基于决策性能评估的多波束低地球轨道卫星网络资源分配算法

王朝炜[1]，庞明亮[1]，王粟[2]，赵玲莉[1]，高飞飞[3]，崔高峰[1]，王卫东[1]
（1.北京邮电大学电子工程学院，北京 100875；2.中国移动通信集团有限公司，北京 100032；3.清华大学自动化系，北京 100084）

摘　要：为了解决多波束低地球轨道（LEO）卫星波束间同频干扰、频谱短缺、业务量分布不均等问题，针对单一决策网络缺乏自我修正能力、容易陷入局部最优解、无法充分考虑长期影响等弊端，提出了一种基于决策性能评估的资源分配算法。该算法引入不同用户的业务满足指数来衡量系统的公平性，在考虑公平性的前提下优化系统的吞吐量性能，并将该优化问题建模为多目标优化问题。将具有时间相关性的连续资源分配过程建模为马尔可夫过程，提出基于决策性能评估的网络资源分配算法来解决该问题。所提算法可以根据评估网络的评估结果调整决策网络参数，从而优化资源分配方案，同时更新评估网络自身参数。通过迭代优化的方式，实现决策网络的准确预测。仿真结果表明，所提算法在吞吐量性能和公平性方面优于传统资源分配算法。

关键词：多波束卫星；深度强化学习；多目标优化；资源管理
中图分类号：TN927
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2024040

Resource allocation algorithm for multi-beam LEO satellite based on decision performance evaluation

WANG Chaowei[1], PANG Mingliang[1], WANG Su[2], ZHAO Lingli[1],
GAO Feifei[3], CUI Gaofeng[1], WANG Weidong[1]
1. School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100875, China
2. China Mobile Communications Corporation, Beijing 100032, China
3. Department of Automation, Tsinghua University, Beijing 100084, China

Abstract: To address challenges such as co-frequency interference, spectrum scarcity, and uneven traffic distribution in multi-beam LEO satellites, a resource allocation algorithm based on decision performance evaluation was proposed. The system fairness was measured by a user satisfaction index and the system throughput was optimized while considering fairness. The optimization problem was modeled as a multi-objective optimization. The continuous resource allocation process with temporal correlation was modeled as a Markov decision process, and a decision-evaluation dual-network algorithm was proposed to solve it. The decision network parameters were adjusted based on evaluation network results to optimize resource allocation and update the evaluation network parameters. Through iterative optimization, the decision network achieved accurate predictions. Simulation results show that the proposed algorithm outperforms traditional resource allocation algorithms in terms of throughput and fairness.

Keywords: multi-beam satellite, deep reinforcement learning, multi-objective optimization, resource management

收稿日期：2023–10–24；修回日期：2024–02–02
通信作者：庞明亮，pangmingliang@bupt.edu.cn
基金项目：重庆市自然科学基金资助项目（No.CSTB2023NSCQ-LZX0118）；北京邮电大学博士生创新基金资助项目（No.CX2023139）
Foundation Items: The Natural Science Foundation of Chongqing (No.CSTB2023NSCQ-LZX0118), BUPT Excellent Ph.D. Students Foundation (No.CX2023139)

0 引言

地面通信系统已经取得了显著进展[1-2]，但其仍然存在一些局限性。卫星通信系统克服了地面通信系统在覆盖范围和可及性方面的不足，现已成为国家信息网络中不可或缺的一部分。由于星上资源的稀缺性和高成本，提高资源利用率成为满足日益增长的通信需求的关键[3]。此外，多波束卫星的提出为上述问题提供了创新的解决方案[4-5]，促进了高吞吐量卫星的发展，从而为满足日益增长的业务需求、实现可靠和灵活的连接以及卫星从广播到宽带任务的转变提供了有利条件。

目前的多波束卫星通信系统广泛采用4色、7色等频率复用方案[6-8]，相邻波束使用不同的频带以避免用户间的同频干扰。近年来伴随着地面通信系统的飞速发展，出现了新的频率复用技术，即同频组网技术，所有波束可以使用系统的全部频带。文献[9]提出了一种多波束卫星同频组网的方案设想，该方案虽然在一定程度上提升了系统的容量，但使波束间的同频干扰大大增加。

因此，如何综合考虑多波束间同频干扰、业务量分布不均和终端位置情况等因素，合理地为接入用户分配功率和带宽资源，提出一个适合于卫星通信系统的动态资源分配方案对整个卫星系统的性能提升具有非常重要的理论意义和现实价值。

1 相关工作

国内外学者对多波束卫星通信系统中的资源分配进行了广泛的研究。在带宽分配方面，Park等[10]为了防止多波束下行链路固有的限制（如功率、带宽，以及使用点波束）造成不必要的资源浪费而提出了一项调整每个波束带宽的资源管理技术，使流量需求与分配容量之间的差异最小，进而提出了灵活的波束管理算法，最终通过启发式搜索的方法来获得最优波束带宽。Ma等[11]为满足时变的流量需求设计了一个使用邻近策略优化的动态带宽分配框架，最大化利用率并保证多波束卫星系统中用户的服务质量（QoS）。文献[12-14]提出基于信息中心网络的缓存策略，旨在多波束卫星系统场景下，最小化用户终端的内容访问时延或降低带宽消耗。廖卫东[15]对卫星缓存资源的限制问题进行了研究，并提出了一种提高缓存利用率的策略，通过设置缓存阈值，并在卫星缓冲区达到阈值时调整发送到卫星的传输数据速率，进而合理地进行带宽资源的分配。

在功率分配方面，Srivastava等[16]提出一种新型动态功率分配算法来处理10 GHz及以上的多波束宽带卫星的动态功率分配问题，采用雨衰随机模型，通过贪婪方法形成具有相似功率需求的用户组，与以往提出的算法相比，该算法可以服务更多的用户，并且可以在很大的范围内最优化利用可用的计算资源。Yang等[17-18]提出了一种自适应功率分配策略，在保障用户动态服务质量的前提下提升了资源利用率和系统性能。Destounis等[19]针对为固定地面终端服务的卫星通信系统，提出了一种使用数学模型进行雨衰预测的动态功率分配算法，最小化未接收到所需服务质量的用户数量，解决了在天线的波束之间平衡功率分配的问题。Jia等[20]将卫星通信中星载资源不足的问题归纳为凸优化问题，结合波束间干扰问题，提出了一种基于黄金分割理论和次梯度迭代的联合功率和带宽资源分配方案。在此基础上，通过拉格朗日对偶理论和次梯度迭代得到最优解。评估结果表明，该方案在时延、带宽利用率方差、容量和公平性方面均优于对比方法，并且在实际场景中具有较好的鲁棒性。Chen等[21]将多波束卫星系统流量匹配中的功率分配问题归纳为一个优化问题，并将凸优化框架分解为不同的子问题，每个子问题由一个给定的智能体执行，提出基于猜想的多智能体Q-learning算法用于搜索最优功率分配方案，在低复杂度的前提下提高了系统的通信满意度和公平性。

在多维资源联合优化方面，Abdu等[22]考虑了载波和功率分配来匹配目标波束需求，简化优化问题，提出2个子问题。首先，估计每个波束所需的相邻频率载波数量；然后，根据先前的载波优化功率分配。与基准方法相比，这种载波和功率分配的方法在满足需求的同时，可以减少系统的总功率和载波数量，平均波束需求匹配和平均功耗方面的性能都有了很大的提升。文献[23-25]考虑到传统的频率复用模式，以有限的频谱效率损耗为代价，最小化波束间的干扰，研究了动态功率分配问题。此外，Kisseleff等[26]针对地球同步卫星多波束卫星系统缓存资源受限条件下载波资源分配的动态资源分配问题进行研究，以最大限度满足用户的总流量需求和频谱效率

为目标，提出改进算法来求解多目标优化问题。张沛等[27]针对多波束卫星系统中资源分配序列决策的多目标优化问题，提出了一种基于深度强化学习的多目标优化算法，对动态变化的系统环境和用户到达模型建模，以归一化处理后的频谱效率、能量效率和业务满意度指数的加权和作为优化目标，实现了系统和用户累计性能的优化。Ding等[28-30]首先提出了一种动态路由算法提高了能量效率和吞吐量，然后基于马尔可夫链提出了一种节能路由算法，将网络寿命延长了2倍以上，同时能保持更好的能量效率。Liu等[31]针对无线通信多用户功率分配问题，提出了一种多目标无线宽带定位系统的最优功率分配策略，首先，将实际的功率分配问题表述为理论多目标优化模型；然后，利用最短距离理想点法将上述模型转化为单目标优化问题；最后，利用遗传算法对优化问题进行求解，得到最优的功率分配结果。实验结果表明，与传统的定位算法相比，复杂度更低的多用户最优功率分配策略能有效地提高定位性能。

本文研究下行多波束低地球轨道（LEO）卫星同频组网资源配置问题，目标是在考虑公平性的前提下优化系统吞吐量。进一步地，将卫星系统总吞吐量和用户的公平性优化问题转化成多目标优化问题，将具有时间相关性的连续资源分配过程建模为马尔可夫过程，并利用深度强化学习进行求解。具体而言，为克服单一决策网络进行资源分配时存在的弊端，提出决策性能评估算法，根据评估网络的评估结果，调整决策网络参数从而优化资源分配方案，同时迭代更新评估网络自身参数以便进行决策网络的准确预测，从而实现多波束LEO卫星网络公平性和系统吞吐量的均衡优化。本文对各种“网络”的定义如表1所示。

2　系统模型

在多波束LEO卫星系统中，考虑用户下行通信场景，如图1所示。在每个波束中随机分布相同数量的用户，每个用户随机生成一个业务请求，该请求通过信关站上传到卫星，然后由卫星下载到目标用户。卫星为满足用户业务需求进行带宽和功率联合分配。由于卫星场景下地面用户大多是静止或缓慢移动的，移动距离与卫星的波束半径相比可以忽略不计，因此在本文中假设用户位置是不变的。由于卫星的星上资源受限，为了提高带宽利用率，在本文中每个波束都可以利用系统的全部带宽资源，但所有用户的和功率会受到卫星最大发射功率的限制。

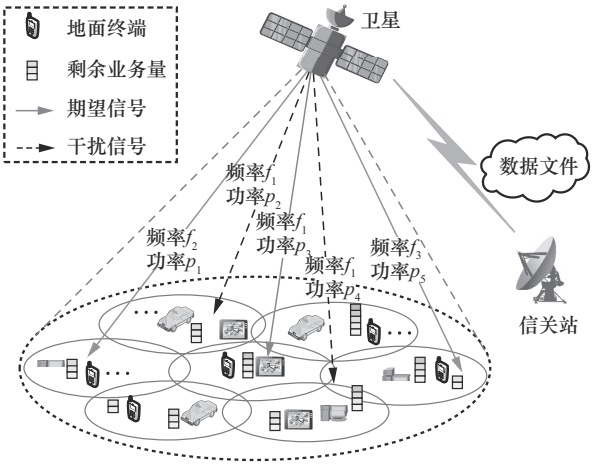

图1　多波束LEO卫星系统用户下行通信场景

考虑单颗多波束LEO卫星，卫星有M个波束，每个波束下有N个用户，第m个波束下第n个用户表示为(m,n)，位于$\Omega_{m,n}$处，每个用户初始化业务量大小为$D^0_{m,n}$。卫星系统的总带宽为B_{total}，被均匀划分为N_B个子信道；卫星最大发射功率为P_{total}。

t时刻卫星向用户(m,n)以功率$P_{m,n,t}$传输数据，用$[m,n,t]$表示该通信链路，链路带宽为$B_{m,n,t}$，星地链路增益为$h_{m,n}$。假定多波束LEO卫星信道矩阵$\boldsymbol{H}$为

$$\boldsymbol{H}=\begin{bmatrix} h_{1,1} & \cdots & h_{1,N} \\ \vdots & h_{m,n} & \vdots \\ h_{m,1} & \cdots & h_{M,N} \end{bmatrix} \tag{1}$$

$$h_{m,n}=G_t(\theta)+\text{PL}+G_r(\varphi) \tag{2}$$

其中，$G_t(\theta)$为链路的发射天线增益，θ为用户偏

表1　本文对各种“网络”的定义

表述	含义
卫星网络	由卫星作为中继节点的通信网络
决策–评估网络	本文提出的基于决策性能评估的资源分配神经网络
决策网络	决策–评估网络中负责生成资源分配决策的网络，由主网络和目标网络构成
评估网络	决策–评估网络中负责对分配决策进行评估的网络，由主网络和目标网络构成

离所在波束天线轴线的角度；PL为由信道环境引起的信号功率的损耗和衰落；$G_r(\varphi)$为链路的接收天线增益，φ为接收信号方向偏离接收天线轴线的角度。

考虑同频干扰问题，某时刻用户(m,n)的频率为f，受到的同频干扰$I_{m,n,t}$为

$$I_{m,n,t}=\sum_{\varphi\in\Xi_f,\varphi\neq[m,n,t]}P_\varphi h_\varphi \tag{3}$$

其中，Ξ_f表示频率为f的所有同频干扰信道，P_φ表示该链路的数据传输功率，h_φ表示该链路的信道增益。链路的信干噪比$\mathrm{SINR}_{m,n,t}$为

$$\mathrm{SINR}_{m,n,t}=\frac{P_{m,n,t}h_{m,n}}{I_{m,n,t}+N_0B_{m,n,t}} \tag{4}$$

其中，N_0为高斯白噪声功率谱密度。t时刻用户(m,n)与卫星间通信链路的通信速率$C_{m,n}^t$为

$$C_{m,n}^t=B_{m,n,t}\mathrm{lb}(1+\mathrm{SINR}_{m,n,t}) \tag{5}$$

$t+1$时刻用户(m,n)的剩余业务量为

$$D_{m,n}^{t+1}=D_{m,n}^0-\sum_{i=1}^{t}C_{m,n}^i \tag{6}$$

T时间段内系统总吞吐量为

$$C_{\mathrm{total}}=\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{t=1}^{T}C_{m,n}^t \tag{7}$$

用户公平性由Jain公平指数表示为

$$F=\frac{\left(\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{t=1}^{T}\frac{C_{m,n}^t}{D_{m,n}^0}\right)^2}{MN\sum_{m=1}^{M}\sum_{n=1}^{N}\left(\sum_{t=1}^{T}\frac{C_{m,n}^t}{D_{m,n}^0}\right)^2} \tag{8}$$

其中，$\frac{C_{m,n}^t}{D_{m,n}^0}$表示用户$(m,n)$的业务满意指数。

完整的优化问题为

$$\begin{aligned}&\mathrm{P1:}\ \max C_{\mathrm{total}}\\&\mathrm{P2:}\ \max F\\&\text{s.t. C1:}\ C_{m,n}^t\leqslant D_{m,n}^t\\&\mathrm{C2:}\ \sum_{m=1}^{M}\sum_{n=1}^{N}P_{m,n}\leqslant P_{\mathrm{total}}\\&\mathrm{C3:}\ \sum_{n=1}^{N}B_{m,n}\leqslant B_{\mathrm{total}}\ \forall m\end{aligned} \tag{9}$$

其中，P1表示最大化系统的总吞吐量；P2表示最大化系统的公平性，这是因为如果没有公平性的限制，在资源受限情况下，可能有用户始终不会分配到任何资源，这显然是不合理的。C1表示t时刻用户的吞吐量不大于该时刻的业务量；C2表示所有用户的总功率不大于卫星最大发射功率；C3表示同一波束内所有用户的总带宽不大于系统总带宽。

3　基于决策性能评估的网络资源分配算法

式(9)所示的优化问题属于序列决策问题，使用传统算法难以求解。因此本文提出了基于决策性能评估的资源分配网络，框架如图2所示。

传统的决策–评估网络只涉及决策和评估2个神经网络，而且每次都是在连续状态中更新参数，每次参数的更新前后都存在相关性，导致神经网络只能片面地看待问题，甚至会出现更坏的情况，即神经网络学习不到东西。为了提高神经网络的准确性和有效稳定性，本文所提决策–评估网络采用

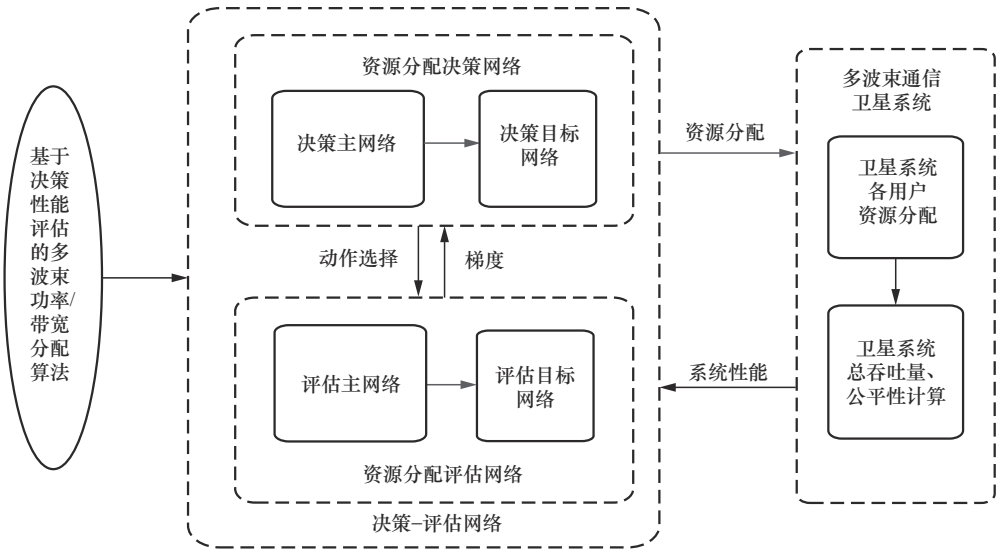

图2　基于决策性能评估的资源分配网络框架

DQN（deep Q network）的双神经网络结构来进行研究，分别是决策主网络、评估主网络、决策目标网络和评估目标网络。

策略梯度的基本思想是根据状态输出动作或者动作的概率。基于值的算法的基本思想是根据当前的状态计算采取每个动作的价值，然后根据价值贪心地选择动作。决策–评估网络则是上述2种思想的结合，决策网络采用策略梯度的算法设计来学习策略，评估网络则负责策略评估的值函数。也就是说，一方面决策网络不断地学习策略，其更新依赖评估网络的值函数；另一方面评估网络估计值函数，而值函数又是关于策略的函数。二者相互依赖，在训练过程中不断地迭代优化直至收敛，最终得到想要的最优结果。

决策–评估网络能够对多波束功率/带宽分配策略直接进行建模和学习，针对资源分配决策网络进行多波束功率/带宽分配的场景，设计评价决策性能优劣的评估网络，通过评估网络的评估结果对决策网络的参数进行调整，进而优化资源分配决策。同时评估网络进行自我评价以实现自身网络参数的更新，通过决策–评估网络迭代优化实现功率/带宽的最优分配。

首先，定义资源分配动作选择策略。该策略为了探索潜在的更优资源分配方案，在动作的选择上引入随机噪声 $\Re_t$，这样资源分配动作的决策机制从确定性过程变成了一个随机过程，再从这个随机过程中采样得到资源分配方案。

在多波束LEO卫星通信系统运行的第 t 个时刻，资源分配决策网络根据资源分配动作选择策略选择相应的功率/带宽分配决策 a_t，表示为

$$a_t = \mu(s_t|\theta^\mu) + \Re_t \tag{10}$$

其中，μ 为最优资源分配方案选择策略。μ 策略结合随机奥恩斯坦–乌伦贝克（OU, Ornstein-Uhlenbeck）噪声后形成了一个关于资源分配方案的随机过程，从中采样即 t 时刻资源分配决策网络选择的功率/带宽分配方案。随着训练过程的不断增加，该随机OU噪声的标准差逐步呈现下降趋势，具体来说，它依靠一个常数 α 来构建指数下降的趋势。这种方式使该决策网络达到先探索后开发的效果。

资源分配决策网络将 t 时刻选择的功率/带宽分配方案传送给多波束LEO卫星通信系统，由该分配方案得到系统与用户通信的实际总吞吐量值（或时延、能耗等其他反映系统性能的参数）以及 t+1时刻系统的业务需求。用户业务需求作为 t+1时刻的系统状态 s_{t+1}，吞吐量值作为环境的返回奖励值 r_t，联合 t 时刻的业务需求和资源分配决策 (s_t,a_t,r_t,s_{t+1}) 存入经验池中，作为训练决策/评估主网络的训练数据。

为解决资源分配决策–评估网络学习过程中学习样本相关性（连续的资源分配）导致的片面性学习问题，决策–评估网络从经验池中随机采用 N 个记忆向量训练决策主网络和评估主网络。

利用决策目标网络和评估目标网络分别计算下一时刻的资源分配动作 a_{t+1} 和 Q 值 Q'_{t+1}，得到训练的目标值 y_i 为

$$y_i = r_i + \gamma Q'\left(s_{i+1},\mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'}\right) \tag{11}$$

利用资源分配决策主网络和评估主网络计算 Q 值，获得评估网络的损失函数值为

$$L = \frac{1}{N}\sum_i \left(y_i - Q(s_i,a_i|\theta^Q)\right)^2 \tag{12}$$

在此阶段，使用均方误差函数作为评估网络的损失函数。通过误差反向传播，获取资源分配评估主网络的梯度，然后采用自适应矩估计优化器更新资源分配评估网络的参数。

完成评估网络的更新后，计算资源分配决策网络的策略梯度为

$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum\nolimits_i \nabla_a Q(s,a|\theta^Q)|_{s=s_i,a=\mu(s_i)} \nabla_{\theta^\mu}\mu(s|\theta^\mu)|_{s_i} \tag{13}$$

通过自适应矩估计优化器对资源分配决策网络的参数进行更新。前述的参数更新过程均指决策主网络和评估主网络参数更新，而目标网络参数（θ^μ 和 θ^Q）则采用软更新的方式进行更新，可以表示为

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'} \tag{14}$$

$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'} \tag{15}$$

其中，τ 为更新系数，通常取0.001。硬更新是指每隔一定步数完全复制参数，软更新则是指每一步都对目标网络参数进行更新，并且每次更新并非完全复制主网络的参数，而是根据新旧参数的衰减比例关系进行更新。这种更新方式使目标网络的参数随时间变化但变化较小，有助于网络进行稳定学习。基于决策性能评估的网络资源分配算法如算法1所示。

算法 1　基于决策性能评估的网络资源分配算法

随机初始化系统性能评估网络 $Q(s,a|\theta^Q)$，决

策网络$\mu(s|\theta^{\mu})$，权重θ^{Q}和θ^{μ}；

1) 初始化目标网络Q'和μ'，权重分别为$\theta^{Q'}\leftarrow\theta^{Q}$，$\theta^{\mu'}\leftarrow\theta^{\mu}$；

2) 初始化经验池R；

3) 对于episode从1至M，循环执行：

4) 初始化一个随机过程σ，用于资源分动作探索；

5) 初始化用户业务需求s_1；

6) 对于t从1至T，循环执行：

7) 根据当前的策略和探索的噪声选择动作$a_t=\mu(s_t|\theta^{\mu})+\sigma_t$；

8) 执行功率、带宽分配决策a_t，获得系统吞吐量r_t和下一时刻的业务需求s_{t+1}；

9) 将(s_t,a_t,r_t,s_{t+1})存入经验池R中；

10) 在R中随机抽取N个minibatch的样本(s_t,a_t,r_t,s_{t+1})，作为决策主网络和评估主网络的训练数据；

11) 决策目标网络根据下时刻状态s_{i+1}产生动作$\mu'(s_{i+1}|\theta^{\mu'})$，并将其传递给评估目标网络；

12) 评估目标网络根据式(11)，利用t时刻的系统奖励r_i和下一时刻动作$\mu'(s_{i+1}|\theta^{\mu'})$计算目标值$y_i=r_i+\gamma Q'(s_{i+1},\mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$，并传递给评估主网络；

13) 评估主网络计算实际Q值$Q(s_i,a_i|\theta^{Q})$，并传递给决策主网络；

14) 评估主网络更新梯度，根据式(12)通过最小化loss函数来更新评估主网络；

15) 决策主网络根据式(13)计算策略梯度以更新策略；

16) 利用式(14)和式(15)更新目标网络参数$\theta^{Q'}$和$\theta^{\mu'}$。

4　决策–评估网络结构

本文采用决策–评估网络来解决多目标资源分配问题，如图3所示。本文提出的基于决策–评估的网络资源分配算法的性能评估方案由基于策略梯度的决策网络和基于值函数的评估网络结合而成。多波束卫星通信系统环境输出系统状态与决策–评估网络进行交互。决策–评估网络根据环境状态选择资源分配决策。

资源分配决策主网络负责根据用户业务需求，选择当前资源分配动作a，用于和多波束LEO卫星通信系统交互生成系统总吞吐量、系统公平性指数和状态，输入为多波束LEO卫星通信系统环境状态，输出为资源分配动作，并且该决策网络根据评估网络返回的资源分配决策–评估结果修正网络参数。

4.1　状态空间

状态是对外界环境的描述，决策–评估网络需要借助状态进行后续的决策，定义决策网络中的状态为$\boldsymbol{s}$。状态随时间变化而改变，t时刻网络状态向量为$\boldsymbol{s}_t=\{\boldsymbol{H}^t,\boldsymbol{I}^t,\boldsymbol{C}^t,\boldsymbol{D}^t\}$，其中$\boldsymbol{H}^t$表示系统内各个用户的信道增益；$\boldsymbol{I}^t$表示用户接收到的同频干扰强度；$\boldsymbol{C}^t$表示用户的吞吐量；$\boldsymbol{D}^t$表示用户的业务量。

4.2　动作空间

动作是决策–评估网络的输出参数，用来调整

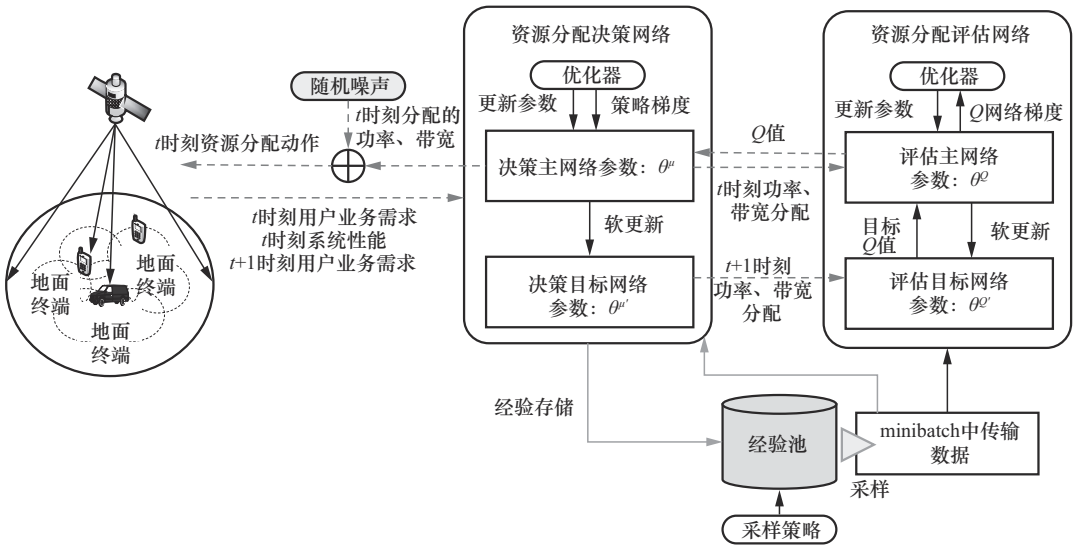

图3　决策–评估网络

系统环境中的可变信息，定义决策网络中的动作为a。网络动作a是针对下一时刻预测情况进行的资源分配决策，需要实施到真实系统中对资源变量进行调整。决策网络动作主要包括各波束的分配带宽、功率资源，这些资源参数的可行解组成该决策–评估网络的动作空间$\boldsymbol{A}$。对于$t \in \{1,2,3,\cdots\}$，系统将在考虑同频干扰的影响下对有业务需求的用户分配功率和带宽资源。动作集表示为$\boldsymbol{A}^t = \{\boldsymbol{P}^t, \boldsymbol{B}^t\}$，其中$\boldsymbol{P}^t$表示系统为用户分配的数据传输功率，$\boldsymbol{P}^t = \{P_{1,1,t}, P_{1,2,t}, \cdots, P_{M,N,t}\}$；$\boldsymbol{B}^t$表示系统为用户分配的数据传输带宽，$\boldsymbol{B}^t = \{B_{1,1,t}, B_{1,2,t}, \cdots, B_{M,N,t}\}$。

系统性能评估主网络根据决策网络输出的当前资源分配决策计算当前Q值，并通过计算损失函数进行网络参数更新。

4.3　奖励

评估网络的奖励值需体现决策网络做出的资源分配决策性能对系统性能的影响。对于每个时刻，环境根据当前状态、当前状态下的动作以及下一状态设计系统奖励值。奖励值的设计应与资源分配决策的目标有关。

考虑到优化目标P1，将系统总吞吐量作为第一个奖励R_1，表示为

$$R_1 = \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{t=1}^{T} C_{m,n}^{t} \tag{16}$$

考虑到优化目标P2，将系统公平性系数作为第二个奖励R_2，表示为

$$R_2 = \frac{\left(\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{t=1}^{T}\frac{C_{m,n}^{t}}{D_{m,n}^{0}}\right)^2}{N_{\text{total}}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(\sum_{t=1}^{T}\frac{C_{m,n}^{t}}{D_{m,n}^{0}}\right)^2} \tag{17}$$

其中，N_{total}表示卫星通信系统中的所有用户数，即$N_{\text{total}} = MN$。

为了加快模型的收敛速度，设置辅助奖励R_3为

$$R_3 = \sum_{t=1}^{T}\left[\sum_{m=1}^{M}\sum_{n=1}^{N} x_{mn}^{t} + \sum_{m=1}^{M}\sum_{n=1}^{N} u_{mn}^{t} + z^{t}\right] \tag{18}$$

其中

$$x_{mn} = \begin{cases} 5, & \text{用户资源分配不合理} \\ 0, & \text{其他} \end{cases} \tag{19}$$

$$u_{mn} = \begin{cases} 10, & \text{用户}t\text{时刻吞吐量大于业务量} \\ 0, & \text{其他} \end{cases} \tag{20}$$

$$z^{t} = \begin{cases} 80, & \text{所有用户功率大于总功率} \\ 0, & \text{其他} \end{cases} \tag{21}$$

资源分配不合理是指决策–评估网络给用户分配了带宽（或功率），但是没有给用户分配功率（或带宽）。

总的奖励设计为

$$R = \omega_1 R_1 + \omega_2 R_2 - \omega_3 R_3 \tag{22}$$

其中，ω_1、ω_2和ω_3表示3种奖励的权重，其取值范围为(0,1)；R_3表示惩罚。

5　仿真与分析

为了验证本文所提基于决策性能评估的网络资源分配算法的有效性，在不同系统频带资源下进行了仿真。仿真工具为PyCharm 2021.2.3（专业版），运行环境为塔式服务器，具体配置为CPU Intel(R) Core(TM) i7-10700F CPU @ 2.90 GHz，内存为16 GB。场景环境为python 3.9，TensorFlow版本为2.5.0。设置用户平均业务量为300 Mbit/s，最小业务量为200 Mbit/s，最大业务量为400 Mbit/s，业务类型为时延不敏感业务，用户每隔时间T更新业务需求。具体参数设置如表2所示。

表2　参数设置

参数	含义	取值
场景参数	卫星轨道高度x_k/km	600
	工作频段σ_k	Ka
	波束内用户数$p^{\max}$	5
	系统总用户数N_{total}	35
	系统波束数M	7
	系统总带宽B_{total}/ MHz	400
	带宽资源块总量N_B	10
	系统总功率p_{tot}/ W	1 000
	卫星最大发射天线增益G_t/ dBi	38.5
	用户终端最大接收天线增益G_r/ dBi	0
	高斯白噪声功率谱密度N_0/ (dBm·Hz^{-1})	–174
决策–评估网络参数	最大训练次数max_episodes	100 000
	Actor网络学习率α_a	0.001
	Critic网络学习率α_c	0.000 1
	经验回放池容量RB	10 000
	minibatch容量Ω	256
	软更新系数τ	0.001
	折扣因子γ	0.99
	噪声方差Var	0.2
	噪声折扣因子γ_N	0.999 94

5.1　算法收敛性分析

本文所提基于决策性能评估的网络资源分配算法具有决策和评估 2 个网络，所以本文用 2 个 loss 函数来衡量网络的收敛性，其中，A_loss 衡量决策网络的收敛性，C_loss 衡量评估网络的收敛性。

决策网络和评估网络的损失曲线如图 4 所示。随着训练次数的增加，决策网络的损失逐渐减小，最终趋于相对较低的水平，这表明决策网络已经通过不断学习逐渐提高了其性能，并且达到了稳定状态，不再出现显著的损失下降。评估网络的损失曲线也呈类似趋势。随着训练次数的增加，评估网络的损失逐渐减小，并最终保持在较低水平。这反映了评估网络的有效性和稳定性，表明它已经成功地学习了数据的特征和模式。

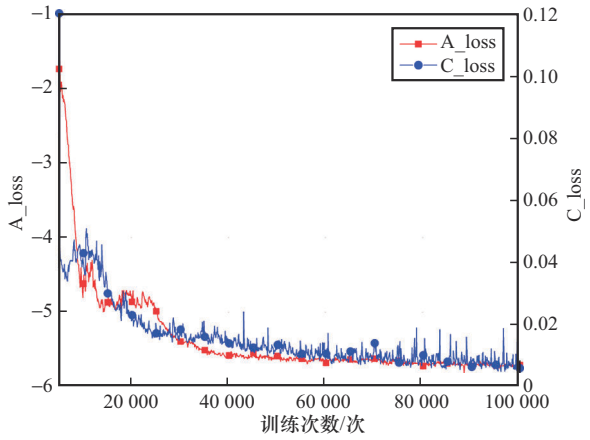

图 4　决策网络和评估网络的损失曲线

系统奖励随训练次数的变化如图 5 所示。随着训练的不断进行，系统的奖励逐渐上升，并最终保持在较高水平。这表示系统的奖励函数趋于稳定，系统在处理任务时获得了良好的性能。

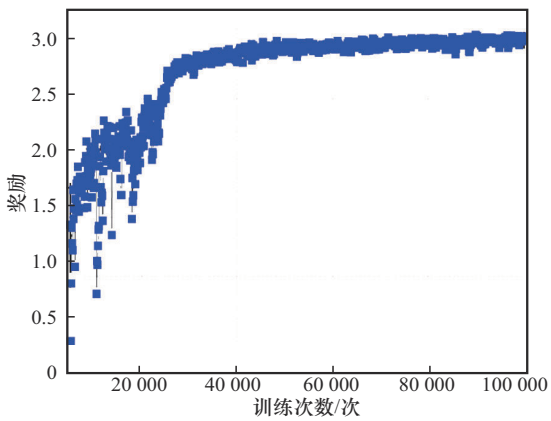

图 5　系统奖励随训练次数的变化

图 6 和图 7 分别给出了系统吞吐量和用户公平性随训练次数的变化，2 种优化目标均随训练次数逐渐上升，表明本文所提算法在多目标联合优化方面表现出良好的性能。

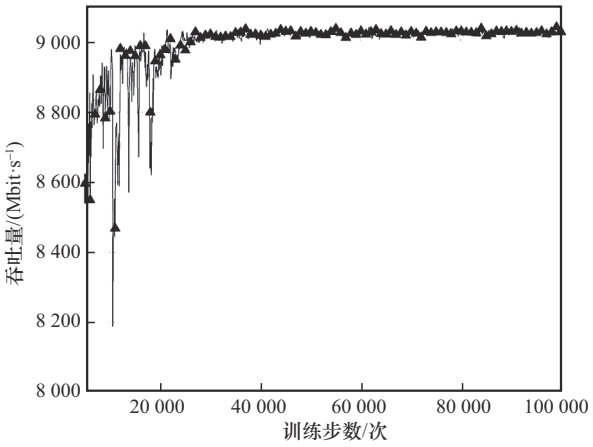

图 6　系统吞吐量

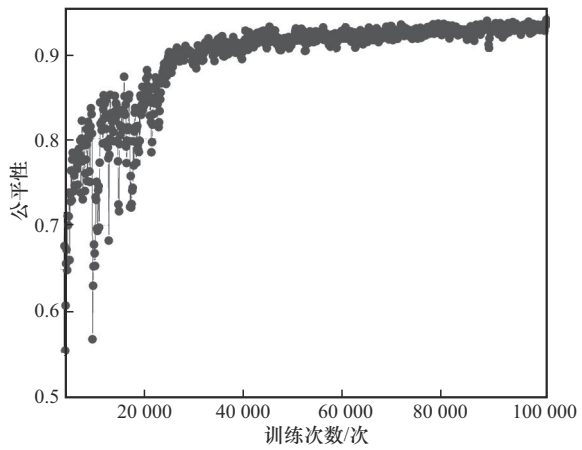

图 7　用户公平性

从损失曲线、奖励曲线以及系统性能曲线的观察中可以明显看出，决策网络和评估网络在训练中都表现出了出色的性能，其损失逐渐减小并趋于稳定，而系统的奖励也达到了一个较高的水平。这些趋势的同时出现表明决策–评估网络整体的收敛性较好，成功地学习了任务并实现了稳定的性能。

5.2　系统吞吐量对比

吞吐量性能是通信系统的重要性能指标之一，表示系统单位时间的业务处理能力。基于决策性能评估的网络资源分配算法、资源平均分配算法、资源随机分配算法和传统四色复用算法的吞吐量对比如图 8 所示。

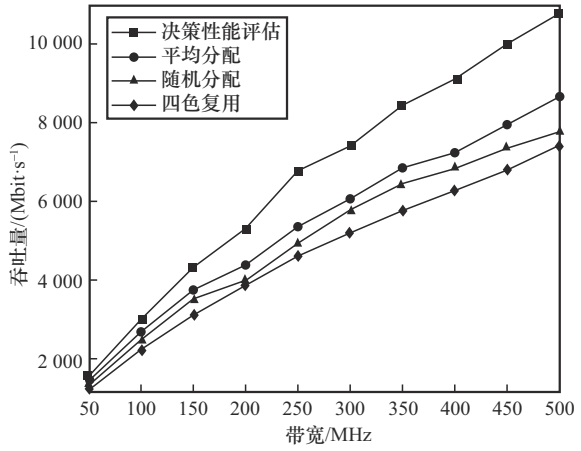

图8　不同算法的吞吐量对比

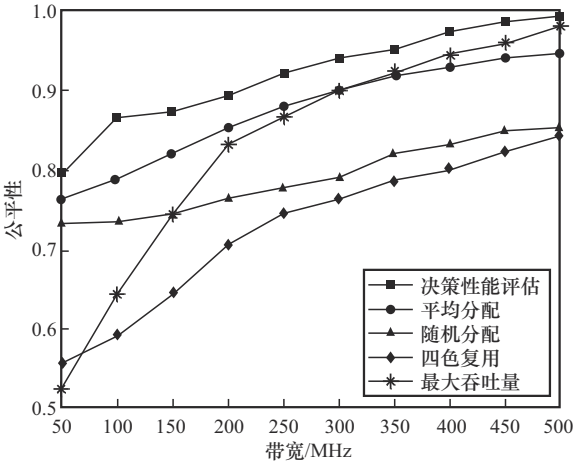

图9　不同算法的公平性对比

系统吞吐量随系统可用带宽的增加而逐渐增加，并且在带宽范围为50～500 MHz时，基于决策性能评估的网络资源分配算法的吞吐量性能优于资源平均分配算法、资源随机分配算法以及传统四色复用算法。基于决策–评估的网络资源分配算法吞吐量性能较资源平均分配算法平均提升20.602%，较资源随机分配算法平均提升29.735%，这是因为本文所提算法可以实现吞吐量和公平性的联合优化，并且可以根据不同的用户业务需求实时动态调整资源分配决策，而传统优化算法的实时性较差，不能根据业务需求进行动态调整；较四色复用算法平均提升40.913%，这是因为本文所提算法不同波束间使用的是相同的频带，用户的可用带宽相较于四色复用有较大的提升。综上，本文提出的基于决策性能评估的网络资源分配算法能在一定程度上提高系统的总吞吐量。

5.3　用户公平性对比

仅优化系统的吞吐量容易导致系统将大部分功率和带宽资源分配给信道状态较好的用户，这会造成部分信道状态较差的用户的业务需求得不到满足，使系统的公平性降低。为了避免上述情况，本文同时优化了系统的吞吐量和公平性。

为了更好地将用户吞吐量、用户业务需求和系统公平性结合，本文对于系统的公平性计算参考了Jain公平指数，并将其中的用户吞吐量替换为用户业务满足指数。基于决策性能评估的网络资源分配算法、资源平均分配算法、资源随机分配算法和传统四色复用算法的公平性对比如图9所示。其中，最大吞吐量算法表示本文所提决策性能评估的网络资源分配算法在仅考虑吞吐量最大化时所达到的公平性。

当系统带宽较小时，用户能达到的吞吐量小于业务需求，因此吞吐量对于优化目标的增益大于公平性，此时智能体优先把资源分配给信道状态好的用户以提升系统多目标奖励，导致用户公平性较低；当系统带宽增大时，用户能达到的吞吐量可能会接近甚至超过业务需求，因此公平性对于优化目标的增益会变大，从而公平性会随着带宽增加而呈现出升高的趋势。在带宽范围为50～500 MHz时，基于决策性能评估的网络资源分配算法的公平性优于资源平均分配算法、资源随机分配算法以及四色复用算法。基于决策性能评估的网络资源分配算法系统公平性较平均分配算法平均提升5.2%，较随机分配算法平均提升16.5%，较四色复用算法平均提升26.6%。这是因为传统算法不能同时兼顾系统吞吐量性能和用户公平性，同时决策性能评估算法可以适应用户的不同业务量需求，并根据业务大小动态调整资源分配决策。

同时，本文提出的吞吐量和公平性联合优化相较于仅考虑吞吐量的单一优化，在带宽受限时能显著提升用户的公平性。当带宽较小时该提升比较明显，这是由于此时最大吞吐量算法把资源优先分配给信道状态好的用户而忽略了公平性，而本文所提算法能在一定程度上兼顾用户公平性；当带宽较大时二者性能接近，这是由于此时信道较好用户的业务需求已经完成，要达到最大化吞吐量的目的需要将资源分给信道较差的用户，因此在一定程度上提高了公平性。这表明本文提出的基于决策性能评估的网络资源分配算法能在提高系统总吞吐量的同时兼顾系统的公平性。

6 结束语

本文对多波束LEO卫星同频组网中的资源配置问题进行研究。综合考虑了多波束间同频干扰、业务量分布不均和终端位置等多重因素，通过建立资源配置模型，以系统公平性和吞吐量性能的联合优化为目标，采用深度强化学习方法，实现了动态资源分配方案的优化。针对单一决策网络资源分配场景，提出了决策性能评估算法，引入了决策–评估网络资源分配技术，通过评估网络的结果来调整决策网络参数，以提高资源分配效果。仿真结果表明，本文所提算法具有较好的鲁棒性和系统性能。

参考文献：

[1] YANG Z, HUSSEIN J A, XU P, et al. A novel hybrid successive interference cancellation for uplink wireless power transfer NOMA in Internet of things[J]. IEEE Transactions on Vehicular Technology, 2023, 72(5): 6090-6102.

[2] YANG Z, XU P, CHEN G J, et al. Performance analysis of IRS-assisted NOMA networks with randomly deployed users[J]. IEEE Systems Journal, 2023, 17(2): 1853-1864.

[3] 许国良, 谭峰, 冉泳屹, 等. 面向多波束卫星系统的波束跳变与覆盖控制联合优化算法[J]. 通信学报, 2023, 44(4): 78-86.
XU G L, TAN F, RAN Y Y, et al. Joint beam hopping and coverage control optimization algorithm for multibeam satellite system[J]. Journal on Communications, 2023, 44(4): 78-86.

[4] JOROUGHI V, VÁZQUEZ M Á, PÉREZ-NEIRA A I. Generalized multicast multibeam precoding for satellite communications[J]. IEEE Transactions on Wireless Communications, 2017, 16(2): 952-966.

[5] LIN Z, AN K, NIU H H, et al. SLNR-based secure energy efficient beamforming in multibeam satellite systems[J]. IEEE Transactions on Aerospace and Electronic Systems, 2023, 59(2): 2085-2088.

[6] RAMÍREZ T, MOSQUERA C. Resource management in the multibeam NOMA-based satellite downlink[C]//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE Press, 2020: 8812-8816.

[7] STOREK K U, KNOPP A. Fair user grouping for multibeam satellites with MU-MIMO precoding[C]//Proceedings of the IEEE Global Communications Conference. Piscataway: IEEE Press, 2017: 1-7.

[8] ZHONG K Y, CHENG Y J, YANG H N, et al. LEO satellite multibeam coverage area division and beamforming method[J]. IEEE Antennas and Wireless Propagation Letters, 2021, 20(11): 2115-2119.

[9] 尹展, 孙晨华. 多波束卫星移动通信系统的同频干扰研究[J]. 无线电通信技术, 2016, 42(2): 23-26.
YIN Z, SUN C H. Research of co-frequency interference in multi-beam satellite mobile communication system[J]. Radio Communications Technology, 2016, 42(2): 23-26.

[10] PARK U, KIM H W, OH D S, et al. Flexible bandwidth allocation scheme based on traffic demands and channel conditions for multibeam satellite systems[C]//Proceedings of the IEEE Vehicular Technology Conference. Piscataway: IEEE Press, 2012: 1-5.

[11] MA S J, HU X, LIAO X L, et al. Deep reinforcement learning for dynamic bandwidth allocation in multi-beam satellite systems[C]//Proceedings of the IEEE 6th International Conference on Computer and Communication Systems. Piscataway: IEEE Press, 2021: 955-959.

[12] WU H, LI J, LU H C, et al. A two-layer caching model for content delivery services in satellite-terrestrial networks[C]//Proceedings of the IEEE Global Communications Conference. Piscataway: IEEE Press, 2016: 1-6.

[13] LIU S J, HU X, WANG Y P, et al. Distributed caching based on matching game in LEO satellite constellation networks[J]. IEEE Communications Letters, 2018, 22(2): 300-303.

[14] DORO S, GALLUCCIO L, MORABITO G, et al. SatCache: a profile-aware caching strategy for information-centric satellite networks[J]. Transactions on Emerging Telecommunications Technologies, 2014, 25(4): 436-444.

[15] 廖卫东. LEO卫星系统中数据存储与转发策略研究[D]. 北京: 北京邮电大学, 2018.
LIAO W D. The research on data storage and forwarding strategy in LEO satellite system[D]. Beijing: Beijing University of Posts and Telecommunications, 2018.

[16] SRIVASTAVA N K, CHATURVEDI A K. Flexible and dynamic power allocation in broadband multi-beam satellites[J]. IEEE Communications Letters, 2013, 17(9): 1722-1725.

[17] YANG Z, HUSSEIN J A, XU P, et al. Power allocation study for non-orthogonal multiple access networks with multicast-unicast transmission[J]. IEEE Transactions on Wireless Communications, 2018, 17(6): 3588-3599.

[18] YANG Z, XU P, HUSSEIN J A, et al. Adaptive power allocation for uplink non-orthogonal multiple access with semi-grant-free transmission [J]. IEEE Wireless Communications Letters, 2020, 9(10): 1725-1729.

[19] DESTOUNIS A, PANAGOPOULOS A D. Dynamic power allocation for broadband multi-beam satellite communication networks[J]. IEEE Communications Letters, 2011, 15(4): 380-382.

[20] JIA M, ZHANG X M, GU X M, et al. Interbeam interference constrained resource allocation for shared spectrum multibeam satellite communication systems[J]. IEEE Internet of Things Journal, 2019, 6(4): 6052-6059.

[21] CHEN R, HU X, LI X H, et al. Optimum power allocation based on traffic matching service for multi-beam satellite system[C]//Proceedings of the 5th International Conference on Computer and Communication Systems. Piscataway: IEEE Press, 2020: 655-659.

[22] ABDU T S, LAGUNAS E, KISSELEFF S, et al. Carrier and power assignment for flexible broadband GEO satellite communications system[C]//Proceedings of the IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications. Piscataway: IEEE Press, 2020: 1-7.

[23] ZHANG P, WANG X H, MA Z G, et al. Joint optimization of satisfaction index and spectrum efficiency with cache restricted for resource allocation in multi-beam satellite systems[J]. China Communications, 2019, 16(2): 189-201.

[24] EFREM C N, PANAGOPOULOS A D. Dynamic energy-efficient power allocation in multibeam satellite systems[J]. IEEE Wireless Communications Letters, 2019, 9(2): 228-231.

[25] ARAVANIS A I, SHANKAR M R B, ARAPOGLOU P D, et al. Power allocation in multibeam satellite systems: a two-stage multi-objective optimization[J]. IEEE Transactions on Wireless Communications,

2015, 14(6): 3171-3182.
[26] KISSELEFF S, SHANKAR B, SPANO D, et al. A new optimization tool for mega-constellation design and its application to trunking systems [C]//Proceedings of the Advances in Communications Satellite Systems. Proceedings of the 37th International Communications Satellite Systems Conference (ICSSC-2019). London: IET, 2019: 1-15.[LinkOut]
[27] 张沛, 刘帅军, 马治国. 基于深度增强学习和多目标优化改进的卫星资源分配算法[J]. 通信学报, 2020, 41(6): 51-60.
ZHANG P, LIU S J, MA Z G. Improved satellite resource allocation algorithm based on DRL and MOP[J]. Journal on Communications, 2020, 41(6): 51-60.
[28] DING Z M, SHEN L F, CHEN H Y, et al. Energy-efficient relay-selection-based dynamic routing algorithm for IoT-oriented software-defined WSNs[J]. IEEE Internet of Things Journal, 2020, 7(9): 9050-9065.
[29] DING Z M, SHEN L F, CHEN H Y, et al. Energy-efficient topology control mechanism for IoT-oriented software-defined WSNs[J]. IEEE Internet of Things Journal, 2023, 10(15): 13138-13154.
[30] DING Z M, SHEN L F, CHEN H Y, et al. Residual-energy aware modeling and analysis of time-varying wireless sensor networks[J]. IEEE Communications Letters, 2021, 25(6): 2082-2086.
[31] LIU F, YAN J. Optimal power allocation strategy for multi-target wireless wideband localization system via genetic algorithm[C]//Proceedings of the 2019 International Conference on Computer, Information and Telecommunication Systems (CITS). Piscataway: IEEE Press, 2019: 1-4.

[作者简介]

王朝炜（1982– ），男，陕西西安人，博士，北京邮电大学副教授、博士生导师，主要研究方向为下一代移动通信技术、无线传感器与IoT技术等。

庞明亮（1998– ），男，山东济南人，北京邮电大学博士生，主要研究方向为卫星通信、IoT和信号处理。

王粟（1982– ），男，安徽霍邱人，中国移动通信集团有限公司高级工程师，主要研究方向为5G移动通信网络规划与优化等。

赵玲莉（2000– ），女，河北保定人，北京邮电大学硕士生，主要研究方向为卫星通信、无线通信和资源管理。

高飞飞（1980– ），男，陕西西安人，博士，清华大学长聘教授、博士生导师，主要研究方向为大规模MIMO、通感一体化等。

崔高峰（1987– ），男，河南驻马店人。博士，北京邮电大学副教授、博士生导师，主要研究方向为空天通信、卫星通信、卫星网络资源管理。

王卫东（1967– ），男，内蒙古包头人。博士，北京邮电大学教授、博士生导师，主要研究方向为卫星通信、移动通信、物联网等。