Modeling and Simulation 建模与仿真, 2025, 14(2), 142-156
Published Online February 2025 in Hans. https://www.hanspub.org/journal/mos
https://doi.org/10.12677/mos.2025.142139

Hans汉斯

基于Linformer和多关系解码器的异构车队路径规划模型研究

李顺龙

上海理工大学管理学院，上海

收稿日期：2025年1月24日；录用日期：2025年2月17日；发布日期：2025年2月25日

摘　要

异构有容量限制的车辆路径规划问题(Heterogeneous Capacitated Vehicle Routing Problem, HCVRP)因其广泛的实际应用和复杂的约束条件，成为现代物流优化中的重要研究课题。然而，现有方法在处理异构车队多目标优化任务时，仍存在计算复杂度高和泛化能力不足的问题。针对上述挑战，本文提出了一种基于深度强化学习(Deep Reinforcement Learning, DRL)的新型HCVRP求解框架。首先，引入低秩注意力机制的Linformer模型，显著降低了传统Transformer在大规模问题中的计算复杂度。其次，设计多关系节点选择解码器，通过结合动态节点特征实时更新和车辆状态建模，有效提升了路径优化的解质量。在随机生成的数据集上，通过与多种经典启发式算法和现有深度强化学习方法的对比实验，验证了所提方法的性能。结果表明，本文方法在解质量和计算效率方面均具有显著优势，尤其在复杂约束和大规模实例中表现出更高的适用性。本文为解决异构车队路径规划问题提供了新的理论方法和实践工具，具备广泛的应用前景。

关键词

异构车辆路径规划，深度强化学习，低秩注意力机制，多关系解码器，马尔可夫决策过程

Research on Heterogeneous Capacitated Vehicle Routing Models Based on Linformer and Multi-Relational Decoder

Shunlong Li

School of Management, University of Shanghai for Science and Technology, Shanghai

Received: Jan. 24th, 2025; accepted: Feb. 17th, 2025; published: Feb. 25th, 2025

文章引用: 李顺龙. 基于 Linformer 和多关系解码器的异构车队路径规划模型研究[J]. 建模与仿真, 2025, 14(2): 142-156. DOI: 10.12677/mos.2025.142139

Abstract

The Heterogeneous Capacitated Vehicle Routing Problem (HCVRP) is a critical research topic in modern logistics optimization due to its extensive real-world applications and complex constraints. However, existing methods often face challenges such as high computational complexity and limited generalization ability when handling heterogeneous fleets and multi-objective optimization tasks. To address these issues, this paper proposes a novel HCVRP solution framework based on Deep Reinforcement Learning (DRL). First, a Linformer model with a low-rank attention mechanism is introduced, significantly reducing the computational complexity of traditional Transformers in large-scale problems. Second, a multi-relational node selection decoder is designed to enhance solution quality by dynamically updating node features and modeling vehicle states in real time. Extensive experiments on randomly generated datasets demonstrate the performance of the proposed approach compared with various classical heuristic algorithms and existing DRL methods. The results show that the proposed framework achieves significant advantages in both solution quality and computational efficiency, especially in scenarios with complex constraints and large-scale instances. This study provides a new theoretical methodology and practical tool for solving heterogeneous fleet routing problems, offering broad application prospects.

Keywords

Heterogeneous Vehicle Routing, Deep Reinforcement Learning, Low-Rank Attention Mechanism, Multi-Relational Decoder, Markov Decision Process

1. 引言

现代物流和城市配送需求的快速增长，使异构车队配送在实际场景中展现出高效的适用性，逐渐成为车辆路径规划问题(Vehicle Routing Problem, VRP)的重要研究方向。其中，异构有容量限制的车辆路径规划问题旨在优化不同类型车辆(如容量和速度差异)的配送路径，以实现运输成本、时间和资源利用率的综合优化。然而，由于车辆异质性和复杂约束的存在，HCVRP 属于典型的 NP 难题，其高效求解具有重要的理论意义和实际价值。

传统的 HCVRP 求解方法可分为两大类：精确算法和启发式算法。Jabali 等[1]通过混合整数非线性规划推导 HVRP 的上下界。但其时间和空间复杂度随问题规模增大而指数增长，限制了实际应用。启发式算法(如模拟退火[2]、粒子群优化[3]和遗传算法[4])通过启发式规则探索解空间，能够在有限时间内找到接近最优解。然而，这些方法往往依赖于特定问题的启发式设计，难以适应动态环境，且在多目标优化任务中的泛化能力有限。

近年来，深度学习和强化学习的快速发展为复杂路径规划问题提供了全新的解决方案。深度强化学习结合了深度学习的特征提取能力和强化学习的决策优化能力，在高维决策空间和动态环境中表现出色。谷歌团队[5]利用 Transformer 模型解决旅行商问题和 CVRP，展示了基于注意力机制的节点特征提取能力。Kool 等[6]进一步结合监督学习和强化学习，显著提升了解的准确性。然而，目前大部分方法主要聚焦于同质车队场景，缺乏对异构车队复杂特性和多目标约束的全面建模和高效求解。

针对上述研究现状与挑战，本文提出了一种基于深度强化学习的新型 HCVRP 求解框架，以高效解决异构车队的路径优化问题。本文的主要创新点如下：

(1) 引入 Linformer 模型[7]：通过低秩近似的注意力机制，大幅降低传统 Transformer 模型的计算复杂度，从而提升在大规模问题中的效率和性能。

(2) 设计多关系节点选择解码器：结合动态节点向量的实时更新机制，根据车辆状态和路径历史优化节点选择，显著提高解的质量。

为验证所提方法的有效性，本文在多组随机生成的数据集上进行了实验，并与现有启发式算法和深度强化学习方法进行了对比分析。实验结果表明，本文方法在解质量和计算效率方面均具有显著优势，为复杂物流场景中的 HCVRP 问题提供了新视角和有效工具。

2. 问题描述与模型构建

在本节中，我们针对异构有容量限制的车辆路径规划问题(HCVRP)提出了问题模型，并进一步将其重构为强化学习模型。表 1 总结了 HCVRP 模型中涉及的参数与变量定义，以便清晰描述模型结构。

Table 1. Parameter and variable descriptions
表 1. 参数和变量描述

符号	描述
n	客户节点数量
m	异构车辆数量
X	节点集，$X=\left\{x^i\right\}_{i=0}^{n}$，包括客户节点和配送中心，其中 x^0 表示配送中心
x^i	节点 i 的坐标，$x^i\in R^3$，$x^i=\left\{\left(s^i,d^i\right)\right\}$，是一个三维向量
X'	客户节点集，$X\setminus\left\{x^0\right\}$，去除了配送中心
s^i	每个节点的二维坐标标识
d^i	节点 i 的需求量，配送中心的需求为0
V	异构车辆集，$V=\left\{v^i\right\}_{i=1}^{m}$
Q^i	每辆车的容量
y_{ij}^v	如果车辆 v 直接从客户 x^i 到 x^j 那么这个值为1；反之为0
l_{ij}^v	车辆 v 在从客户 x^i 行驶到客户 x^j 之前的剩余容量
f,f_v	所有车辆统一速度/车辆 v 的速度
$D\left(x^i,x^j\right)$	节点 x^i 和节点 x^j 之间的欧几里得距离

2.1. 问题描述

HCVRP 问题可以描述为：本文所提到的 HCVRP 可以描述为：在一个配送区域内，包含一个配送中心 x^0 和若干客户节点 $X'=\left\{x^i\middle|i=1,2,\cdots,n\right\}$。配送中心内拥有一支由 m 辆异构车辆 $V=\left\{v_1,v_2,\cdots,v_m\right\}$ 组成的车队，每辆车具有不同的容量 Q_v 和 f_v。每个客户节点 x^i 具有特定的需求量 d^i，所有车辆从配送中心出发，依次访问客户节点完成配送任务，并最终返回配送中心，其目标是最小化车队中的车辆的最大行驶时间(MM-HCVRP)或最小化车辆所有车辆的总行驶时间(MS-HCVRP)。

2.2. 模型构建

为简化分析，我们假设所有车辆的行驶速度相同，这种设定易于扩展到具有不同速度的车辆场景。HCVRP 的数学模型定义如下：

MM-HCVRP 目标函数：

$$\min \max_{v\in V}\left(\sum_{i\in X}\sum_{j\in X}\frac{D\left(x^i,x^j\right)}{f}y_{ij}^v\right) \tag{1}$$

MS-HCVRP 目标函数：

$$\min \sum_{v\in V}\sum_{i\in X}\sum_{j\in X}\frac{D\left(x^i,x^j\right)}{f_v}y_{ij}^v \tag{2}$$

受以下约束条件的限制：

$$\sum_{v\in V}\sum_{j\in X}y_{ij}^v=1,\quad i\in X' \tag{3}$$

$$\sum_{i\in X}y_{ij}^v-\sum_{k\in X}y_{jk}^v=0,\quad v\in V,j\in X' \tag{4}$$

$$\sum_{v\in V}\sum_{i\in X}l_{ij}^v-\sum_{v\in V}\sum_{k\in X}l_{jk}^v=d^j,\quad j\in X' \tag{5}$$

$$d^j y_{ij}^v\le l_{ij}^v\le\left(Q^v-d^i\right)\cdot y_{ij}^v,\quad v\in V,i\in X,j\in X \tag{6}$$

$$y_{ij}^v=\{0,1\},\quad v\in V,i\in X,j\in X \tag{7}$$

$$l_{ij}^v\ge 0,d^i\ge 0,\quad v\in V,i\in X,j\in X \tag{8}$$

其中约束条件(2)和(3)确保每个客户仅被访问一次，并且每条路线由同一辆车完成。约束条件(4)通过确保在服务客户前后，货物负载的差异等于该客户的需求，来保证需求的满足。约束条件(5)和(6)确保车辆不超过其最大容量，并且决策变量是二进制的。

2.3. 马尔可夫决策模型

强化学习特别适用于处理复杂的序列决策问题，在本文中，HCVRP 可以看成是一个序贯决策的马尔可夫决策过程(markov decision process, MDP)，MDP 由元组 $M=\{S,A,\rho,R\}$ 表示，其中：

(1) 状态空间(S)：时间步骤 t 下的状态 S_t 由车辆状态 V_t 和节点状态 X_t 组成：

车辆状态：$V_t=\left\{v_t^1,v_t^2,\cdots,v_t^m\right\}$，其中 $v_t^i=\left(o_t^i,T_t^i,G_t^i\right)$。

o_t^i 是车辆 v^i 在步骤 t 时的剩余容量。在此模型中，每个步骤的时间间隔定义为所选车辆完成当前操作所需的时间，即访问下一个客户或配送中心。此定义确保每个决策步骤对应于车辆的实际操作，保持了在现实场景中车辆异步操作的实用性。

T_t^i 是车辆 v^i 在步骤 t 时的累计行驶时间。

$G_t^i=\left\{g_0^i,g_1^i,\cdots,g_t^i\right\}$ 是车辆 v^i 在步骤 t 时的部分路径，其中 g_j^i 表示车辆 v^i 在步骤 j 访问的节点。所有车辆的部分路径(路径中节点的数量)的维度保持不变，这意味着如果车辆 v^i 在步骤 t 被选中服务节点 x^j，则其他车辆将继续访问各自的最后服务节点。

在初始状态(即 $t=0$)时，初始车辆状态设置为 $V_0=\left\{\left(Q^1,0,\{0\}\right),\left(Q^2,0,\{0\}\right),\cdots,\left(Q^m,0,\{0\}\right)\right\}$，其中 Q^i 是车辆 v^i 的最大容量。

节点状态： $X_t=\left\{x_t^0,x_t^1,\cdots,x_t^n\right\}$，其中 $x_t^i=\left(s^i,d_t^i\right)$。

s^i 表示节点 i 的位置，是一个二维向量。

d_t^i 表示节点 i 的需求，在节点被服务后，需求 d_t^i 设为零。

(2) 动作空间(A)：动作 $a_t\in A$ 涉及选择一辆车和一个客户或配送中心进行访问。具体来说，$a_t=\left(v_i^t,x_j^t\right)$，表示在时间步 t 时，车辆 v^i 将服务节点 x_j。每个步骤只选择一辆车。

(3) 状态转移规则(ρ)：状态转移规则 ρ 将根据执行的动作 $a_t=\left(v_t^i,x_t^j\right)$ 将状态 s_t 转移到下一个状态 s_{t+1}。车辆状态 V_{t+1} 的元素更新如下：

$$o_{t+1}^k=\begin{cases}o_t^k-d_t^i, & \text{if } k=i\\ o_t^k, & \text{otherwise}\end{cases} \tag{9}$$

$$T_{t+1}^k=\begin{cases}T_t^k+\dfrac{D\left(g_t^k,x\right)}{f}, & \text{if } k=i\\ T_t^k, & \text{otherwise}\end{cases} \tag{10}$$

$$G_{t+1}^k=\begin{cases}\left[G_t^k,x^j\right], & \text{if } k=i\\ \left[G_t^k,g_t^k\right], & \text{otherwise}\end{cases} \tag{11}$$

其中，g_t^k 是 G_t^k 的最后一个元素，即车辆 v^k 在步骤 t 时访问的最后一个客户，[·，·，·]表示连接操作。节点状态 X_{t+1} 的元素更新如下：

$$d_{t+1}^l=\begin{cases}0, & \text{if } l=j\\ d_t^l, & \text{otherwise}\end{cases} \tag{12}$$

其中，任何未被访问的需求将保持为 0。

(4) 奖励函数(R)：对于 MM-HCVRP，目标是最小化车队中车辆的最大行驶时间，奖励定义为该值的负数，奖励通过在每个步骤中逐步累积每辆车的行驶时间来计算。奖励表示为 $R=-\max\limits_{v\in V}\left\{\sum\limits_{t=0}^{T}\boldsymbol{r}_t\right\}$，其中 r_t 是所有车辆在步骤 t 时的增量行驶时间。对于 MS-HCVRP，奖励定义为所有车辆的总行驶时间的负值，即 $R=-\sum\limits_{i=1}^{m}\sum\limits_{t=1}^{T}\boldsymbol{r}_t$。具体来说，假设在步骤 t 和 $t+1$ 时选择了节点 x^j 和 x^k，并且这两个节点将由车辆 v^i 服务，奖励 r_{t+1} 被表示为一个 m-维向量，如下所示：

$$\begin{aligned}\boldsymbol{r}_{t+1}&=r\left(s_{t+1},a_{t+1}\right)=r\left(\left(V_{t+1},X_{t+1}\right),\left(v_{t+1}^i,x_{t+1}^k\right)\right)\\&=\left\{0,\cdots,0,D\left(x^j,x^k\right)\Big/f,0,\cdots,0\right\}\end{aligned} \tag{13}$$

其中，$D\left(x^j,x^k\right)\Big/f$ 是车辆 v^i 从节点 x^j 到节点 x^k 所花费的时间，奖励向量 $r\left(s_{t+1},a_{t+1}\right)$ 中的所有其他元素均为 0。

3. 求解算法设计

在本章中，我们提出了一种基于深度强化学习的模型来解决 HCVRP。解决过程如图 1 所示。首先，我们展示了基于编码器－解码器结构的 DRL 模型，用于表示策略，其中在每个决策步骤中都会进行车辆和节点的选择，如图 2 所示。最后，我们描述了用于训练该模型的策略网络。

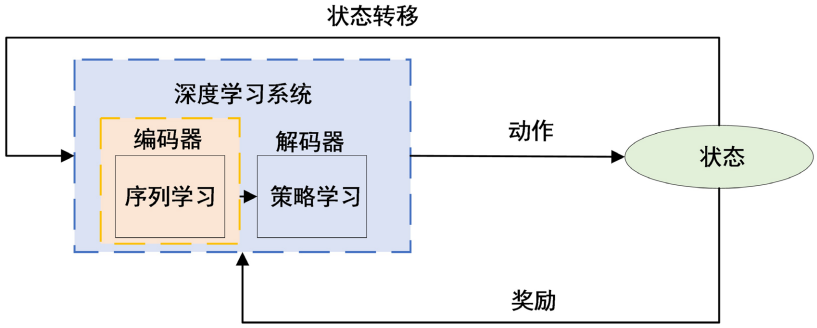

Figure 1. Solution process diagram
图 1. 解法过程图

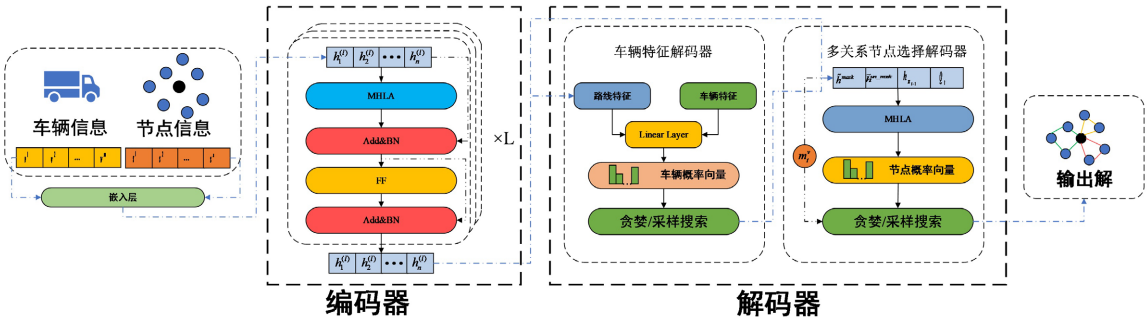

Figure 2. Model architecture diagram
图 2. 模型架构图

在所采用的模型中，每个节点和配送中心的信息，以及车辆信息，输入到编码器中。编码器的输出与上下文信息一起被传递到解码器，以生成一系列信息。我们的目标是选择异构车辆访问每个节点(即选择一个最优序列)，确保所有节点都被服务，同时最小化目标函数值。本文基于马尔可夫过程构建了解决方案，该过程包含若干元素，如状态 s 、动作 a 和策略 π 。在初始时刻，状态被初始化为 s_0 。策略 π_θ 用于选择下一时间步的动作 a_1 ，并将状态更新为 s_1 ，直到达到终止状态 s_τ 在时间 $t=\tau$ 时为止。通过训练神经网络模型，优化训练参数 θ ，使得模型学习到一个能够最小化随机策略的策略 π_θ 。本文定义了一个随机选择策略 P ，其定义如下：

$$p\left(s_\tau \middle| s_o\right)=\sum \pi_\theta\left(a_t \middle| s_t\right) p\left(s_{t+1} \middle| s_t, a_t\right) \tag{14}$$

3.1. 嵌入层

嵌入层提取当前解的两个部分特征，包括每个节点的个体特征，并通过线性变换得到 h_i 。节点的个体特征是直观的，包括与问题实例相关的静态特征和与当前解高度相关的动态特征。静态特征包括：节点 i 的 X 轴坐标 x_i 、 Y 轴坐标 y_i 和需求 d_i 。动态特征包括：在节点 i 处，车辆 m 的负载 c_i 和该车辆的容量 Q^m 。最终输入到 HCVRP 问题的是 $s_i=\left[x_i, y_i, d_i, \frac{c_i}{Q^1}, \frac{c_i}{Q^2}, \cdots, \frac{c_i}{Q^m}\right]$ ，这些特征通过线性层转换为隐藏向量 h_i 。

3.2. 编码器

如图所示，编码器的结构由 L 个线性注意力层组成。每个线性注意力层由多头线性注意力机制和前馈网络组成。编码器的输入是每个节点的隐藏特征 h_i ，输出是每个节点的编码特征以及图的特征。

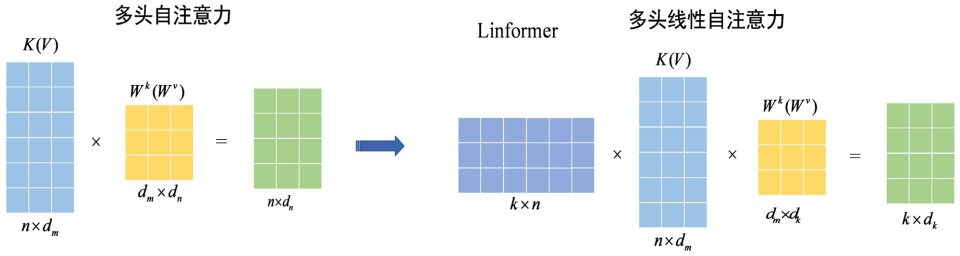

Figure 3. MHA and MHLA legend
图 3. MHA 和 MHLA 图例

首先，节点嵌入向量 h_i 被传递到具有 l 层的编码器中，其中编码器的每一层主要由多头线性注意力(MHLA) (如图 3)和前馈层(FF)组成。线性自注意力机制的核心思想是在计算 Key 和 Value 时引入两个线性投影矩阵 $E_i, F_i \in \mathbb{R}^{n\times k}$，将原始的 KW_i^K 和 VW_i^V 从 $n\times d$ 维度变换为 $k\times d$ 维度。随后，使用缩放点积注意力计算一个 $n\times k$ 维的注意力矩阵 $\overline{P}$，并利用 $\overline{P}\cdot\left(F_iVW_i^V\right)$ 计算每个注意力头的输出。计算过程如下：

$$\begin{aligned}\text{head}_i &= \text{Attention}\left(QW_i^{Q,c,l}, E_iKW_i^{K,c,l}, F_iVW_i^{V,c,l}\right)\\ &= \underbrace{\text{softmax}\left(\frac{QW_i^Q\left(E_iKW_i^k\right)^{\mathrm{T}}}{\sqrt{d_k}}\right)}_{\overline{P}:n\times k}\cdot\underbrace{F_iVW_i^V}_{k\times d}\end{aligned} \tag{15}$$

$$\text{MHLA}\left(h_i^l\right) = \text{Concat}\left(\text{head}_i^{1,l}, \text{head}_i^{2,l}, \cdots, \text{head}_i^{c,l}\right)W_i^{O,L} \tag{16}$$

其中，$W_i^{Q,c,l}, W_i^{Q,c,l} \in \mathbb{R}^{Y\times d\times d_k}, W_i^{Q,c,l} \in \mathbb{R}^{Y\times d\times dv}, W_i^{Q,c,l} \in \mathbb{R}^{d\times d}$ 是 MHA 第 l 层中的可学习参数。最后，将每个注意力头获得的注意力特征连接起来，以实现更好的特征表示。在每个注意力层的计算过程中，仅需要 $O(n\times k)$ 的时间和空间复杂度。如果选择一个非常小的投影维度 k，例如 $k \ll n$，可以显著减少内存和空间消耗。MHA 和 MHLA 的区别如图 4 所示。随后，使用前馈神经网络、残差连接和批量归一化来处理第 l 层多头注意力的输出，如公式(17)所示：

$$M^l = BN\left(h_i^l + MHLA^l\left(h_i^l\right)\right) \tag{17}$$

$$h_i^{l+1} = BN\left(M_i^l + FF\left(M_i^l\right)\right) \tag{18}$$

3.3. 解码器

3.3.1. 车辆选择解码器

车辆选择解码器用于生成选择特定车辆的概率分布。该概率分布主要依赖于两种嵌入：车辆特征嵌入和路径特征嵌入。这些嵌入反映了车辆和路径的特性，从而帮助模型做出有效的车辆选择决策。

(1) 车辆特征嵌入：为了捕捉每辆车在当前步骤的状态，我们定义车辆特征上下文 $C_t^V = \mathbb{R}^{1\times 3m}$ 在步骤 t 的形式如下：

$$C_t^V = \left[\tilde{g}_{t-1}^1, T_{t-1}^1, \tilde{g}_{t-1}^2, T_{t-1}^{=2}, \cdots, \tilde{g}_{t-1}^m, T_{t-1}^{=m}\right] \tag{19}$$

其中，$\tilde{g}_{t-1}^1$ 表示车辆 v_i 在步骤 $t-1$ 时部分路径上的最后一个节点 g_{t-1}^i 的位置，T_{t-1}^1 表示车辆 v_i 到步骤 $t-1$ 为止的累计行驶时间。然后，车辆特征上下文通过可训练的参数 W_1 和 b_1 进行线性投影，并通过一个具有 ReLU 激活函数的 512 维前馈全连接层进一步处理，从而得到时间步 t 时的车辆特征嵌入 H_t^V，表示如下：

$$H_t^V = FF\left(W_1C_t^V + b_1\right) \tag{20}$$

(2) 路径特征嵌入：路径特征嵌入从所有车辆的现有部分路径中提取信息，帮助策略网络从先前步骤中访问的节点中学习。对于每辆车 v_i 在步骤 t 时，我们将其路径特征上下文 $\tilde{C}_t^i$ 定义为其部分路径 G_{t-1}^i 中节点的嵌入排列(即， h_N^k 是节点 x^k 的节点嵌入)。具体来说，每辆车的路径特征上下文 $\tilde{C}_t^i$ 定义如下：

$$\tilde{C}_t^i = \left[\tilde{h}_0^i, \tilde{h}_1^i, \cdots, \tilde{h}_{t-1}^i\right] \tag{21}$$

其中， $\tilde{C}_t^i \in \mathbb{R}^{t\times \dim}$ (第一维的大小为 t ，因为在步骤 t 时， G_{t-1}^i 应包含 t 个元素)， $\tilde{h}_j^i$ 表示车辆 v_i 在其部分路径 G_{t-1}^i 中第 j 个节点的对应节点嵌入 h_N 。然后，所有车辆的路径特征上下文通过最大池化和连接操作进行聚合，以获得整个车队的路径特征 $\tilde{C}_4^R$ 。该特征随后通过具有可训练参数 W_2 和 b_2 的线性投影以及一个512 维的前馈全连接层进行处理，以获得时间步 t 时的路径特征嵌入 H_t^R ，表示如下：

$$\tilde{C}_t^i = \max\left(\tilde{C}_t^i\right) \tag{22}$$

$$\tilde{C}_t^R = \left[\tilde{C}_t^1, \tilde{C}_t^2, \cdots, \tilde{C}_t^m\right] \tag{23}$$

$$H_t^R = FF\left(W_2\tilde{C}_t^R + b_2\right) \tag{24}$$

最后，车辆特征嵌入 H_t^V 和路径特征嵌入 H_t^R 被连接在一起，并通过具有参数 W_3 和 b_3 的线性投影进行处理，然后通过 softmax 函数计算出概率向量，如下所示：

$$H_t = W_3\left[H_t^V, H_t^R\right] + b_3 \tag{25}$$

$$p_t = \text{softmax}\left(H_t\right) \tag{26}$$

其中， $p_t \in \mathbb{R}^m$ 及其元素 p_t^i 表示在时间步 t 选择车辆 v_i 的概率。根据策略，可以通过贪婪地选择最高概率的车辆，或者根据概率向量 p_t 进行采样来选择车辆。选择的车辆 v_i 将作为节点选择解码器的输入。

3.3.2. 多关系节点选择解码器

我们设计了一个多关系节点选择解码器(如图 4)，以更好地捕捉节点之间的动态关系。给定来自编码器的节点嵌入和来自车辆选择解码器的选定车辆 v_i ，节点选择解码器输出一个针对所有未服务节点的概率分布 $\tilde{p}_t$ (先前步骤中已服务的节点会被屏蔽)，用于确定选定车辆应该访问的节点。解码过程持续进行，直到车辆达到其容量限制，最终目标是通过学习的策略最大化奖励。

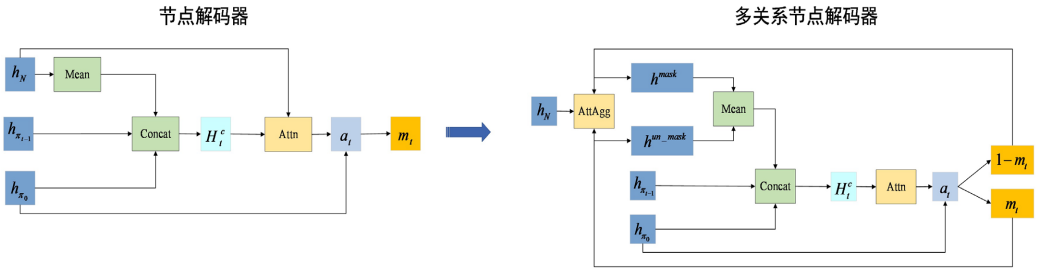

Figure 4. Node decoder and Multi-relational node decoder legend
图 4. 节点解码器和多关系节点解码器图例

在基于注意力的强化学习模型中，上下文嵌入由图嵌入 $\overline{h} = \sum_{i=1}^{N} h_N$ 、第一个选定节点 h_{π_0} 和最后一个选定节点 $h_{\pi_{t-1}}$ 组成。

$$h_t^c = \begin{cases} \left[\overline{h}, h_{\pi-1}, D_{m,t}\right], & t>1 \\ \left[\overline{h}, h_{\pi_0}, D_{m,t}\right], & t=1 \end{cases} \tag{27}$$

由于图嵌入是通过对所有节点嵌入进行平均静态生成的，因此唯一随时间变化的组件是最后选定的节点，这不足以捕捉状态转移的动态特征。在多关系节点选择解码器中，为了丰富上下文表示，上下文信息进一步修正如下：

$$h^{mask} = \text{softmax}\left(W_m h_N\right) \odot m_t^v \odot h_N \tag{28}$$

$$\tilde{h}^{mask} = \frac{1}{n}\sum_{i=1}^{N} h^{mask} \tag{29}$$

$$h^{un\_mask} = \text{softmax}\left(W_{um} h_N\right) \odot \left(1 - m_t^v\right) \odot h_N \tag{30}$$

$$\tilde{h}^{un\_mask} = \frac{1}{n}\sum_{i=1}^{N} h^{un\_mask} \tag{31}$$

其中，W_m 和 W_{um} 是可训练参数，m_t^v 是已访问节点的掩码矩阵，h_N 表示节点嵌入。这里，h^{mask} 表示与已访问节点相关的节点嵌入，h^{un_mask} 表示与未访问节点相关的节点嵌入。通过计算已掩蔽关系 h^{mask} 和 h^{un_mask} 的平均值，得到已访问图嵌入 $\tilde{h}^{mask}$ 和未访问图嵌入 $\tilde{h}^{un\_mask}$。这些嵌入与最后访问的节点 $h_{\pi-1}$ 结合，形成增强的多关系上下文特征：

$$H_t^c = \begin{cases} \left[\tilde{h}^{mask}, \tilde{h}^{un\_mask}, h_{\pi_{t-1}}, D_{m,t}\right], & t>1 \\ \left[\tilde{h}^{mask}, \tilde{h}^{un\_mask}, h_{\pi_0}, D_{m,t}\right], & t=1 \end{cases} \tag{32}$$

多关系上下文特征 H_t^c 表示当前时间步 t 时节点和车辆的总特征。然后，上下文向量被输入到多头线性注意力机制中。节点解码器和多关系节点解码器的结构如图 4 所示。与编码器架构中的 MHLA 不同，这里 Query 来自上下文向量，而 Key 和 Value 来自节点嵌入。其定义如下：

$$\hat{H}_t^c = MHLA\left(H_t^c W_c^Q, E_c h_N W_c^K, F_c h_N W_c^V\right) \tag{33}$$

其中，W_c^Q，W_c^K，W_c^V，E_c 和 F_c 是可训练参数，类似于编码器。然后，通过将增强的上下文 $\hat{H}_t^c$ 与 h_N 进行比较，生成概率分布：

$$u_t = C \times \tanh\left(\frac{\left(W_Q H_t^c\right)^{\mathrm{T}}\left(W_K h_N\right)}{\sqrt{d_K}}\right) \tag{34}$$

这里，W_Q 和 W_k 是可训练参数，我们使用 $C=10$ 来将 u_i 截断在 $[-C, C]$ 范围内。解码器对客户点进行掩蔽操作。每个节点的选择概率使用 softmax 函数进行归一化：

$$p_t = \text{softmax}\left(u_t\right) = \frac{e^{u_t}}{\sum_j e^{u_j}} \tag{35}$$

在训练阶段，我们采用基于解码器输出概率 $p_{i,t}$ 的采样解码方法。在测试阶段，我们使用贪婪解码方法，选择最高概率值 $p_{i,t}$。

3.4. 训练策略

本文使用策略梯度方法来训练模型。目标函数 $L\left(\theta|s\right)$ 是期望奖励，基于参数 θ 进行评估：

$$\nabla_\theta L(\theta|s) \approx E_{p_\theta(a|s)}\left[\left(R(a|s)-R^{BL}(s)\right)\nabla \log p_\theta(a|s)\right] \tag{36}$$

在训练过程中，我们使用两个网络来表示：

(1) 策略网络 $R(a|s)$，根据概率分布 $p_{i,t}$ 解码以获得样本的总成本；

(2) 基准网络 $R^{BL}(s)$，用于评估训练过程中的性能，并通过 $R^{BL}(s)$ 来消除训练的方差。在每个训练周期中，通过每次更新中的采样策略获得平均目标 L，并使用 Adam 优化器更新参数 $\nabla_\theta L(\theta|s)$。

4. 实验验证

本节展示了我们提出的深度强化学习方法在解决 HCVRP 中的实验评估，该问题涉及具有不同容量的车辆车队。车辆从配送中心出发，穿越客户节点以满足需求，目标是最小化所有车辆的总行驶时间或最大行驶时间。

4.1. 实验设置

我们设计了实验，旨在全面评估基于 DRL 的解决方案，特别是在涉及异构车队的场景下。车辆从中央配送中心出发，按照预定路线随机分布的客户需求进行配送。为了确保严格的评估，问题实例的生成和实验参数设置如下。

仓库和客户位置的坐标在单位正方形 $[0,1]\times[0,1]$ 内随机抽样。客户需求在集合 $\{1,2,\cdots,9\}$ 中随机分配，配送中心的需求固定为零。我们评估了两种车队组成，V3 和 V5，V3 的容量为 20、25 和 30，V5 的容量为 20、25、30、35 和 40。为了评估可扩展性，每个车队的客户数量分别为：V3 为 40、60、80 和 100，V5 为 80、100、120 和 140。

我们进行了实验，以评估两种 HCVRP 配置：MM-HCVRP 和 MS-HCVRP。在 MM-HCVRP 中，所有车辆的标准速度为 1.0，以便在不同车辆容量的路径之间进行公平比较。在 MS-HCVRP 中，车辆的速度与其容量成反比，以防止大容量车辆的过度使用，并最小化总行驶时间。

训练过程中使用动态生成的实例，以确保模型对不同客户需求的鲁棒性和适应性。每个训练周期包括 1,280,000 次迭代，分为 2,500 个小批量。由于超过 50 个周期后收益递减，训练限制为 50 个周期。节点和车辆特征嵌入到 128 维空间中，并通过车辆和节点选择解码器处理。使用 Adam 优化器，初始学习率为 10^{-4}，每个周期衰减 0.995。为了稳定训练，梯度范数被裁剪为 3.0，并应用了 0.05 的衰减系数。

验证使用了每个问题规模 1,280 个实例，以确保实验之间的一致性。所有实验都在相同的硬件上进行，以消除计算差异带来的不一致性，确保公平的性能评估。

4.2. 比较分析

为了评估我们提出的基于 DRL 的方法的有效性，我们进行了比较分析，涉及几种经典的启发式算法和一种先进的 DRL 方法。由于获得 MM-HCVRP 的最优解的计算复杂性，尤其是在较大的实例中，启发式方法被用作基准，以提供实际的参考标准。为确保实验公平性，所有方法在相同数据集、硬件条件和时间限制下运行，并统一设置超参数和随机种子。此外，基准方法的启发式设计和深度学习模型均经过调整，以适应实验场景。

本研究中的基准方法包括：字符串移除的松弛诱导法(SISR)、蚁群优化算法(ACO)、萤火虫算法(FA)和基于深度强化学习的注意力模型(AM)。SISR [8]在解决 CVRP 及其变种时表现出色，通常在目标值和最优间隙方面优于 LKH3 启发式算法。ACO 方法[9]专门为带时间窗的异构车辆路径规划问题进行了修改，通过并行构造所有蚂蚁的解来加速计算，显著减少计算时间。FA 方法[10]是传统萤火虫算法的增强版本，旨在更有效地解决异构固定车队路径规划问题。最后，AM [6]是最先进的 DRL 方法，通过强化学

习在决策过程中学习节点选择策略，用于构建 TSP 和 CVRP 问题的解。

为了确保比较的一致性，我们调整了基准配置，包括目标函数和相关参数，使其与 MM-HCVRP 的设置一致。具体来说，ACO 和 FA 的迭代设置根据问题规模线性扩展，以增强性能，偏离了通常使用固定迭代次数的做法。对于 SISR，我们遵循其原始的迭代策略，允许迭代次数随问题复杂度增加，从而确保在不同问题规模下公平比较算法性能。对于 MS-HCVRP 场景，使用与 MM-HCVRP 实验相同的启发式基准方法。

除了经典的基准方法，我们还使用两种不同的动作选择策略来评估我们的基于 DRL 的方法。第一种是贪婪动作选择策略，完全依赖于策略网络的输出。在每个决策步骤中，选择概率最高的车辆–节点对。我们在结果中将此策略表示为 AM (Greedy)和 Paper (Greedy)。第二种策略是采样动作选择，它基于解码器输出的概率分布进行采样，而不是选择具有最高概率的动作。每个动作的采样是根据其概率进行的，引入了解的多样性。概率是通过对解码器输出进行 softmax 操作定义的，每个动作采样 S 次，从采样集中选择最佳解。在我们的实验中，S 被设置为 1,280 和 12,800，相应的结果分别表示为 AM (S = 1280)、Paper (S = 1280)和 AM (S = 12800)、Paper (S = 12800)。

为了确保评估的可靠性，我们在 MM-HCVRP 和 MS-HCVRP 场景下，使用三辆车和五辆车的车队进行比较。结果呈现在表 2、表 3、表 4 和表 5 中，包括每种设置的关键指标，如平均目标值(Obj.)、最优间隙(Gap)和计算时间(Time)。由于优化 MM-HCVRP 的计算成本较高，最优间隙是通过将每个方法的目标值与所有试验中表现最好的方法进行比较计算的。报告的结果是三次独立运行的平均值，确保我们的发现具有可靠性和可重复性。

我们的分析表明，所提出的基于 DRL 的方法在提供竞争性解的质量的同时，显著减少了计算时间，特别是相比于传统的启发式方法。这些结果突出了我们方法在高效解决大规模异构车辆路径规划问题中的潜力。

从表 2 的实验结果可以看出，对于具有三辆车的 MM-HCVRP 和 MS-HCVRP 实例，精确求解器(SISR)在小规模问题(V3-C40 和 V3-C60)中实现了最小的平均目标值和最优间隙，并且计算时间相对较短。然而，随着问题规模的增大(V3-C80 和 V3-C100)，SISR 的计算时间呈指数增长，导致其在大规模问题中的计算效率降低，使其在现实应用中不具备实用性。

我们的 DRL 方法(Paper)在不同策略下(贪婪和采样(S = 1280 和 S = 12800))表现出色。特别是在 V3-C100 的 MM-HCVRP 实例中，Paper (S = 12800)达到了最低的平均目标值(9.00)和间隙(1.24%)，优于其他基准方法。这证明了采样策略在提高解质量方面的有效性。增加采样大小理论上能带来更好的解，尽管计算时间较长。虽然 S = 12800 的计算时间(6.44 秒)略长于 S = 1280 (3.19 秒)和贪婪策略(1.28 秒)，但仍显著短于 SISR (1135 秒)。

对于 MS-HCVRP，Paper (S = 12800)提供了具有竞争力的结果，分别为 127.29 和 2.15%，同时显著减少了与 SISR 和其他启发式算法相比的计算时间。这表明，尽管我们的 DRL 方法的计算时间可能稍高于 AM 算法，但它提供了更优的解质量，特别是在大规模问题中。

Table 2. Comparison of the DRL method and the benchmark method for three vehicles
表 2. 三辆车的 DRL 方法与基准方法比较

	方法	V3-C40		V3-C60		V3-C80		V3-C100	
		Obj.	Gap	Obj.	Gap	Obj.	Gap	Obj.	Gap
Min-max	SISR	4.00	0%	5.58	0%	7.27	0%	8.89	0%
	ACO	4.31	7.75%	6.18	10.75%	8.14	11.97%	10.05	13.05%

续表

	FA	4.49	12.25%	6.30	12.90%	8.32	14.44%	10.11	13.72%
	AM (Greedy)	4.85	21.25%	6.57	17.74%	8.32	14.44%	9.98	12.26%
	AM (S = 1280)	4.36	9.00%	5.99	7.39%	7.73	6.33%	9.36	5.29%
	AM (S = 12800)	4.31	7.75%	5.92	6.09%	7.66	5.36%	9.28	4.39%
	Paper (Greedy)	4.33	8.25%	5.97	6.45%	7.71	6.05%	9.38	5.51%
	Paper (S = 1280)	4.14	3.50%	5.74	2.89%	7.45	2.48%	9.04	1.69%
	Paper (S = 12800)	4.10	2.50%	5.69	1.97%	7.40	1.79%	9.00	1.24%
Min-sum	Exact-solver	55.43	0%	78.47	0%	102.42	0%	124.61	0%
	SISR	55.79	0.65%	79.12	0.83%	103.41	0.97%	126.19	1.27%
	ACO	60.11	8.44%	86.05	9.66%	113.75	11.06%	140.61	12.84%
	FA	59.94	8.14%	85.36	8.78%	112.81	10.14%	138.92	11.48%
	AM (Greedy)	66.54	20.04%	91.19	16.21%	117.22	14.45%	141.14	13.27%
	AM (S = 1280)	60.95	9.96%	85.74	9.26%	111.78	9.14%	135.61	8.83%
	AM (S = 12800)	60.26	8.71%	84.96	8.27%	110.94	8.32%	134.72	8.11%
	Paper (Greedy)	58.68	5.86%	82.91	5.66%	107.92	5.50%	130.98	5.11%
	Paper (S = 1280)	56.93	2.71%	80.50	2.59%	104.64	2.22%	127.94	2.67%
	Paper (S = 12800)	56.82	2.51%	80.23	2.24%	104.51	2.04%	127.29	2.15%

Table 3. Comparison of the calculation time of the DRL method and the benchmark method for three vehicles
表 3. 三辆车的 DRL 方法与基准方法的计算时间比较

	方法	V3-C40	V3-C60	V3-C80	V3-C100
		Time	Time	Time	Time
Min-max	SISR	245 s	468 s	752 s	1135 s
	ACO	209 s	317 s	601 s	878 s
	FA	168 s	285 s	397 s	522 s
	AM (Greedy)	0.37 s	0.54 s	0.82 s	1.07 s
	AM (S = 1280)	0.88 s	1.19	1.81 s	2.51 s
	AM (S = 12800)	1.35 s	2.46 s	3.67 s	5.17 s
	Paper (Greedy)	0.65 s	0.74 s	0.98 s	1.28 s
	Paper (S = 1280)	1.20 s	1.32 s	1.97 s	3.19 s
	Paper (S = 12800)	1.54 s	2.84 s	4.32 s	6.44 s
Min-sum	Exact-solver	71 s	214 s	793 s	2512 s
	SISR	254 s	478 s	763 s	1140 s
	ACO	196 s	302 s	593 s	859 s
	FA	164 s	272 s	388 s	518 s
	AM (Greedy)	0.49 s	0.83 s	1.01 s	1.23 s
	AM (S = 1280)	0.92 s	1.17 s	1.79 s	2.49 s

续表

	AM (S = 12800)	1.35 s	2.31 s	3.61 s	5.19 s
	Paper (Greedy)	0.54 s	0.92 s	1.05 s	1.46 s
	Paper (S = 1280)	1.09 s	1.36 s	2.18 s	3.27 s
	Paper (S = 12800)	1.52 s	2.77 s	4.41 s	6.45 s

在五辆车的实例中，我们的 DRL 方法的有效性得到了进一步验证。在 V5-C140 MM-HCVRP 实例中，从表 4 和表 5 的实验结果中显示，Paper (S = 12800)达到了最低的平均目标值(6.44)和间隙(1.74%)，同时显著减少了计算时间(10.56 秒)，与 SISR(1863 秒)相比。尽管 AM 算法速度更快(8.73 秒)，但其解的质量不如我们的 DRL 方法。

对于 MS-HCVRP，五辆车的结果进一步验证了我们方法的有效性。在 V5-C140 实例中，Paper (S = 12800)达到了最低的平均目标值(174.75)和间隙(1.08%)，并且计算时间(10.91 秒)表现良好。

Table 4. Comparison of the DRL method and the benchmark method for five vehicles
表 4. 五辆车的 DRL 方法与基准方法比较

	方法	V5-C80		V5-C100		V5-C120		V5-C140	
		Obj.	Gap	Obj.	Gap	Obj.	Gap	Obj.	Gap
Min-max	SISR	3.90	0%	4.72	0%	5.48	0%	6.33	0%
	ACO	4.50	15.38%	5.56	17.80%	6.47	18.07%	7.52	18.80%
	FA	4.61	18.21%	5.62	19.07%	6.58	20.07%	7.60	20.06%
	AM (Greedy)	4.84	24.10%	5.70	20.76%	6.57	19.89%	7.49	18.33%
	AM (S = 1280)	4.32	10.77%	5.18	8.75%	6.03	10.04%	6.93	9.48%
	AM (S = 12800)	4.25	8.97%	5.11	8.26%	5.95	8.58%	6.86	8.37%
	Paper (Greedy)	4.29	10.00%	5.12	8.47%	5.88	7.30%	6.71	6.00%
	Paper (S = 1280)	4.05	3.85%	4.89	3.60%	5.62	2.55%	6.48	2.38%
	Paper (S = 12800)	4.00	2.56%	4.85	2.75%	5.58	1.82%	6.44	1.74%
Min-sum	Exact-solver	102.42	0%	124.63	0%	-	-	-	-
	SISR	103.49	1.04%	126.35	1.38%	149.18	0%	172.88	0%
	ACO	118.58	15.78%	146.51	17.56%	171.82	15.18%	200.73	16.11%
	FA	116.13	13.39%	142.39	14.25%	167.87	12.53%	196.48	13.65%
	AM (Greedy)	128.31	25.28%	152.91	22.69%	177.39	18.91%	201.85	16.76%
	AM (S = 1280)	119.41	16.59%	144.23	15.73%	168.95	13.25%	193.65	12.01%
	AM (S = 12800)	118.04	15.25%	142.79	14.57%	167.45	12.25%	192.13	11.13%
	Paper (Greedy)	107.82	5.27%	130.88	5.01%	153.69	3.02%	177.88	2.89%
	Paper (S = 1280)	105.21	2.72%	127.52	2.32%	151.01	1.23%	174.86	1.15%
	Paper (S = 12800)	104.64	2.17%	127.06	1.95%	150.69	1.01%	174.75	1.08%

Table 5. Comparison of the calculation time of the DRL method and the benchmark method for five vehicles
表 5. 五辆车的 DRL 方法与基准方法的计算时间比较

	方法	V5-C80	V5-C100	V5-C120	V5-C140
		Time	Time	Time	Time
Min-max	SISR	727 s	1091 s	1572 s	1863 s
	ACO	612 s	890 s	1285 s	2081 s
	FA	412 s	541 s	682 s	822 s
	AM (Greedy)	1.08 s	1.31 s	1.74 s	1.93 s
	AM (S = 1280)	1.88 s	2.64 s	3.38 s	4.47 s
	AM (S = 12800)	3.7 s	5.19 s	6.94 s	8.73 s
	Paper (Greedy)	1.22 s	1.56 s	2.26 s	2.24 s
	Paper (S = 1280)	2.41 s	3.45 s	4.88 s	6.22 s
	Paper (S = 12800)	4.93 s	6.80 s	9.42 s	10.56 s
Min-sum	Exact-solver	1787 s	6085 s	-	-
	SISR	735 s	1107 s	1580 s	1881 s
	ACO	608 s	865 s	1269 s	1922 s
	FA	401 s	532 s	677 s	801 s
	AM (Greedy)	0.82 s	1.28 s	1.45 s	1.69 s
	AM (S = 1280)	1.84	2.66 s	3.63 s	4.68 s
	AM (S = 12800)	3.74 s	5.20 s	7.02 s	8.93 s
	Paper (Greedy)	1.11 s	1.54 s	1.99 s	2.83 s
	Paper (S = 1280)	2.53 s	4.02 s	5.07 s	6.58 s
	Paper (S = 12800)	5.16 s	7.44 s	8.86 s	10.91 s

通过比较三辆车和五辆车的结果，我们提出的 DRL 方法在不同规模和目标下表现出一致的性能。虽然精确求解器(SISR)在小规模问题中表现良好，但在大规模问题中的计算时间急剧增加，限制了其实际应用。而我们的 DRL 方法在显著减少计算时间的同时，保持了较高的解质量，特别是在大规模问题中。

与传统的启发式算法如 ACO 和 FA 相比，我们的方法在解的质量和计算效率上均展现出显著优势。使用采样策略(S = 1280 和 S = 12800)进一步提升了解的质量，验证了其有效性。尽管在某些情况下，计算时间可能略高于 AM 算法，但解的质量仍然优越，且计算时间处于可接受范围内。

总之，我们的方法在解决 HCVRP 问题方面展示了显著的优势。尽管本研究中进行的实验集中于较小规模的问题，但模型的架构设计——特别是 Linformer 和多关系节点选择解码器的集成——提供了固有的可扩展性优势。Linformer 通过低秩近似有效管理了计算复杂性，使得该模型能够应用于更大的数据集，而计算负担不会显著增加。

5. 结论

本文提出了一种基于深度强化学习的新型 HCVRP 求解框架，通过引入低秩注意力机制的 Linformer 模型和动态节点特征建模，显著提升了解的质量和计算效率。本文的主要结论包括以下几点：(1) 计算复杂度的优化：Linformer 模型通过低秩近似的注意力机制，有效降低了传统 Transformer 模型在处理大规

模 HCVRP 问题时的计算复杂度。(2) 解质量的提升：动态节点特征建模结合多关系解码器，能够在复杂场景下优化车辆路径，显著提升了解的质量，尤其是在多目标优化任务中的表现。

参考文献

[1] Jabali, O., Gendreau, M. and Laporte, G. (2012) A Continuous Approximation Model for the Fleet Composition Problem. *Transportation Research Part B*: *Methodological*, **46**, 1591-1606. https://doi.org/10.1016/j.trb.2012.06.004

[2] Ilhan, İ. (2021) An Improved Simulated Annealing Algorithm with Crossover Operator for Capacitated Vehicle Routing Problem. *Swarm and Evolutionary Computation*, **64**, Article ID: 100911. https://doi.org/10.1016/j.swevo.2021.100911

[3] Wang, F., Liao, F., Li, Y., Yan, X. and Chen, X. (2021) An Ensemble Learning Based Multi-Objective Evolutionary Algorithm for the Dynamic Vehicle Routing Problem with Time Windows. *Computers & Industrial Engineering*, **154**, Article ID: 107131. https://doi.org/10.1016/j.cie.2021.107131

[4] Sadati, M.E.H. and Çatay, B. (2021) A Hybrid Variable Neighborhood Search Approach for the Multi-Depot Green Vehicle Routing Problem. *Transportation Research Part E*: *Logistics and Transportation Review*, **149**, Article ID: 102293. https://doi.org/10.1016/j.tre.2021.102293

[5] Vaswani, A., Shazeer, N., Parmar, N., *et al*. (2017) Attention Is All You Need. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.

[6] Kool, W., van Hoof, H. and Welling, M. (2018) Attention, Learn to Solve Routing Problems!

[7] Wang, S., Li, B.Z., Khabsa, M., *et al*. (2020) Linformer: Self-Attention with Linear Complexity.

[8] Christiaens, J. and Vanden Berghe, G. (2020) Slack Induction by String Removals for Vehicle Routing Problems. *Transportation Science*, **54**, 417-433. https://doi.org/10.1287/trsc.2019.0914

[9] Palma-Blanco, A., González, E.R. and Paternina-Arboleda, C.D. (2019) A Two-Pheromone Trail Ant Colony System Approach for the Heterogeneous Vehicle Routing Problem with Time Windows, Multiple Products and Product Incompatibility. In: Paternina-Arboleda, C. and Voß, S., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 248-264. https://doi.org/10.1007/978-3-030-31140-7_16

[10] Matthopoulos, P.P. and Sofianopoulou, S. (2019) A Firefly Algorithm for the Heterogeneous Fixed Fleet Vehicle Routing Problem. *International Journal of Industrial and Systems Engineering*, **33**, 204-224. https://doi.org/10.1504/ijise.2019.102471