

基于最大熵逆强化学习面向双分支特征的车辆路径选择模型

黄明玥¹ 夏大文^{1*} 张 瀚¹ 何文琳¹

HUANG Mingyue XIA Dawen ZHANG Han HE Wenlin

摘 要

针对交通系统的路径选择行为建模面临可行路径多、轨迹数据稀疏和个体偏好不确定等问题,文章提出了一种基于最大熵逆强化学习面向双分支特征的车辆路径选择模型,实现路径选择行为的高效精准建模。该模型面向环境特征分支与上下文特征分支,分别对路网拓扑序列信息和驾驶员偏好进行建模,采用最大熵逆强化学习方法将双分支特征融合至奖励函数,以增强驾驶员个性化决策的表征能力。同时,构建图注意力网络(GAT)与插入式Transformer协同的特征增强模块,GAT增强链路表征以解决轨迹稀疏问题,插入式Transformer有效捕捉轨迹上下文的依赖关系。基于上海市GPS轨迹数据集的实验结果表明,所提模型在路径选择的准确性与轨迹拟合度方面均优于可比较方法,可以有效解决因忽略驾驶员个性化偏好与实际轨迹数据稀疏问题,能为个性化路径选择建模提供理论方法和技术支持。

关键词

路径选择模型;最大熵逆强化学习;插入式Transformer;图注意力网络

doi: 10.3969/j.issn.1672-9528.2025.12.035

0 引言

车辆路径选择行为建模是构建智能交通系统的关键内容,在智能导航、交通管理和城市规划中具有重要意义^[1]。虽然该领域已有众多的研究积累,但是规划路径与实际行驶路径之间仍存在显著偏差。

早期研究通常基于传统离散选择模型。Fosgerau等^[2]提出递归Logit(R-Logit)模型,将路径选择建模为动态离散选择框架的链路选择序列。邱树涵等^[1]开发基于数据-知识融合驱动的递归Logit模型,以实现路段转移效用的有效建模。随着深度学习的发展,越来越多的研究逐步探索车辆路径选择建模应用。Wang等^[3]将循环神经网络与图注意力网络相结合,以提升传统A*搜索算法的性能。Choi等^[4]基于一种生成对抗模仿学习框架,提出生成与真实轨迹分布高度一致的合成车辆轨迹。Huang等^[5]进一步提出一种融合神经网络条件预测模型与最大熵逆强化学习评分机制的预测式行为规划方法,有效提升路径生成的鲁棒性和多样性。

近年来,基于深度逆强化学习的路径选择研究取得显著进展。Ziebart等^[6]构建最大熵逆强化学习(MaxEnt-IRL)框架,将奖励函数表示为特征线性组合并应用于路径偏好建模。Liu等^[7]结合图神经网络与注意力机制引入动态交通特征与时序特征,并估算预期状态访问频率(SVF)。Zhang等^[8]采用

图注意力网络增强的分治路径规划框架,以提升轨迹拟合精度。Li等^[9]通过神经网络分别建模个体静态特征与动态建成环境特征,并结合最大熵逆强化学习方法,以增强模型对特征交互关系的解释能力。然而,现有研究仍然面临挑战:缺乏能够有效捕捉链路之间时空依赖关系的建模机制,忽视驾驶员个性化偏好对决策行为的影响。此外,受限于真实轨迹数据的稀疏性,模型在路径选择预测的准确性与泛化能力有待提升。

针对上述问题,本文提出一种基于最大熵逆强化学习面向双分支特征的车辆路径选择模型(如图1所示)。

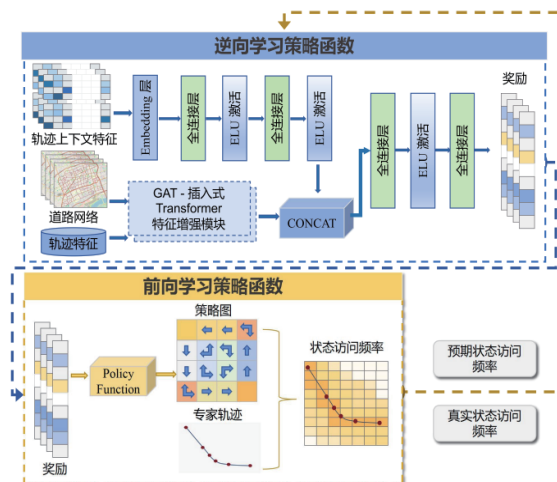


图1 基于最大熵逆强化学习面向双分支特征学习的车辆路径选择模型

该模型通过环境特征与轨迹上下文双分支刻画路网结构

1. 贵州民族大学数据科学与信息工程学院 贵州贵阳 550025
[基金项目] 高维数据统计分析与智能计算重点实验室项目
(GZMUSYS〔2021〕04)

与个性化偏好，并在奖励函数中融合双分支特征以增强决策表征能力。同时，引入 GAT-插入式 Transformer 特征增强模块，GAT 通过邻域聚合增强链路表示以缓解数据稀疏，插入式 Transformer 捕捉链路间时序依赖以实现更为精准的时空建模。

1 模型描述

1.1 问题定义

路径选择问题可被视为马尔可夫决策过程 (MDP) [5]，可以有效地对出行者的决策过程提供数学建模框架。在该框架中，将路径选择问题表示为 (S, A, T, R, ρ_0) 的 MDP 元组，其中， S 表示状态空间的集合； A 表示出行者采取行动的集合； ρ_0 表示初始状态的概率分布； $T(s|s, a)$ 表示在状态 s 中采取动作 a 后转移到状态 s' 的概率； $R(s, a)$ 表示奖励函数，根据当前状态给予奖励，以捕捉出行者的路径选择偏好。

本文的输入数据是一组包含 N 个观测得到的车辆轨迹集合，记为 $X = \{\tau_1, \tau_2, \dots, \tau_n\}$ 。每条轨迹 τ 表示为一个状态-动作序列： $\tau = ((s_1, a_1), (s_2, a_2), \dots, (s_m, a_m))$ ，其中， $s_i \in S$ 表示状态（如道路路段）； $a_i \in A$ 表示动作（如转向决策）。

1.2 GAT-插入式 Transformer 增强链路表示

为了解决轨迹稀疏问题，本文引入 GAT-插入式 Transformer 特征增强模块 [8]，以强化链路特征表示（如图 2 所示）。轨迹稀疏使得部分链路在训练数据时出现频率较低而导致特征学习不足，从而影响路径选择模型的整体性能。

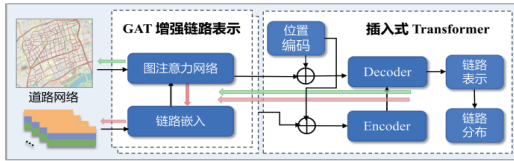


图 2 GAT-插入式 Transformer 特征增强模块

为了实现轨迹中每个决策点的建模，本文为每条链路 s_i 构建丰富的特征表示。首先，对于链路 s_i ，分别设置环境嵌入向量 w_i 与环境特征嵌入向量 w'_i （如链路长度、道路类型）。针对时间窗口 t ，设置上下文嵌入向量 v_t 与上下文特征嵌入向量 v'_t （如天气、通行时间、车速）。链路 s_i 初始特征通过拼接上述 4 个向量获得：

$$s_i = [w_i \parallel w'_i \parallel v_t \parallel v'_t] \quad (1)$$

式中： $\parallel$ 表示拼接操作。

其次，采用 GAT 对链路的环境特征进行增强。增强后的链路表征维度设为 d' 。在 GAT 中，首先对空间相邻的链路进行线性变换，并计算注意力系数：

$$a_{ij} = \frac{\exp(\text{LeakyReLU}(a^T [Hs_i \parallel Hs_j]))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(a^T [Hs_i \parallel Hs_k]))} \quad (2)$$

式中： $a \in \mathbb{R}^{2d'}$ 和 $H \in \mathbb{R}^{d' \times d}$ 表示将表征转换为关联系数的权重参数； N_i 表示链路 s_i 的相邻链路集合。最后，通过注意力系数对邻居特征进行加权聚合，得到增强后的链路表征：

$$s'_i = \sigma \left(\sum_{j \in N_i} a_{ij} Hs_j \right) \quad (3)$$

式中： σ 为非线性激活函数。

在此基础上，本文进一步引入插入式 Transformer 模块，以捕捉链路序列间的依赖关系。首先，对 Transformer 的输入表示进行处理，为了充分利用序列中元素的相对或绝对位置信息，采用正弦和余弦函数计算位置编码 PE：

$$\text{PE}_{\text{pos}, 2i} = \sin(\text{pos}/10000^{2i/p'}) \quad (4)$$

$$\text{PE}_{\text{pos}, 2i+1} = \cos(\text{pos}/10000^{(2i)/d'}) \quad (5)$$

式中： pos 是路径序列中链路的位置； i 是维度索引。其次，将位置编码的 PE 值相加得到 X 值，并将结果值 X_{PE} 馈送到编码器和解码器。

最后，本文模型的编码器与解码器均采用基于缩放点积的多头注意力机制。对于输入序列向量集合 $[x_1, x_2, \dots, x_n]$ ，其中每个向量 x_i ，分别计算查询向量 $q_i = x_i W_Q$ 、键向量 $k_i = x_i W_K$ 和值向量 $v_i = x_i W_V$ ：

$$Q = XW_Q, K = XW_K, V = XW_V \quad (6)$$

式中： $W_Q, W_K, W_V \in \mathbb{R}^{d \times d}$ 分别表示查询、键和值的权重。随后，通过缩放点积计算注意力权重，并对其进行归一化，接着以此对值向量进行加权求和，得到最终的注意力输出：

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d'}} \right) V \quad (7)$$

1.3 基于双分支特征学习的最大熵逆强化学习

在车辆路径选择过程中，驾驶员的决策不仅受到路网环境属性的影响，而且受到轨迹上下文特征的约束与引导。为了充分刻画这种决策行为中的异质性特征，本文引入双分支特征学习机制，在路径选择建模中分别学习环境特征与上下文特征双分支的奖励参数，以更好地反映驾驶员在复杂交通场景下的个体化偏好 [9]。为了提升奖励建模能力，本文将轨迹奖励定义为各状态奖励的累加：

$$\begin{aligned} R_\phi(\tau) &= \sum_{(s_i, a_i) \in \tau} [\lambda_\gamma f_\gamma(s_i, a_i) + \lambda_\theta f_\theta(c)] \\ &= \sum_{(s_i, a_i) \in \tau} r_\phi(s_i, a_i, c), \end{aligned} \quad (8)$$

式中： ϕ 表示奖励函数的神经网络参数； $r_\phi(s_i, a_i, c)$ 表示在状态 s_i 执行动作 a_i 时的奖励； c 表示轨迹上下文信息； f_γ 和 f_θ 分别表示参数 γ 和 θ 的神经网络。

在最大熵框架下，得到轨迹分布：

$$p_{\phi}(\tau) = \frac{\exp(R_{\phi}(\tau))}{Z_{\phi}}, \quad Z_{\phi} = \sum_{\zeta \in \Gamma} \exp(R_{\phi}(\zeta)) \quad (9)$$

式中： Z_{ϕ} 表示配分函数； Γ 表示从相同起始状态出发的所有可能轨迹的集合； ζ 表示一条候选轨迹。为了学习最优的奖励参数 ϕ ，最大化观测轨迹集合 T 的对数似然可得：

$$L_r(\phi) = \sum_{\tau \in T} \log p_{\phi}(\tau) = \sum_{\tau \in T} (R_{\phi}(\tau) - \log Z_{\phi}) \quad (10)$$

对参数 ϕ 求导并利用，可得：

$$\begin{aligned} \nabla L_r(\phi) &= \sum_{\tau \in T} \left(\frac{\partial R_{\phi}(\tau)}{\partial r_{\phi}} - \frac{\partial \log Z_{\phi}}{\partial r_{\phi}} \right) \\ &= (\mu_{\text{real}}[T] - \mu_{\text{exp}}[T_i]) \cdot \frac{\partial r_{\phi}}{\partial \phi} \end{aligned} \quad (11)$$

式中： $\mu_{\text{real}}[T_i]$ 表示期望状态访问频率； $\mu_{\text{exp}}[T_i]$ 表示真实状态访问频率； $\partial r_{\phi} / \partial \phi$ 表示奖励函数 r_{ϕ} 关于神经网络参数 ϕ 的导数。

2 实验结果与分析

2.1 数据集

本文采用由 Zhao 等^[10] 发布的上海市出租车 GPS 轨迹数据集，包含 10 609 条轨迹，涉及 10 000 余辆车。该路网由 319 个交叉口和 713 个路段构成，每条 GPS 记录包含车辆 ID、日期时间、速度、经纬度坐标、载客状态等。轨迹数据采样间隔约为 10 s，并采用地图匹配算法将 GPS 数据映射至路网。

对于模型性能评估，首先对轨迹数据进行清洗与划分，剔除含循环路径的行程。为了评估训练数据规模对模型性能的影响，本文设置 100、1 000、10 000 条轨迹 3 种不同规模的训练集。

2.2 评价指标

为了评估模型性能，本文采用编辑距离 (ED)、BLEU 得分 (BLEU)^[4]、Jensen-Shannon 距离 (JSD) 和平均对数概率 (LP) 4 个评价指标以衡量真实轨迹与预测轨迹之间的相似性：

$$\text{ED} = \frac{1}{N} \sum_{i=1}^N \min \left(\frac{\text{Edit}(\hat{x}_i, x_{i,\text{ref}})}{T_{i,\text{ref}}}, 1 \right) \quad (12)$$

$$\text{BLEU}_n = \frac{1}{N} \sum_{i=1}^N \min \left(1, \frac{T_i}{T_{i,\text{ref}}} \right) \left(\prod_{j=1}^n P_j \right)^{\frac{1}{n}} \quad (13)$$

$$d_{js} = \sqrt{\left(D_{\text{KL}} \left(p \parallel \frac{p+q}{2} \right) + D_{\text{KL}} \left(q \parallel \frac{p+q}{2} \right) \right) / 2} \quad (14)$$

式中： N 表示轨迹数量； $\hat{x}_i$ 表示第 i 条预测轨迹； $x_{i,\text{ref}}$ 表示具有相同起始点的真实参考轨迹； $T_{i,\text{ref}}$ 表示参考轨迹长度； P_j 表示生成的预测轨迹与真实参考轨迹在 j -gram 片段上的匹配程度； D_{KL} 表示 KL 散度。

2.3 参数设置

在实验中，数据集按照 8:2 随机划分为训练集和测试集。同时，模型基于 PyTorch 实现，训练阶段采用 GPU 加速和多线程数据加载以提高效率。本文模型的参数设置如表 1 所示。

表 1 参数设置

参数类别	设置内容
批大小	64
优化器	Adam
学习率	1e-3
迭代次数	1 000, 2 000, 3 000

2.4 实验结果

为了评估模型在不同训练数据规模下的性能，在 100、1 000 和 10 000 条训练数据上，将本文模型与 5 种比较模型进行对比实验，包括 PSL^[10]、DNN-PSL^[10]、R-Logit^[2]、Maxent-IRL^[5]、Deep-IRL^[11]。实验结果如表 2 所示。

表 2 基于不同训练数据规模的模型性能比较

模型	训练数据量=100			训练数据量=1 000			训练数据量=10 000		
	ED	BLEU	JSD	ED	BLEU	JSD	ED	BLEU	JSD
PSL	0.519	0.351	0.566	0.473	0.592	0.484	0.479	0.673	0.382
DNN-PSL	0.528	0.331	0.576	0.471	0.581	0.486	0.485	0.652	0.384
R-Logit	0.514	0.387	0.516	0.425	0.612	0.509	0.223	0.641	0.469
Maxent-IRL	0.308	0.633	0.412	0.271	0.659	0.379	0.245	0.674	0.351
Deep-IRL	0.246	0.647	0.396	0.243	0.685	0.351	0.218	0.708	0.348
Ours	0.251	0.651	0.401	0.231	0.721	0.329	0.183	0.746	0.315

随着训练数据规模的增加，比较模型的性能总体提升，但是在大规模数据下，本文模型在 ED 和 JSD 指标上均表现最优，以及 BLEU 分数明显优于其他模型。实验结果表明，通过提取更为复杂的模式与特征，在轨迹匹配和概率分布拟合方面具有显著优势，并保持稳健的路径预测能力。

为了验证模型结果的不确定性，计算各模型在 3 次重复实验中的标准偏差，并以误差形式进行呈现（如图 3 所示）。实验结果显示，PSL、DNN-PSL 和 R-Logit 模型的标准偏差接近于零，主要原因在于它们依赖于路径生成方法，并在训练过程中通过极大似然估计收敛至稳定的最优解；Maxent-IRL 和 Deep-IRL 在不同数据规模则存在一定的波动，其中前者对数据规模更为敏感。相比之下，本文模型在各数据规模下均保持较高的稳定性，即使在数据有限的情况下，依然可以输出一致可靠的预测结果，具有良好的适应性与稳健性。

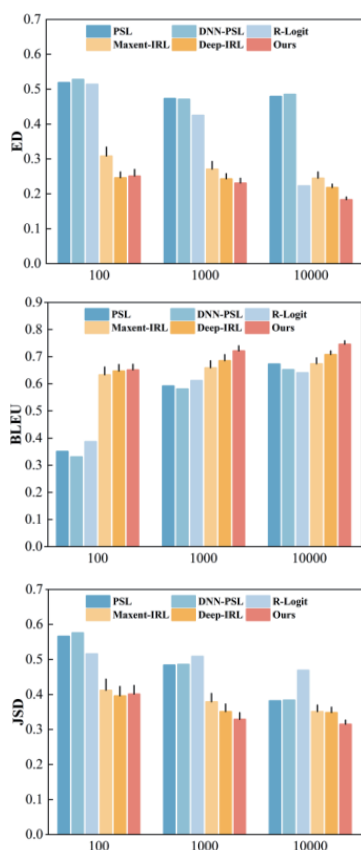


图 3 基于 ED、BLEU 和 JSD 指标的模型性能比较

3 结语

本文提出一种基于最大熵逆强化学习面向双分支特征的车辆路径选择模型。该模型通过环境与上下文双分支分别建模路网拓扑信息和驾驶员个性化偏好，融合图注意力网络与插入式 Transformer 以构建特征增强模块，有效提升路径序列的表征能力。在真实轨迹数据上的实验结果表明，所提模型在路径选择的准确率与轨迹拟合度方面均优于可比较方法，具有良好的泛化性与实用性。不仅能为驾驶员路径选择行为建模提供融合时空依赖与个性化偏好的新思路，而且能为智能导航系统的个性化路径规划与交通行为分析提供技术支撑。

参考文献：

- [1] 邱树涵, 陈雪剑, 秦国阳, 等. 数据-知识融合驱动的车辆路径选择行为建模 [J/OL]. 中国公路学报, 1-19[2025-11-03]. <https://link.cnki.net/urlid/61.1313.U.20250808.1044.004>.
- [2] FOSGERAU M, FREJINGER E, KARLSTROM A. A link based network route choice model with unrestricted choice set [J]. Transportation research part B: methodological, 2013, 56: 70-80.
- [3] WANG J Y, WU N, ZHAO W X, et al. Empowering A* search algorithms with neural networks for personalized route recommendation [C]//Proceedings of the 25th ACM SIGKDD

International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019: 539-547.

- [4] CHOI S, KIM J, YEO H. TrajGAIL: generating urban vehicle trajectories using generative adversarial imitation learning[J]. Transportation research part C: emerging technologies, 2021, 128: 103091.
- [5] HUANG Z Y, LIU H C, WU J D, et al. Conditional predictive behavior planning with inverse reinforcement learning for human-like autonomous driving[J]. IEEE transactions on intelligent transportation systems, 2023, 24(7): 7244-7258.
- [6] ZIEBART B D, MAAS A, ANDREW BAGNELL J, et al. Maximum entropy inverse reinforcement learning [C]//AAAI'08: Proceedings of the 23rd national conference on Artificial intelligence. Palo Alto: AAAI Press, 2008: 1433-1438.
- [7] LIU S, JIANG H, CHEN S P, et al. Integrating Dijkstra's algorithm into deep inverse reinforcement learning for food delivery route planning[J]. Transportation research part E: logistics and transportation review, 2020, 142: 102070.
- [8] ZHANG P J, LIU S, SHI J, et al. Route planning using divide-and-conquer: a GAT enhanced insertion transformer approach[J]. Transportation research part E: logistics and transportation review, 2023, 176: 103176.
- [9] LI B Y, ZHANG W J. Enhancing pedestrian route choice models through maximum-entropy deep inverse reinforcement learning with individual covariates (MEDIRL-IC)[J]. IEEE transactions on intelligent transportation systems, 2024, 25(12): 20446-20463.
- [10] ZHAO Z, LIANG Y B. A deep inverse reinforcement learning approach to route choice modeling with context-dependent rewards [J]. Transportation research part C: emerging technologies, 2023, 149: 104079.
- [11] WULFMEIER M, RAO D, WANG D Z, et al. Large-scale cost function learning for path planning using deep inverse reinforcement learning [J]. International journal of robotics research, 2017, 36(10): 1073-1087.

【作者简介】

黄明玥 (1998—), 女, 贵州安顺人, 硕士研究生, 研究方向: 海量数据统计与分析、大数据分析、人工智能。

夏大文 (1982—), 通信作者 (email: dwxia@gzmu.edu.cn), 男, 湖南永州人, 博士, 教授、硕士生导师, 研究方向: 大数据、人工智能、数据挖掘、机器学习。

(收稿日期: 2025-10-16 修回日期: 2025-12-09)