

金友龙, 夏大文. 基于领域自适应深度强化学习的跨分布车辆路径求解方法[J]. 智能计算机与应用, 2025, 15(12): 17-22.
DOI: 10.20169/j.issn.2095-2163.25101405

基于领域自适应深度强化学习的跨分布车辆路径求解方法

金友龙, 夏大文

(贵州民族大学 数据科学与信息工程学院, 贵阳 550025)

摘要: 现有基于神经网络的车辆路径求解方法通常假设训练和测试实例服从相同分布(即, 均匀分布), 从而导致跨分布场景的泛化能力弱。为此, 本文提出一种基于领域自适应深度强化学习的跨分布车辆路径求解方法, 核心在于领域自适应。该方法设计一种基于分布引导的自适应策略网络(DGATP), 并将其嵌入端到端深度强化学习框架(DRL), 以解决跨分布车辆路径求解问题。具体的是, 首先, 构建分布识别模块, 以感知源域与目标域的差异并进行特征提取和分布识别; 其次, 建立门控融合网络, 以自适应地加权与融合不同分布的特征; 最后, 设计感知注意力解码器, 以生成路由策略。基于两个代表性深度模型的实验结果表明, 与传统方法相比, DGATP 性能在跨分布场景下取得显著提升, 展现优异的泛化性和通用性。

关键词: 车辆路径问题; 深度强化学习; DGATP; 门控融合网络; 领域自适应

中图分类号: TP181

文献标志码: A

文章编号: 2095-2163(2025)12-0017-06

Across-distribution vehicle routing method based on domain-adaptive deep reinforcement learning

JIN Youlong, XIA Dawen

(School of Data Science and Information Engineering, Guizhou Minzu University, Guiyang 550025, China)

Abstract: Most existing neural network-based vehicle routing methods usually assume identical distributions (i. e., a uniform distributions) across instances, which makes them difficult to generalize to cross-distribution scenarios. To this end, this paper proposes a cross-distribution vehicle routing approach based on domain-adaptive reinforcement learning, with domain adaptation as its core principle. The proposed method involves designing a Distribution-Guided Adaptive Policy Network (DGATP) and embedding it within an end-to-end deep reinforcement learning (DRL) framework to address cross-distribution vehicle routing problems. Specifically, a distribution recognition module is first constructed to perceive differences between the source and target domains by performing feature extraction and distribution identification. Then, a gated fusion network is established to adaptively weight and fuse features from different distributions. Finally, a perception-aware attention decoder is designed to generate routing policies. Experimental results based on two representative deep learning models demonstrate that the DGATP approach significantly outperforms traditional methods in cross-distribution scenarios, exhibiting excellent generalization and universality.

Key words: vehicle routing problems; deep reinforcement learning; DGATP; gate-controlled fusion network; domain adaptation

0 引言

随着物流需求的日益增长, 开发高性价比的物流系统已成为提高竞争力的关键因素, 物流支出的份额成为评估商业效率的关键指标。特别是, 供应链的最后一英里已成为运输总成本的重要因素, 约占总支出的50%^[1]。因此, 车辆路径问题(VRP)已得到政界、学术界和工业界的广泛关注。

VRP在货运配送^[2]、最后一英里物流^[3]和网约车^[4]等领域得到广泛应用, 是经典的组合优化问题, 涌现出众多精确或近似的启发式算法。传统启发式算法虽然计算效率高, 但是其解的质量通常受到限制。近年来, 基于深度学习的模型可以在确保期望解的同时进一步降低计算代价。然而, 在复杂的现实环境中, 现有模型主要是集中在单一分布(如均匀分布)^[5,6]的数据上进行训练, 容易导致跨

基金项目: 贵州省高等学校大数据分析与应用重点实验室(黔教技[2023]012号)。

作者简介: 金友龙(1999—), 男, 硕士研究生, 主要研究方向: 大数据分析, 人工智能。Email: 1242637026@qq.com。

通信作者: 夏大文(1982—), 男, 博士, 教授, 硕士生导师, 主要研究方向: 大数据, 人工智能, 数据挖掘等。Email: dwxia@gzmu.edu.cn。

收稿日期: 2025-10-14

哈尔滨工业大学主办 ◆ 学术研究与应用

分布现实场景下的泛化能力显著下降。鉴于实际场景的 VRP 实例通常服从多样分布甚至未知分布,如何使得深度模型在保证解质量的同时具有跨分布的泛化能力,已成为推动实际应用的关键挑战。

1 研究现状

近年来,由于传统算法^[7]的手工规则设计复杂且难以动态适应,导致车辆路径问题求解逐渐进入神经网络领域^[8]。特别是,基于深度模型的神经网络方法已引起广泛关注,该类方法大致可以分为神经构造模型和神经改进模型。在神经构造模型方面,利用深度(强化)学习以端到端的方式自回归地构建解策略。Vinyals 等^[9]使用指针网络结合有监督学习解决旅行商问题(TSP),Bello 等^[10]将该方法扩展至强化学习领域,为求解 VRP 奠定基础^[11]。随后,进一步探索图拓扑信息^[12],利用图卷积神经网络(GCN)计算边出现的概率以求解 VRP。随着注意力机制(AM)的发展,基于编码器-解码器结构的学习构造模型在路径优化中展现出明显优势^[13]。此外,启发式算法在 VRP 的广泛应用(如 LKH3^[14]算法)及确定性问题上效果显著,但是在动态需求的 VRP 上表现得还不够稳定。

同时,利用深度(强化)学习结合局部搜索,以迭代改进初始解。Chen 等^[15]提出 NeuReWriter,通过改进局部重构解决方案。Hottung 等^[16]利用深度神经网络学习实现大型邻里搜索。Ma 等^[17]进一步优化 Wu 等的深度模型,为双向协同 Transformer(DACT)交替解决车辆路径问题,以构建性能更优的深度模型。总体上看,改进方法比构造方法耗费更长的推理时间,以获得更高的解质量。

上述神经网络方法通常在同一实例分布下训练深层模型,性能虽然优于传统启发式算法,但是难以将学习到的策略迁移至新的实例分布,为实际应用带来严峻挑战。因此,研究者们开始关注交叉泛化问题。Xin 等^[18]提出一种基于生成式对抗性网络(GAN)框架,生成用于模型训练的难解实例。Wang 等^[19]利用博弈论,提出策略空间响应 Oracle(PSRO)框架,并学习可训练求解器和实例生成器。Zhang 等^[20]设计一种硬度自适应的 TSP 实例生成器,并采用课程学习方法对模型进行训练,以侧重于数据扩充或仅限于 TSP。与其不同的是,Bi 等^[21]采用知识蒸馏方法,针对不同的分布进行训练得到不同的教师模型,以产生综合性更强的学生模型。然而,上述方法存在手工设定规则、模型经历多次参数

化学习以及串行式处理数据的局限性,从而导致推理时间长、求解精度低等问题。

综上所述,大多数现有的基于深度模型的方法在解决车辆路径问题时存在分布泛化能力不足、推理时间长和求解精度低等问题。为此,本文提出一种基于领域自适应强化学习的跨分布车辆路径求解方法。

2 问题定义

2.1 路网图

在车辆路径求解问题中,旨在满足所有客户需求的同时,实现成本最低。在车辆路径问题中,成本被定义为路线距离。在此,定义城市路网为 G ,如图 1 为不同分布对应的路网节点,整个道路网络由顶点和边构成。拓扑定义为:

$$G = (V, E)$$

其中, $V = \{x^i \mid s_i, i = 0, 1, 2, \dots, n\}$ 表示节点集合且每一个节点 $x^i = (s_i, d_i)$, 该节点由坐标位置 s_i (静态信息)和客户需求 d_i (动态信息)组成, V 为顶点集, E 为边集, $E = \{e_{ij} \mid e_{ij} = (v_i, v_j)\}$, e_{ij} 表示节点 v_i 与节点 v_j 之间的连边。

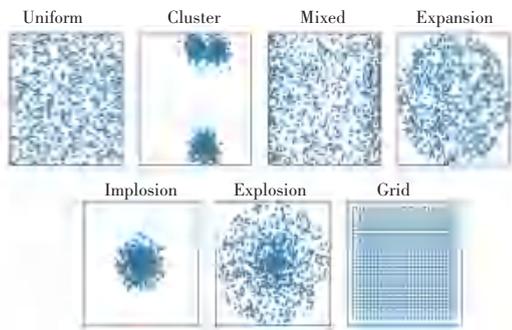


图 1 分布实例 $G(N=1\ 000)$

Fig. 1 Distribution instance $G(N=1\ 000)$

2.2 车辆路径规划

为了确定总出行成本最小化目标函数,需要确保以下约束条件成立:(1)确保每个节点(除仓库外)仅能被访问一次。(2)确保流量守恒,即车辆到达一个点并从该点出发。(3)确保车辆的路径从起始点开始并在终点结束。(4)确保车辆的容量没有超过最大载重量 Q 。基于方向约束(即 VRPs 的约束条件)和 Binary(0-1)约束,若车辆 k 在访问客户点 i 之后移动到客户点 j , 则 e_{ij}^k 为 1, 否则为 0。为此,构建动态目标路径规划模型如下:

$$\min L = \sum_{k=0}^m \sum_{i=0}^n \sum_{j=0}^n c_{ij} \cdot e_{ij}^k \quad (1)$$

$$\text{s. t. } \sum_{k=1}^m \sum_{j=1}^n e_{ij}^k = 1, \forall i = 1, \dots, n \quad (2)$$

$$\sum_{i=0}^N e_{ij}^k - \sum_{j=0}^N e_{ji}^k = 0, \forall i \in V \quad (3)$$

$$\sum_{j=1}^n e_{0j}^k = 1, \forall k = 1, \dots, m \quad (4)$$

$$\sum_{j=1}^n e_{j0}^k = 1, \forall k = 1, \dots, m \quad (5)$$

$$\sum_{j=1}^n d_j \cdot e_{ij}^k \leq Q, \forall i = 1, \dots, n, \forall k = 1, \dots, m \quad (6)$$

$$e_{ij}^k \in \{0, 1\}, \forall i = 0, \dots, n, \forall j = 0, \dots, n, \forall k = 0, \dots, m \quad (7)$$

其中, 目标函数(1)以距离为单位, 使得总出行成本最小化。式(2)确保每一个节点(除厂库外)仅能被访问一次。式(3)确保流量守恒, 即车辆到达一个点并从该点出发。式(4)和式(5)确保车辆的路径从起始点开始并在终点结束, 式(6)确保车辆容量没有超过最大载重量。式(7)为 Binary(0-1)约束, 若车辆 k 在访问客户点 i 后移动到客户点 j , 则为 1, 否则为 0。

3 DGATP 方法

如图 2 所示, DGATP 方法基于卷积神经网络的特征提取器识别数据分布, 生成分布特征向量, 设计门控融合机制, 并通过分布特征动态加权不同分支。接着, 提出感知注意力机制方法, 使得分布特征直接参与 VRP 的节点选择决策。最后, 该方法被集成到 DRL 框架, 以实现动态迭代式调优。与传统方法固定处理不同分布的方式不同, 本文通过 DGATP 融合端到端学习的方法以实现分支自适应融合, 利用生成的分布特征, 直接参与 VRP 的节点选择决策, 实现并行精准处理, 使得解码过程动态适应不同分布的数据特性。

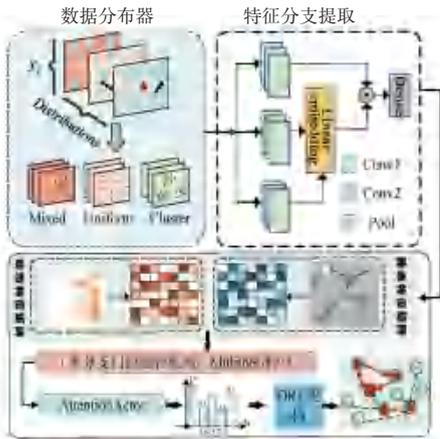


图 2 DGATP 方法

Fig. 2 DGATP method

3.1 动态分布识别与自适应模块

该模块主要由数据特征识别提取和自适应融合组成。首先, 通过动态识别功能实时感知输入 VRP 实例的空间分布特性。其次, 基于识别的分布类型, 激活最匹配的特征提取分支。最后, 融合分布先验知识与当前状态信息, 生成最优的路径决策策略。

3.1.1 数据识别特征提取

如图 2 所示, 在 DGATP 中, 特征分支提前以客户点坐标矩阵 X 作为输入, 采用轻量级卷积神经网络(CNN)以捕捉输入的全局空间统计特征(如, 点密度、分散度、聚类程度、网格化程度)。主要包含两个卷积层和一个池化层, 计算如下:

$$H^{(1)} = \sigma_{\text{ReLU}}(XW^{(1)} + b^{(1)}) \quad (8)$$

$$H^{(2)} = \sigma_{\text{ReLU}}(XW^{(2)} + b^{(2)}) \quad (9)$$

其中, $H^{(1)}, H^{(2)}$ 分别是初级特征和高级特征; $W^{(1)}$ 和 $W^{(2)}$ 是卷积核权重; $b^{(1)}$ 和 $b^{(2)}$ 是偏置。

对提取的特征 $H^{(2)}$ 进行全局平均池化, 以得到固定长度的分布特征向量 f 。该向量编码输入实例的整体分布特性:

$$f = \frac{1}{L} \sum_{i=1}^L H_i^{(2)} \quad (10)$$

其中, L 是 $H^{(2)}$ 的总元素数量。

最后, 该分布特征向量 f , 通过全连接层(FC)和 Softmax 激活函数映射为分类别的概率分布 P :

$$P = \sigma_{\text{Softmax}}(W_c f + b_c) \quad (11)$$

其中, W_c, b_c 为可学习参数。

注: 在本文中, 除非另有说明, 权重矩阵与向量的并置(如 $W_c t$) 均表示矩阵与向量的乘法。

3.1.2 特征自适应融合

结合所提取特征, 对应设计开关函数(Switch Function), 根据识别的分布概率类型进行选择并激活对应的特征提取分支, 以生成分布感知节点嵌入 E 。分支库包含独立设计的特征提取分支 $\{B_U, B_M, \dots, B_C\}$, 每个分支针对一种特定的空间分布模式予以优化。开关函数的选择过程为:

$$E = B_d(X) \quad (12)$$

其中, E 是分布感知嵌入, $B_d(\cdot)$ 是分支库中结合第 d 类空间分布模式优化的特征提取分支($B_d(\cdot)$ 为线性嵌入层)。

其次, 分布感知嵌入 E 与提供的分布特征向量 f 进行深度融合以动态调整策略网络的决策偏好, 并用于输出节点选择概率(动作策略) π 。具体的是, 设计一种门控机制以控制分布特征 f 对节点嵌入 E

的影响强度 g , 计算如下:

$$g = \sigma(\mathbf{W}_g[\mathbf{f}; \text{pool}(E) + \mathbf{b}_g]) \quad (13)$$

其中, σ 是 Sigmoid 激活函数; “;” 表示两个向量的拼接; pool 是池化函数以压缩信息。

接着, 通过一个门控融合单元生成增强后的节点嵌入:

$$\tilde{E} = g \odot (\mathbf{W}_f \mathbf{f} \oplus E) + (1 - g) \odot E \quad (14)$$

其中, $\mathbf{W}_g, \mathbf{W}_f$ 是可学习权重矩阵; \mathbf{b}_g 是偏置向量; \odot 表示逐元素乘法; \oplus 表示广播加法。门控 g 自适应地决定在很多程度上通过分布知识调整原始节点嵌入, 以生成增强嵌入 \tilde{E} 。

3.2 感知注意力 Actor 网络

结合增强嵌入 \tilde{E} 和分布特征 \mathbf{f} , 并将其注入到基于注意力机制的 Actor 网络以形成感知注意力网络。同时, 计算来自不同分布图上的节点概率分布, 并依次确定节点的选择。为此, 解决传统固定式无适数据特性的局限, 使得解码过程动态适应数据特性, 并增强区分来自不同信息的效益。具体的是, 首先, 本文以 LSTM 网络作为 DRL 架构的解码器, 并引用 Xia 等^[22] 的自定义掩码方案, 基于生成的增强嵌入 \tilde{E} 和 LSTM 隐藏状态 h_t , 以构建动态决策状态 $\Psi(\tilde{E}, h_t)$ 。其次, 将 $\Psi(\tilde{E}, h_t)$ 和客户需求特征 d_i 作为感知注意力网络的输入, 并通过多层卷积变换, 将所提取的特征划分至多头并行处理并合并输出结果。最后, 将其与训练参数通过运算操作以生成节点概率分布的基值:

$$a = a(\tilde{E}, h_t) = \text{Softmax}(v_a \tanh(W_a[\tilde{E}; h_t])) \quad (15)$$

$$\mathbf{c} = \sum_i a \tilde{E} \quad (16)$$

$$P(y_{t+1} | Y_t, X_t) = \text{Softmax}(v_c \tanh(W_c[\tilde{E}; \mathbf{c}])) \quad (17)$$

其中, a 是下一个解码步骤 t 中与当前输入节点信息的相关程度; “;” 表示两个张量的拼接; \mathbf{c} 为上下文向量; $P(y_{t+1} | Y_t, X_t)$ 表示下一个节点的概率; W_a, W_c, v_a 和 v_c 是可训练参数。

3.3 深度强化学习

本文提出的 DGATP 模型采用强化学习架构进行循环优化。如图 3 所示, 首先, 环境数据生成器生成具有多种分布的路网图数据, 待完成初始化后, 将环境状态的初始值 (记为 s_t , 其中 $t = 0$) 输入 DGATP (Agent) 模型。同时, DGATP 模型生成客户节点的概率基值分布, 并通过解码策略选择相应节点, 以得到动作 a_t 。该动作结果被反馈至环境, 并触发环境状态更新。该过程属于参数化, 其中感知注意力网络的参数记为 θ , 强化学习调优策略的参数记为

φ 。具体形式表达如下:

$$d\theta \leftarrow \frac{1}{N} \sum_{n=1}^N (R^n - V(X_0^n; \varphi)) \nabla_{\theta} \log P(Y^n | X_0^n) \quad (18)$$

$$d\varphi \leftarrow \frac{1}{N} \sum_{n=1}^N \nabla_{\varphi} (R^n - V(X_0^n \varphi))^2 \quad (19)$$

其中, 上标表示第 n 个实例的变量; N 表示问题样本; $V(X_0^n; \varphi)$ 是强化学习调优策略网络计算的奖励近似值; R 是实际奖励值。

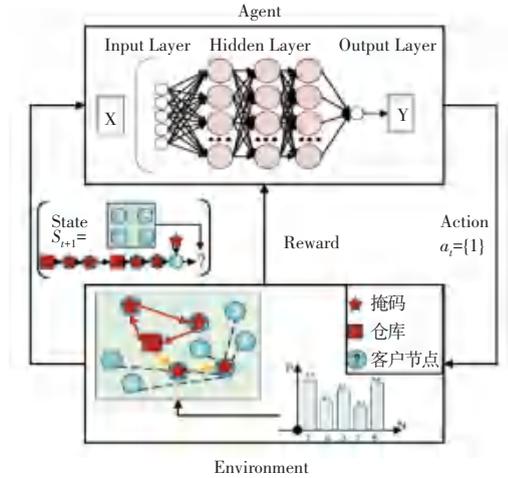


图3 DRL 框架

Fig. 3 DRL Framework

4 实验与结果分析

4.1 实验数据

实验数据来源于 4 个不同大小的数据集, 并参照 DRL^[11]、DACT^[17] 和 MHAAC^[22] 方法生成 4 种节点规模的随机实例。具体生成过程是, 在 $[0, 1] \times [0, 1]$ 区间内, 基于特定分布类型随机采样客户坐标 $s_i = (x_i, y_i) \in [0, 1] \times [0, 1]$, 并设置一个坐标为仓库中心 $s_0 = (x_0, y_0) \in [0, 1] \times [0, 1]$ 。每个客户的需求值 $d_i \in R^+$ 在区间 $[1, 10]$ 范围内随机生成。节点数量 N 分别设定为 10、20、50 和 100。客户节点和仓库节点通过在 $[0, 1] \times [0, 1]$ 区域内利用分布函数而采样生成, 并随机选定一个节点作为仓库的库节点。

4.2 评估指标与基准

参照文献[9]、[15]和[20], 本文采用平均巡回长度 (Mean)、已知最优值的差距 (Gap)、优异程度 (ΔGap) 3 个评估指标衡量模型性能。上述指标定义如下:

$$\text{Mean} = \frac{L}{N} \quad (20)$$

$$\text{Gap} = \frac{Z - Z^*}{Z^*} \times 100\% \quad (21)$$

$$\Delta\text{Gap} = \text{Gap}_{z_1} - \text{Gap}_{z_2} \quad (22)$$

其中, L 表示 N 个实例的巡回长度总和; Z 表示不同方法的解成本; Z^* 表示最优解成本; ΔGap 表示方法 z_1 优于 z_2 的百分点。同时, 本文选择 DRL^[11] 和 MHAAC^[22] 作为基准。

4.3 结果分析

为了验证方法的有效性, 首先在同分布数据上开展训练, 并与基准模型进行泛化性能对比。其次, 在未知数据集上进行测试, 进一步分析泛化能力。最后, 与同类方法进行比较, 验证本文方法的优越性。

在表 1 和表 2 中, DP-DRL 和 DP-MC 分别为 DGATP 融入两种不同的深度强化学习架构, 前者为 Nazari 等的 DRL^[11], 后者为 Xia 等的 MHAAC^[22]。表中 M_U 、 M_C 、 M_M 分别表示测试数据分布为 Uniform、Cluster、Mixed 上的平均巡回长度。从表 M_U 列可以发现, 本文方法的结果较传统方法略差, 那是由于传统方法依赖于 Uniform 分布。同时, 基于模型在未知数据集 M_C 和 M_M 列的结果可以得出, 本文方法均优于传统方法, 以及 3 种分布的均值效果均优于传统方法。值得注意的是, 随着节点数量的增加, 效果更为显著(见表 2)。

表 1 10 个节点的实验结果

Table 1 Experimental results on 10 nodes

Method		CVRP10, Cap20			
		M_U	M_C	M_M	Avg
Greedy	DRL	4.78	1.02	2.87	2.89
	DP-DRL	4.80	0.87	2.84	2.83
	MHAAC	4.75	1.05	2.89	2.89
	DP-MC	4.79	0.87	2.87	2.84
Beam Search	DRL	4.64	0.93	2.75	2.77
	DP-DRL	4.64	0.82	2.72	2.72
	MHAAC	4.60	0.96	2.77	2.77
	DP-MC	4.60	0.82	2.75	2.72

表 2 100 个节点的实验结果

Table 2 Experimental results on 100 nodes

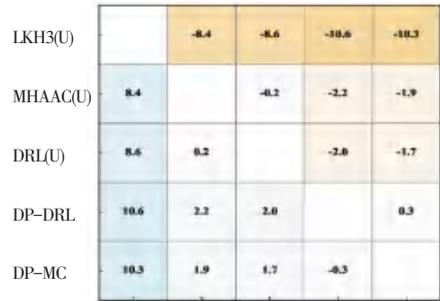
Method		CVRP100, Cap50			
		M_U	M_C	M_M	Avg
Greedy	DRL	16.97	6.34	11.65	11.65
	DP-DRL	17.85	3.77	10.76	10.69
	MHAAC	16.92	6.92	11.71	11.85
	DP-MC	17.64	3.72	10.63	10.66
Beam Search	DRL	16.73	6.17	11.45	11.45
	DP-DRL	17.57	3.68	10.58	10.61
	MHAAC	16.69	6.49	11.50	11.56
	DP-MC	17.47	3.61	10.49	10.53

根据表 3 所示, LKH3 是 VRP 的最优求解器^[20], 为此选其作为 Z^* 。如图 4(a)~(c)所示, 本文方法在两个未知数据集上优于其它方法, 改进性能最优可达 48.5%。同时, 基于传统分布数据集 Uniform, 略差于传统求解器。

表 3 50 个节点的实验结果

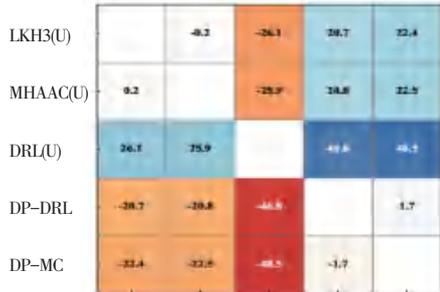
Table 3 Experimental results on 50 nodes

Methods	Uniform		Cluster		Mixed	
	Mean	Gap/%	Mean	Gap/%	Mean	Gap/%
LKH3_U	10.38	0	2.95	0	7.06	0
MHAAC	11.25	8.40	3.43	16	7.30	3.40
DRL(U)	11.27	8.60	3.72	26.10	7.43	5.24
DP-DRL	11.48	10.59	2.34	-20.67	6.86	-2.83
DP-MC	11.45	10.30	2.29	-22.37	6.81	-3.54



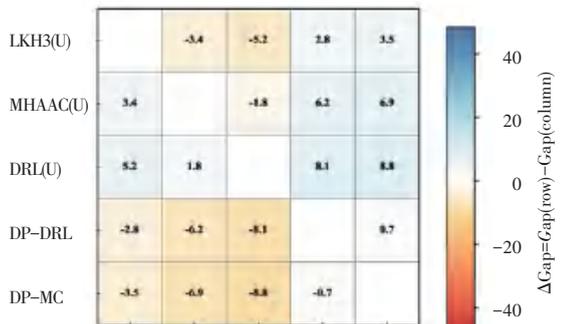
LKH3(U) MHAAC(U) DRL(U) DP-DRL DP-MC

(a) Uniform problems



LKH3(U) MHAAC(U) DRL(U) DP-DRL DP-MC

(b) Cluster problems



LKH3(U) MHAAC(U) DRL(U) DP-DRL DP-MC

(c) Mixed problems

图 4 横向与纵向方法在 ΔGap 指标上的对比 (%)

Fig. 4 Comparison of horizontal and vertical methods on the ΔGap metric (%)

5 结束语

本文提出一种基于领域自适应深度强化学习的跨分布车辆路径求解方法。该方法的核心是构建多分布引导的自适应策略网络(DGATP),并通用于两种不同的端到端深度强化学习框架(DRL)。首先,设计基于卷积神经网络的分布识别模块以感知源域与目标域的差异,并进行特征提取和分布识别。其次,构建门控融合网络,通过多分支并行结构对不同分布特征进行自适应加权融合。接着,设计感知注意力层以整合特征信息,基于深度学习模型生成节点概率分布,依次实现完整路径解构造,并结合强化学习策略实现解的迭代调优。最后,在两个具有代表性的深度模型上,与传统神经网络进行比较,实验结果验证了本文方法的有效性。

参考文献

- [1] YUAN Y, CATTARUZZA D, OGIER M, et al. Last mile delivery problem; The one-vehicle case [C]//Proceedings of the 2018 ODYSSEUS Conference on Logistics and Transportation. Athens, Greece: ODYSSEUS, 2018;45-52.
- [2] DUAN L, ZHAN Y, HU H, et al. Efficiently solving the practical vehicle routing problem: a novel joint learning approach [C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020; 3054-3063.
- [3] KONSTANTAKOPOULOS G, GAYIALIS S, KE CJAGIAS E. Vehicle routing problem and related algorithms for logistics distribution: A literature review and classification[J]. Operational Research, 2022, 22(3): 2033-2062.
- [4] QIN Z, TANG X, JIAO Y, et al. Ride-hailing order dispatching at DiDi via reinforcement learning [J]. INFORMS Journal on Applied Analytics, 2020, 50(5): 272-286.
- [5] LI J, MA Y, CAO Z, et al. Learning feature embedding refiner for solving vehicle routing problems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(11): 15279-15291.
- [6] BOGYRBAYEVA A, MERALIYEV M, MUSTANKHOV T, et al. Machine learning to solve vehicle routing problems: A survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(6): 4754-4772.
- [7] 顾艳鑫, 赖红波. 模糊时间窗下考虑动态需求的生鲜路径优化[J]. 智能计算机与应用, 2023, 13(2): 53-59.
- [8] 杨威, 周义棚. 基于 LSTM 网络预测智能车辆变道的路径规划研究[J]. 智能计算机与应用, 2023, 13(7): 144-149.
- [9] VINYALS O, FORTUNATO M, JAITLY N. Pointer networks [C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2015; 2692-2700.
- [10] BELLO I, PHAM H, LE V, et al. Neural combinatorial optimization with reinforcement learning [EB/OL]. arXiv: 1611.09940, 2016. DOI:10.48550/arxiv.1611.09940
- [11] NAZARI M, OROOJLOOY A, SNYDER L, et al. Deep reinforcement learning for solving the vehicle routing problem [C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. NeurIPS, 2018; 9861-9871.
- [12] JOSHI C, LAURENT T, BRESSON X. An efficient graph convolutional network technique for the travelling salesman problem [J]. arXiv: 1906.01227, 2019. DOI: 10.48550/arXiv.1906.01227
- [13] LI J, XIN L, CAO Z, et al. Heterogeneous attentions for solving pickup and delivery problem via deep reinforcement learning [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(3): 2306-2315.
- [14] HELSGAUN K. An extension of the Lin-Kernighan-Helsgaun TSP solver for constrained traveling salesman and vehicle routing problems; RUC-DCS-TR-17-01 [R]. Roskilde, Denmark: Department of Computer Science, Roskilde University, 2017.
- [15] CHEN X, TIAN Y. Learning to perform local rewriting for combinatorial optimization [C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. NeurIPS, 2019; 6281-6292.
- [16] HOTTUNG A, TIERNEY K. Neural large neighborhood search for the capacitated vehicle routing problem [C]//Proceedings of the 24th European Conference on Artificial Intelligence. Cham: Springer, 2020; 443-450.
- [17] MA Y, LI J, CAO Z, et al. Learning to iteratively solve routing problems with dual-aspect collaborative transformer [C]//Proceedings of the 35th International Conference on Neural Information Processing Systems. NeurIPS, 2021; 11096-11107.
- [18] XIN L, SONG W, CAO Z, et al. Generative adversarial training for neural combinatorial optimization models [EB/OL]. (2022-09-22) [2025-10-28]. <https://openreview.net/forum?id=nJuzV-izmPJ>
- [19] WANG C, YANG Y, SLUMBERS O, et al. A game-theoretic approach for improving generalization ability of TSP solvers [C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI, 2021; 12432-12440.
- [20] ZHANG Z, ZHANG Z, WANG X, et al. Learning to solve travelling salesman problem with hardness-adaptive curriculum [C]//Proceedings of the 36th AAAI Conference on Artificial Intelligence. AAAI, 2022; 9136-9144.
- [21] BI J, MA Y, WANG J, et al. Learning generalizable models for vehicle routing problems via knowledge distillation [C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. NeurIPS, 2022; 31226-31238.
- [22] XIA D, JIN Y, HUANG M, et al. A multi-head adaptive actor-critic algorithm for solving vehicle routing problems [J]. Applied Intelligence, 2025, 55(13): 1-21.