的调度结果. 可以看出, 采用基于原场景训练的 Q 函数的调度结果, 与基于新 Q 函数的较为接近, 能达到其收益的90%以上, 且与对比算法相比仍有明显优势.

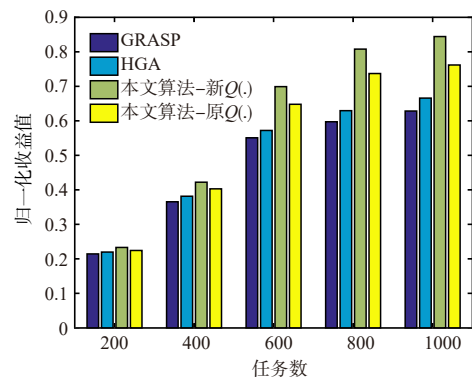


图5 训练结果适应性验证

5 结束语

本文结合传统数学模型与机器学习方法的优点, 提出了一种基于图学习的密集空间网络传输资源调度方法. 所提方法将传输资源调度问题转化为冲突图上的最大权重独立集问题, 以此为基础, 结合图论与强化学习理论设计了基于图学习资源调度问题的求解算法, 该算法由基于图结构嵌入的图信息参数化方法以及基于深度 Q 学习的训练算法两部分组成. 最后, 仿真结果验证了所提方法的资源调度总收益与训练收敛性以及训练结果的一般性.

参考文献

- [1] Zhang Z, Zhang W, Tseng F H. Satellite mobile edge computing: Improving QoS of high-speed satellite-terrestrial networks using edge computing techniques[J]. IEEE Network, 2019, 33(1): 70 - 76.

- [2] 姜会林, 付强, 赵义武, 刘显著. 空间信息网络与激光通信发展现状及趋势[J]. 物联网学报, 2019, 3(2): 1 – 8.
Jiang H L, Fu Q, Zhao Y W, LIU X Z. Development status and trend of space information network and laser communication[J]. Chinese Journal on Internet of Things, 2019, 3(2): 1 – 8. (in Chinese)
- [3] Rojanasoonthon S, Bard J. A GRASP for parallel machine scheduling with time windows[J]. INFORMS Journal on Computing, 2005, 17(1): 32 – 51.
- [4] Wang L, Jiang C, Kuang L, et al. Mission scheduling in space network with antenna dynamic setup times[J]. IEEE Transactions on Aerospace and Electronic Systems, 2019, 55(1): 31 – 45.
- [5] Baek S, Han S, Cho K, et al. Development of a scheduling algorithm and GUI for autonomous satellite missions[J]. Acta Astronautica, 2011, 68(7-8): 1396 – 1402.
- [6] Barbulescu L, Watson J P, Whitley L D, et al. Scheduling space-ground communications for the air force satellite control network[J]. Journal of Scheduling, 2004, 7(1): 7 – 34.
- [7] Chen M, Challita U, Saad W, et al. Artificial neural networks-based machine learning for wireless networks: A tutorial[J]. IEEE Communications Surveys & Tutorials, 2019, 21(4): 3039 – 3071.
- [8] Liang L, Ye H, Yu G, et al. Deep-learning-based wireless resource allocation with application to vehicular networks [J]. Proceedings of the IEEE, 2019, 108(2): 341 – 356.
- [9] Sun H, Chen X, Shi Q, et al. Learning to optimize: Training deep neural networks for interference management[J]. IEEE Transactions on Signal Processing, 2018, 66(20): 5438 – 5453.
- [10] Deng B, Jiang C, Yao H, et al. The next generation het-

erogeneous satellite communication networks: Integration of resource management and deep reinforcement learning [J]. IEEE Wireless Communications, 2019, 27(2): 105 – 111.

- [11] Meng X, Wu L, Yu S. Research on resource allocation method of space information networks based on deep reinforcement learning[J]. Remote Sensing, 2019, 11(4): 448.
- [12] Qiu C, Yao H, Yu F R, et al. Deep q-learning aided networking, caching, and computing resources allocation in software-defined satellite-terrestrial networks[J]. IEEE Transactions on Vehicular Technology, 2019, 68(6): 5871 – 5883.
- [13] Song B, Yao F, Chen Y, et al. A hybrid genetic algorithm for satellite image downlink scheduling problem[J]. Discrete Dynamics in Nature and Society, 2018: 1531452.

作者简介



刘润滋 女, 1988年生, 山东潍坊人. 现为西安建筑科技大学信息与控制工程学院副教授, 主要研究方向为无线自组织网络、空间信息网络智能组网技术等.
E-mail: rzliu@xauat.edu.cn



吴伟华(通信作者) 男, 1988年生, 河北石家庄人. 现为西安电子科技大学通信工程学院讲师, 主要研究方向为无线通信网络、资源管理方法设计、随机网络优化等.
E-mail: whwu@xidian.edu.cn