

本文引用格式: 黄烁,张学习,谢兴旺.等.强化图注意力网络模型在选址路径问题中的应用[J].自动化与信息工程,2025,46(3):43-51.

HUANG Shuo, ZHANG Xuexi, XIE Xingwang, et al. Application of reinforced graph attention network model for location routing problem[J]. Automation & Information Engineering, 2025,46(3):43-51.

强化图注意力网络模型在选址路径问题中的应用*

黄烁 张学习 谢兴旺 张涛

(广东工业大学, 广东 广州 510006)

摘要: 针对传统注意力网络模型在选址路径问题中无法有效保留图结构信息的问题, 提出一种强化图注意力网络模型。首先, 在保留图结构信息的前提下, 编码器通过注意力机制提取图结构中的节点信息, 以获得节点的高维特征表示及整体图特征信息; 然后, 解码器利用门控循环单元有效捕获节点序列中的时间依赖关系, 并通过逐步解码的方式获取完整解; 最后, 引入额外的价值网络评估每个动作的价值, 以引导策略更新, 提升训练效率。实验结果表明, 该强化图注意力网络模型能够快速获取选址路径问题的高质量解。

关键词: 选址路径问题; 强化学习; 图注意力网络; 价值网络; 图结构

中图分类号: TP391

文献标志码: A

文章编号: 1674-2605(2025)03-0007-09

DOI: 10.12475/aie.20250307

开放获取

Application of Reinforced Graph Attention Network Model for Location Routing Problem

HUANG Shuo ZHANG Xuexi XIE Xingwang ZHANG Tao

(Guangdong University of Technology, Guangzhou 510006, China)

Abstract: To address the limitation of traditional attention network models in effectively preserving graph structural information for Location Routing Problems, this paper proposes a reinforced graph attention network model. First, while retaining graph structural information, the encoder extracts node information from the graph structure through attention mechanisms to obtain high-dimensional feature representations of nodes and global graph feature information. Then, the decoder utilizes Gated Recurrent Units to effectively capture temporal dependencies in node sequences and acquires complete solutions via step-by-step decoding. Finally, an auxiliary Value Network is introduced to evaluate the value of each action, guiding policy updates to enhance training efficiency. Experimental results demonstrate that this reinforced graph attention network model can rapidly obtain high-quality solutions for LRP.

Keywords: location routing problem; reinforcement learning; graph attention network; value network; graph structure

0 引言

选址路径问题 (location routing problem, LRP) 是一个组合优化问题, 旨在同时优化设施位置与车辆运输路径, 以最小化总成本, 广泛应用于物流、供应链管理等领域。

随着交通物流产业的持续发展, LRP 逐渐成为运筹优化领域的研究热点, 同时衍生了众多改进的数学

模型。文献[1]最早提出了 LRP, 但由于其复杂性和动态变化的特点, 求解相对困难。为此, 学者们提出了多种运筹优化算法, 主要分为精确算法^[2-6]和启发式算法^[7-9]两类。文献[10]构建了 LRP 的整数规划模型, 并利用分支定界法进行求解, 这是采用精确算法求解 LRP 的初步尝试。随后, 许多学者采用基于 0-1 线性模型的分支定界法^[11], 求解基于切割和定界算法^[12]的

* 基金项目: 广东省基础与应用基础研究基金自然科学基金项目 (2024A1515011652)

LRP。但精确算法仅适用于解决中小规模的 LRP，越来越多的学者开始关注能够处理大规模 LRP 的启发式算法。文献[13]设计并实现了一种基于非支配排序的离散萤火虫算法来求解 LRP，该算法在多样性指标上表现更好，且更具鲁棒性。文献[14]提出一种基于树的搜索算法用于位置决策，并结合禁忌搜索算法，通过分裂、交换、搬迁等操作进行路径规划，能够在较短时间内获得更好的解决方案。

随着深度强化学习在组合优化问题求解中^[15]的成功应用，越来越多的学者尝试利用其来解决 LRP。文献[16]首次采用端对端的方式将深度学习应用于旅行商问题。随后，文献[17]利用强化学习替代监督学习，求解车辆路径问题，并取得了接近最优解的性能。文献[18]基于 Transformer 模型架构提出一种基于注意力机制的网络模型，并采用带基线的策略梯度算法进行网络训练，提高了车辆路径等组合优化问题的求解效率。在文献[18]的基础上，文献[19]提出一种基于强化学习的两阶段算法，分别设计了用于解决选址问题和车辆路径问题的注意力模型，联合求解 LRP。然而，基于注意力机制的模型虽然能捕捉节点间的关联关系，但无法保留图结构信息，导致关键特征信息丢失。另外，以往的模型训练大多采用整体学习策略，难以提供有效的动作价值。

为此，本文提出一种用于求解 LRP 的强化图注意力网络（enhanced graph attention network, EGAT）模型，在有效保留图结构信息的基础上，通过注意力机制自动学习图结构中节点间的重要特征，增强模型对关键特征的感知能力。此外，还设计了额外的价值网络，用于评估每个动作的价值，以引导策略更新。

1 问题与模型

1.1 问题描述

LRP 的目标是在已知客户位置和需求的情况下，从多个仓库中选择最佳仓库，并确定最优配送方案和车辆运输路径，以实现总成本（仓库运营成本、车辆

固定成本、车辆运输成本）最低，LRP 示意图如图 1 所示。

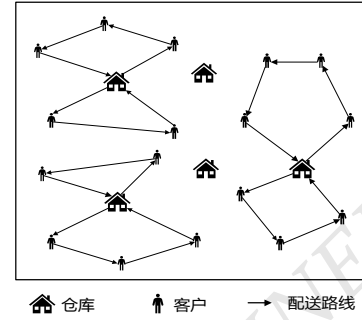


图 1 LRP 示意图

为了便于分析和研究，对 LRP 做出以下假设：

- 1) 所有车辆的车型、运输成本相同；
- 2) 客户需求小于车辆容量上限；
- 3) 每个客户有且仅有一个仓库为其服务；
- 4) 仓库之间不允许存在运输路径；
- 5) 每辆车从一个仓库出发且返回同一仓库。

1.2 LRP 数学模型

假设有 n 个节点，每个节点代表仓库或者客户，节点集合为 $N = N_d \cup N_c = \{1, 2, \dots, n\}$ 。其中，有 D 个仓库，仓库集合为 $N_d = \{1, 2, \dots, D\}$ ；有 C 个客户，客户集合为 $N_c = \{D+1, D+2, \dots, D+C\}$ ；车辆路径集合为 $R = \{1, 2, \dots\}$ ，其他参数如表 1 所示。

表 1 参数符号及其定义

参数符号	定义
s_v	节点 v 的二维坐标， $v \in N_d \cup N_c$
c_i	仓库 i 的容量， $i \in N$
o_i	仓库 i 的运营成本， $i \in N$
d_j	客户 j 的需求， $j \in N$
P	单位距离的路径成本
F	每辆车的固定成本
L	每辆车的容量
x_i	如果仓库 i 被选择，值为 1，否则为 0
u_{ij}	如果客户 j 被仓库 i 服务，值为 1，否则为 0
e_{ijk}	如果在路径 k 上，客户 j 被仓库 i 服务，值为 1，否则为 0

LRP 的数学模型为

$$z_{\min} = \sum_{i \in N_d} x_i o_i + \sum_{i \in N} \sum_{j \in N} \sum_{k \in R} P \|s_i - s_j\|_2 e_{ijk} + \sum_{i \in N_d} \sum_{j \in N_c} \sum_{k \in R} F e_{ijk} \quad (1)$$

式中： $z_{\min}$ 为最小化总成本的目标函数， $\sum_{i \in N_d} x_i o_i$

为仓库运营成本， $\sum_{i \in N} \sum_{j \in N} \sum_{k \in R} P \|s_i - s_j\|_2 e_{ijk}$ 为车辆运输

成本， $\sum_{i \in N_d} \sum_{j \in N_c} \sum_{k \in R} F e_{ijk}$ 为车辆固定成本。

$$\text{s.t.} \quad \sum_{j \in N} e_{ijk} - \sum_{j \in N} e_{jik} = 0, \forall i \in N, k \in R \quad (2)$$

公式(2)确保车辆驶入、驶出每个节点。

$$e_{ijk} + e_{jik} \leq 1, \forall i, j \in N, k \in R \quad (3)$$

公式(3)确保车辆不会重复驶入相同的路径。

$$e_{ijk} = 0, \forall i = j \in N, k \in R \quad (4)$$

公式(4)确保从某节点出发的车辆不会驶回该节点。

$$e_{ijk} = 0, \forall i, j \in N_d, k \in R \quad (5)$$

公式(5)确保仓库之间不存在运输路径。

$$\sum_{i, j \in N} \sum_{k \in R} e_{ijk} \leq |S| - 1, 2 \leq |S| \leq n - 1, S \subseteq N \quad (6)$$

公式(6)防止每条路径中产生子路径。

$$\sum_{i \in N} \sum_{j \in N_c} d_j e_{ijk} \leq L, \forall k \in R \quad (7)$$

公式(7)确保每条路径上的客户总需求不超过车辆容量上限。

$$\sum_{j \in N_c} d_j u_{ij} \leq y_i c_i, \forall i \in N_d \quad (8)$$

公式(8)确保仓库服务的客户总需求不超过仓库容量上限。

$$\sum_{i \in N_d} u_{ij} = 1, \forall j \in N_c \quad (9)$$

公式(9)确保每个客户有且仅有一个仓库为其

服务。

$$\sum_{i \in N_d} \sum_{j \in N_c} e_{ijk} \leq 1, \forall k \in R \quad (10)$$

$$\frac{1}{|N_c|} \sum_{i \in N} \sum_{j \in N} e_{ijk} \leq \sum_{i \in N_d} \sum_{j \in N} e_{ijk}, \forall k \in R \quad (11)$$

公式(10)、(11)确保每条路径上有且仅有一个仓库。

$$\frac{1}{|N_c|} \sum_{i \in N_c} \sum_{k \in R} e_{ijk} \leq y_i \leq \sum_{j \in N_c} \sum_{k \in R} e_{ijk}, \forall i \in N_d \quad (12)$$

公式(12)通过仓库与客户之间的路径，限制仓库的开放数量。

2 算法描述

LRP 可被视为一个顺序决策过程，能够采用深度强化学习算法进行求解。

2.1 EGAT 模型

本文提出的 EGAT 模型主要由编码器、解码器和价值网络组成，结构如图 2 所示。

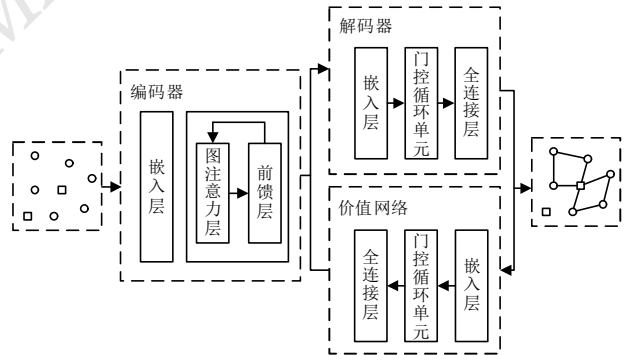


图 2 EGAT 模型结构

编码器在保留原有输入图结构的情况下，提取输入图结构中的节点信息，以获得节点的高维特征表示及整体图特征信息。

解码器接收编码器输出的节点特征表示和图特征信息，并采用逐步解码的方式，在每个时间步选择动作并更新状态，直至获取完整解。

价值网络在训练过程中提供动作的价值，辅助 EGAT 模型学习更精细的动作，使策略更新更稳定。

2.2 编码器

编码器由嵌入层和图注意力网络层组成，结构如

图 3 所示。其中，图注意力网络层由 3 个图注意力层和 1 个前馈层构成。

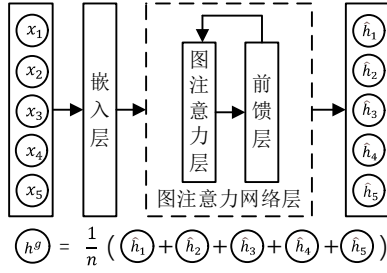


图 3 编码器结构

2.2.1 嵌入层

嵌入层的输入 $X = \{x_1, x_2, \dots, x_n\}$ 为仓库和客户的节点信息。其中， $x_i = [s_{x_i}; s_{y_i}; d_i]$ ， s_{x_i} 、 s_{y_i} 为节点 i 的二维坐标， d_i 为节点 i 的需求（仓库需求为 0）。嵌入层将每个输入 x_i 映射到节点嵌入特征 h_i^0 （特征维度为 128）：

$$h_i^0 = W^x \times x_i + b^x \quad (13)$$

式中： W^x 、 b^x 分别为可学习的权值矩阵和偏置。

2.2.2 图注意力层

将节点嵌入特征 $h^0 = \{h_1^0, h_2^0, \dots, h_n^0\}$ 输入到图注意力层，以捕获节点间的关联关系和重要性。

首先，根据节点嵌入特征 h_i^0 、邻接节点的嵌入特征 h_j^0 来计算注意力权重 α_{ij} ：

$$\alpha_{ij} = \frac{\exp\{\text{LeakyReLU}(a^T [W^g h_i \parallel W^g h_j])\}}{\sum_{k \in N_i} \exp\{\text{LeakyReLU}(a^T [W^g h_i \parallel W^g h_k])\}} \quad (14)$$

式中： W^g 为特征变换矩阵， a 为注意力权重的学习参数， $\parallel$ 为特征拼接操作， N_i 为节点 i 的邻接节点集合。

然后，根据注意力权重对邻接节点的特征进行加权求和，更新节点嵌入特征为 $\tilde{h}_i$ ：

$$\tilde{h}_i = \sigma \left(\sum_{j \in N_i} \alpha_{ij} W h_j^0 \right) \quad (15)$$

式中： σ 为 ReLU 激活函数， W^a 为可训练权重

矩阵。

通过多头注意力机制，将多个注意力头的输出拼接或平均，进一步丰富节点的表达能力。

最后，图注意力层输出更新后的节点嵌入特征 $\tilde{h} = \{\tilde{h}_1, \tilde{h}_2, \dots, \tilde{h}_n\}$ 。

2.2.3 前馈层

将更新后的节点嵌入特征输入到前馈层，进一步提取节点的高维非线性特征，以提升模型的表达能力。

$$\hat{h}_i = W_2^F \times \text{ReLU}(W_1^F \times \tilde{h}_i + b_1^F) + b_2^F \quad (16)$$

式中： W_1^F 和 b_1^F 为第一个全连接层的网络参数， W_2^F 和 b_2^F 为第二个全连接层的网络参数。

最后， $\hat{h}_i$ 整体图特征 h^g 为

$$h^g = \frac{1}{n} (\hat{h}_1 + \hat{h}_2 + \dots + \hat{h}_n) \quad (17)$$

2.3 解码器

本文采用强化学习框架将 LRP 的求解建模为一个马尔可夫决策过程。

首先，LRP 的求解状态表示为 $S = \{S_g, S_a\}$ ，其中 S_g 为编码器输出的整体图特征信息； $S_a = \{\hat{h}_{t-1}, L_t\}$ 为智能体状态， $\hat{h}_{t-1}$ 为智能体上一步选择的节点， L_t 为当前车辆的剩余容量；

然后，动作 π_t 为智能体在当前时间步 t 所选择的节点；

最后，回报 R 为总成本的负数。

解码器可充当强化学习框架下的智能体，主要由嵌入层、门控循环单元和全连接层组成，结构如图 4 所示。

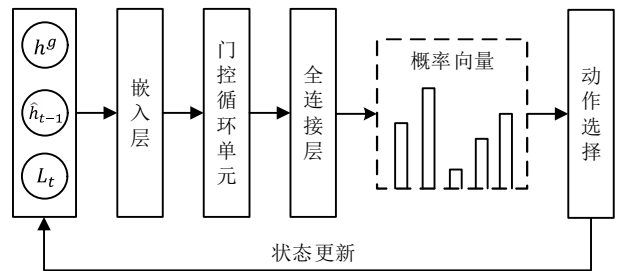


图 4 解码器结构

2.3.1 嵌入层

嵌入层将智能体的状态特征 h^c 映射为上下文特征向量 $\hat{h}^c$ ：

$$h^c = \text{concat}(h^g, \hat{h}_{t-1}, L_t) \quad (18)$$

$$\hat{h}^c = W^c \times h^c + b^c \quad (19)$$

式中： W^c 和 b^c 为解码器嵌入层的网络参数。

2.3.2 门控循环单元

将上下文特征向量输入到门控循环单元，以捕获节点间的动态依赖关系和交互信息。

首先，通过公式(20)~(23)更新上一个时间步 $t-1$ 的输入节点嵌入特征 $\hat{h}_{t-1}$ 和上下文特征 $\hat{h}_{t-1}^c$ 的隐藏状态：

$$z_t = \sigma(W^z \times \hat{h}_{t-1} + U^z \times \hat{h}^c + b^z) \quad (20)$$

$$r_t = \sigma(W^r \times \hat{h}_{t-1} + U^r \times \hat{h}^c + b^r) \quad (21)$$

$$\hat{h}_t = \tanh(W^h \times \hat{h}_{t-1} + r_t \odot U^h \times \hat{h}_{t-1}^c + b^h) \quad (22)$$

$$\hat{h}_t^c = (1 - z_t) \odot \hat{h}_{t-1}^c + z_t \odot \hat{h}_{t-1} \quad (23)$$

式中： z_t 为更新门， W^z 、 U^z 和 b^z 为更新门可训练的权重参数矩阵， r_t 为重置门， W^r 、 U^r 和 b^r 为重置门可训练的权重参数矩阵， W^h 、 U^h 和 b^h 为隐藏状态权重参数， $\odot$ 为元素级乘法。

2.3.3 全连接层

全连接层接收门控循环单元输出的隐藏状态序列 $\hat{h}_t^c$ ，先将其映射到输出动作空间 u_t ，再通过 softmax 函数输出概率分数 p_t ：

$$u_t = W^u \times \hat{h}_t^c + b^u \quad (24)$$

$$p_t = \text{softmax}(u_t) = \frac{e^{u_t}}{\sum_{j \in N} e^{u_{j,t}}} \quad (25)$$

式中： W^u 和 b^u 为可训练的权重参数矩阵， $u_{j,t}$

为 u_t 第 j 行参数。

根据概率分数 p_t 进行多项式采样，获得动作 π_t 。重复以上步骤直到访问完所有客户，生成完整解 $\pi = \{\pi_1, \pi_2, \dots, \pi_T\}$ 。

2.4 价值网络及训练算法

价值网络采用与解码器相同的网络结构，其输入为解码器输出动作所选择的节点，输出为一维的评估分数，用以评估编码器输出动作的优劣。

本文结合策略优化和价值评估的训练方法，能够有效平衡策略梯度算法的高方差和基于值函数方法的偏差问题。对于一个给定的实例 s ，将 EGAT 模型的编码器-解码器结构组合为策略网络 θ ，通过策略网络 θ 输出策略的动作概率分布 $p_\theta(\pi|s)$ 采样得到动作 π_t ，即 $\pi_t = \text{sample}(p_\theta(\pi|s))$ 。同时，将价值网络估计状态的状态基准值 $V_\phi(s)$ 作为策略梯度更新的基准，以降低估计方差。

根据策略梯度定理，策略网络的期望累计回报为

$$L(\theta) = \mathbb{E}_{\pi_\theta}[R(\pi)] \quad (26)$$

式中： $R(\pi)$ 为策略解 $\pi = \{\pi_1, \pi_2, \dots, \pi_T\}$ 的累计回报。

策略梯度的计算公式为

$$\nabla_\theta L(\theta) = \mathbb{E} \left[\sum_{t=1}^T \nabla_\theta \log \pi_\theta(\pi_t|s_t) \cdot A_t \right] \quad (27)$$

式中： $A_t = R_t - V_\phi(s)$ 为优势函数，用于衡量当前动作的实际回报 R_t 与状态基准值 $V_\phi(s)$ 的偏差。

在每次训练迭代过程中，通过梯度上升法更新策略网络参数 θ ，通过最小化状态值的均方误差更新价值网络参数 ϕ ：

$$\theta \leftarrow \theta + \alpha_\theta \cdot \nabla_\theta L(\theta) \quad (28)$$

$$\phi \leftarrow \phi - \alpha_\phi \cdot \nabla_\phi (R_t - V_\phi(s))^2 \quad (29)$$

式中： α_θ 和 α_ϕ 为学习率。

EGAT 模型训练算法的伪代码如下：

EGAT 模型训练算法

输入：节点集 N ，训练轮数 E ，批次 B ，步数限制 T ，策略网络参数 θ

输出：位置-路径构建策略

```

1: 随机初始化策略网络参数  $\theta$ 
2: for epoch = 1, 2, ..., E do
3:   for batch = 1, 2, ..., B do
4:     获取当前批次数据，设置  $t \leftarrow 0$ 
5:     while  $t < T$  do
6:       根据策略网络参数  $\theta$  选择节点  $\pi_t$ 
7:        $t \leftarrow t + 1$ 
8:     end
9:     根据公式(26)和(27)计算梯度  $\nabla_{\theta} L(\theta)$ 
10:    更新策略网络参数  $\theta \leftarrow \theta + \alpha_{\theta} \cdot \nabla_{\theta} L(\theta)$ 
11:    更新价值网络参数  $\varphi \leftarrow \varphi - \alpha_{\varphi} \cdot \nabla_{\varphi} (R_t - V_{\varphi}(s))^2$ 
12:  end
13: end

```

首先，调整相关的训练超参数，并初始化策略网络参数 θ ；

然后，根据策略网络参数 θ ，对批次中的每个实例逐步生成动作 π_t ，直到选择完所有节点；

最后，根据公式(26)、(27)计算策略梯度 $\nabla_{\theta} L(\theta)$ ，用于更新策略网络参数 θ 和价值网络参数 φ 。

3.1 实验设置

本实验在 NVIDIA A30、Intel Xeon Silver 4310 CPU (2.10 GHz) 和 128 GB RAM 上进行，通过 Python 3.10 和 PyTorch 2.1.0 实现 EGAT 模型训练。

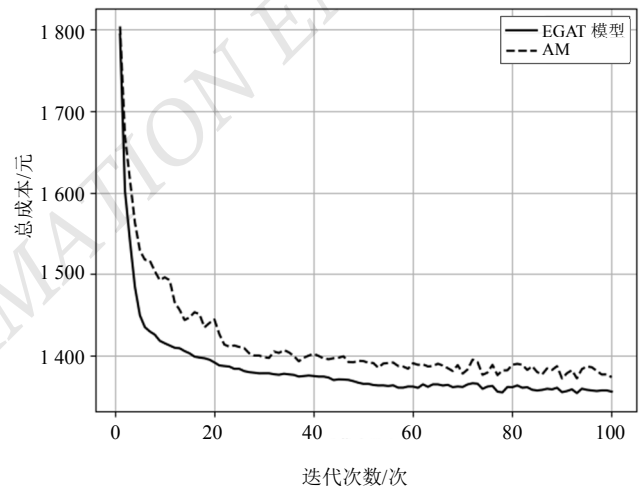
EGAT 模型训练参数设置：训练轮数为 100 轮，每轮的批次为 6 400 次，每批算例为 640 例；采用 Adam 优化器来优化策略网络参数，学习率为 1×10^{-4} 。

通过 3 种不同规模的实例来评估本文提出的 EGAT 模型性能：小规模实例（20 个客户，5 个仓库，车辆容量 30 吨）、中规模实例（50 个客户，5 个仓库，车辆容量 50 吨）、大规模实例（100 个客户，10 个仓库，车辆容量 50 吨），节点在 $[0, 100] \times [0, 100]$

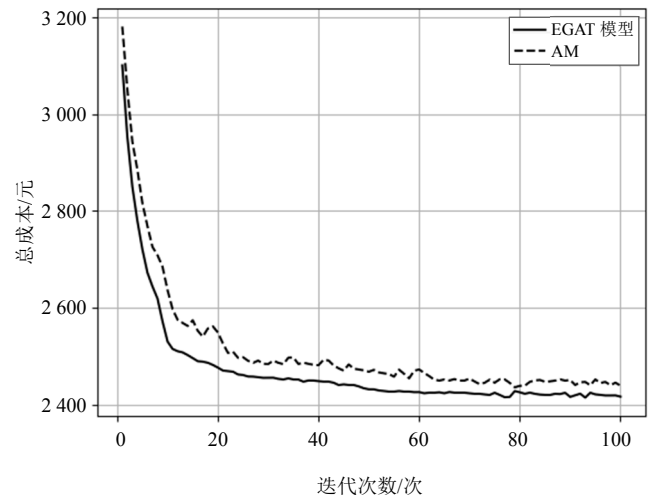
上均匀分布。在 3 种不同规模的实例上分别测试 1 000 次，并将实例平均求解成本和平均求解时间作为模型性能的评价指标。平均成本越低，代表模型性能越好；平均求解时间越短，代表模型求解效率越高。

3.2 对比实验

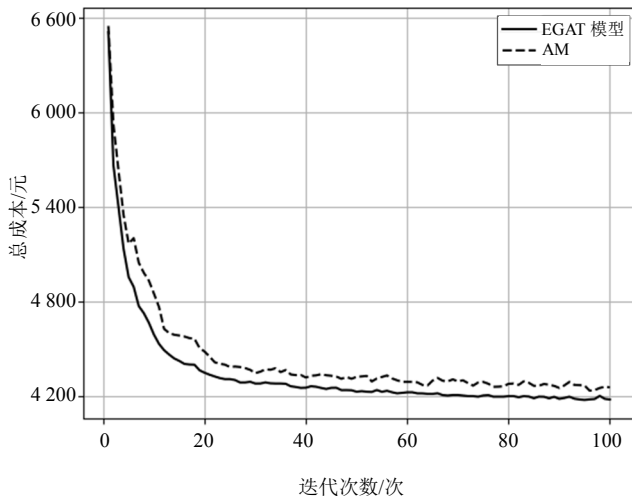
为验证本文提出的 EGAT 模型的有效性，将其与文献[8]的注意力模型（attention model, AM）^[18]进行对比实验。相较于 AM 采用的强化学习训练算法^[20]，本文 EGAT 模型采用价值网络来引导策略更新，减少了策略梯度估计的方差，提升了训练效率。EGAT 模型与 AM 训练曲线如图 5 所示。



(a) 小规模实例



(b) 中规模实例



(c) 大规模实例

图5 EGAT模型与AM训练曲线

由图5可知：EGAT模型与AM训练曲线具有一致的下降趋势，即总成本在开始时迅速下降，然后逐渐趋于稳定；EGAT模型的最终稳态值和训练稳定性均优于AM，表明引入价值网络加速了模型收敛过程，提升了训练稳定性和策略的最终表现。

3.3 消融实验

为了进一步评估价值网络对模型的影响，将未采用价值网络和采用价值网络的图注意力模型在大规模实例上进行训练，训练曲线如图6所示。

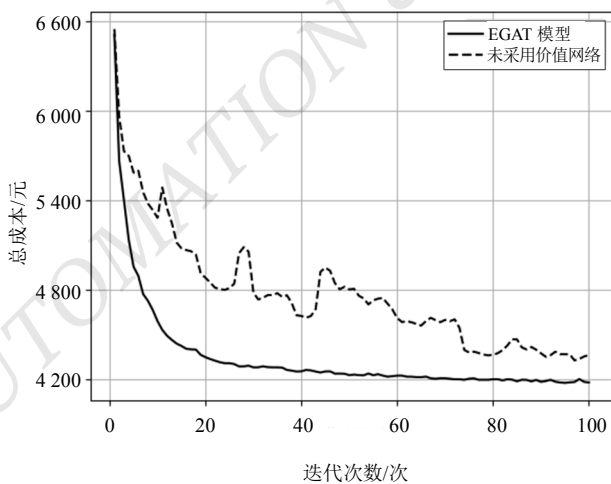


图6 未采用价值网络和采用价值网络的图注意力模型训练曲线

由图6可知，采用价值网络的图注意力(EGAT)模型的训练曲线下降较为平滑且连续，下降幅度较大且波动较小；未采用价值网络的图注意力模型的训练曲线波动较大且存在震荡，表明价值网络能够有效提升模型的训练效果和性能。

3.4 模型性能评估

为进一步验证EGAT模型的性能，将其与AM及求解LRP的常用算法：精确算法（采用Gurobi求解器）、模拟退火（simulated annealing, SA）算法^[21]、遗传算法（genetic algorithm, GA）^[22]和灰狼优化（gray wolf optimization, GWO）算法^[23-24]在求解成本、求解时间两个评价指标上进行对比，结果分别如表2、3所示。

表2 EGAT模型与其他算法/模型求解成本对比 单位：元

算法/模型	问题规模		
	20-5 小规模	50-5 中规模	100-10 大规模
Gurobi	1 313.5	2 451.4	4 420.2
SA	1 361.2	2 505.3	4 621.7
GA	1 401.7	2 567.1	4 821.5
GWO	1 414.3	2 636.2	4 752.4
AM	1 374.3	2 447.2	4 260.8
EGAT	1 357.8	2 416.0	4 173.8

表3 EGAT模型与其他算法/模型求解时间对比 单位：s

算法/模型	问题规模		
	20-5 小规模	50-5 中规模	100-10 大规模
Gurobi	2 461.72	3 600.00	3 600.00
SA	59.32	109.22	256.25
GA	61.21	134.21	298.34
GWO	74.37	144.41	314.58
AM	0.18	0.45	1.52
EGAT	0.15	0.31	1.02

由表2、3可知：本文提出的EGAT模型在20-5和50-5规模问题上，求解成本与Gurobi、SA、GA和GWO算法相近，但求解时间显著快于上述算法；在100-10规模问题上，其求解成本和求解时间均优于上述算法。具体而言：Gurobi在小规模问题上能够获得

最优解，但其求解时间随着规模问题的增大而显著增加；GA 和 SA 算法在中小规模问题上表现尚可，但在大规模问题上求解成本与求解时间均上升；GWO 算法求解成本较低，但求解时间较长。相较于同样采用深度强化学习框架的 AM，EGAT 模型的求解成本和求解时间均表现得更好，这主要得益于强化图注意力模型的快速推理能力，能够在较短的时间内生成高质量的解。实验结果有效证明了本文提出的 EGAT 模型求解 LRP 的实用性和有效性。

4 结论

本文提出一种 EGAT 模型，相较于传统的 Transformer 模型架构，可有效保留图结构信息，增强对关键特征的感知能力。此外，通过引入额外的价值网络来评估每个动作的价值，引导策略更新，减少策略梯度估计的方差，提升训练效率。实验结果表明，本文提出的 EGAT 模型在小规模问题上的求解成本与传统算法相当，但效率更高；在中大规模问题上的求解成本和求解时间均优于传统算法，证明了其在 LRP 求解中的实用性和高效性。后续研究将考虑模型在更复杂现实场景下的泛化能力和更大规模问题上的求解能力。

©The author(s) 2024. This is an open access article under the CC BY-NC-ND 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

参考文献

- [1] COOPER L. The transportation location problem[J]. *Operations Research*, 1972,20(1):94-108.
- [2] BERGER R, COULLARD C, DASKIN M. Location routing problems with distance constraints[J]. *Transportation Science*, 2007,41(1):29-43.
- [3] FEILLET D, DEJAX P, GENDREAU M, et al. An exact algorithm for the elementary shortest path problem with resource constraints: Application to some vehicle routing problems[J]. *Networks: An International Journal*, 2004,44(3):216-229.
- [4] AKCA Z, BERGER R, RALPHS T. A branch-and-price algorithm for combined location and routing problems under

- capacity restrictions[C]//*Operations Research and Cyber-Infrastructure*. Springer US, 2009:309-330.
- [5] BALDACCIO R, MINGOZZI A, WOLFLER CALVO, R.. An exact method for the capacitated location-routing problem[J]. *Operations Research*, 2011,59(5):1284-1296.
- [6] CONTARDO C, CORDEAU J F, GENDRON B. An exact algorithm based on cut-and-column generation for the capacitated location-routing problem[J]. *INFORMS Journal on Computing*, 2014,26(1):88-102.
- [7] MANIEZZO V, STÜTZLE T, VOß S. *Matheuristics*, volume 10 of *annals of information systems*[J]. New York: Springer, 2010.
- [8] CONTARDO C, CORDEAU J F, GENDRON B. A GRASP+ ILP-based metaheuristic for the capacitated location-routing problem[J]. *Journal of Heuristics*, 2014,20:1-38.
- [9] ALVIM A, TAILLARD E. POPMUSIC for the world location-routing problem[J]. *EURO Journal on Transportation and Logistics*, 2013,2(3):231-254.
- [10] LAPORTE G, NOBERT Y. An exact algorithm for minimizing routing and operating costs in depot location[J]. *European Journal of Operational Research*, 1981,6(2):224-226.
- [11] BELENGUER J M, BENAVENT E, PRINS C, et al. A branch-and-cut method for the capacitated location-routing problem[J]. *Computers & Operations Research*, 2011,38(6):931-941.
- [12] CONTARDO C, CORDEAU J F, GENDRON B. An exact algorithm based on cut-and-column generation for the capacitated location-routing problem[J]. *INFORMS Journal on Computing*, 2014,26(1):88-102.
- [13] SHAHSAVARI-POUR N, BAHRAM-POUR N, KAZEMI M. Fuzzy multi-objective location and routing problem[J]. *Journal of Intelligent & Fuzzy Systems*, 2020,39(3):3259-3273.
- [14] SCHNEIDER M, LOFFER M. Large composite neighborhoods for the capacitated location-routing problem[J]. *Transportation Science*, 2019,53(1):301-318.
- [15] BENGIO Y, LODI A, PROUVOST A. Machine learning for combinatorial optimization: a methodological tour d' horizon [J]. *European Journal of Operational Research*, 2021,290(2):405-421.
- [16] VINYALS O, FORTUNATO M, JAITLY N. Pointer networks [C]. *The 29th Conference on Neural Information Processing Systems*. Montreal, 2015:2692-2700.
- [17] BELLO I, PHAM H, LE Q V, et al. Neural combinatorial optimization with reinforcement learning[J]. *arXiv: 1611.09940*, 2016.
- [18] KOOL W, VAN HOOFF H, WELLING M. Attention, learn to

- solve routing problems[J]. arXiv: 1803.08475,2018.
- [19] WANG S, ZHOU J, LIANG H, et al. A new approach for solving location routing problems with deep reinforcement learning of emergency medical facility[C]// Proceedings of the 8th ACM SIGSPATIAL International Workshop on Security Response using GIS, 2023:50-53.
- [20] WILLIAMS R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning[J]. Machine Learning, 1992,8:229-256.
- [21] YU X, ZHOU Y, LIU X F. The two-echelon multi-objective location routing problem inspired by realistic waste collection applications: The composable model and a metaheuristic algorithm[J]. Applied Soft Computing, 2020,94:106477.
- [22] HIASSAT A, DIABAT A, RAHWAN I. A genetic algorithm approach for location-inventory-routing problem with perishable products[J]. Journal of Manufacturing Systems, 2017,42: 93-103.
- [23] 冯苗苗,崔敏,吕淑清,等.混合遗传灰狼算法在装配车间调度中的应用[J].机电工程技术,2023,52(9):118-122.
- [24] TIRKOLAE E B, GOLI A, MARDANI A. A novel two-echelon hierarchical location-allocation-routing optimization for green energy-efficient logistics systems[J]. Annals of Operations Research, 2023,324(1):795-823.

作者简介:

黄烁,男,1999年生,在读硕士研究生,主要研究方向:运筹优化与深度强化学习。E-mail: 2112204091@mail2.gdut.edu.cn

张学习,男,1978年生,博士研究生,副教授,主要研究方向:机器人与智能控制。E-mail: zxxnet@gdut.edu.cn

谢兴旺,男,2001年生,在读硕士研究生,主要研究方向:运筹优化与深度强化学习。E-mail: 1570475816@qq.com

张涛,男,2004年生,在读本科生,主要研究方向:运筹优化与深度强化学习。E-mail: 2430515400@qq.com