引用本文：王寒凝，王江，齐钰，等. 意图驱动光网络的冲突解决方法[J]. 光通信技术，2024，48(1)：74-79.

意图驱动光网络的冲突解决方法

王寒凝[1]，王 江[1]，齐 钰[1]，滕 云[2]，杨 辉[2*]

（1.61932 部队，北京 100000；2.北京邮电大学 信息光子学与光通信国家重点实验室，北京 100876）

摘要：在意图驱动光网络中，由于同时下发的意图数量突然增多或底层物理资源的限制，部分用户的意图请求无法连接成功，导致意图冲突。提出一种基于强化学习的意图驱动光网络的冲突解决方法，根据强化学习输出的用户意图回退策略，更新用户意图对应的网络需求指标，在保障用户意图都能够成功连接的前提下使当前网络中的平均服务质量（QoS）值最大。实验结果表明，提出的方法可以有效提升当前网络环境中用户的平均 QoS 值，降低网络阻塞率。

关键词：意图驱动光网络；意图冲突；强化学习；用户协商；平均服务质量

中图分类号：TN256　**文献标志码**：A　**文章编号**：1002-5561(2024)01-0074-06

DOI：10.13921/j.cnki.issn1002-5561.2024.01.014　**开放科学(资源服务)标识码(OSID)**：

Conflict resolution method of intention-driven optical network

WANG Hanning[1], WANG Jiang[1], QI Yu[1], TENG Yun[2], YANG Hui[2*]

(1. Unit 61932, Beijing 100000, China; 2. State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: In intent-driven optical networks, due to the sudden increase in the number of intents sent at the same time or the limitation of underlying physical resources, some users' intent requests cannot be successfully connected, resulting in intent conflicts. In this paper, an intent-driven optical network conflict resolution method based on reinforcement learning is proposed. According to the user intention rollback policy output by reinforcement learning, the network demand index corresponding to the user intention is updated, and the average quality of service (QoS) value in the current network is maximized on the premise of ensuring that the user intention can be successfully connected. The experimental results show that the proposed method can effectively improve the average QoS value of users in the current network environment and reduce the network blocking rate.

Key words: intention-driven optical network, intention-conflict, reinforcement learning, user negotiation, average quality of service

0 引言

随着软件定义光网络(SDON)和网络功能虚拟化(NFV)的快速发展，能够实现智能运维的定义光网络将成为下一个研究热点。与传统的 SDON 不同，意图驱动的光网络采用基于意图的北向接口，将网络资源管理抽象为一个“黑盒系统”，自动转译业务对网络能力的需求，将传统的人工配置方式逐步升级为自适应的智能管控。用户无需提供具体的网络指标(如带宽、时延)[1]，只需使用自然语言表达网络需求，即可将业务部署在网络中。但是，意图驱动光网络在用户意图转译和验证后下发至网络设备层的过程中，由于当前时刻网络中的用户意图数量突然增加或底层物理资源的限制，会导致多用户之间的意图冲突[2]。

目前，针对网络资源分配冲突问题已有广泛的研究。文献[3]提出当网络中出现阻塞时，通过对用户降级来减少阻塞的方案，但该方案会影响用户的服务质量(QoS)，甚至导致低优先级用户请求无法成功连接。在网络资源紧张时期，文献[4]将时延不敏感业务请求错峰提供，但互联网服务提供商和上层应用并不属于同一方利益实体，因此不能进行联合优化。

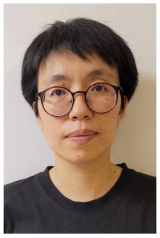
在多用户意图冲突方面，文献[5]提出了一种解决方案，用于检测当前网络中承载的意图是否存在冲

收稿日期：2023-05-01。

基金项目：国家自然科学基金面上项目(62271075)资助。

作者简介：王寒凝(1978—)，女，博士，高级工程师。

*** 通信作者**：杨辉(1987—)，男，博士，教授，主要研究方向为意图驱动光网络、多波段光网络、区块链光网络。



突，但检测出意图冲突后如何自动解决，目前没有相应有效的研究。之前提出的资源冲突解决方案也无法直接应用在意图驱动光网络这种新场景中。

随着人工智能的迅速发展，强化学习在资源分配领域被广泛应用[6]。强化学习能在实践中不断学习，通过不断地试错，并根据反馈的结果不断调整解决方案，从而寻找出能够产生最大效益的习惯性策略。因此，强化学习模型适合意图冲突问题中训练数据无标签的场景，即不存在用户意图到冲突解决策略的映射数据；同时，强化学习中的智能体可以根据网络性能指标的反馈不断寻找更优的意图回退策略[7]。因此，为解决意图驱动光网络中多用户之间的意图冲突问题，本文基于强化学习提出一种意图驱动光网络的冲突解决方法。

1 意图驱动光网络冲突解决方法

意图驱动光网络架构如图 1 所示。该网络框架包含意图层和网络层。意图层包含意图的转译和验证，网络层包含网络设备的自动化配置和网络策略的保障与自优化[8]。在意图层，智能体从知识库中获取转译后的用户意图，并生成相应的策略动作下发到网络层，随着数据库中用户意图的变化动态更新知识库。网络层中，不同节点之间的数字表示节点距离。当意图层往网络层下发策略时，可能会产生意图冲突(图 1 中感叹号位置)。网络层根据下发的策略在无线设备、物联网(IOT)设备和传输设备上自动配置相应的网络变更。当用户下发网络连接请求后，意图驱动光网络架构会形成一个闭环，自动建立一个保障用户需求的网络连接。例如，当用户陈述一个意图“从 A 到 C 建立一条高性能连接”，其转译策略会首先选取路由路径，然后选择其中符合带宽资源要求的路径进行传输，最后进行波长资源分配。

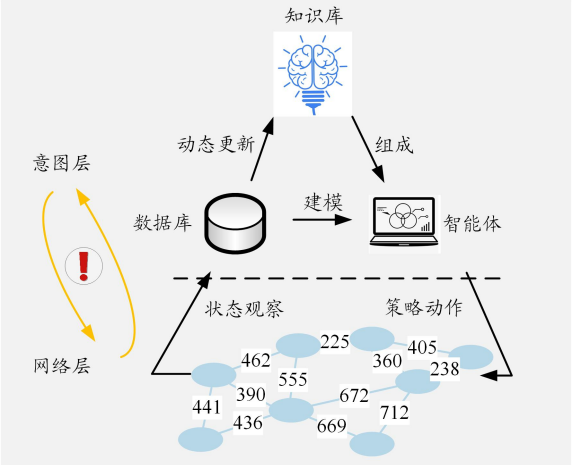


图 1 意图驱动光网络架构图

1.1 数据预处理

意图驱动光网络将用户使用自然语言表达的连接请求转译为具体的网络指标要求，包括带宽、时延、抖动、丢包率等多维度网络指标[9]，但不同指标之间的数值表示差距较大。为了后续更方便地进行模型训练，防止模型出现无法收敛的情况，在数据预处理阶段，本文将用户意图的转译结果进行归一化处理，提高意图冲突协商模型的准确度。

在意图驱动光网络中，由于不同用户的连接请求差异较大，当出现意图冲突时，无法为每一用户制定意图冲突的解决方案。因此，文献[10]将意图驱动光网络中的用户意图按照请求类型划分为会话类业务、视频类业务、游戏交互类业务和网页浏览类业务 4 种业务类型。其中，会话类业务时延容忍度较低，带宽要求适中；视频类业务的带宽容忍度较低，但时延容忍度较高；游戏交互类业务带宽和时延容忍度较低；网页浏览类业务对时延要求较低，但对丢包率要求高。在不同的意图冲突场景中，光网络可以根据冲突意图的类型选择适当的回退策略。

1.2 强化学习模型

与传统有监督和无监督学习相比，强化学习不存在分类标签的训练数据，也没有输出值，只包含数据特征[11]，因而被应用在游戏、资源调度和自动控制等领域中。强化学习有 2 种求解方式：基于策略的迭代和基于值的迭代，其代表性算法包含 Q 学习(Q-Learning)算法、深度 Q 网络(DQN)算法、深度确定性策略梯度(DDPG)算法等。强化学习模型原理结构图如图 2 所示。该模型包含 3 个关键参数：1)环境状态 s，即当前环境的状态，通常用一组向量空间表示；2)动作 a，即时刻智能体采取的动作；3)奖励 r，即当前时刻智能

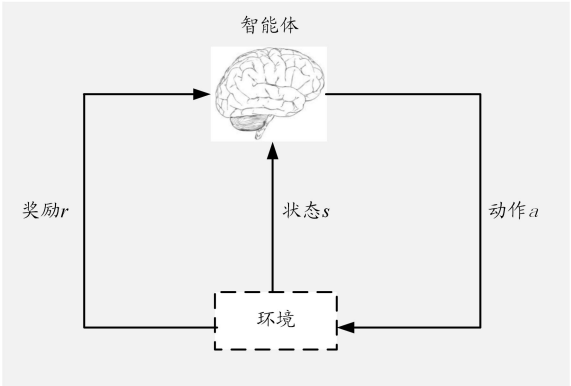


图 2 强化学习模型原理结构图

体在环境状态 s 的情况下采取动作 a 之后，在下一时刻所获得的奖励值[12]。图 2 中，智能体起初不存在任何可供参考的训练数据，通过初始动作 a 环境的交互，并根据环境状态 s 对当前选择的动作 a 的奖励 r 反馈，逐步调整动作，在不断试错的实践过程中迭代出针对不同环境状态下最优的动作映射。

本文采用基于策略迭代的强化学习模型解决意图驱动光网络中多用户之间的意图冲突问题。由于在意图驱动光网络中，用户请求的连接意图只说明了想要达到的网络连接效果，并未指明网络指标的具体要求，其相关的网络指标只是意图转译的结果，因此当网络环境中出现资源紧缩，存在不能成功配置的意图请求时，可在存在意图冲突的用户之间进行协商，意图协商方式如图 3 所示。蓝色的圆柱体表示路由器，R_1、R_2、R_3 表示不同路由器的编号。不同颜色虚线表示当前网络中承载的不同业务；红色实线表示当前到达的用户意图的路由路径，其中有 1 条路由路径被阻塞。虽然通过将产生冲突的用户意图转译结果进行适当的回退操作[13]可以解决意图驱动光网络中意图冲突问题，但在已有的研究成果中，不存在从意图冲突场景到用户意图回退方式的映射数据，需要根据产生冲突时刻的链路资源自适应地不断调整回退方式，而这种情景恰好符合强化学习模型的训练方式。

本文将出现意图冲突的意图网络场景中用户进行回退操作的过程抽象成马尔科夫决策过程，建模一个强化学习模型[14]。即将划分的每一类型的用户建模为一个智能体，通过智能体在产生冲突的意图网络环境状态中选择不同类型的用户意图回退方式，链路可用资源发生变化，并对当前智能体做出的动作反馈奖励值，智能体根据这个奖励值不断调整回退的动作，找到最优的意图协商策略解决意图冲突。在此场景中，强化学习模型参数情况如下：

1)动作 a。产生冲突的用户意图带宽改变量。

2)环境状态 s。产生冲突时刻，意图冲突用户所处链路的可用资源。

3)奖励 r。奖励函数中的奖励部分为当前时刻网络中可用资源变化量，惩罚部分为用户意图对应的网络指标归一化之后的值，并根据用户类型为各指标赋予不同的权重。

4)状态-动作空间。由于状态改变空间是确定的，所以状态-动作转移概率设定为 1。

5)折扣因子。针对此细粒度复杂任务，将折扣因子设定为 0.7，适当考虑往前的动作。

1.3 DDPG 的求解算法

在建立完强化学习模型后，本文采用 DDPG 算法训练强化学习模型。由于基于值迭代的算法，如离线学习的 Q-Learning 的动作空间不适合高维连续的情景，会导致动作空间爆炸；基于策略梯度(PG)的算法则在经历一次完整的回合后才进行一次参数更新，效率较低，且在本文意图驱动光网络冲突解决的场景中，模型的动作空间为带宽改变量，是一段连续的区间。因此，本文选择融合了基于策略迭代和基于值迭代 2 种方式的 DDPG 算法求解强化学习模型。DDPG 网络结构图如图 4 所示。

DDPG 算法结合了 Actor-Critic 算法和 DQN 算法。其中，DQN 改进了使用 Q 表的形式存储状态-动作映射，借鉴深度学习的方法，将状态-动作映射表示为神经网络结构；基于值迭代的 Q-Learning 算法虽然是单步更新，但仅适合离散的低维状态动作空间；基于策略迭代的算法可以在连续的动作空间中选取当前状态下的最优动作，但 PG 则是回合更新，学习效率较低。Actor-Critic 算法融合了基于策略迭代和基于值迭

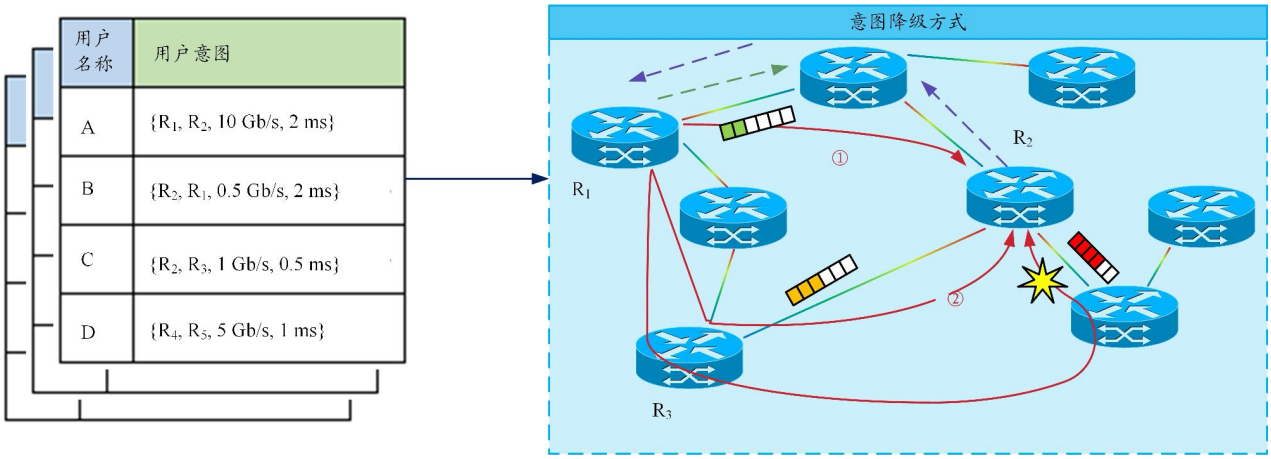


图 3 意图协商方式

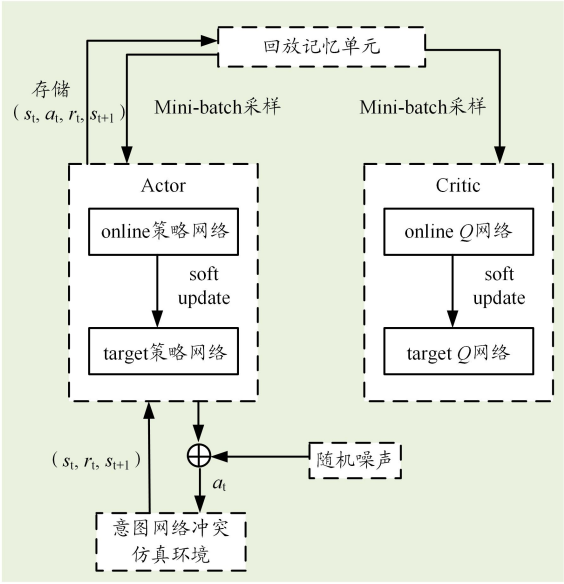


图 4 DDPG 网络结构图

代这 2 种方式,Actor 是一个基于策略迭代的网络,无需估算值函数,通过训练神经网络更新策略参数寻找最佳策略,Critic 是一个基于值迭代的网络,通过估算每个状态下的值函数寻找最佳策略。这 2 种方式的结合避免了模型的学习速率较慢的问题。Actor 不断迭代输入不同状态下选择每种可能的动作概率值,Critic 同时也不断迭代,对每个状态下选择的动作打分,输出奖励值。基于 DDPG 算法训练建立意图网络冲突强化学习模型,其 DDPG 网络参数如表 1 所示。

表 1 DDPG 网络参数

参数	数值	说明
LR_A	0.001	Actor 学习速率
LR_C	0.002	Critic 学习速率
DAMMA	0.9	奖励折扣率
TAU	0.01	软更新速率
MEMORY_CAPACITY	10 000	回放记忆单元大小

由于意图网络中发生冲突时的环境状态较多,若使用 DQN 算法计算产生的冲突时刻,在当前链路可用的资源环境状态下,对应的解决冲突则需采用每一回退动作的概率值,这导致模型的计算开销较大,同时要预先准备许多的训练样本数据才能得到准确的概率值。DDPG 算法改进上述随机策略,采用确定性策略,即选取同一状态下概率最大的动作,策略定义如式(1)所示。

$$\pi_\theta(s)=a \tag{1}$$

DDPG 网络基于 Q 值的确定性 PG 的梯度计算公式如式(2)所示。

$$\nabla_\theta J(\pi_\theta)=E_{S\sim\rho^\pi}[\ \nabla_a \pi_\theta(s)\ \nabla_\theta Q_\pi(s,a)\big|a=\pi_\theta(s)\ ] \tag{2}$$

其中,ρ^π 为状态的采样空间,$Q_\pi(s,a)\big|a=\pi_\theta(s)$ 表示回报 Q 函数对动作的导数。

DQN 是直接将评估网络中的参数赋值过去,而 DDPG 使用了软更新的方式,即每次对 target 网络中的参数只更新一小步,却能大大提高模型的稳定性,如式(3)所示。

$$\begin{cases} w'\leftarrow\tau w+(1-\tau)\,w' \\ \theta'\leftarrow\tau\theta+(1-\tau)\,\theta' \end{cases} \tag{3}$$

其中,w' 和 w 分别表示目标和当前价值网络的参数,θ' 和 θ 分别表示目标和当前策略网络的参数,τ 表示更新系数(默认值一般取 0.1 或 0.01)。

由图 4 可知,模型训练的初始状态来自于回放记忆单元的 Mini-batch 随机采样。但在意图网络冲突协商问题中,使用随机采样的初始化状态,可能由于初始值过小,在迭代更新时陷入局部最优值,网络较难收敛。此处,采用回放记忆单元中能满足冲突用户意图的最低标准连接请求。I_r 作为模型训练的初始状态,如式(4)所示。

$$\left\{\begin{array}{l} S_t^{\text{init}}=\min\limits_{I_r,\text{where}\,I_r\in C}\big|I_r\big| \\ C=\{\,I_r\,\big|\,(I_r,V_{I_r})\in\text{memory},V_{I_r}=1\,\} \end{array}\right\} \tag{4}$$

其中,C 表示在所有 (I_r,V_{I_r}) 意图冲突的历史数据中,V_{I_r} 为 1 时 I_r 所组成的集合。通过比较历史数据的模长,选择其中的最小值作为初始状态 S_t^{init}。基于强化学习的意图驱动光网络的冲突解决算法流程如算法 1 所示。

算法 1:基于强化学习的意图驱动光网络的冲突解决算法

输入 意图网络架构中用户意图的数据集 X(包含连接请求的源节点、目的节点、带宽、时延、抖动、丢包率等网络指标要求),发生冲突的网络链路资源环境。

输出 发生冲突时的不同状态下,用户选择的最优回退策略。

1)将不同用户的意图数据集进行数据预处理。

2)将数据集中的不同用户意图划分为组。

3)将意图网络冲突协商问题建模为一个强化学习模型。

4)采用 DDPG 算法对强化学习模型训练。

①初始化状态 s;

②在 Actor 当前网络基于状态 s 选择动作 a;

③执行动作 a，计算当前环境状态下执行动作 a 后的奖励值 r,环境状态更新为 s',将更新后的五元组存储进回放记忆单元集合;

④从回放记忆单元集合中通过 Mini-batch 采样,计算当前目标 Q 值;

⑤使用神经网络误差反向传播的方式更新 Critic 当前价格网络参数 w 和 Actor 当前策略网络参数 θ;

⑥若未到达终止态,则转到步骤②,否则当前轮迭代完毕,当智能体寻找到最优策略组合,算法结束。

2 仿真结果

本文从阻塞率和用户平均 QoS 这 2 个方面,对所提基于强化学习的意图光网络冲突解决方法的性能进行评估,包括模型收敛速度和网络运营商的利润趋势。在存在用户冲突的意图光网络场景中，测试平台基于 12 核 2.50GHz Intel Core I5-7200 CPU，双核 NVIDIA GTX TITAN XP GPU 和 64GB RAM。服务器运行环境为 Ubuntu16.04,代码基于深度学习框架 Keras2.1.0。基于上述环境，为了更真实地模拟意图冲突时网络拥塞环境的特征,随机生成用户意图,采用的网络拓扑如图 5 所示。节点之间的数字标识代表距离,单位为 km。其中，用户平均 QoS 被定义为网络负载中所有业务 QoS 的平均值[15]。

本文采用经典的贪婪算法和基于遗传变异的遗传算法与所提意图冲突解决算法进行对比，结果如图 6 所示。可以看出,与贪婪算法和遗传算法相比,随着网络中下发的用户意图数量增加，所提算法具有更低的网络阻塞率。这是因为此模型生成意图冲突回退策略时,增加了对回退策略进行评估的过程,在尽可能保证用户平均 QoS 的前提下,对意图转译的结果进行回退。

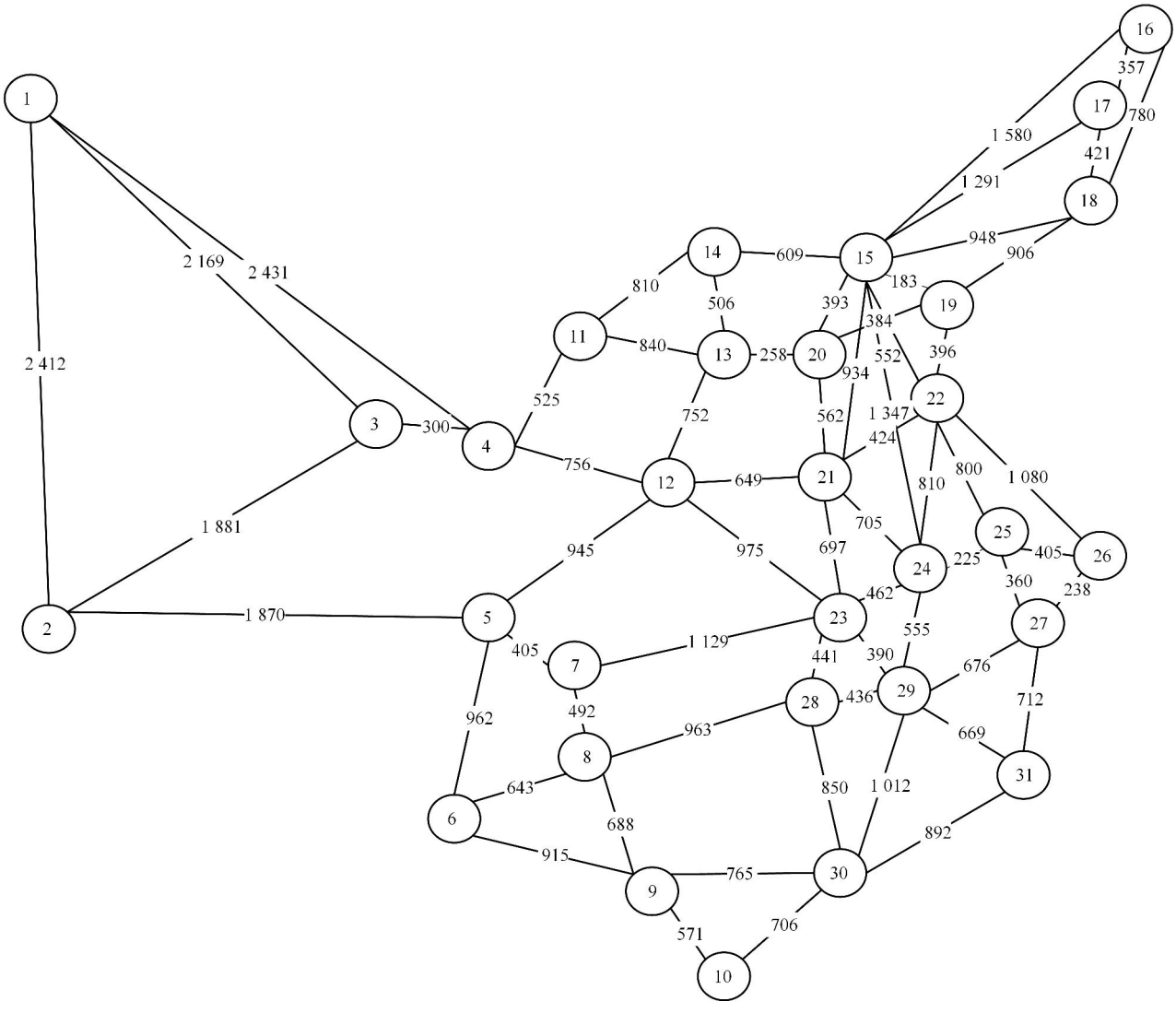


图 5 网络拓扑图

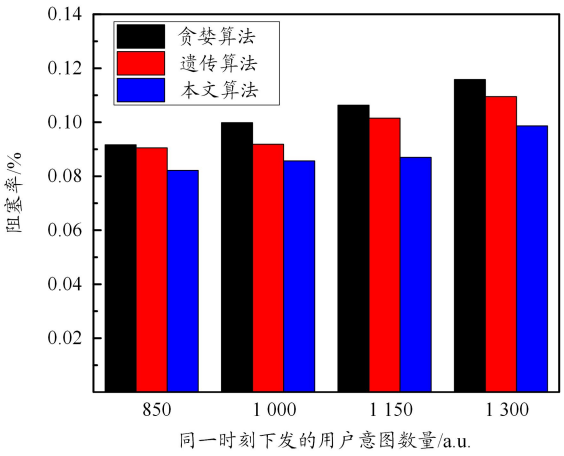


图 6　3 种算法的阻塞率变化图

贪婪算法、基于遗传变异的遗传算法与所提算法的用户平均 QoS 变化情况，如图 7 所示。可以看出，与遗传算法和贪婪算法的资源冲突解决方案的用户平均 QoS 值相比，本文算法的用户平均 QoS 值分别高 30%和 18%。这是因为其它资源冲突解决方案只考虑在资源紧缩时期如何解决冲突，导致了部分低优先级用户可能直接被阻塞，而提出的算法可以根据不同类型用户的带宽、时延容忍度的不同，合理安排冲突用户的回退范围，让用户请求都能连接成功，从而提高发生冲突链路的用户平均 QoS。

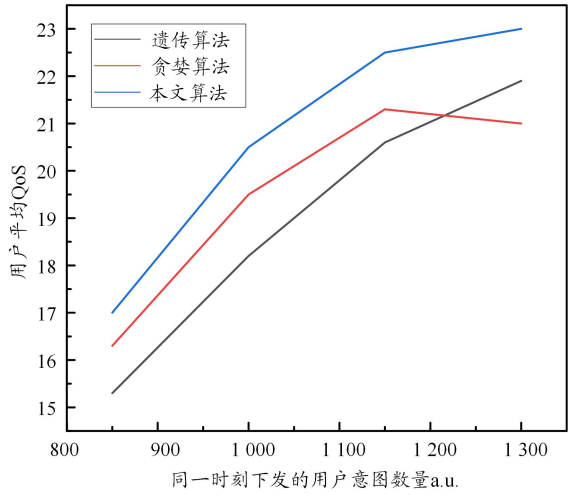


图 7　3 种算法的用户平均 QoS 变化图

3 结束语

本文针对智能运维的多用户意图光网络，提出了一种基于强化学习的意图驱动光网络的冲突解决方法。仿真结果表明：本文所提方法能够实现高精度的模型训练，获得较高的用户平均 QoS 和较低的网络阻塞率，在保障用户意图都能够成功连接的前提下使当前网络中的用户平均 QoS 值最大。光网络智能控制已经实现了新的优化和升级，从过去固化的关注网络传输指标，到如今更关注用户的体验质量，未来光网络管控技术将向自适应的方向发展。

参考文献：

[1] 胡世红. 边缘计算中资源动态调度的 QoS 优化技术研究 [D]. 无锡：江南大学，2021.
[2] 方娟，叶志远，张梦媛，等. 边云协同场景下基于强化学习的精英分层任务卸载策略研究[J]. 物联网学报，2022，6(1)：1–11.
[3] 王晓昌，吴璠，孙彦赞，等. 基于深度强化学习的车联网资源管理[J]. 工业控制计算机，2021，34(9)：31–33，36.
[4] 张影，龚亮亮，胡阳，等. 基于深度强化学习的智能电网 RAN 切片策略[J]. 计算机系统应用，2021，30(8)：293–299.
[5] 王金予，魏欣然，石文磊，等. 强化学习在资源优化领域的应用[J]. 大数据，2021，7(5)：131–149.
[6] 唐伦，贺小雨，王晓，等. 基于异步优势演员–评论家学习的服务功能链资源分配算法[J]. 电子与信息学报，2021，43(6)：1733–1741.
[7] 王磊，张丹，陈洁. 面向 5G 网络的无线意图驱动技术研究与实例[J]. 电信工程技术与标准化，2021，34(10)：28–33.
[8] 江巍，朱江. 基于深度强化学习反向散射网络资源分配机制[J]. 电讯技术，2022，62(10)：1–14.
[9] 俞文武，杨晓亚，李海昌，等. 面向多智能体协作的注意力意图与交流学习方法[J]. 自动化学报，2022，47(9)：1–16.
[10] 周洋程. 意图驱动无线接入网络的转译和自优化技术 [D]. 北京：北京邮电大学，2020.
[11] 龙瑞明. 面向应用 QoS 的意图网络部署[D]. 成都：电子科技大学，2019.
[12] 王海宁，贾鹏，曹宇诗，等. 基于 IBN 的 5G 网络管理系统的调度算法研究[J]. 电子技术应用，2019，45(10)：5–10.
[13] KHAN, LAREB Z, JOÃO P, et al. Data augmentation to improve performance of neural networks for failure management in optical networks[J]. Journal of Optical Communications and Networking, 2023, 15(1): 57–67.
[14] VELASCO L, BARZEGAR S, TABATABAEIMEHR F, et al. Intent-based networking and its application to optical networks [Invited Tutorial] [J]. Journal of Optical Communications and Networking, 2022, 14 (1): A11–A22.
[15] YANG H, ZHAN K, YAO Q, et al. Intent defined optical network with artificial intelligence-based automated operation and maintenance [J]. Science China Information Sciences, 2020, 63(6): 55–66.