

基于需求密度预测的网约车集约化调度方法

郭羽含^{1,2}, 丁文婧¹

(1. 辽宁工程技术大学软件学院, 辽宁葫芦岛 125105; 2. 浙江科技学院理学院, 浙江杭州 310023)

摘要: 为提升网约车接单率和利润率、实现全局供需平衡, 提出一种基于需求密度预测的网约车集约化调度方法。首先, 根据历史数据设计基于多层混合感知野的深度时空残差感知网络结构, 该网络基于需求频度划分历史时空数据, 并通过卷积指数线性网络及残差单元对不同时空数据进行差异化处理。结合基于门控机制的融合及求和融合方法动态聚合时间、空间和外部特征, 实现了对需求密度的准确预测, 从而预估网约车需求密度集群效益。其次, 基于网约车经济效益和需求密度集群效益, 建立调度数学模型, 设计传感邻域限制调度范围, 提升搜索效率。将遗传算法与匈牙利算法相结合, 提高算法寻优能力, 避免基因缺失, 通过改进选择和变异算子, 增强遗传算法的局部随机搜索能力, 规避早熟风险, 从而得到网约车与乘客的最佳匹配结果, 保证了全局供需平衡和总体盈利能力。最后, 基于大规模真实数据集对预测模型的性能和调度算法的有效性进行验证, 实验结果表明, 预测模型精度可达到97%, 调度算法的求解质量可达最优解的99%, 可为网约车平台提供调度策略, 保障交通系统稳定。

关键词: 智能交通系统; 车辆调度; 网约车需求密度预测; 遗传算法; 匈牙利算法; 深度神经网络

DOI: 10.11907/rjtk.231463

开放科学(资源服务)标识码(OSID):

中图分类号: TP302

文献标识码: A

文章编号: 1672-7800(2024)004-0021-10



Intensive Scheduling Method of Ride-Sharing Based on Demand Density Prediction

GUO Yuhan^{1,2}, DING Wenjing¹

(1. School of Software, Liaoning Technical University, Huludao 125105, China;

2. School of Science, Zhejiang University of Science and Technology, Hangzhou 310023, China)

Abstract: An intense scheduling strategy for online vehicles based on demand density prediction is suggested to increase the order acceptance rate and profitability of these vehicles as well as attain worldwide supply-demand equilibrium. The first step is to design a deep spatiotemporal residual perception network structure based on a multilayer hybrid perception field using historical data. This structure divides historical spatiotemporal data based on demand frequency and separates different types of spatiotemporal data using a convolutional exponential linear network and residual units. Accurate demand density prediction is achieved by combining the fusion and summation fusion methods based on gating mechanisms to dynamically aggregate temporal, spatial, and external variables. This method also predicts the advantage of demand density clustering of online cars. Second, a scheduling mathematical model is developed, and the sensing neighborhood is created to reduce the scheduling range and increase search efficiency. This is based on the economic benefits and demand density clustering benefits of online cars. To increase the search capacity of the algorithm and prevent gene deficiencies, the genetic algorithm is combined with the Hungarian algorithm. Additionally, the local random search capacity of the genetic algorithm is improved by enhancing the selection and variation operators to reduce the risk of premature maturation and to achieve the best match between online taxi and passenger, which ensures the equilibrium of supply and demand globally and overall profitability. Finally, using sizable real data sets, the performance of the prediction model and the efficiency of the scheduling technique are confirmed. According to the experimental findings, the prediction model's accuracy can reach 97%, and the scheduling algorithm's solution quality can reach 99% of the best possible result, which can be used to develop scheduling plans for online taxi platforms and guarantee the stability of the transportation system.

Key Words: intelligent transportation systems; vehicle scheduling; ride-sharing demand density forecast; genetic algorithm; Hungarian algorithm; deep neural network

收稿日期: 2023-05-04

基金项目: 国家自然科学基金项目(61404069)

作者简介: 郭羽含(1983-), 男, 博士, 辽宁工程技术大学软件学院副教授、硕士生导师, 研究方向为智能搜索算法、车辆调度问题、供应链优化问题; 丁文婧(1998-), 女, 辽宁工程技术大学软件学院硕士研究生, 研究方向为车辆调度问题。本文通讯作者: 丁文婧。

0 引言

城市交通的快速发展导致了交通拥堵、资源浪费和出行困难等问题,因此缓解交通压力、合理规划出行路线成为城市发展所要应对的重要挑战。网约车需求密度预测是典型的时间序列预测问题,可通过建模历史需求密度数据对未来某一时段或多个时段的需求密度进行预测。准确的网约车需求密度预测可辅助相关部门采取针对性措施,有效缓解交通压力、提高路网运行能力和效率。车辆调度是构建智慧城市高效交通运输系统的关键。近年来,由于快速增长的交通量与低效的调度操作之间的矛盾,传统的交通系统在为人类提供舒适便利出行条件的同时,也面临着严峻挑战。由于当前智能交通系统高度依赖未来需求模式,因此需要了解并预测不同时间间隔内不同空间区域的车辆需求分布,各平台再根据准确预测的车辆分布信息估算网约车未来需求密度集群效益,使系统智能地进行车辆分配。基于此,运输资源的利用率将有所提高,服务效率、平台收益和乘客等待时间都可以得到改善。

1 相关工作

针对预测问题,Wu等^[1]通过分析上车点乘客订单视图、下车点乘客订单视图、兴趣点视图、空间视图和天气视图,提出一种多视角深度时空网络预测网约车需求,同时考虑了需求的非线性和随机性;Ye等^[2]提出一种基于注意力机制的长短期记忆(Long Short-term Memory, LSTM)组合预测模型,考虑了短时需求密度预测中孤立点对模型效果的影响,可捕获众多可区分的时间特征,有效抵消噪声和离群点。针对调度问题,Abid等^[3]考虑了客户偏好,将车辆调度系统建模为多准则决策问题,不仅兼顾了乘客与车辆的距离,而且将客户偏好纳入给定的乘客出行请求中,在司机合理接单的同时满足乘客期望;Liu等^[4]提出一种情境感知出租车调度方法(Context-aware, COX),将丰富的情境信息整合到深度强化学习(Deep Reinforcement Learning, DRL)模型中,实现对出租车的高效分配。上述方法能够为网约车需求密度预测和调度提供技术支撑,然而,现有研究仍存在以下问题:①在建立预测模型时,网约车需求密度具有复杂的时空相关性,无法较为高效对这些特征进行提取;②需求密度受个别极值点影响,会使得最终结果误差较大;③神经网络深层结构易影响模型预测精度;④需求密度与天气、节假日等外部因素有关,这些因素很难被识别和捕捉;⑤住宅区、商业区、铁路和地铁站以及行政机构的建设与搬迁动态地改变了需求密度,很难找到有效的特征揭示这种变化与需求密度之间的相关性;⑥在进行调度时,缺乏对网约车下客区接单率的考虑,导致网约车因缺少后续订单而长时间空置在下车点;⑦网约车和

乘客的空间分布差距较大,存在某些地区车辆无法满足当地乘客需求的情况;⑧网约车与乘客的匹配存在随机性,很难保证网约车收益以及后续接单率,无法为网约车平台或司机个人提供合理、高效的调度决策;⑨求解数学模型时,传统遗传算法精度较低,寻优能力一般,局部搜索能力较弱,易出现早熟现象,且容易造成基因缺失。如何更为合理地结合预测模型和调度策略,使其应用于更广泛的场景,使司机获取更高收益,并提升乘客满意度,是目前亟需解决的问题。

针对上述问题和挑战,本文设计一种基于需求密度预测的网约车集约化调度方法,主要贡献包括:

(1)提出基于多维时空数据的需求密度预测问题模型和基于集群效益的车辆调度问题模型。预测模型综合考虑了时间、空间和外部特征,消除了需求密度对某一特征的过度依赖。调度模型可将网约车及时调度到具有高收益潜力的区域,兼顾了全局供需平衡与总体盈利能力。

(2)提出基于多层混合感知野的深度时空残差感知网络模型,充分考虑了需求密度的时空特性。针对网约车需求密度受个别极值点影响使得最终结果误差较大的问题,提出一种基于需求频度的划分方法,以平衡空间特征中差距较大的数据;设计了CELU网络结构捕捉不同区域间的时空依赖性,以确保模型预测精度不受神经网络深层结构的影响;引入兴趣点作为外部特征,结合气象、节假日等属性,解决了因模型影响因素考虑不周全导致预测结果不准确的问题。

(3)提出集约化调度方法,同时考虑网约车接送乘客所获利润和后续接单率两方面影响,建立调度模型,解决了因影响因素单一而难以获得高质量解的问题;设计了传感邻域,对调度范围加以限制,解决了因网约车空驶距离和乘客等待时间过长而导致调度方法效率过低的问题;为防止遗传操作造成基因缺失而无法获得最优解,提出基于匈牙利算法的去重和恢复操作,以提高解的质量;为避免遗传算法陷入局部收敛而无法得到全局最优解,设计了混合选择算子,既保存了优良个体,又给较差个体机会;设计了改进变异算子,以维持种群多样性,防止出现早熟和基因意外缺失的现象。

(4)通过在北京和海口的规模真实数据集上进行大量实验,证明本文预测模型在预测空间需求密度方面具有较高精度,并且验证了用于车辆调度的算法具有较高效率,且能获得较高收益。

2 问题定义

2.1 预测模型相关定义

定义1 空间网格^[5]:取地理区域范围的最小经度和纬度分别为 lng_s 、 lat_s ,最大经度和纬度分别为 lng_b 、 lat_b ,以步长 $o_{lng} = (lng_b - lng_s)/M$ 和 $o_{lat} = (lat_b - lat_s)/N$ 将该地理

区域等距均分为 $M \times N$ 个空间网格, 网格中的单元 $Z_{m,n}$ 表示一个区域, $m \in [1, M] \cap \mathbb{Z}, n \in [1, N] \cap \mathbb{Z}$ 。

定义2 时段: 将给定的待处理时域按照时间周期均等分割为 T 个时段, 每个时段由 t_i 表示, $i \in [1, T]$ 。

定义3 需求密度: 需求 $demand_p = (time_p, lng_{p_{start}}, lat_{p_{start}})$ 为乘客 p 的网约车出行订单, 其中 $time_p$ 为乘客需要服务的时间, $lng_{p_{start}}, lat_{p_{start}}$ 为乘客出发地的经度和纬度。若 $lng_s + o_{lng}(m-1) \leq lng_{p_{start}} \leq lng_s + o_{lng}m$ 且 $lat_s + o_{lat}(n-1) \leq lat_{p_{start}} \leq lat_s + o_{lat}n$, 则需求 $demand_p$ 属于区域 $Z_{m,n}$ 。对于每个区域 $Z_{m,n}$ 和时段 t_i , 其相应需求密度 $G_{t_i}^{m,n}$ 可以定义为区域 $Z_{m,n}$ 中在时段 t_i 范围内的需求数量。另外, 用 $X_{t_i}^{m,n}$ 表示当前时段真实的需求密度。

问题1 需求密度预测: 给定含有 κ 个数据的测试集 $\{(G_z, E_z), (\hat{G}_z, \hat{E}_z)\}_{z=1}^{\kappa}$, 将其作为输入数据映射到完成训练的预测模型中, 经预测模型处理得到预测值 \tilde{X} 。输出数据与输入数据关系见式(1)。

$$\tilde{X} = \mathfrak{h} \left(\left\{ (G_z, E_z), (\hat{G}_z, \hat{E}_z) \right\}_{z=1}^{\kappa} \right) \quad (1)$$

其中, \mathfrak{h} 为预测模型函数, $E_z = \{E_{hol}, E_{wrk}, E_{wee}^c, E_{met}, E_{poi}\}$ 。 E_{hol} 为节假日向量; E_{wrk} 为工作日向量; E_{wee}^c 为星期向量, c 为星期数序号, $E_{wee}^c = \{E_{wee}^1, E_{wee}^2, \dots, E_{wee}^7\}$ 包含一周的7天; E_{met} 为气象向量; E_{poi} 为兴趣点向量。

2.2 调度算法相关定义

定义4 网约车集: 网约车集 $R = \{r_1, r_2, \dots, r_s\}$, r_s 代表一个网约车个体, $r_s = \langle r_{id}, r_{departure} \rangle$, r_{id} 为网约车标识, $r_{id} = s, s \in [1, S] \cap \mathbb{Z}$, S 为网约车总数, $r_{departure} = (lng_{r_{departure}}, lat_{r_{departure}})$ 为网约车出发地, $lng_{r_{departure}}, lat_{r_{departure}}$ 为网约车出发地经度和纬度。

定义5 乘客集: 乘客集 $P = \{p_1, p_2, \dots, p_d\}$, p_d 代表一个乘客个体, $p_d = \langle p_{id}, p_{start}, p_{arrive} \rangle$, p_{id} 为网约车标识, $p_{id} = d, d \in [1, D] \cap \mathbb{Z}$, D 为乘客总数, $p_{start} = (lng_{p_{start}}, lat_{p_{start}})$ 为乘客出发地, $p_{arrive} = (lng_{p_{arrive}}, lat_{p_{arrive}})$ 为乘客目的地, $lng_{p_{arrive}}, lat_{p_{arrive}}$ 为乘客目的地经度和纬度。

定义6 经济效益: 定义 r_s 与 p_d 匹配后, p_d 在 t_i 时段到达 p_{arrive} 时应收车费与全程油耗之差为经济效益 $I_{s,d}$, 表达式见式(2)。

$$I_{s,d} = \left(\theta + \frac{\phi}{v_{mean}^{t_i}} - \psi \right) \cdot B_{s,d} - \psi \cdot A_{s,d} \quad (2)$$

其中, θ 为网约车里程费; ϕ 为网约车时长费; $v_{mean}^{t_i}$ 为 t_i 时段网约车平均速度, 其受不同时段路况的影响而不断变化; ψ 为油耗指数, 其值随着平均车速而变化^[6]; $B_{s,d}$ 为乘客

出发地与乘客目的地的测地距离。

定义7 需求密度集群效益: 考虑下客区需求密度可避免网约车因缺少后续订单而长时间空置。因此, 定义需求密度集群效益 $C_{s,d}$ 表达式见式(3)。

$$C_{s,d} = O_{s,d} + O_{s,d}^{\theta_{r,s}} - O_{avg} \quad (3)$$

其中, $O_{s,d}$ 为预测模型预测的下客区需求密度; $O_{s,d}^{\theta_{r,s}} = \sum_{\gamma=1}^8 O_{s,d}^{\theta_{r,\gamma}}$ 为预测模型预测的下客区一阶传感邻域需求密度; O_{avg} 为 t_i 时段全局需求密度均值, 用来修正不同时段全局需求密度的差异对当前调度的影响, 以保证调度策略的公平性。

问题2 集约化调度: 优化目标为求得整体经济效益 $\sum_{s,d} I_{s,d}$ 和整体需求密度集群效益 $\sum_{s,d} C_{s,d}$ 最大的匹配方案。因此, 将调度数学模型的目标函数设置为最大化全局 $I_{s,d}$ 和 $C_{s,d}$ 标准化加权和, 即:

$$f = \max \sum_{s,d} \varpi L_{s,d} + \tau V_{s,d} \quad (4)$$

$$L_{s,d} = \frac{I_{s,d} - I_{min}}{I_{max} - I_{min}} \quad (5)$$

$$V_{s,d} = \frac{C_{s,d} - C_{min}}{C_{max} - C_{min}} \quad (6)$$

$$p_d \in \{Z_r, \vartheta_r, 8\alpha\} \quad (7)$$

$$\varpi + \tau = 1 \quad (8)$$

$$J_{s,d}, A_{s,d}, O_{s,d}, O_{s,d}^{\theta_{r,s}}, O_{avg} \geq 0 \quad (9)$$

其中, 式(4)为目标函数, 式(5)和式(6)分别表示 $I_{s,d}$ 、 $C_{s,d}$ 的标准化, I_{max} 、 I_{min} 分别为经济效益最大值和最小值, C_{max} 、 C_{min} 分别为需求密度集群效益最大值和最小值; 式(7)为 r_s 匹配范围约束; 式(8)中的 ϖ 、 τ 为 $I_{s,d}$ 和 $C_{s,d}$ 权重约束; 式(9)为非负约束。

3 基于多层混合感知野的深度时空残差感知网络

本文提出的基于多层混合感知野的深度时空残差感知网络 (Deep Spatio-temporal Residual Perception Network, DST-PN) 由两个模块组成, 分别是时空特征模块和外部特征模块, 整体框架如图1所示。时空特征模块按照前 l_{cdo} 个时段、前 l_{day} 天、前 l_{week} 周和前 l_{mon} 月共4种时间尺度划分为邻近层 $G_{l_{cdo}} = [G_{t_i - l_{cdo}}, G_{t_i - (l_{cdo} - 1)}, \dots, G_{t_i - 1}]$ 、周期层 $G_{l_{day}} = [G_{t_i - l_{day} \cdot T}, G_{t_i - (l_{day} - 1) \cdot T}, \dots, G_{t_i - T}]$ 、近趋势层 $G_{l_{week}} = [G_{t_i - l_{week} \cdot 7T}, G_{t_i - (l_{week} - 1) \cdot 7T}, \dots, G_{t_i - 7T}]$ 和远趋势层 $G_{l_{mon}} = [G_{t_i - l_{mon} \cdot 28T}, G_{t_i - (l_{mon} - 1) \cdot 28T}, \dots, G_{t_i - 28T}]$, 其中这4个时间层对应4组相互独立的神经网络, 每个子网络又依据需求频度边界划分成3个感知区域, 分别对其进行建模。外部特征模块提取 E_{met} 、 E_{hol} 、 E_{poi} 、 E_{wrk} 、 E_{wee}^c , 按照其特性分别通过

卷积层和全连接层进行特征提取与优化。最后综合两个模块的输出,得到最终预测结果。

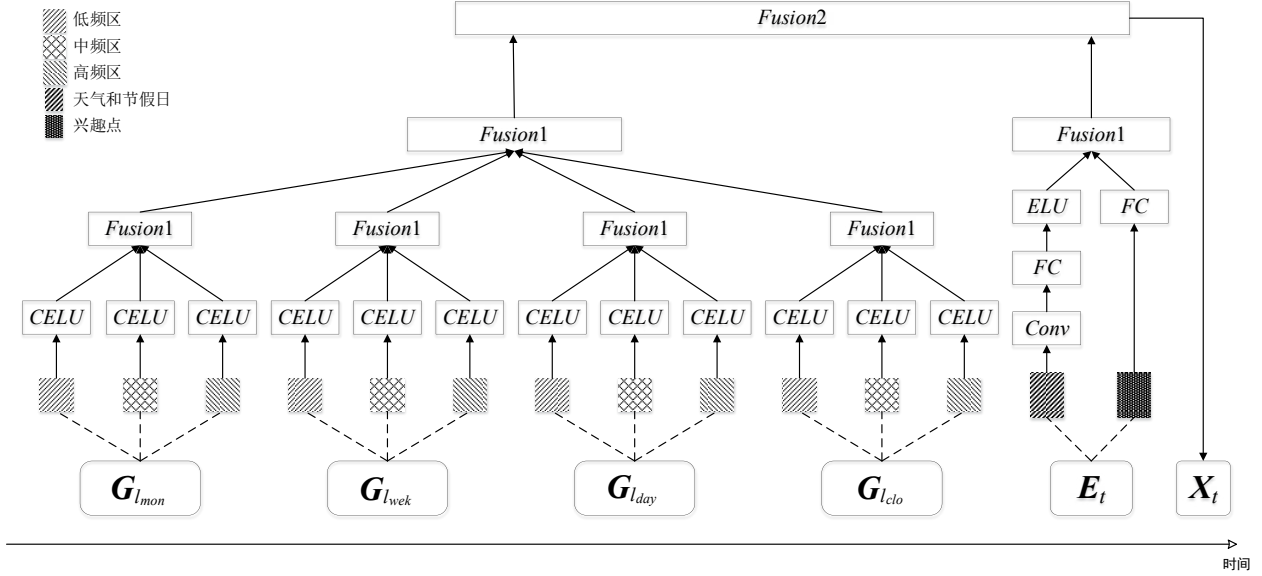


Fig. 1 Structure of DST-PN model

图1 DST-PN模型结构

3.1 时空特征模块

为较为准确地反映需求密度局部关联和全局关联的空间依赖性,并避免其陷入局部极值,设计如下需求频度划分方法,将需求密度依照高中低3种需求频度进行筛选判断,见式(10)。

$$G_{l,\lambda}^{(0)} = \begin{cases} G_{l,0}^{(0)}, flow \leq \varepsilon_1 \\ G_{l,1}^{(0)}, \varepsilon_1 < flow < \varepsilon_2 \\ G_{l,2}^{(0)}, flow \geq \varepsilon_2 \end{cases} \quad (10)$$

其中, $l_\lambda \in \{l_{clo}, l_{day}, l_{wek}, l_{mon}\}$, (0)表示当前未经过任何网络层, $flow$ 为需求密度, $\varepsilon_1, \varepsilon_2$ 为动态确定的需求密度边界。

针对4个时间层,充分考虑不同时段内附近区域和较远区域的时空依赖性^[7],设计出由前后两层卷积和中间L层残差单元堆叠而成的CELU网络结构捕获全范围时空相关性,从而保证了模型在网络深度加深情况下的拟合性能。残差单元结构如图2所示。激活函数ELU可为模型增加非线性因素,其与现有模型中经常采用的ReLU不同,其所含的负指数项可防止静默神经元出现,从而防止导数收敛为0,可达到提高效率、提升模型预测性能的效果。CELU网络结构计算公式见式(11)。

$$G_{l,\lambda}^{K+2} = \rho \left(W_{l,\lambda}^{K+2} * \left(W_{l,\lambda}^{(1)} * G_{l,\lambda}^{(0)} + b_{l,\lambda}^{(1)} + \xi \left(G_{l,\lambda}^K; \mu_{l,\lambda}^K \right) \right) + b_{l,\lambda}^{K+2} \right) \quad (11)$$

其中,*为卷积算子, ξ 为残差函数, $G_{l,\lambda}^K$ 为堆叠K层残差单元的变量, $\mu_{l,\lambda}^K, W_{l,\lambda}^{(1)}, W_{l,\lambda}^{K+2}, b_{l,\lambda}^{(1)}, b_{l,\lambda}^{K+2}$ 为需学习的参数。

3.2 外部特征模块

现有研究大多考虑气象和节假日等外部特征,而忽略

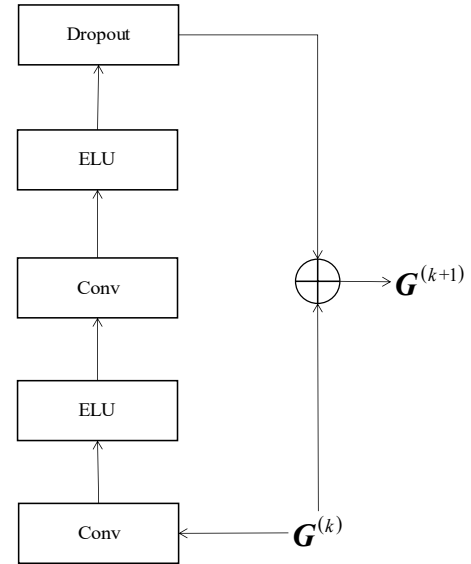


Fig. 2 Structure of the residual unit

图2 残差单元结构

了兴趣点对需求密度的影响。因此,本文在考虑 $E_{met}, E_{wee}^c, E_{wrk}, E_{hol}$ 的基础上,添加了 E_{poi} 特征,为需求密度预测的空间相关性建模引入一个新的维度,从而完善外部特征的影响。根据外部特征的不同内在特点,采用全连接层对不同时段特征向量 $E_{met}, E_{wee}^c, E_{wrk}, E_{hol}$ 拼接成的矩阵 E_{hmv} 进行特征提取,记为 E_{st} ,具体计算公式见式(12)。因 E_{poi} 较其他外部特征的数据量更庞大,因此在对其进行特征提取时,首先采用卷积神经网络以减小训练过程中的参数训练数量,再通过全连接层规范其规格。

$$E_{st} = \rho \left(W_{st} \times E_{hmv} + b_{st} \right) \quad (12)$$

其中, \times 表示矩阵乘法, W_{st}, b_{st} 为可学习参数。

3.3 特征模块融合

需求密度在日常生活中呈周期性变化,而在节日期间会出现激增现象。只对原图像卷积可能会放大值域范围较大但重要性较低的点,而忽略值域范围较小但重要性较高的点。因此,采用基于参数矩阵的融合方式,实现从4个不同的时间层上挖掘丰富的时空信息^[8]。时空模块输出 $X_{t,TC}$ 计算公式如下:

$$X_{t,TC} = \sum_{l_s} W_{l_s} \otimes \left(\sum_{j=0}^2 W_{l_s,j} \otimes G_{l_s,j}^{K+2} \right) + b_{t,TC} \quad (13)$$

其中, \otimes 为 Hadamard 乘积, $W_{l_s,j}$ 为3个需求频度感知域的权重矩阵, W_{l_s} 为4个时间层的权重矩阵, $b_{t,TC}$ 为偏置。其均为网络训练中需学习的参数。

外部特征模块采用与时空特征模块相同的方法进行融合,融合后的变量由 X_{Ex} 表示。考虑到外部特征可为需求密度预测提供全局空间表示,因此将时空特征模块与外部特征模块融合,使二者进行协同训练以获取不同领域的的数据特点。由于不同外部因素可能以不同方式改变需求密度,因此利用两种不同的聚合方法,采用求和融合方法 ($X_{t,TC} + X_{Ex}$) 表征逐步变化的需求密度,而对于突然的变化,则应用基于门控机制的融合方法 ($\sigma(X_{Ex}) \otimes X_{t,TC}$)^[9]。最后通过 Tanh 函数将聚合映射到 $[-1, 1]$, 以便在训练过程中得到更快的收敛速度。

4 集约化调度方法

为减少网约车空驶距离和乘客等待时间,设计传感邻域,以网约车 r_s 所在区域 Z_{r_s} 为中心,定义其外围 8α 个区域为 α 阶传感邻域 $\mathcal{D}_{r_s,8\alpha}$, $\alpha \in \mathbb{Z}$ 。规定 r_s 只在其所在区域和传感邻域内进行调度,该调度范围决定了网约车与乘客的匹配距离 $A_{s,d}$ 。乘客所在传感邻域阶数越低,与网约车距离则越近, $A_{s,d}$ 也即越低。反之, $A_{s,d}$ 越高。因此,设定该值随传感邻域阶数呈线性增长,见式(14)。

$$A_{s,d} = \begin{cases} J_{s,d} & \alpha = 1 \\ 2^{\alpha-1} \cdot J_{s,d} & \alpha > 1 \end{cases} \quad (14)$$

其中, $J_{s,d}$ 为网约车当前位置与乘客出发地的测地距离。

由此引出基因组与基因组个体概念,将网约车所在区域所有网约车和乘客及该区域对应的一阶传感邻域内的乘客定义为一个基因组,一个基因组内网约车和乘客数量为 D_{gen} 与 S_{gen} 。在基因组内执行编码和初始化操作形成的染色体称为基因组个体。

4.1 生成效用矩阵

$L_{s,d}$ 与 $V_{s,d}$ 加权求和为:

$$U_{s,d} = \omega L_{s,d} + \tau V_{s,d} \quad (15)$$

建立效用矩阵 U 为:

$$U = \begin{matrix} & P_1 & P_2 & \cdots & P_d \\ \begin{matrix} r_1 \\ r_2 \\ \vdots \\ r_s \end{matrix} & \begin{bmatrix} U_{11} & U_{12} & \cdots & U_{1d} \\ U_{21} & U_{22} & \cdots & U_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ U_{s1} & U_{s2} & \cdots & U_{sd} \end{bmatrix} \end{matrix} \quad (16)$$

其中, $s = S, d = D$ 。

4.2 算法描述

本文所述网约车和乘客匹配属于即时匹配问题,对时间敏感性高,且数据规模大,采用数学解法或精确算法求解效率低,无法在短时间内求解大规模算例,因此选用启发式算法。启发式算法虽然无法保证得到最优解,但可在较短的求解时间内获得高质量的解,具有通用性、稳定性以及较快的收敛性。

将种群分为 Y 个基因组(基因组数量 Y 将根据网约车位置动态决定),各组分别形成 g 个基因组个体, g 值见式(17),从各基因组中分别随机选取基因组个体拼合形成 Q 条带有基因组标记的完整染色体。对 Q 条染色体进行去重和遗传操作,经历 v 次迭代,选出最优染色体即为匹配结果。

$$g = \begin{cases} \max\{D_{gen}, S_{gen}\} & \min\{D_{gen}, S_{gen}\} < 10 \\ 10 & \min\{D_{gen}, S_{gen}\} \geq 10 \end{cases} \quad (17)$$

(1) 编码。采用实数编码方式,即 $[x_1, y_1, x_2, y_2, \dots, x_h, y_h]$ 。其中, x_a 为网约车 r_{id} , y_a 为乘客 p_{id} , $[x_a, y_a]$ 表示网约车与乘客匹配,也表示染色体内的一个基因。

(2) 初始化。在基因组内,当 $D_{gen} - S_{gen} > 0$ 时,固定 r_{id} , 不重复地随机生成 p_{id} ; 当 $D_{gen} - S_{gen} < 0$ 时,随机选取数量为 D_{gen} 的网约车,固定其对应的 r_{id} , 不重复地随机生成 p_{id} 。

(3) 基于匈牙利算法的去重与修复。各基因组内个体随机拼合后会导致染色体中基因出现重复,因此设计新的去重方法。在去重过程中,从基因组编号差集中随机选取元素替换原重复元素,无法得到最优匹配结果,而匈牙利算法在求解最值问题上具有优势。但因其在处理大规模数据时效率低且计算时间长,不适合对本文数据直接求解,因此该去重方法针对基因组中小规模数据引入匈牙利算法,将基因组与匈牙利算法结合可以有效规避上述缺陷。具体操作见图3。

图3中加粗数字表示网约车标识,加粗+下划线数字表示乘客标识。因上述去重方法会导致基因缺失而无法得到最优解,故设计修复方法对染色体进行修复。首先获取网约车、乘客总编号集合与染色体中网约车、乘客编号的差集 R_{DS} 和 P_{DS} , 其次将 R_{DS} 和 P_{DS} 汇总并形成二分图,采用匈牙利算法得到最优匹配基因,最后将其添加到当前染色体尾部。

(4) 适应度函数。目标函数为求最大值,故个体适应度定义为:

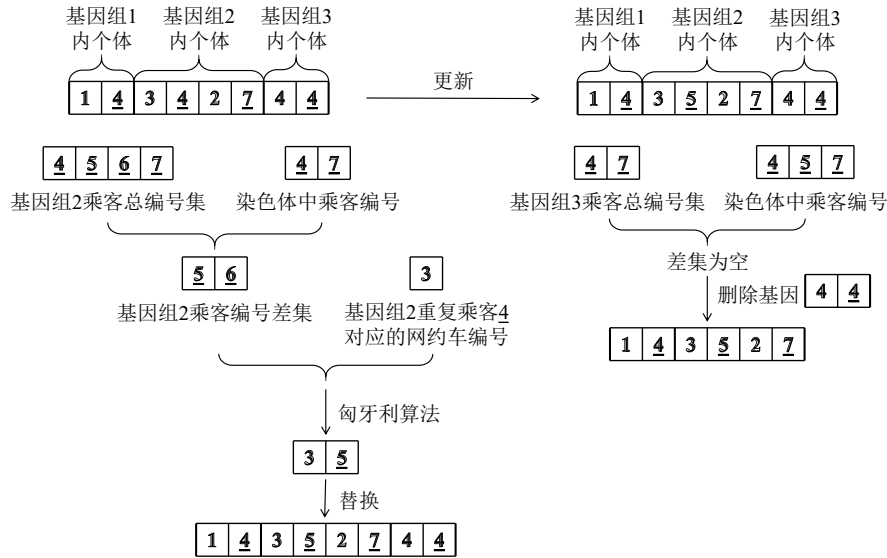


Fig. 3 Specific operation of the deduplication method

图3 去重方法具体操作

$$f' = \sum_{a=1}^h U_{x_a, y_a} \quad (18)$$

(5)遗传操作。

混合选择:一个种群中的最优个体并不一定在全局最优附近。因此,为避免陷入局部收敛而无法得到全局最优,设计最优个体选择和轮盘赌选择相结合的混合选择算子,将适应度排在前 $\lfloor \frac{Q}{8} \rfloor$ 的个体直接复制到下一代而不进行交叉,剩余 $Q - \lfloor \frac{Q}{8} \rfloor$ 个个体采用轮盘赌选择法选择个体进行交叉,在保存优良个体的同时给相对不好的个体机会。

交叉:图4显示了不同交叉概率对解的影响,曲线整体呈上升趋势。当交叉概率等于0.9时,得到最高平均解,可知交叉概率越大,染色体之间越容易进行交叉操作,从而有机会获得更多优良个体,产生更高的平均解。因此,设置交叉概率 $F=0.9$ 。当满足交叉概率时,进行部分匹配交叉,随机在乘客编码位后产生两个交叉点。定义两点间区域为匹配区域,使用位置交换操作交换两个染色体的匹配区域。

变异:变异概率会对算法的收敛性以及最终的最优解取值产生较大影响。当变异概率较小时,解的稳定性较高,但很容易陷入局部最优解,并且难以跳出局部最优解的区间,而变异概率较大时,整个搜索过程又会退化为一个随机搜索过程。图5显示了不同变异概率对解的影响,从图中可以看到变异概率过大或过小都不会得到最优解。因此,结合实验结果,选择0.2的变异概率能够获得较优解,从而保证算法效果。

设置变异概率 $H=0.2$,在满足变异概率时进行变异。为增强遗传算法的局部随机搜索能力,加速向最优解收敛,维持种群多样性,防止出现未成熟收敛现象,避免因意

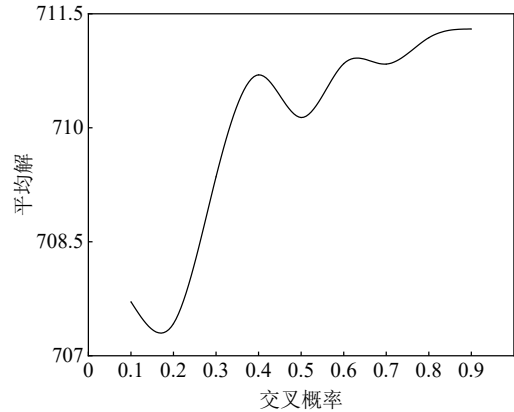


Fig. 4 Effect of crossover probability

图4 交叉概率的影响

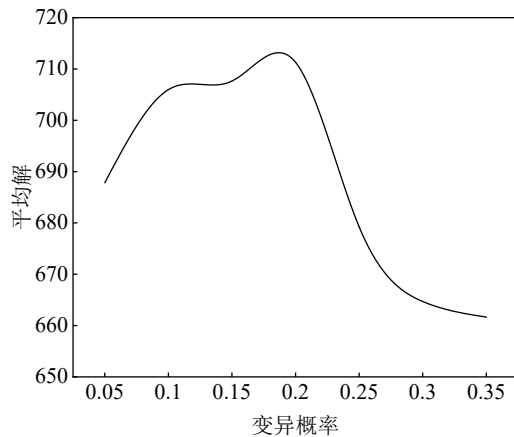


Fig. 5 Effect of mutation probability

图5 变异概率的影响

外导致基因缺失而无法获得最优值的问题,设计改进变异算子,流程如下:

步骤1:遍历网约车总编号集,若其中有未匹配到乘客编号的网约车编号,将其汇总为集合 R_{DS} 。针对 R_{DS} 中所有

网约车编号, 分别获取其所在基因组中未匹配网约车的乘客集 P_{ds} 。若 $P_{ds} \neq \emptyset$, 则需在其中找到与网约车编号匹配后使染色体适应度值变大的乘客编号, 将该网约车和乘客编号对插入到染色体中。如果没有可使染色体适应度值变大的乘客编号或 $P_{ds} = \emptyset$ 或 $R_{DS} = \emptyset$, 直接执行步骤 2。

步骤 2: 若 $R_{DS} \neq \emptyset$ 或乘客总编号集与染色体中乘客集的差集 $P_{DS} \neq \emptyset$, 对染色体中网约车和乘客依次执行图 6 中的操作。否则, 直接执行步骤 3。

步骤 3: 从染色体网约车集中第一个元素开始, 找到其后与之基因组相同且调换位置后可使染色体适应度值增大的元素, 更新染色体, 直到将网约车集中所有元素遍历完毕。

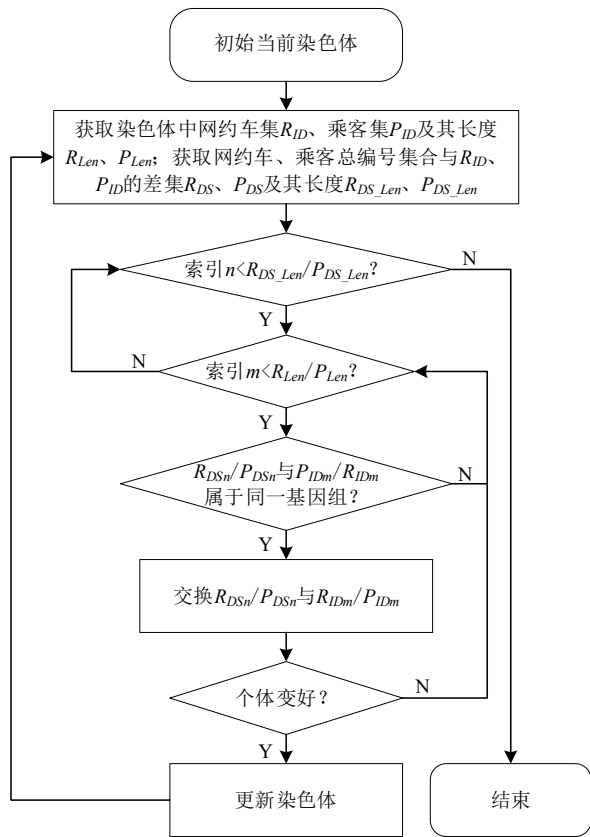


Fig. 6 Flow of mutation operation step 2
图 6 变异操作步骤 2 流程

5 实验与分析

基于 Tensorflow 开源架构实现 DST-PN 模型和 ISA 算法。实验数据源基于滴滴盖亚数据开放计划, 采用海口市和北京市城市区域网约车订单数据, POI 数据通过百度地图应用程序编程接口 (Application Programming Interface, API) 获取, 气象数据从中国气象数据网和天气网获取。

5.1 预测模型评估

预测模型以最小化预测矩阵与真实矩阵之间的均方误差为训练目标, 获得预测模型中全部权值 W 和偏差 b 的最优取值。采用反向传播算法进行模型训练, 并配合 Adam 优化算法提高训练效率^[10]。

预测模型中的可学习参数使用 Keras 中的默认参数, 通过均匀分布对其进行初始化, 应用早停法选择最优模型参数, 选择 90% 的数据作为训练集, 10% 的数据作为测试集, 并设置学习率为 0.000 2。实验数据参数如表 1 所示。

Table 1 Experimental data parameters

表 1 实验数据参数

参数	数值
区域网格数 (M, N)	(32, 32), (16, 16)
时隙/min	30
卷积核数目	64
卷积核大小	(3, 3)
残差单元	8
批尺寸	32
l_{clo}	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
l_{day}	1, 2, 3, 4
l_{week}	1, 2, 3
l_{mon}	1, 2

为验证模型预测结果, 本文将 DST-PN 与其他几种基准算法进行对比, 各算法参数设置标准为最优预测精度。由于文献[11]–[16]中模型与本文中使用的特征及数据存在差异, 为确保对比的公平性, 对文献中模型进行了部分修改以保证与本文算法的对比条件一致。本文选用均方根误差 (RMSE)、平均绝对误差 (MAE)、决定系数 (R^2) 以及解释方差 (EVS) 作为模型的评价指标。对比结果见表 2。

Table 2 Comparison experiment results

表 2 对比实验结果

实验模型	北京数据集				海口数据集			
	EVS	R^2	MAE	RMSE	EVS	R^2	MAE	RMSE
LR	0.852 8	0.787 4	30.602 0	54.078 6	0.853 6	0.843 4	4.744 3	9.523 8
DTR	0.898 7	0.834 0	24.316 5	47.783 1	0.930 0	0.923 0	2.975 2	6.678 0
KNN	0.925 8	0.880 1	20.424 3	40.614 4	0.943 5	0.938 7	2.648 4	5.960 0
LSTM ^[11]	0.938 7	0.928 8	18.931 5	31.306 4	0.877 3	0.871 9	4.290 9	8.615 3
GRU ^[12]	0.962 6	0.956 0	15.055 8	24.608 2	0.920 0	0.918 1	3.317 1	6.886 9
RNN ^[13]	0.969 5	0.964 2	13.389 2	22.186 6	0.936 3	0.934 2	3.164 0	6.174 2
CNN ^[14]	0.972 9	0.965 3	12.907 1	21.847 4	0.955 3	0.953 6	2.631 2	5.185 2
DeepST ^[15]	0.973 5	0.966 6	12.672 4	21.426 5	0.953 8	0.951 5	2.735 2	5.303 0
ConvLSTM ^[16]	0.973 6	0.969 3	11.633 1	20.068 1	0.934 7	0.932 2	3.092 4	6.267 3

如表2所示,与其他基准方法相比,本文模型在上述4种指标中均表现最佳。相较于基于深度学习的LSTM模型,本文模型效果有所提升,充分说明了空间依赖性对需求预测的重要程度。

在时空方向上,模型 R^2 高于ConvLSTM和CNN, $RMSE$ 与 MAE 均大幅下降,说明本文模型通过划分需求密度感知域,不但可保留对全局需求密度信息的捕获能力,而且可在此基础上更加精细地捕获不同的需求密度范围信息,以获得更优的预测效果。本文模型相比于采用激活函数ReLU的DeepST模型效果略微提升,说明本文使用的激活函数ELU可使模型拥有更强的学习能力,可达到提升模型预测性能的目的。LR在所有模型中表现最差,无法准确预测需求密度。

在实际应用中,预测未来多时段需求密度的重要性高于单个时段。网约车平台可根据未来多个时段的需求密度,提前制定出合理的车辆调度方案。如图7所示,说明本文模型在单步预测中的表现优于多步预测,但相较于其他对比模型,本文模型在多步预测中仍表现最优。同时, R^2 下降缓慢,表明该模型具有适应多步预测任务的能力。

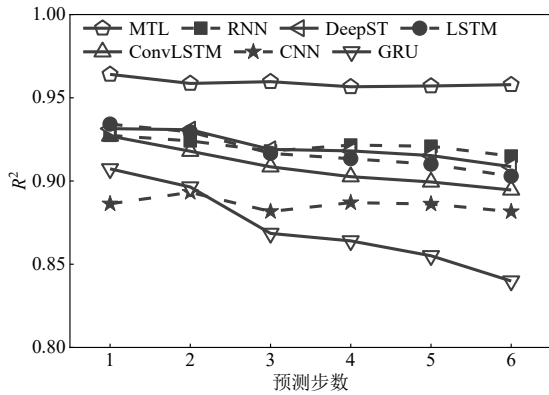


Fig. 7 Sensitivity analysis of multi-step prediction
图7 多步预测敏感性分析

5.2 调度策略评估

本文模拟了10个实例,各实例参数见表3。对10个实例执行ISA和Kuhn-Munkres(KM)算法,并将ISA的解、平均运行时间与KM算法最优解、平均运行时间作比较,得到最大质量比、最小质量比、平均质量比和时间比。实验结果见表4。

Table 3 Instance parameters

参数	默认值	取值范围
网约车数	1 000	500, 600, 700, 800, 900, 1 000, 1 100, 1 200, 1 300, 1 400
乘客数	1 000	500, 600, 700, 800, 900, 1 000, 1 100, 1 200, 1 300, 1 400
地理区域经度	/	[110.144, 110.696]
地理区域纬度	/	[19.685, 20.131]
种群规模 Q	40	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
ν	90	10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150
ϖ	0.5	/
τ	0.5	/
$\theta/\text{元}/\text{km}$	1.8	/
$\phi/\text{元}/\text{min}$	0.8	/

如表4所示,无论数据规模如何增大,质量比均可保持在99%以上,说明ISA算法可获得高质量的解,证明了其在处理较大规模数据时的可靠性。时间比随乘客规模的增加逐渐减小,甚至低于10%,表明ISA的时间效率优于KM算法,证明了该算法在处理较大规模数据时的高效性。KM算法的平均时间随乘客规模的增加迅速上升,ISA的平均时间随乘客规模的增加缓慢上升,且显著低于KM算法,表明ISA在处理较大规模数据时相对稳定,在运行时间上具有优势。综上,引入传感邻域和在去重部分结合匈牙利

Table 4 Experimental results

表4 实验结果

实例	网约车数 (乘客数)	ISA			平均 时间/s	KM算法		最大比/%	最小比/%	平均比/%	时间比/%
		最大解	最小解	平均解		最优解	平均 时间/s				
1	500	335.6	335.4	335.5	9.47	335.9	45.15	99.91	99.85	99.88	20.97
2	600	399.9	399.7	399.8	12.74	400.3	76.53	99.90	99.85	99.88	16.65
3	700	469.7	469.4	469.6	16.65	470.2	119.83	99.89	99.83	99.87	13.89
4	800	544.8	544.4	544.6	21.10	545.4	176.97	99.89	99.82	99.85	11.92
5	900	625.4	624.9	625.2	26.10	626.2	249.87	99.87	99.79	99.84	10.44
6	1 000	711.3	710.7	711.0	31.65	712.2	340.46	99.87	99.79	99.83	9.30
7	1 100	802.6	802.0	802.3	37.75	803.6	450.64	99.88	99.80	99.84	8.38
8	1 200	877.8	877.1	877.5	44.38	878.9	582.33	99.87	99.80	99.84	7.62
9	1 300	937.6	936.8	937.2	51.57	938.8	737.46	99.87	99.79	99.83	6.99
10	1 400	1 012.3	1 011.4	1 011.9	59.31	1 013.6	917.94	99.87	99.78	99.83	6.46

算法的集约化调度方法的求解质量和求解效率都较为优异, 可达到提高网约车收益的效果, 且在上述大量不同规模的算例上验证表明其具有较好的鲁棒性。

迭代次数决定了算法何时终止以及所求解何时稳定。在其他参数保持不变的情况下, 在不同迭代次数下执行实例 1。图 8 展示了迭代次数对 ISA 平均解的影响, 图中平均解随迭代次数逐渐增大, 最后基本趋于稳定, 表明该算法收敛速度较快, 可在短时间内获得稳定的最优解。

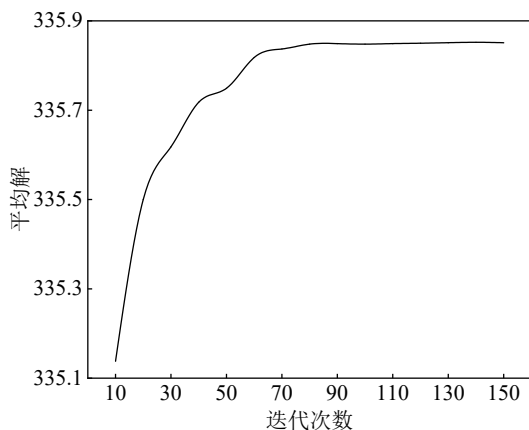


Fig. 8 Influence of iteration times on solution

图 8 迭代次数对解的影响

为证明引入传感邻域和在去重部分结合匈牙利算法能够获得更高质量的解, 分别对其进行敏感度分析。图 9 展示了在去重部分结合匈牙利算法对平均解的影响, 结合匈牙利算法得到的平均解比未结合匈牙利算法的平均解提高了 3.3%, 证明对每个基因组中的小规模数据应用匈牙利算法去重可达到提高网约车收益的效果。

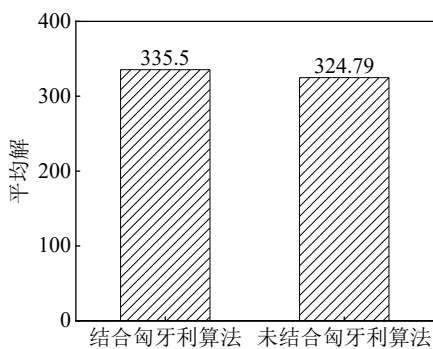


Fig. 9 Influence of Hungarian algorithm on solution

图 9 匈牙利算法对解的影响

图 10 展示了引入传感邻域对平均解的影响, 利用传感邻域进行分组, 平均解可提高 3.33%。该结果证明了传感邻域的有效性。

4 结语

本文构建基于多层混合感知野的深度时空残差感知

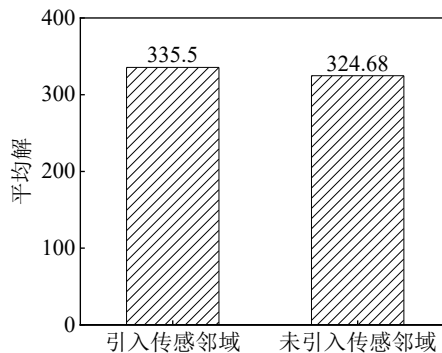


Fig. 10 Influence of sensing neighborhood on solution

图 10 传感邻域对解的影响

网络模型 (DST-PN) 对网约车需求密度进行预测, 设计了基于需求频度的划分方法和 CELU 网络结构, 引入兴趣点作为外部特征, 并结合了气象、节假日等属性; 提出集约化调度方法, 设计传感邻域, 并综合考虑网约车接送乘客所获利润和后续接单率两方面影响, 建立数学模型, 在去重和恢复操作中结合匈牙利算法对选择和变异算子进行改进, 设计改进遗传算法用于求解数学模型, 最终实现高收益调配。实验结果表明, DST-PN 模型的预测结果优于对比实验中的其他算法, 相较于基于深度学习的 LSTM 模型提高了 8%, R^2 较对比模型中表现最优的 ConvLSTM 提高了 0.2%, $RMSE$ 与 MAE 分别降低了 6% 和 11%, 说明该模型在需求密度预测方面具有优势。ISA 在以相同效用矩阵作为输入条件时, 求解质量可达到 KM 算法最优解的 99% 以上。在处理大规模数据时, 求解时间可达到 KM 算法的 10% 以下。说明 ISA 可在较短时间内得到高质量解, 证明了 ISA 在调度问题上的可靠性和高效性。未来研究将针对预测模型引入更多外部因素, 如交通事故数量、空气污染指数等, 以提升模型的预测精度和鲁棒性。针对调度问题, 将结合 Hopcroft - Karp 算法降低时间复杂度、加快匹配时间, 以及时、高效地完成车辆调度。

参考文献:

- [1] WU Y H, ZHANG H Y, LI C, et al. Urban ride-hailing demand prediction with multi-view information fusion deep learning framework[J]. Applied Intelligence, 2022, 53(8): 879-897.
- [2] YE X F, YE Q M, YAN X C, et al. Demand forecasting of online car-hailing with combining LSTM + attention approaches [J]. Electronics, 2021, 10(20): 2480.
- [3] ABID A, NAWAZNAEEM A, SHOAB F M, et al. Taxi dispatch optimization in smart cities using topsis[J]. Security and Communication Networks, 2022, 72(1): 33-36.
- [4] LIU Z D, LI J Z, WU K S. Context-aware taxi dispatching at city-scale using deep reinforcement learning [J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 99: 1-14.
- [5] LIAO H R, YANG Y. Predicting citywide traffic flow using attention-based spatial-temporal neural network[J]. Application Research of Com-

- puters, 2021, 38(10): 2935-2940.
- 廖挥若, 杨燕. 基于注意力的时空神经网络城市区域交通流量预测[J]. 计算机应用研究, 2021, 38(10): 2935-2940.
- [6] XU H L, FANG L Y, ZHANG Z D, et al. Research on the influence of WLTC standard speed curve on vehicle fuel consumption [J]. Shanghai Auto, 2023(3): 29-34.
- 徐红林, 方卢耀, 张振栋, 等. WLTC标准车速曲线对车辆油耗影响的研究[J]. 上海汽车, 2023(3): 29-34.
- [7] AHMAD A, YANMIN Z, MUHAMMAD Z. Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction[J]. Neural Networks, 2022, 145: 233-247.
- [8] LIU S Z, QIN Y J, WANG C X, et al. Traffic flow prediction algorithm based on deep residual long short-term memory network [J]. Journal of Computer Applications, 2021, 41(6): 1566-1572.
- 刘世泽, 秦艳君, 王晨星, 等. 基于深度残差长短记忆网络交通流量预测算法[J]. 计算机应用, 2021, 41(6): 1566-1572.
- [9] SUN J K, ZHANG J B, LI Q F, et al. Predicting citywide crowd flows in irregular regions using multi-view graph convolutional networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34: 2348-2359.
- [10] KANAPARTHI T, RAMESH S, YARRAB-OTHU S R. K-means cluster-based interference alignment with adam optimizer in convolutional neural networks [J]. International Journal of Information Security and Privacy, 2022, 16(2): 1-18.
- [11] ZHENG J H, HUANG M F. Traffic flow forecast through time series analysis based on deep learning [J]. IEEE Access, 2020, 8: 2562-2570.
- [12] PENG T X, HAN Y, WANG C, et al. Hybrid deep learning model for short-term subway passenger flow prediction [J]. Computer Engineering, 2022, 48(5): 297-305.
- 彭桐歆, 韩勇, 王程, 等. 面向短时地铁客流量预测的混合深度学习模型[J]. 计算机工程, 2022, 48(5): 297-305.
- [13] BOUARARA A H. Recurrent neural network (RNN) to analyse mental behaviour in social media [J]. International Journal of Software Science and Computational Intelligence, 2021, 13(3): 1-11.
- [14] KISHORE K R, RAO K S. A novel approach to unsupervised pattern discovery in speech using convolutional neural network [J]. Computer Speech & Language, 2021, 71(1): 101-259.
- [15] ZHANG J B, ZHENG Y, QI D K, et al. DNN-based prediction model for spatio temporal data [C]// California: The 24th ACM SIGSPATIAL International Conference, 2016.
- [16] LU X J, MA C X, QIAO Y H. Short-term demand forecasting for online car-hailing using ConvLSTM networks [J]. Physica A: Statistical Mechanics and its Applications, 2021, 570: 125838.
- [17] CHEN J Y, YUAN P C. Ride-sharing scheduling optimization model considering passenger risk level [J]. Software Guide, 2023, 22(2): 108-114.
- 陈静怡, 袁鹏程. 考虑乘客风险等级的网约车拼车调度优化模型[J]. 软件导刊, 2023, 22(2): 108-114.
- [18] CHEN R H, FANG H L. Research on joint optimization of order allocation and vehicle routing [J]. Software Guide, 2022, 21(12): 69-75.
- 陈荣虎, 方红丽. 订单分配与车辆路径联合优化研究[J]. 软件导刊, 2022, 21(12): 69-75.
- [19] DING W J. Research on online car hailing scheduling and matching methods based on effective travel demand prediction [D]. Fuxin: Liaoning University of Engineering and Technology, 2023.
- 丁文婧. 基于有效出行需求预测的网约车调度和匹配方法研究[D]. 阜新: 辽宁工程技术大学, 2023.
- [20] LI Y J, YUAN P C. Optimization model of electric network carpooling scheduling under the prevention and control of COVID-19 [J]. Journal of East China Jiaotong University, 2022, 39(5): 52-60.
- 李咏洁, 袁鹏程. 新冠疫情防控下的电动网约车拼车调度优化模型[J]. 华东交通大学学报, 2022, 39(5): 52-60.

(责任编辑: 黄健)