

基于资源流网络和图卷积策略网络的多元物资存储空间保障仿真中的资源重调度方法研究

黄荟宇¹, 袁博文², 王文杰³, 吕启斌⁴, 田开顺⁵

(1. 北京工业大学 新校区建设办公室, 北京 100124; 2. 中国人民解放军 95899 部队, 北京 100085;
3. 北京工业大学 计算机学院, 北京 100124; 4. 国防大学 联合装备保障系, 北京 100036;
5. 北京航天测控技术有限公司, 北京 100041)

摘要: 针对大型存储基地物资种类多样、数量庞大的特点, 为实现高效管理, 需按设备类别划分存储空间 (如重设备、轻设备、电子设备等), 并依据各类空间的容量、物资属性与保障需求差异, 制定相应的清扫、检修等保障措施; 保障任务在时间周期与事件类型方面存在显著差异, 需以保障任务序列为基础, 在满足各项任务资源需求的前提下, 尽可能减少资源配置的调整幅度; 首先对保障资源调度问题进行了详细描述, 并建立相应数学模型; 其次, 针对现有资源流网络模型在复杂多元存储空间保障任务关系表示中的不足, 对其进行改进, 并将保障资源重调度问题转化为完备资源流网络生成问题; 进而, 结合并扩展图表示学习、强化学习与生成对抗网络, 提出一种基于图卷积策略网络 (GCPN) 的完备资源流网络生成方法; 最后, 通过仿真实验验证了所提保障资源调度模型的有效性与可行性。

关键词: 资源流网络; 图卷积策略网络; 多元物资存储; 保障仿真; 资源重调度

Research on Resource Rescheduling Method in Multi-Resource Storage Space Support Simulation Based on Resource Flow Network and Graph Convolutional Policy Network

HUANG Huiyu¹, YUAN Bowen², WANG Wenjie³, LÜ Qibin⁴, TIAN Kaishun⁵

(1. Beijing University of Technology, New Campus Construction Office, Beijing 100124, China;
2. The 95899 Troop of People's Liberation Army, Beijing 100085, China;
3. Beijing University of Technology, School of Computer Science, Beijing 100124, China;
4. National Defence University, Department of Joint Equipment Support, Beijing 100036, China;
5. Beijing Aerospace Measurement & Control Technology Co., Ltd., Beijing 100041, China)

Abstract: Given the diverse types and large quantities of materials in large-scale storage bases, efficient management requires categorizing storage spaces according to equipment types (e. g., heavy equipment, light equipment, electronic equipment, etc.) and formulating corresponding maintenance measures such as cleaning and inspection based on the capacity, material attributes, and support requirements of each space. Maintenance tasks vary significantly in terms of time cycles and event types. Based on a sequence of maintenance tasks, the objective is to minimize adjustments in resource allocation while ensuring that the resource demands of each task are satisfied. This paper first provides a detailed description of the maintenance resource scheduling problem and establishes a corresponding mathematical model. Next, to address the shortcomings of existing resource flow network models in representing complex and diverse maintenance task relationships in storage spaces, an improved model is proposed, transforming the maintenance resource rescheduling problem into a complete resource flow network generation problem. Furthermore, by integrating and extending graph representation learning, reinforcement learning, and generative adversarial networks, a method for generating complete resource flow networks based on a Graph Convolutional Policy Network (GCPN) is introduced. Finally, simulation experiments are conducted to validate the

收稿日期:2025-08-14; 修回日期:2025-09-24。

作者简介:黄荟宇(1989-),女,博士,助理研究员。

引用格式:黄荟宇,袁博文,王文杰,等. 基于资源流网络和图卷积策略网络的多元物资存储空间保障仿真中的资源重调度方法研究[J]. 计算机测量与控制, 2025, 33(11): 299-307, 315.

effectiveness and feasibility of the proposed maintenance resource scheduling model.

Keywords: resource flow network; graph convolutional policy network (GCPN); multi-resource storage; support simulation; resource rescheduling

0 引言

本文研究的多元物资存储空间（以下简称：空间）保障资源重新调度问题，是指在空间保障计划不变，与之对应的空间保障任务序列也未发生改变条件下，当部分空间保障任务序列保障资源缺失时，需要对原空间保障计划中的空间保障资源配置进行重新调整。项目在实际执行过程中，由于各种不确定性因素的影响，造成项目活动的实际开始时间偏离基准调度计划^[1]。可能出现此类情况的原因有两种：一是原空间保障计划中空间保障任务执行失败需要重新执行；二是原空间保障计划中执行空间保障任务的保障资源出现故障。

面临资源调度问题，现有求解方法因控制方式以及求解方式不同而各有侧重^[2]，常规随机调度方法可划分为静态策略^[3-4]与动态策略^[5-6]两大类。其中，动态策略适用于解决任务和可执行任务的资源未知或者可能发生改变的问题。国内外专家学者针对不同的动态调度问题进行了研究：针对多无人机智能体联盟动态任务分配问题，提出一种基于分布式拍卖机制的动态目标分配方法^[7]。针对时间窗口可能发生变化的时敏目标任务分配问题，提出了一种时敏任务动态分配算法^[8]。基于分阶段贪心规划算法对分发资源问题进行研究^[9]。针对新目标发现导致的初始计划失效问题，设计了一种基于匹配策略的局部任务重分配方法^[10]。

在对原有空间保障资源进行调整过程中，由于空间保障任务之间存在复杂的逻辑关系，资源配置需要重点考虑时间约束和能力约束等条件。对于优化目标，空间保障任务执行效能是重调度的关键指标。但是，频繁的空间保障资源调度可能会降低其他空间保障任务的执行效能和调度的实时性，对其鲁棒性带来不利影响，调度计划的鲁棒性通常由实际调度计划与基准调度计划的偏离程度来体现^[11]。有学者提出采用简单的并行调度生成机制构建可行的资源流网络^[12]；有学者提出三种基于整数规划的启发式算法（Min-EA、Max-PF 和 Min-ED）生成合理的资源流网络^[13]。部分学者也采用时间缓冲添加方法处理不确定情形的项目调度问题^[14-15]。由于问题规模、任务复杂关系和需要进行力量调度的任务数量等区别，现有的调度方法主要存在以下不足：1) 多任务之间的存在复杂关系给资源调度产生新的约束条件，需要重点考虑；2) 现有文献多针对单一任务分配单类型资源的重调度问题开展研究，而对于多任务且需要分配缺少不同类型力量的重调度问题研究较少。

本文研究的空间保障资源重调度，需要考虑多种约

束条件和多项优化目标，重点把握时间约束和能力约束等约束条件，空间保障任务执行效能和资源调度率等优化指标。

1 空间保障资源重调度建模

假定原有空间保障计划为 $Plan_{pre}$ ，与其对应的行为树为 BT_{pre} 。原有空间保障计划的资源清单为 $ForList = \{For_1, For_2, \dots, For_N\}$ ，资源数目为 N ，根据能力属性可分为 b 类。空间保障任务为 $T_{pre} = (TA_{pre}, TR_{pre})$ ，空间保障任务集合为 $TA_{pre} = \{Ta_1, Ta_2, \dots, Ta_M\}$ ，空间保障任务数目为 M 。需要强调的是，空间保障任务并不是按照标号顺序依次单独执行，而是依据时序关系和逻辑关系等，由不同空间保障资源协同执行。空间保障任务关系集合 $TR_{pre} = \{TR_{j'j} | Ta_j, Ta_{j'} \in TA_{pre}, j \neq j'\}$ 。通过空间保障资源预调度，原有空间保障计划中的空间保障资源配置可以表示为：

$$X_{pre} = \begin{pmatrix} x_{11} & \cdots & x_{1M} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NM} \end{pmatrix} \quad (1)$$

其中： $x_{ij} \in \{1, 0\}$ 为空间保障资源调度的决策变量， $x_{ij} = 1$ 表示空间保障资源 For_i 参与执行空间保障任务 Ta_j ，否则 $x_{ij} = 0$ 。

假定在某一时刻 T ，需要进行空间保障资源重调度。此时，尚未执行的空间保障任务集合为 $TA_{rest} = \{Ta_1^{rest}, Ta_2^{rest}, \dots, Ta_{M_T}^{rest}\}$ ，关系集合为 $TR_{rest} = \{TR_{j'j} | Ta_j \in TA_{rest}, Ta_{j'} \in TA_{rest}, j \neq j'\}$ 。其中，没有空间保障资源执行的空间保障任务集合为 $TA_{bug} = \{Ta_1^{bug}, Ta_2^{bug}, \dots, Ta_{m_T}^{bug}\}$ ，并有 $TA_{bug} \subseteq TA_{rest}$ ， $m_T < M_T$ ， $m_T < M$ 。需要通过重调度模型计算新的空间保障资源配置：

$$X_{now} = \begin{pmatrix} x_{11} & \cdots & x_{1M_T} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NM_T} \end{pmatrix} \quad (2)$$

相关变量定义参照式 (1)。

1.1 目标函数

一方面，空间保障资源调度方案应保证空间保障任务实施获得最高的效能，即空间保障任务实施效能最大化。另一方面，为避免过多的空间保障资源配置调整对原计划的鲁棒性产生影响，应尽量使空间保障资源调整最少，即空间保障资源调度率最小化。因此，以空间保障任务执行效能和空间保障资源调度率为优化指标，构设下列目标函数：

$$\max G = \frac{f_4}{f_5} \quad (3)$$

其中: G 是空间保障资源重调度的优化目标; f_4 是空间保障资源调整的期望效能; f_5 是空间保障资源调整数量占空间保障资源总数量的比例。

假定在剩余的 M_T 项空间保障任务中, 存在 K 项空间保障任务涉及空间保障资源调整。其中, m_T 项空间保障任务缺失空间保障资源保障, $K - m_T$ 项空间保障任务需要对原定保障空间保障资源进行调整。空间保障资源调整期望效能主要取决于涉及空间保障资源调整的空间保障任务执行效能, 而空间保障任务执行效能是由空间保障资源空间保障任务的重要程度和空间保障资源空间保障任务执行概率决定的。因此, 空间保障资源调整效能为:

$$f_4 = \sum_{j \in K} \Delta E_j = \sum_{j \in m_T} E_j^{\text{now}} + \sum_{j \in K - m_T} E_j^{\text{now}} - E_j^{\text{pre}} \quad (4)$$

其中: E_j 表示空间保障任务 Ta_j 的实施效能, 其表达式如式 (5) 所示:

$$E_j = e^{\lambda_j r_j} * \prod_{i, x_{ij} = 1} p_{ij} \quad (5)$$

其中: $e^{\lambda_j r_j}$ 是空间保障任务 Ta_j 的重要程度; r_j 是 Ta_j 的空间保障任务目标 Tar^j 的排序结果, λ_j 是目标 Tar^j 相关空间保障任务重要程度系数, 当只存在一项空间保障任务与 Tar^j 相关时, $\lambda_j = 1$ 。 p_{ij} 是空间保障资源 For_i 参与实施空间保障任务 Ta_j 的成功概率。

空间保障资源调度率为:

$$f_5 = \frac{\sum_{j \in K} \text{Count}(For^{Ta_j})}{\sum_{j \in M_T} \text{Count}(For^{Ta_j})} \quad (6)$$

其中: For^{Ta_j} 表示实施空间保障任务 Ta_j 的空间集合; $\text{Count}(\cdot)$ 表示对集合求元素数。

1.2 约束空间

在空间保障资源重调度过程中, 综合考虑时间约束和能力约束, 构建问题的主要约束空间。

1.2.1 时间约束

为了满足空间保障资源执行空间保障任务过程中的时间约束, 每份空间保障资源一次只能分配保障一项空间保障任务。如果需要一份空间保障资源先后服务保障多项空间保障任务, 那么空间保障资源的准备时间应满足空间保障任务之间转换的时间要求, 即:

$$\begin{cases} \sum_{j=1}^{j=M_T} x_{ij}^{\Delta t} \leq 1 \\ \left(\frac{D(Ta_j^i, Ta_{j'}^i)}{v(For_i)} + \text{ready}(For_i) \right) \cdot x_{ij} \cdot x_{ij'} \leq \\ te(Ta_j) - ts(Ta_{j'}) \end{cases} \quad (7)$$

其中: $x_{ij}^{\Delta t} = 1$ 表示空间保障资源 For_i 在 Δt 参与执行空间保障任务 Ta_j ; $te(Ta_j)$ 表示空间保障任务 Ta_j 的

结束实施时间; $ts(Ta_{j'})$ 表示空间保障任务 $Ta_{j'}$ 的开始实施时间; $\text{ready}(For_i)$ 表示空间保障资源 For_i 实施空间保障任务前的准备时间; $D(Ta_j^i, Ta_{j'}^i)$ 表示实施空间保障任务 Ta_j 与实施空间保障任务 $Ta_{j'}$ 之间的空间距离; $v(For_i)$ 表示空间保障资源 For_i 的机动速度。

1.2.2 能力约束

为了满足空间保障资源服务保障空间保障任务过程中的能力约束, 对于单项空间保障任务, 用于服务保障空间保障任务的空间保障资源能力总量应不小于空间保障任务的能力需求, 即:

$$\begin{cases} \sum_{i=1}^{i=N} (x_{ij}^{\text{now}} \cdot \text{ForPVal}_h^i) \geq \text{TaPVal}_h^j \\ 1 \leq h \leq H \end{cases} \quad (8)$$

其中: x_{ij}^{now} 取自式 (2); ForPVal_h^i 表示空间保障资源 For_i 的第 h 项能力取值, 与空间保障资源 For_i 的属性 Pro_h 相关; TaPVal_h^j 表示实施空间保障任务 Ta_j 对第 h 项能力需求; H 表示整个空间保障计划涉及的能力总数。

2 基于资源流网络和图卷积策略网络的空间保障资源重调度方法

由上一部分问题描述与建模可知, 空间保障资源重调度是规划已有空间保障资源如何在空间保障任务之间流转的过程。资源流网络是一种用来表征可更新资源在空间保障任务之间的流动和传递关系的网状模型, 适用于空间保障计划的资源配置优化。因此, 如果将空间保障资源看作不同的资源, 空间保障资源重调度问题即可转换为资源流网络生成问题。针对该问题, 本文利用并扩展图表示学习、强化学习和生成对抗网络, 构建了图卷积策略网络 (GCPN, graph convolutional policy network)^[16], 用于生成符合目标导向的资源流网络。

2.1 资源流网络改进

结合前文所述空间保障资源重调度问题特点, 将空间保障资源看作不同的资源, 采用资源流网络方法对其进行求解 (若无特别说明, 下文所述的资源均指代空间保障资源)。

2.1.1 现有资源流网络简介

定义 1 【资源流】^[17]:

假设分配执行空间保障任务 Ta_{j_i} 的资源集合为 $TaFor^{j_i}$, 当空间保障任务 Ta_{j_i} 完成后, 其占用的这些资源若部分或全部分配执行空间保障任务 Ta_{j_i} , 则称空间保障任务 Ta_{j_i} 与空间保障任务 Ta_{j_i} 之间存在资源流 $l_{j_i}^{j_i}$, $l_{j_i}^{j_i} \subseteq TaFor^{j_i}$ 。

在传统的节点式有向网络 $G(N, A)$ 中, N 表示所有空间保障任务节点的集合, A 表示所有有向线的集合, 若存在从空间保障任务 Ta_{j_i} 到空间保障任务 Ta_{j_i} 有向

线 (j_1, j_2) ，则表示空间保障任务 Ta_{j_2} 必须在空间保障任务 Ta_{j_1} 完成后才能开始。下面在节点式有向网络 $G(N, A)$ 的基础上给出资源流网络 $RG(N, L)$ 的定义。

定义 2 【资源流网络】^[16]：

在 $G(N, A)$ 上，若空间保障任务 Ta_{j_1} 与空间保障任务 Ta_{j_2} 之间存在资源流 $l_{j_1}^{j_2}$ ，则存在一条从节点 j_1 到节点 j_2 的资源弧。由所有空间保障任务节点集合 N 以及资源弧集合 L 构成的网络，称为资源流网络 $RG(N, L)$ 。其中， N 表示所有空间保障任务节点的集合， L 表示所有有向弧的集合。

下面通过一个简单的算例对现有资源流网络进行说明。

假设在如图 1 (a) 所示的空间保障任务图中，每项空间保障任务只需要一种资源，每项空间保障任务 Ta_{j_i} 的处理时间 $TaTime^{j_i}$ 和执行资源集合需求 $TaFor^{j_i}$ 在空间保障任务节点的上方标出。已知可用资源总量为 3，一个由图 1 (a) 所示的空间保障任务图得到的资源流网络如图 1 (b) 所示，有向弧上的数字表示转移的资源数。

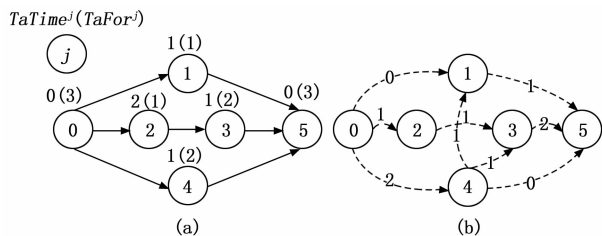


图 1 现有的资源流网络图生成过程

2.1.2 面向空间保障任务关系的资源流网络改进

现有资源流网络模型只考虑了空间保障任务之间的顺序关系 TR_Seq 和与关系 TR_And ，而对于或关系 TR_Or 和条件关系 TR_Cnd 表示具有一定的局限性。其中，或关系 TR_Or 表示应根据当前态势，决定执行多项备选任务中的一项；条件关系 TR_Cnd 表示应根据当前态势，决定是否执行某一项任务。为此，需要对现有资源流网络模型进行改进，以适应空间保障任务之间存在其他关系的情况。

2.1.2.1 或关系 TR_Or

假设存在空间保障任务 Ta_{j_1} 与空间保障任务 Ta_{j_2} 满足 $TR_Or(Ta_{j_1}, Ta_{j_2})$ ，则在构建资源流网络时进行如下处理。

步骤 1：在空间保障任务 Ta_{j_1} 与空间保障任务 Ta_{j_2} 之前增加或判断虚拟空间保障任务节点 Ta_{j_s} ；

步骤 2：将 Ta_{j_s} 执行时间记为 $TaTime^{j_s} = 0$ ，计算其执行资源需求记为 $TaFor^{j_s} = TaFor^{j_1} \cup TaFor^{j_2}$ ， Ta_{j_s} 与 Ta_{j_1} 之间的资源流为 $l_{j_s}^{j_1} = TaFor^{j_1}$ ， Ta_{j_s} 与 Ta_{j_2} 之间的资源流为 $l_{j_s}^{j_2} = TaFor^{j_2} - TaFor^{j_1}$ ；

步骤 3：由于 Ta_{j_1} 与 Ta_{j_2} 之间只有一项空间保障任务会被执行，在对资源流进行计算时，资源被占用的时间是 $TaTime^{j_s} = TaTime^{j_1} = \max(TaTime^{j_1}, TaTime^{j_2})$ 。在空间保障任务执行过程中，若 Ta_{j_1} 不被执行，则立刻将资源 $TaFor^{j_s} \cap TaFor^{j_2}$ 部分由 Ta_{j_s} 转移至 Ta_{j_2} 。

至此，在占用时间最短、资源数量最少的条件下，保证了具有或关系 TR_Or 的空间保障任务执行。具有或关系 TR_Or 的多项空间保障任务参照上述方式处理。或关系处理过程如图 2 所示。

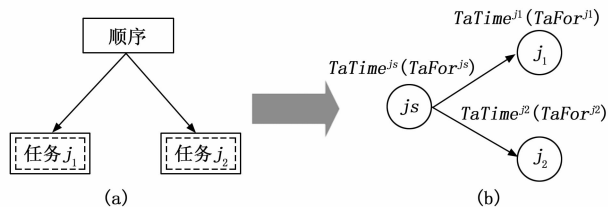


图 2 或关系处理过程示意图

2.1.2.2 条件关系 TR_Cnd

假设存在空间保障任务 Ta_{j_1} 与空间保障任务 Ta_{j_2} 满足 $TR_Cnd(Ta_{j_1}, Ta_{j_2}, Cnd)$ ，则在构建资源流网络时进行如下处理。

步骤 1：在空间保障任务 Ta_{j_1} 之前增加条件判断虚拟空间保障任务节点 Ta_{j_s} ，其执行时间记为 $TaTime^{j_s} = 0$ ；

步骤 2：将 Ta_{j_s} 执行资源需求记为 $TaFor^{j_s} = TaFor^{j_1}$ 。并且， Ta_{j_s} 与 Ta_{j_1} 之间的资源流为 $l_{j_s}^{j_1} = TaFor^{j_1}$ ；

步骤 3：由于 Ta_{j_1} 是否会被执行具有不确定性，在对资源流进行计算时，资源被占用的时间保持不变。在空间保障任务执行过程中，若 Ta_{j_1} 不被执行，则立刻将资源 $TaFor^{j_s}$ 释放。

至此，在占用时间最短、资源数量最少的条件下，保证了具有条件关系 TR_Cnd 的空间保障任务执行。条件关系 TR_Cnd 处理过程如图 3 所示。

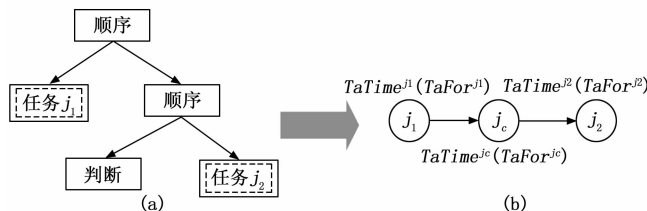


图 3 条件关系处理过程示意图

资源流是资源在空间保障任务之间流转的抽象表征，其目的是保证空间保障任务的正常执行，应满足空间保障任务执行的时间约束和能力约束。为此，本文给出完备资源流网络与非完备资源流网络的定义。

定义 3 【完备资源流网络】：

对于一个资源流网络，若资源在空间保障任务之间

的流转方案使得所有空间保障任务的时间约束和能力约束均被满足, 则称之为完备资源流网络。

定义 4 【非完备资源流网络】:

对于一个资源流网络, 若资源在空间保障任务之间流转方案使得存在部分空间保障任务的时间约束和能力约束未被满足, 则称之为非完备资源流网络。

定义 5 【完备节点和非完备节点】:

对于一个资源流网络中的空间保障任务节点, 若流转入该空间保障任务节点的资源等于流转出该空间保障任务节点的资源, 并且等于该空间保障任务节点的资源需求, 则称该空间保障任务节点为完备节点。否则, 称该空间保障任务节点为非完备节点。

下面仍以一个简单的算例对完备资源流网络与非完备资源流网络进行说明。在图 4 所示的两个资源流网络 RG_A 和 RG_B 中, 每项空间保障任务只需要一支空间保障资源, 可用空间保障资源总数为 5。在图 4 (a) 所示的资源流网络 RG_A 中, 包含 7 个普通空间保障任务和 2 个虚拟空间保障任务, 并且每项空间保障任务时间约束和能力约束均被满足。因此, 资源流网络 RG_A 为完备资源流网络, 所有节点均为完备节点。当 RG_A 中某支空间保障资源出现故障、损坏等意外情况时 (图 4 (a) 中叉标记), 得到如图 4 (b) 所示的资源流网络 RG_B 。显然, RG_B 存在空间保障任务节点 j_c 、 j_i 和 j_s 未满足能力约束, 为非完备节点, RG_B 为非完备资源流网络。

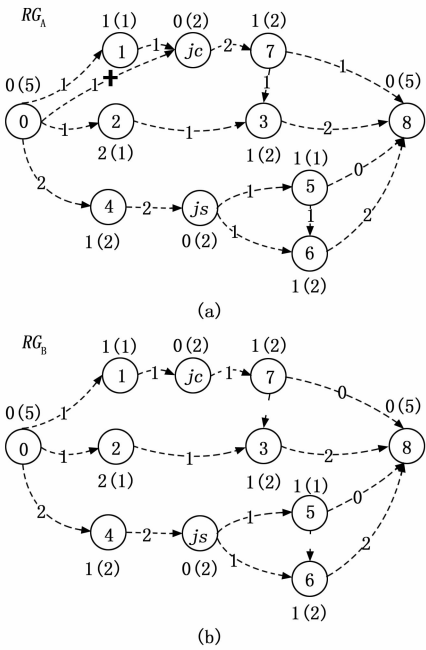


图 4 完备资源流网络与非完备资源流网络示例

2.2 基于 GCPN 的完备资源流网络生成

在原有空间保障计划中, 各项空间保障任务的服务空间保障资源已经确定。根据上一节内容可知, 完整的

空间保障计划可转化为一个完备资源流网络, 部分空间保障任务缺少空间保障资源执行的空间保障计划可以转化为一个非完备资源流网络。因此, 空间保障资源重调度问题可以转化为以下完备资源流网络生成问题:

假定原有一个完备资源流网络 RG_{pre}^C 。在执行空间保障计划过程中, 发生非预期态势变化, 出现部分空间保障任务缺少空间保障资源执行的情况, 得到非完备资源流网络 RG_{rest}^{Inc} 。如何根据 RG_{rest}^{Inc} 生成新的完备资源流网络 RG_{now}^C ?

针对这一完备资源流网络生成问题, 本文利用“图表示学习”方法获取图状态的向量表示, 并构建策略网络来预判不同图状态下输出不同行动的概率分布, 以实现端到端的完备资源流网络生成。在网络学习过程中, 利用历史空间保障计划构建的完备资源流网络进行生成对抗性训练, 以提高学习效率。

2.2.1 图卷积网络与策略梯度

图卷积网络 (GCN, graph convolutional networks) 是一种能够对图数据进行表示学习的方法, 如图 5 所示。与深度学习类似, GCN 是一种端到端 (end-to-end) 的学习方法, 能够从原始数据中自动提取特征。两者的区别在于, 深度学习等卷积神经网络处理对象限制在有规则的空间结构, 比如图片和语音等。而 GCN 将研究对象扩展至不规则的数据结构, 比如脑信号和分子结构等抽象的图谱。在 GCN 模型中, 通过堆叠图卷积层, 节点属性信息和图结构信息不断被编码学习, 协同式地影响最终节点的表示。

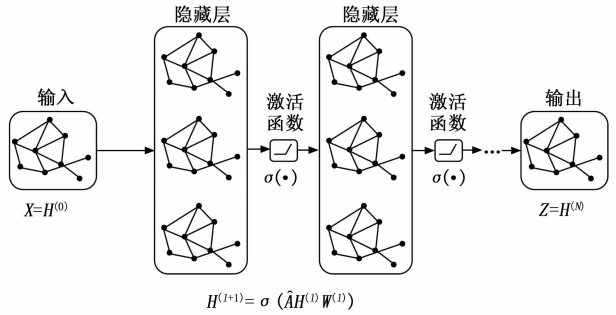


图 5 GCN 网络示意图

策略梯度 (PG, policy gradient) 是强化学习 (RL, reinforcement learning) 中经常使用的一类算法, 如图 6 所示。

在强化学习框架下, 具体的方法主要分为基于值和基于策略两大类。基于值的强化学习方法主要以深度 Q 网络 (DQN, deep Q networks) 系列算法为代表, 这种方法存在三个方面的问题和不足, 一是对连续动作的处理能力不足; 二是对受限状态下的问题处理能力不足; 三是无法解决随机策略问题。针对上述问题 and 不足, 基于策略的强化学习算法^[18]。将策略表示为一个

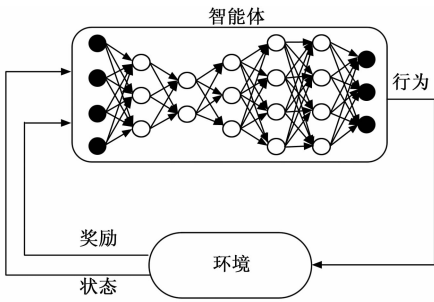


图 6 策略梯度算法示意图

连续函数,并用连续函数的优化方法来寻找最优策略。策略梯度就是一种典型的优化算法。它将动作策略 π_θ 表示为参数化 θ 的神经网络,并通过调整 θ 以使得期望回报最大。即使得最终回报变大的动作出现的概率增加,反之,则减少这个动作出现的概率。参数 θ 调整的关键在于求取回报函数关于 θ 的梯度(即策略梯度),再采用梯度上升的方法对策略梯度进行优化。

2.2.2 资源流网络生成环境

下面介绍资源流网络生成环境,主要包括状态空间、行动空间、状态转移和奖励四个部分。在该环境下,非完备资源流网络通过 GCPN 执行一连串行动,可以得到完备资源流网络。

2.2.2.1 状态空间

资源流网络 $RG(N, L)$ 可以表示为 $(\mathbf{A}, \mathbf{E}, \mathbf{F})$ 。其中, $\mathbf{A} \in \{0, 1\}^{n \times n}$ 是邻接矩阵; $\mathbf{F} \in R^{n \times d}$ 是节点特征矩阵,每个节点有 d 个特征; $\mathbf{E} \in \{0, 1\}^{b \times n \times n}$ 是边类型邻接张量, $\mathbf{E}_{i,j,k} = 1$ 表示节点 j 和节点 k 之间存在类型为 i 的边,并有 $\mathbf{A} = \sum_{i=1}^b \mathbf{E}_i$ 。

假定在时间步长 t 下,中间生成的资源流网络为 RG_t ,对应的环境状态记为 $s_t = RG_t$ 。需要强调的是,环境状态 s_t 对于 RL 代理是完全可观察的,即所有节点和连边的状态均为已知的。并有 $s_0 = RG_0 = RG_{rest}^{inc}$ 。

2.2.2.2 行动空间

在本文所提的学习框架中,需要构建一个维度确定的行动空间。完备资源流网络的构建与网络科学领域的链接预测具有极大的相似性,行动空间可参照其进行设计。在时间步长 t 下,对于一个中间环境状态 $s_t = RG_t$,一个行动可以记为:

$$a_t = \text{CONCAT}(a_{\text{first}}, a_{\text{second}}, a_{\text{edge}}, a_{\text{stop}}) \quad (9)$$

其中: $\text{CONCAT}(\cdot)$ 表示对有关内容进行连接,参照字符串连接操作; a_{first} 表示连接的第一个节点,对应着空间保障任务 Ta_{j_1} ; a_{second} 表示连接的第二个节点,对应着空间保障任务 Ta_{j_2} ; a_{stop} 表示连边结束的概率,用于表征行动结束后网络完备概率; a_{edge} 表示连边的类型。下面进行详细说明。

根据定义 1 可知,资源流中流转的资源取决于执行

空间保障任务的资源集合,不同空间保障任务之间的资源流具有互异性。为了构建一个维度确定的行动空间,若从个体的角度设计连边类型, a_{edge} 的维度应等于计划中资源个体总数 $\|ForList\|$ 。显然,实际空间保障计划涉及的资源数量巨大,这种设计将会导致构建的策略网络结构巨大,参数调整失效,难以收敛;而且训练得到的网络泛化性较差,资源发生变化后,训练得到的策略网络便不再适用。

为此,设计 a_{edge} 只输出资源的类型,不输出具体的资源,即 $a_{\text{edge}} \in \{FT_i | i = 1, 2, \dots, b\}$,其中 FT_i 表示资源的类型。资源流的具体资源采用最大能力或者最大需求的模式确定,即:

假定在空间保障任务 Ta_{j_1} 与空间保障任务 Ta_{j_2} 之间增加资源流 $l_{FT}^{j_1, j_2}$,资源的类型为 FT ,空间保障任务 Ta_{j_1} 尚未计划流输出的 FT 类型资源为 $TaFor_{FT}^{j_1, rest} \subseteq TaFor^{j_1}$,空间保障任务 Ta_{j_2} 尚缺少流入的 FT 类型资源为 $TaFor_{FT}^{j_2, lack} \subseteq TaFor^{j_2}$,则 $l_{FT}^{j_1, j_2} = \min(TaFor_{FT}^{j_1, rest}, TaFor_{FT}^{j_2, lack})$ 。其中, $\min(\cdot)$ 表示求最小值。若空间保障任务节点的能力有剩余,或者需求有缺失,则该空间保障任务节点为非完备节点,需要继续增加资源流。

2.2.2.3 状态转移

状态 s_t 转移到状态 s_{t+1} 主要是产生新的资源流,并且应符合式 (7) 和式 (8) 表示的相关约束。因此,为了避免产生不合理的资源流网络,设定当策略网络提出的行动不符合约束时,行动无效,状态保持不变。

2.2.2.4 奖励

RL 代理行为的奖励由中间奖励和最终奖励组成。最终奖励包括目标函数奖励和对抗性奖励。目标函数奖励由式 (3) 计算求得,对抗性奖励由生成对抗网络 (GAN) 计算求得,后面将详细介绍。中间奖励包括阶梯式的有效性奖励和对抗性奖励。如果连边行动符合式 (7) 和式 (8) 约束,就会分配一个较小的正奖励,否则分配一个较小的负奖励。

为了提升完备资源流网络生成效率,可将已有的大量空间保障计划转化为完备资源流网络库,并据此与策略网络构建生成对抗网络。采用生成对抗网络框架^[19]来定义对抗性奖励:

$$\min_{\theta} \max_{\varphi} V(\pi_{\theta}, D_{\varphi}) = \mathbb{E}_{x \sim p_{\text{data}}} [\log D_{\varphi}(x)] + \mathbb{E}_{x \sim \pi_{\theta}} [\log(1 - D_{\varphi}(x))] \quad (10)$$

其中: x 是由策略网络生成的一个图对象, π_{θ} 是生成的策略网络, D_{φ} 是鉴别器网络, p_{data} 是完备资源流网络库基础数据。由于, x 对参数 φ 不可导, D_{φ} 只能采用随机梯度下降方法训练。

采用 $-V(\pi_{\theta}, D_{\varphi})$ 作为其他奖励的额外奖励,并使用策略梯度方法对总奖励进行梯度优化(见下文第四部

分)。鉴别器网络采用与策略网络相同的结构来计算节点嵌入(见下文第三部分), 然后汇总为图的嵌入, 并最终转换为标量预测。

2.2.3 图卷积策略网络

下面主要介绍 GCPN 网络结构。由前两小节可知, GCPN 网络是由 RL 代理对生成环境进行行动执行学习得到的策略网络。其输入是非完备资源流网络 $s_t = RG_t$, 输出行动 a_t , 表现为产生新的资源流。如图 7 所示, 图卷积策略网络主要包括 GCN、策略网络(PN, policy network)和鉴别器(Discriminator)。

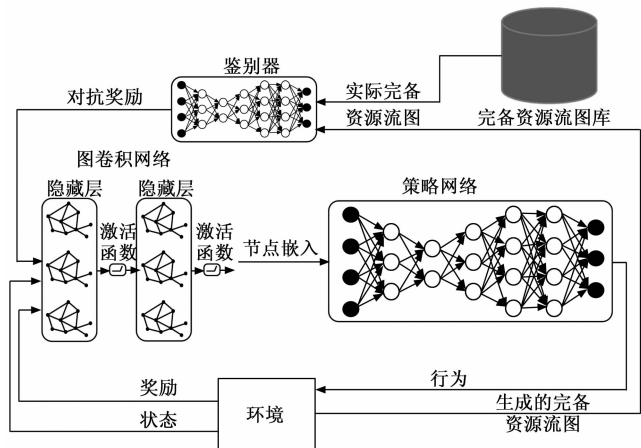


图 7 图卷积策略网络结构

GCN 用于资源流网络节点嵌入 (Node Embedding) 计算, 为 PN 提供图数据特征提取。PN 用于资源流生成行动预测, 向资源流网络输出行动。将 GCN 与 PN 组合在一起, 看作是一个生成器, 与鉴别器共同形成生成对抗网络结构, 并提供对抗奖励。下面对各个网络结构设计进行详细介绍。

2.2.3.1 节点嵌入计算

资源流网络属于不规则图数据, 传统的卷积神经网络难以直接提取图的特征。为此, 在预测连边之前, 先使用 GCN 计算输入图的节点嵌入。由于不仅要预测哪些节点之间连边, 还要预测连边的类型, 因此需要对传统 GCN 的结构进行改进。在 GCN 的第 l 层中, 将不同类型的连边信息进行融合, 作为第 $l+1$ 层节点嵌入的输入。计算公式如下:

$$H^{(l+1)} = \text{AGG}(\text{ReLU}(\{\tilde{D}_i^{-1/2} \tilde{E}_i \tilde{D}_i^{-1/2} H^{(l)} W_i^{(l)}\}, \forall i \in (1, \dots, b))) \quad (11)$$

其中: $H^{(l+1)} \in \mathbb{R}^{n \times k}$ 在第 $l+1$ 层得到节点嵌入, n 是节点数量, k 是嵌入维度; E_i 是第 i 层边类型邻接张量, $\tilde{E}_i = E_i + I$; $\tilde{D}_i = \sum_k \tilde{E}_{ijk}$; $W_i^{(l)}$ 是第 i 层待训练神经网络连边权重矩阵。AGG($\cdot$) 表示一个聚合函数, 可以是 $\{\text{MEAN}, \text{MAX}, \text{SUM}, \text{CONCAT}\}$ 中的一个。应用 L 层 GCN 对网络图进行节点嵌入计算, 最终得到节点

嵌入矩阵 $X = H^{(L)}$ 。

2.2.3.2 行动预测

针对式 (9) 中行动包含的四部分内容, 采用如下方式进行预测:

$$\begin{aligned} f_{\text{first}}(s_t) &= \text{SOFTMAX}(m_f(X)), \\ a_{\text{first}} &\sim f_{\text{first}}(s_t) \in \{0, 1\}^n \\ f_{\text{second}}(s_t) &= \text{SOFTMAX}(m_s(X_{a_{\text{first}}}, X)), \\ a_{\text{second}} &\sim f_{\text{second}}(s_t) \in \{0, 1\}^n \\ f_{\text{edge}}(s_t) &= \text{SOFTMAX}(m_e(X_{a_{\text{first}}}, X_{a_{\text{second}}}), \\ a_{\text{edge}} &\sim f_{\text{edge}}(s_t) \in \{0, 1\}^b \\ f_{\text{stop}}(s_t) &= \text{SOFTMAX}(m_t(\text{AGG}(X))), \\ a_{\text{stop}} &\sim f_{\text{stop}}(s_t) \in \{0, 1\} \end{aligned} \quad (12)$$

其中: $m_f(\cdot)$ 是将 $s_t = RG_t$ 中 n 个节点嵌入 $Z_{0:n} \in \mathbb{R}^{n \times k}$ 映射到 n 维向量的多层感知器 (MLP, multilayer perceptron), 该 n 维向量表示选择每个节点作为第一个节点的概率分布; $m_s(\cdot)$ 是预测第二个节点的 MLP, 其将第一个被选择节点 a_{first} 的节点嵌入 Z_{first} 与 $s_t = RG_t$ 中 n 个节点嵌入结合在一起, 并映射到 n 维向量, 该 n 维向量表示选择每个节点作为第二个节点的概率分布; $m_e(\cdot)$ 是将节点 a_{first} 的节点嵌入 Z_{first} 与节点 a_{second} 的节点嵌入 Z_{second} 映射到 b 维向量的 MLP, 用于预测连边类型; $m_t(\cdot)$ 是计算终止概率的 MLP; 聚合函数 AGG($\cdot$) 用于将节点嵌入聚合为图嵌入。

2.2.4 策略梯度优化

基于策略梯度的方法广泛用于策略网络的优化。本文采用 PPO (Proximal Policy Optimization)^[20] 方法对策略网络进行优化, 其目标函数为:

$$\begin{aligned} \max L^{\text{CLIP}}(\theta) &= H_t[\min(r_t(\theta) \hat{A}_t, \\ &\quad \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon) \hat{A}_t)], \\ r_t(\theta) &= \frac{\pi_\theta(a_t | s_t)}{\pi_{\text{old}}(a_t | s_t)} \end{aligned} \quad (13)$$

其中: $r_t(\theta)$ 是概率比; $\text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)$ 是将 $r_t(\theta)$ 剪切在 $[1 - \epsilon, 1 + \epsilon]$ 范围内, 为 $L^{\text{CLIP}}(\theta)$ 提供了保守策略迭代目标的下限^[21]; $\hat{A}_t$ 是估计优势函数, 与一个用于减少估计差距的学习价值函数 $V_w(\cdot)$ 相关。在本文设计的 GCPN 网络中, $V_w(\cdot)$ 是前文设计用于计算图嵌入的 MLP。

使用专家知识库对策略网络进行预训练, 往往会带来更好的训练稳定性和性能。在 GCPN 的训练优化之前, 可以先使用资源流网络库对网络进行预训练。预训练过程中, 目标函数表示为 $\min L^{\text{EXPERT}}(\theta) = -\log(\pi_\theta(a_t | s_t))$ 。其中, $(a_t | s_t)$ 是由完备资源流网络获得的。具体来说, 从资源流网络库中随机抽取一个完备资源流网络, 并去除其中某一条资源流, 得到非完备资源流网络作为状态 s_t 。在状态 s_t 下, 调整网络参数, 使得输出行动 a_t , 以生成抽样的完备网络图。

3 实验仿真与分析

3.1 实验方案

假设空间保障计划主要分为 2 个阶段，分别是阶段 1：空间保障准备阶段、阶段 2：空间保障实施阶段。计划内容大体如下：

- 1) 4 台 R1 型卫生清扫器 (R1_1, R1_2, R1_3, R1_4) 分 2 批次分别对轻重设备空间进行卫生清扫；
- 2) 1 名 Y 型值班管理人员 (E_1) 对存储区进行空间情况督查，2 名辅助管理人员伴随巡查；
- 3) D 型设备检修人员分别对轻重设备空间进行设备检修服务保障，每名设备检修人员由 2 名辅助管理人员伴随；
- 4) H1 型电气设备 (H1_1, H1_2, H1_3, H1_4) 和 H2 型显示设备 (H2_1, H2_2) 保障重设备

存储空间，B 型投影设备 (B_1, B_2) 保障重设备存储空间；

- 5) 4 型 M2 环境控制设备 (M2_1, M2_2, M2_3, M2_4) 分 2 批次保障轻设备存储空间运行；
- 6) 2 台 R1 型卫生清扫器 (R1_5, R1_6) 对走廊、卫生间等场所进行卫生打扫服务；
- 7) D 型设备检修人员分别对消防设备、逃生设备等设备进行设备检修服务保障，每名设备检修人员由 2 名辅助管理人员伴随；
- 8) B 型投影设备对空间场所的资料室、工作间等场所进行保障；
- 9) 对其他任务进行空间保障。

空间保障计划中包含的空间保障任务如表 1 所示。相对应的空间保障任务图如图 8 所示。其中，T0 与 T31

表 1 空间保障任务列表

保障步骤	设备存储空间保障任务代号	设备存储空间保障任务说明	执行存储空间保障资源	保障步骤	设备存储空间保障任务代号	设备存储空间保障任务说明	执行存储空间保障资源
1	T1	R1 型卫生清扫器对轻设备库房进行卫生打扫	R1_1,R1_2	2	T16	H1 型电气设备保障电子设备库房	H1_1,H1_2,H1_3,H1_4
1	T2	R1 型卫生清扫器对重设备库房进行卫生打扫	R1_3,R1_4	2	T17	H2 型数据设备保障电子设备库房	H2_1,H2_2
1	T3	Y 型管理人员进行空间督查	Y_1	2	T18	B 型高强照明设备保障电子设备库房	B_1,B_2
1	T4	M1 型辅助管理人员伴随值班管理人员 Y_1 巡查	M1_1,M1_2	2	T19	R1 型卫生清扫器对库房的走廊、卫生间进行清扫	R1_5,R1_6
1	T5	D 型机电工具设备检修人员对轻设备库房进行安全检修	D_1	2	T20	D 型机电工具设备检修人员对消防设备进行设备检修	D_1
1	T6	M1 型辅助管理人员伴随检修人员 D_1 进行检修	M1_3,M1_4	2	T21	M1 型辅助管理人员伴随检修人员 D_1 进行检修	M1_11,M1_12
1	T7	D 型机电工具设备检修人员对重设备库房进行安全检修	D_2,D_3,D_4	2	T22	D 型机电工具设备检修人员对逃生设备进行设备检修等	D_2
1	T8	M1 型辅助管理人员伴随检修人员进行检修	M1_5,M1_6,M1_7,M1_8,M1_9,M1_10	2	T23	M1 型辅助管理人员伴随检修人员 D_2 进行检修	M1_13,M1_14
2	T9	H1 型电气设备保障重设备库房运行	H1_1,H1_2,H1_3,H1_4	2	T24	D 型机电工具设备检修人员对安防设备进行设备检修等	D_3
2	T10	H2 型数据设备保障重设备库房运行	H2_1,H2_2	2	T25	M1 型辅助管理人员伴随检修人员 D_3 进行检修	M1_3,M1_4
2	T11	B 型高强照明设备保障重设备库房运行	B_1,B_2	2	T26	B 型高强照明设备保障工作间	B_3,B_4
2	T12	M2 环境控制设备保障轻设备库房运行	M2_1,M2_2,M2_3,M2_4	2	T27	B 型高强照明设备保障资料室	B_3,B_4
2	T13	H1 型电气设备保障数据中心	H1_1,H1_2,H1_3,H1_4	2	T28	B 型高强照明设备保障	B_3,B_4
2	T14	H2 型数据设备保障数据中心	H2_1,H2_2	2	T29	B 型高强照明设备保障工具室	B_3,B_4
2	T15	B 型高强照明设备保障数据中心	B_1,B_2	2	T30	对其他任务进行保障	M2_5—M2_16

为开始空间保障任务与结束空间保障任务, 不指代具体的空间保障任务。

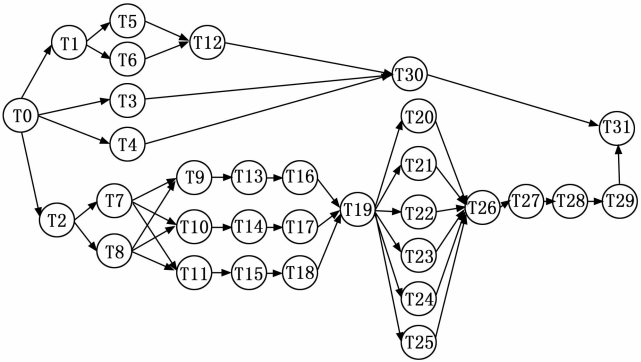


图 8 空间保障任务图

3.2 结果分析

由表中执行空间保障资源可知, 部分空间保障资源只存在单次使用, 不涉及空间保障任务之间的资源流动。为了便于描述, 以 D 型机电工具设备检修人员和 B 型高强照明设备为例, 在空间保障任务图的基础上得到原有完备资源流网络如图 9 所示。其中, 绿色弧线为 D 型机电工具设备检修人员对应的资源弧, 绿色标记的空间保障任务为由 D 型机电工具设备检修人员执行的空间保障任务, 空间保障任务周边标记的绿色文字为空间保障任务执行时间和空间保障资源需求数量; 蓝色弧线、蓝色标记的空间保障任务和蓝色文字与 B 型高强照明设备相关信息相对应, 这里不再赘述。

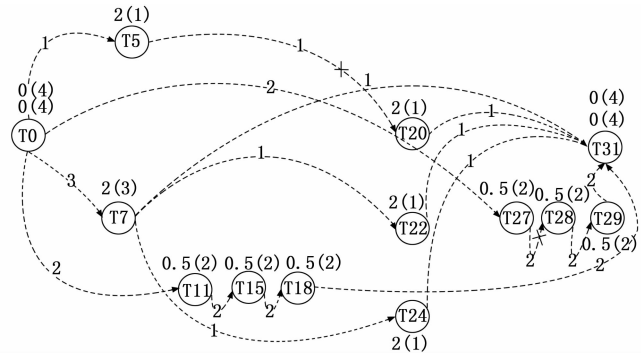


图 9 原有完备资源流网络

假定在空间保障计划执行过程中, 分别出现以下突发情况: (1) D_1 在执行完空间保障任务 T5 后, 因突发病情需要退出任务; (2) B_3 和 B_4 保障任务 T28 中突发故障, 保障任务失败, B_3 和 B_4 已无继续工作能力。此时, 图 9 中叉标记的资源弧故障, 得到非完备资源流网络, T20、T28 和 T29 为非完备节点。故障资源弧断开后, 默认连接至结束空间保障任务节点, 视为已经完成流转的资源弧。

利用预训练的 GCPN 网络对非完备资源流进行处理, 得到新生成的完备资源流网络, 如图 10 所示。需要

说明的是, 在处理过程中, 对于已经完成资源流转的资源弧不会被断开。新生成的完备资源流网络显示, 空间保障任务 T20 由 D_4 设备检修人员在参与执行完成保障任务 T7 后执行; 空间保障任务 T28 和空间保障任务 T29 由 B_1 和 B_2 在执行完成空间保障任务 T18 后执行。由此, 非完备节点得到了满足执行空间保障资源需求的资源输入, 空间保障资源配置得到了合理的调整。

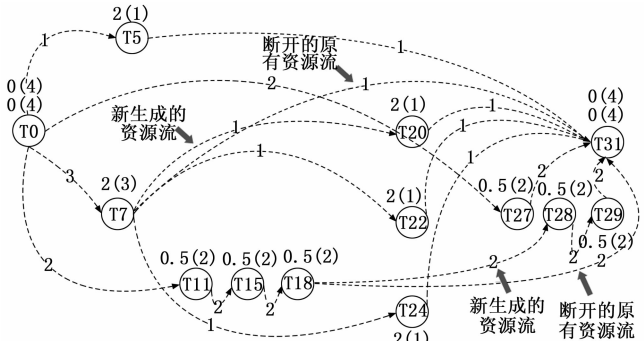


图 10 新生成的完备资源流网络

综上所述, 本文所提的资源流网络改进方法可有效将空间保障资源重调度问题转换为完备资源流网络生成问题, 并可利用预训练的 GCPN 网络进一步处理, 生成新的完备资源流网络。

4 结束语

本文研究了部分空间保障任务执行空间保障资源缺失条件下的空间保障资源重调度问题。首先, 对空间保障资源重调度问题进行了详细描述, 建立了相应的问题模型。其次, 针对现有资源流网络模型在复杂空间保障任务关系表示方面存在的不足, 对该模型进行了改进, 并给出了资源流网络完备性的相关定义, 将空间保障资源重调度问题转换为完备资源流网络生成问题。然后, 利用并扩展图表示学习、强化学习和生成对抗网络, 构建了图卷积策略网络, 用于生成符合目标导向的资源流网络。其中, 利用图卷积网络对非完备资源流网络进行节点嵌入计算, 为策略网络提供数据特征提取。利用策略网络预测资源流生成行动, 向非完备资源流网络输出连边行动。图卷积网络与策略网络组合构成了一个生成器, 与鉴别器共同形成对抗网络结构, 以提供对抗性奖励。最后, 设计原有空间保障计划在执行过程发生突发情况, 对所提的空间保障资源重调度方法进行了实验验证。通过分析实验过程中, 根据空间保障任务图得到的原有完备资源流网络, 以及利用预训练的 GCPN 网络新生成的完备资源流网络可知, 改进后的资源流网络模型可有效将空间保障资源重调度问题进行转换, 并且可利用本文所提的基于 GCPN 的完备资源流网络生成方法进一步处理, 可有效解决空间保障资源重调度问题。

(下转第 315 页)