

DOI:10.3969/j.issn.1003-5060.2024.09.014

# 一种基于深度学习的即时配送时间预测模型

丁翔<sup>1,2</sup>, 倪丽萍<sup>1,2</sup>, 韩露<sup>1,2</sup>

(1. 合肥工业大学管理学院, 安徽合肥 230009; 2. 合肥工业大学过程优化与智能决策教育部重点实验室, 安徽合肥 230009)

**摘要:** 为了提高即时配送服务水平, 文章针对即时配送时间预测问题, 提出一种分特征学习的预测模型。首先考虑不同因素对配送过程的影响, 使用地理哈希、图嵌入等方法对多种特征进行表示; 然后使用多头自注意力模型(multi-head self-attention, MHSA)和残差连接组合的方法学习多种特征间存在的关联关系, 同时利用卷积神经网络(convolutional neural network, CNN)对配送节点间存在的空间关系进行提取, 实现对不同特征的学习; 最后将提取的特征进行融合, 输入多层感知机实现对配送时间的预测。在真实即时配送数据集上的对比实验表明, 该文提出的预测模型能够有效学习各类特征及关联关系, 预测效果更优。

**关键词:** 即时配送; 时间预测; 多头自注意力模型(MHSA); 卷积神经网络(CNN); 深度学习

**中图分类号:** F259.22; TP183 **文献标志码:** A **文章编号:** 1003-5060(2024)09-1248-08

## Instant delivery time prediction model based on deep learning

DING Xiang<sup>1,2</sup>, NI Liping<sup>1,2</sup>, HAN Lu<sup>1,2</sup>

(1. School of Management, Hefei University of Technology, Hefei 230009, China; 2. Key Laboratory of Process Optimization and Intelligent Decision Making of Ministry of Education, Hefei University of Technology, Hefei 230009, China)

**Abstract:** Based on the goal of achieving higher quality of instant delivery services, a model for instant delivery time prediction is constructed with multi-feature learning. Firstly, the impact of different factors on the instant delivery process is fully considered, and multiple features are represented using geographic hashing and graph embedding to facilitate the input of different models subsequently. Then, the association relationships between multiple features are learned by a combination of multi-head self-attention(MHSA) and residual connections, and the spatial relationships between delivery nodes are extracted using convolutional neural network(CNN), thus achieving sufficient learning and extraction of different features. Finally, the features extracted by different modules are fused and input to the multilayer perceptron module to realize the prediction of instant delivery time. The comparative experiments on real instant delivery dataset show that the proposed prediction model can effectively learn various features and association relationships, and the prediction effect is better.

**Key words:** instant delivery; time prediction; multi-head self-attention(MHSA); convolutional neural network(CNN); deep learning

随着传统电子商务市场日趋饱和, 各大电商企业纷纷探索线上与线下相结合的新零售模式, 如饿了么、美团、盒马鲜生等新零售业近年来发展迅速<sup>[1]</sup>。由于即时配送是新零售模式的关键, 而

配送用时作为即时配送服务最直观的评价指标, 直接影响消费者的购物体验, 因此, 准确预估配送时间对于提高即时配送服务水平至关重要, 通过预测结果一方面可以提供顾客合理的等待时间,

收稿日期: 2023-04-12; 修回日期: 2023-08-08

基金项目: 国家自然科学基金资助项目(62276146)

作者简介: 丁翔(1998—), 男, 安徽长丰人, 合肥工业大学硕士生;

倪丽萍(1981—), 女, 安徽合肥人, 博士, 合肥工业大学副教授, 硕士生导师, 通信作者, E-mail: niliping@hfut.edu.cn.

另一方面可以掌握骑手行为,进而合理调度订单,提高整体配送速度。

即时配送时间预测问题是对骑手前往取货到送达整个过程用时的预测,由于骑手在路程上的所需时间占比最高,现有研究倾向于将此类问题看作行程时间预测问题,目前主要解决方法有基于路线和基于始末位置2种。基于路线的方法通过提取完整路线特征预测总通行时间,基于始末位置的方法只关注起点和终点的位置信息,不考虑中间过程<sup>[2]</sup>。2种方法在不同任务环境下均取得一定效果,但基于路线的方法需要对道路信息进行提取,面临数据稀疏的问题。基于始末位置的方法虽然降低了数据稀疏性,但对路线信息关注较少,当即时配送过程存在多个配送节点的情况时,预测精度会降低。

在具体预测模型上,针对行程时间预测问题,已有研究中的预测模型主要有统计模型、机器学习模型、深度学习模型3类<sup>[3]</sup>。统计模型主要有回归<sup>[4]</sup>、卡尔曼滤波<sup>[5]</sup>(Kalman filter, KF)等模型,这类模型有一定预测能力,但当特征较多时,学习非线性关系能力弱加上依赖人为设置的参数,造成预测性能下降。

随着机器学习的发展,机器学习模型在多特征、非线性关系上有较强的学习能力,一定程度上解决了传统统计模型存在的问题,有研究者利用支持向量回归(support vector regression, SVR)<sup>[6]</sup>、梯度提升决策树(gradient boosting decision tree, GBDT)<sup>[7-8]</sup>和XGBoost<sup>[9]</sup>等模型解决此类问题。由于行程信息不仅存在时间相关性,还存在空间关联性,基于机器学习的方法虽然能够对大量特征进行学习,但在提取空间关系上仍存在一定局限性。

近年来,由于神经网络强大的特征学习能力,行程时间预测研究逐渐趋于使用基于深度学习的模型。研究者除了利用长短时记忆模型(long short-term memory, LSTM)<sup>[10]</sup>、门控循环单元(gated recurrent unit, GRU)<sup>[11-12]</sup>等神经网络学习特征中存在的时间相关性,还使用相关模型学习空间相关性。文献[2]提出一种复合的时空注意力机制提取时空特征;文献[13]采用注意力机制和时空卷积构建一种联合时空相关的预测模型。而随着特征的增多,一些研究者使用Wide-Deep<sup>[14]</sup>或者因式分解机<sup>[12]</sup>等方法学习特征间的交互关系。

由于深度学习模型在行程时间预测上的优异

表现,研究者在行程时间预测研究的成果上结合配送任务,提出了基于深度学习的即时配送时间预测模型。文献[15]使用LSTM、Wide-Deep等模型学习不同特征;文献[16]侧重对配送过程的不同阶段进行表示,利用多层感知机(multilayer perceptron, MLP)学习全部环节的信息。虽然这些研究都较好地解决了对应的配送时间预测问题,但仍存在一定局限。在多特征学习方面,现有方法存在多项式拟合限制和可解释性差的问题,没有充分挖掘特征的深层信息。此外,已有方法未考虑节点间空间关系的影响,对特征的空间关系学习不足。

针对上述研究中存在的问题,本文提出一种基于深度学习的即时配送时间预测模型DT-NET,使用多头自注意力模型(multi-head self-attention, MHSA)学习特征交互关系,同时利用Node2vec和卷积神经网络(convolutional neural network, CNN)学习配送节点中存在的空间相关性,最后融合多类特征,实现对配送时间的准确预测。

本文的主要贡献如下:

1) 选择订单的取货点和送货点作为重要节点,用重要节点序列表示订单配送行程,并预测总体用时,不仅降低了数据量,而且保留了关键路线信息,避免了因行程复杂造成的较大误差,同时利用Node2vec和卷积神经网络学习节点间的空间关系,挖掘空间关系对整体用时的影响。

2) 使用多头自注意力模型和残差连接相组合的方法对多种特征进行学习,在学习特征间关联关系的同时保留部分原始信息,有效挖掘不同特征对配送用时的影响。

## 1 模型构建

### 1.1 相关定义

本文考虑当配送员有若干订单等待配送时,在配送顺序确定的情况下,以所经过的节点序列表示即时配送路线,同时融合多种特征,利用提出的深度学习模型预测即时配送时间,下面给出相关定义。

1) 节点网络。使用取货点、送货点等节点的位置信息表示配送路线,将节点的连接网络用1个无向有权图表示,即 $G=(V, E)$ 。其中: $V$ 表示节点集合, $V=\{v_i | 0 \leq i \leq n_v\}$ , $v_i$ 为节点位置信息, $n_v+1$ 为节点数量; $E$ 表示节点间的连接关系集合, $E=\{e_{ij} | 0 \leq i < j \leq n_v\}$ 。当 $v_i$ 与 $v_j$ 不直接

相连时,令  $e_{ij} = 0$ ; 直接相连时,  $e_{ij}$  值为两节点间的距离。

2) 配送行程。即时配送的行程为  $I_k = \{S_k, F_k, T_k\}$ ,  $0 \leq k \leq M$  ( $M$  为配送任务总数)。其中:  $S_k$  为配送路线信息, 是节点组成的序列, 表示为  $S_k = \{v_0, v_1, \dots, v_{n_k}\}$ ,  $n_k + 1$  为配送路线上的节点数量;  $F_k = \{f_k^v, \dots, f_k^o\}$  为配送任务相关的特征, 如天气、时间、骑手信息等;  $T_k$  为完成即时配送任务  $k$  的用时。

3) 配送时间预测。本文将即时配送时间预测看作回归任务, 具体目标是: 给定待预测即时配送任务  $I_r' = \{S_r', F_r'\}$ ,  $0 \leq r \leq M_r$ , 其中  $M_r$  为待预测任务数量, 通过构建预测模型  $F$ , 使用历史数据  $I_k = \{S_k, F_k, T_k\}$  进行模型训练, 最后利用模型预测即时配送时间  $\tilde{T}_r'$ , 并使预测值与实际用时的偏差最小。

## 1.2 特征分类

即时配送过程涉及多方面信息, 但配送时间主要受到参与主体、环境和任务特性等方面的影响, 为充分挖掘相关信息对即时配送用时的影响, 本文归纳几方面因素, 选择以下 4 种特征。

1) 骑手特征。骑手是即时配送任务的关键, 由于不同骑手对配送任务的决策和执行上存在一定差异, 本文考虑骑手对配送用时的影响, 选择骑手编号、平均速度、等级以及最大接单量作为骑手特征。

2) 外部特征。外部特征包括天气情况、时间信息, 在不同天气情况下, 骑手速度会出现变化, 如在暴雨天气时, 骑手为保证安全通常会主动降低速度。时间信息是执行该配送任务时, 当天是周几以及处于何时, 如周一、17:00 等信息。

3) 过程特征。过程特征包括商家信息、送货点信息、总路程以及节点数。用商家 ID 表示商家, 总路程和节点数是完成配送任务所经过的总距离以及任务节点数。

4) 位置特征。使用任务节点坐标序列表示即时配送过程, 同时表示任务节点间的空间位置关系。

## 1.3 模型框架

本文考虑配送路线存在的时空关系以及相关特征对配送时间的影响, 将上文提出的 4 种特征作为模型输入, 使用深度学习方法对特征进行学习, 最后融合多种特征向量进行预测。本文提出的模型框架如图 1 所示, 主要分为以下 3 个部分。

1) 特征表示。针对各类特征, 使用适用的表

示方法构造特征向量, 方便后续模型的输入和特征提取。

2) 特征提取。使用多头自注意力模型和残差连接相组合的模块提取特征关系, 同时使用卷积神经网络提取空间相关性。

3) 特征融合与预测。将提取的特征向量融合进而预测配送时间。

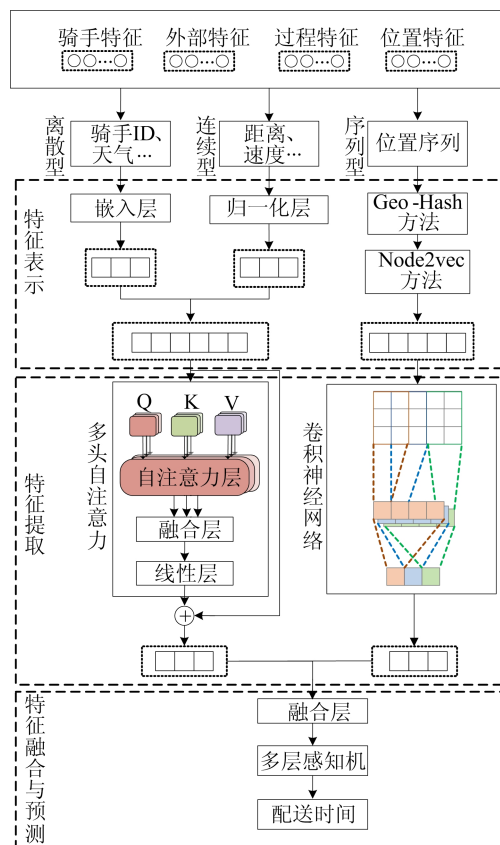


图 1 模型框架

### 1.3.1 特征表示

本文将影响配送时间的特征类型分为离散型、连续型、序列型 3 类, 为保留更多的信息, 本文利用不同方法对 3 类特征进行向量化。

离散型特征主要是一些类别特征, 用来代表某一类别或标签, 如骑手编号、商家编号、天气情况等。针对这类特征, 为避免数据稀疏同时保留关键信息, 本文选择嵌入的方法对离散型特征进行表示。通过构造嵌入矩阵  $E_c$  以及权重矩阵  $W_c$ , 相乘得到嵌入向量  $H_c$ , 即

$$H_c = W_c E_c \quad (1)$$

连续型特征是指数值特征, 如行程距离、速度等。由于不同特征之间的单位、取值范围不同, 直接输入神经网络不利于学习不同特征对目标值的影响, 本文使用极差法对连续型特征进行归一

化,即

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

其中: $x$  为原数值; $x_{\min}$  为该特征的最小值; $x_{\max}$  为该特征的最大值。

序列型特征为任务节点坐标序列,表示骑手完成某一配送任务经过的路线,为了降低路线数据的稀疏性,本文使用 Geo-Hash 方法对节点的经纬度进行编码。Geo-Hash 是一种空间索引算法,利用编码表示不同区域,减少数据量的同时保留了区域间的邻近关系信息<sup>[17]</sup>。本阶段将节点坐标编码为长度为 8 的字符串,将精度缩小到 19 m,使其具有一定代表性。

将节点坐标编码之后,考虑到节点间存在的空间关系,本文使用 Node2vec 方法对位置特征进行嵌入表示。Node2vec 是一种图嵌入方法<sup>[18]</sup>,通过将广度优先搜索和深度优先搜索结合,能够实现对全局信息和局部信息的充分提取。

1) 根据节点间的连接信息构建节点网络  $G=(V,E)$ 。接着使用二阶随机游走的方法提取节点序列,在节点网络中,节点间的距离越近,两者的关联性越强,为捕捉这层关联关系,Node2vec 中使用深度优先搜索和广度优先搜索相结合的方法生成随机游走序列。计算公式为:

$$P(v_i | v_{i-1}) = \frac{\alpha W_{v_{i-1},v_i}}{\beta} \quad (3)$$

$$\beta = \sum_{u \in V_{v_{i-1}}} W_{u,v_{i-1}} \quad (4)$$

$$\alpha = \begin{cases} 1/p, & d_{i,i-2} = 0; \\ 1, & d_{i,i-2} = 1; \\ 1/q, & d_{i,i-2} = 2 \end{cases} \quad (5)$$

其中: $P(v_i | v_{i-1})$  为从节点  $v_{i-1}$  移动到节点  $v_i$  的概率; $W_{v_{i-1},v_i}$  为节点  $v_i$  和  $v_{i-1}$  连接边的权重; $\beta$  为归一化的系数; $V_{v_{i-1}}$  为与节点  $v_{i-1}$  相连的节点集合; $\alpha$  为控制游走策略的参数; $d_{i,i-2}$  为之前节点与现节点的所有近邻节点的最短距离; $p$ 、 $q$  为概率参数。

若两节点直接相连,则  $d_{i,i-2}$  为 0,每增加 1 个中间节点,则距离增加 1,通过改变  $p$ 、 $q$  的值控制游走策略。

2) 利用 Skip-gram 方法将节点序列向量化,构造节点的嵌入矩阵  $E_v$ 。

3) 使用  $E_v$  将节点序列表示为  $E_l \in \mathbf{R}^{n_e \times n_{dl}}$ ,其中: $n_e$  为最大节点数; $n_{dl}$  为嵌入维度。

### 1.3.2 特征提取

将特征向量化表示后,考虑到特征中包含不

同方面的信息,本阶段使用多头自注意力模型和卷积神经网络分别进行特征提取。

1) 多头自注意力模型。自注意力机制通过对序列内部各部分计算注意力概率分布,关注内部的自相关性,能够强化关键特征的影响。而多头自注意力通过设置多个注意力头部,关注不同的区域,实现对不同层次信息的映射表达<sup>[19]</sup>。考虑到多种特征信息之间存在一定的关联关系,本阶段使用多头自注意力模型对离散型和连续型两类特征进行学习。

首先将连续型和离散型特征向量进行拼接,构造特征矩阵  $E_a \in \mathbf{R}^{n_l \times n_{dz}}$ ,其中: $n_l$  为特征数量; $n_{dz}$  为嵌入维度。然后对每个头部模块设置 3 个权重矩阵,分别为  $W_h^a$ 、 $W_h^k$ 、 $W_h^v$ ,  $1 \leq h \leq n_h$ ,  $n_h$  为头部数量。接着将权重矩阵分别与特征矩阵  $E_a$  相乘获得 3 个新矩阵,即

$$Q_h = E_a W_h^a \quad (6)$$

$$K_h = E_a W_h^k \quad (7)$$

$$V_h = E_a W_h^v \quad (8)$$

针对每个头部,根据  $Q_h$ 、 $K_h$ ,使用缩放点积注意力方法计算注意力得分,采用 Softmax 函数得到注意力权重矩阵,最后将权重矩阵与  $V_h$  相乘得到每个头部的输出  $A_h$ ,即

$$A_h = \text{Softmax}\left(\frac{Q_h K_h^T}{\sqrt{d}}\right) V_h \quad (9)$$

其中, $\sqrt{d}$  为缩放因子。

在完成每个头部的注意力计算之后,将所有输出通过融合层进行横向拼接,构造注意力矩阵  $E_z$ ,再输入到线性层。这里初始化 1 个权重矩阵  $W_z$ ,两矩阵相乘得到线性层输出  $E_a'$ ,即

$$E_a' = E_z W_z \quad (10)$$

为了优化模型的训练以及提高泛化能力,在使用多头自注意力模型提取特征时,增加 1 个残差连接<sup>[20]</sup>,保留部分原始特征,将残差连接与多头自注意力模型的输出相加得到最终输出  $E_{fl}$ ,即

$$E_{fl} = \text{ReLU}(E_a' + W_a E_a) \quad (11)$$

其中: $W_a$  为权重参数矩阵;ReLU 为激活函数。

2) 卷积神经网络。本阶段使用卷积神经网络对路线序列中的空间特征进行提取,虽然经过图嵌入,节点间的近邻关系得到表示,但是预测整个配送过程的用时,配送节点序列的空间关系还需要进一步提取。

通过卷积层和池化层,卷积神经网络能够实现对局部和全局信息的全面提取,同时利用参数

共享降低特征学习的复杂度,在对多维特征提取上优势明显<sup>[21]</sup>。根据路线嵌入向量格式,本阶段使用一维卷积神经网络,卷积核大小设置为  $n_c \times n_{dl}$ ,卷积核个数为  $M_c$ ,池化层使用最大池化,步长为 1,池化层大小设置为  $n_p$ 。经过第  $m$  个卷积核,将路线序列嵌入矩阵  $E_1 \in \mathbf{R}^{n_c \times n_{dl}}$  转变为位置特征向量  $F_m$ ,即

$$e_m^g = \text{ReLU}(W_m E_1^{g:g+n_c-1} + b_m) \quad (12)$$

$$F_m = [e_m^1 \ e_m^2 \ e_m^3 \ \cdots \ e_m^{n_c-n_c+1}]^T \quad (13)$$

其中: $e_m^g$  为第  $g$  个滑动窗口计算输出; $1 \leq g \leq (n_c - n_c + 1)$ , $1 \leq m \leq M_c$ ;  $W_m$  为权重参数; $b_m$  为偏置常数。

将所有  $e_m^g$  在同一维度上进行拼接,获得向量  $F_m$ ,然后将所有经过卷积层的输出向量进行拼接得到  $E_1' \in \mathbf{R}^{(n_c - n_c + 1) \times M_c}$ 。随后对  $E_1'$  进行最大池化得到  $E_{l2} \in \mathbf{R}^{(n_c - n_c - n_p + 2) \times M_c}$ ,再接 1 个线性层获得最终输出向量  $V_{l2}$ ,即

$$V_{l2} = W_{l2} E_{l2} + b_{l2} \quad (14)$$

其中: $W_{l2}$  为参数矩阵; $b_{l2}$  为偏置常数矩阵。

### 1.3.3 特征融合与预测

为将上述特征提取模块所学习的不同特征信息充分传递给预测层,本阶段通过特征拼接的方式将 2 个模块的输出  $E_{l1}$ 、 $V_{l2}$  进行特征融合。首先将矩阵  $E_{l1}$  的所有行向量按照顺序连接成一维的特征向量  $V_{l1}$ ,之后将  $V_{l1}$  和  $V_{l2}$  横向拼接得到向量  $V_f = (V_{l1}, V_{l2})$ ,最后将  $V_f$  输入多层感知机得到预测值  $\hat{T}$ ,即

$$\hat{T} = \text{MLP}(V_f) \quad (15)$$

模型训练损失函数使用均方误差(mean square error, MSE),计算公式为:

$$L_{\text{MSE}} = \frac{1}{M} \sum_{k=1}^M (T_k - \hat{T}_k)^2 \quad (16)$$

其中, $\hat{T}_k$  为配送任务  $k$  的预测时间。

## 2 实验与结果分析

### 2.1 数据集

外卖作为最常见的一种即时配送,相关平台已经相当完善,能够收集较完整的配送过程信息,可以满足实验需要,因此本文选择饿了么平台上的外卖配送数据集作为实验数据。该数据集收集地点是大连市某地区,时间为 2020 年 2 月 1 日至 2020 年 2 月 27 日。数据集共包含 254 802 条配送订单数据,将 80% 的数据用于训练,20% 的数据用于测试。

### 2.2 基线方法

本文根据相关研究成果,选择以下 8 种模型与本模型进行性能对比,其中包括 5 种基准模型和 3 种改进模型。

1) 历史平均值方法(history average, HA)以所有骑手的历史配送平均时间作为预测值。

2) SVR 为一种回归算法,将数据映射到高维度空间构造具有边界的超平面,通过优化特征边界值实现误差最小,以寻找到最优预测模型。

3) 随机森林算法(random forest, RF)根据样本特征构造多棵决策树,然后将多棵树进行集成实现最优预测。

4) 深度神经网络(deep neural networks, DNN)通过多个全连接层堆叠,进行特征提取与预测,实验中 DNN 模型设置 3 个隐藏层,隐藏层大小分别设置为 128、64、16。

5) 门控循环单元(gated recurrent unit, GRU)模型为循环神经网络的变体,通过门控机制有选择地保留重要信息,实验中隐藏层个数为 2。

6) WDR<sup>[14]</sup> 为一种行程时间预测模型,模型结合了交叉变换、DNN 和循环神经网络等方法,其思想是通过对不同类型特征采用不同方法进行特征提取,能够实现特征交互以及时空特征的学习。

7) HSETA<sup>[12]</sup> 为一种解决数据异构和稀疏问题的到达时间预测模型,使用因式分解、注意力机制等提高特征学习能力。

8) FMA-ETA<sup>[22]</sup> 为一种结合前馈神经网络和自注意力的到达时间预测模型,对不同路线序列特征分别学习,最后融合所有特征进行预测。

### 2.3 参数设置

本实验模型基于 Pytorch 框架构建,其中多头自注意力模型设置头部数为 2,路线节点图嵌入维度为 128,卷积神经网络的卷积核大小设置为  $5 \times 128$ ,卷积核数量为 256,训练批量大小为 128,学习率为 0.000 6。

### 2.4 评估指标

1) 均方根误差(root mean square error, RMSE)为均方误差的开方,数值越小,模型的精确度越高,计算公式为:

$$L_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (17)$$

其中: $N$  为数据总量; $y_i$  为即时配送订单实际用时; $\hat{y}_i$  为即时配送时间预测值。

2) 平均绝对误差 (mean absolute error, MAE) 可以反映出实际预测误差的情况, 计算公式为:

$$L_{MAE} = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t| \quad (18)$$

3) 平均绝对百分比误差 (mean absolute percentage error, MAPE) 利用比值表示模型的准确度, 计算公式为:

$$L_{MAPE} = \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{y}_t}{y_t} \right| \times 100\% \quad (19)$$

## 2.5 实验结果分析

将对模型和本文所提出的 DTNET 模型分别在数据集上进行训练和测试, 实验结果见表 1 所列。

表 1 实验结果对比

模型	RMSE	MAE/s	MAPE/%
HA	631.966 0	502.645 6	71.05
SVR	585.860 1	437.041 2	51.54
RF	528.431 5	409.661 9	51.17
DNN	512.390 8	385.127 9	47.45
GRU	514.043 8	383.122 4	46.51
FMA-ETA	518.734 7	375.812 8	41.69
WDR	501.794 1	363.851 8	40.72
HSETA	492.741 6	360.847 3	41.78
DTNET	485.288 1	348.770 4	38.51

从表 1 可以看出, 在 3 个评价指标上, DTNET 模型预测效果比其他模型更好。相较于表现较好的 HSETA 模型, 本文模型的 RMSE 值下降了约 1.51%, MAE 值下降了约 3.34%, MAPE 值下降了约 7.82%。

由于 HA 模型是基于全部历史数据进行大概估计, 未考虑相关特征的影响, 导致预测精度较差。支持向量回归和随机森林等机器学习模型考虑了特征对结果的影响, 预测精度有较大幅度的提升, 但是一些复杂的非线性关系未能有效学习。DNN 模型通过全连接层学习非线性关系, GRU 通过门控机制实现重要信息提取, FMA-ETA 模型利用前馈神经网络和自注意力进行特征学习, 与支持向量回归等模型相比, 3 种模型预测性能有一定提升。WDR 模型通过分特征学习, 有效提高特征学习的能力, HSETA 模型通过引入注意力机制, 学习特征间的交叉关系, 相较于 WDR 模型进一步提高了预测能力。本文所提模型针对不同类型数据进行学习, 使用多头自注意力和卷积神经网络等模型, 结合即时配送任务, 有效提取

配送路线上的多种特征, 实现对即时配送时间的精准预测, 预测能力有一定提升。

## 2.6 消融实验

为了解不同特征和结构对于预测结果的影响, 本文设置消融实验, 包括以下变体模型。

1) Without-D 去除特征中的骑手特征, 其余模块保持不变。

2) Without-O 去除外部特征, 其余模块保持不变。

3) Without-P 去除过程特征, 其余模块保持不变。

4) Without-L 去除位置特征, 其余模块保持不变。

5) Without-M 去除多头自注意力模型, 其余模块保持不变。

6) Without-C 去除卷积神经网络模块, 其余模块保持不变。

消融实验结果见表 2 所列, 在去除上述几种特征之后, 预测性能均有一定下降, 其中去除过程特征之后预测精度下降最多, 与原模型相比, 3 个指标分别上升 12.58%、15.31%、27.42%, 说明即时配送路线相关的信息对于提高预测准确度十分重要, 同时也说明在预测即时配送时间时, 考虑骑手、天气、时间以及空间特征对整体用时的影响很重要。

表 2 消融实验结果对比

模型	RMSE	MAE/s	MAPE/%
Without-D	492.056 3	354.625 9	39.10
Without-O	495.534 9	352.480 3	37.71
Without-P	546.352 8	402.157 8	49.07
Without-L	506.754 1	362.802 1	39.04
Without-M	496.737 4	355.506 1	39.71
Without-C	494.949 0	352.293 9	38.08
DTNET	485.288 1	348.770 4	38.51

去除卷积神经网络模型之后, 预测性能有所下降, 说明利用卷积神经网络提取配送节点间的总体空间关系十分必要。在去除多头自注意力模型之后, 预测效果随之变差, 说明多头自注意力部分能够对特征间的关系进行有效学习。

为了解多种特征对即时配送时间的影响, 本文将多个特征的注意力得分可视化, 注意力得分热力图如图 2 所示。从图 2 可以看出, 骑手速度、最大载荷和距离等特征组合对于即时配送时间的影响很大, 除此之外, 骑手等级、节点数量、天气和

时间点等特征组合的较高得分说明在不同天气情况和时刻点,任务复杂度和骑手能力对配送时间有着不同的影响。

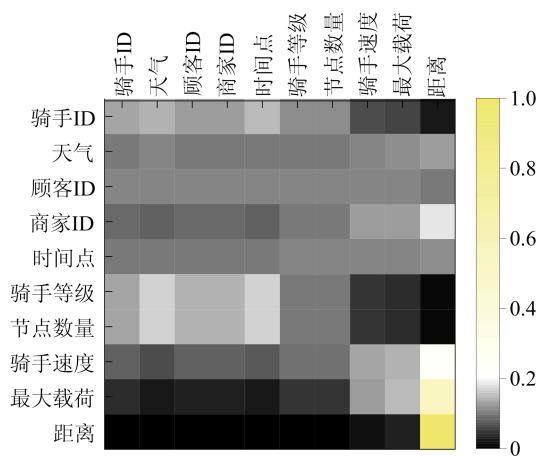


图 2 注意力得分热力图

### 3 结 论

针对即时配送时间预测问题,本文提出一种基于深度学习的即时配送时间预测模型。通过采用分特征学习的方式,将多头自注意力模型和残差连接方法组合,提高了学习多种特征间关联关系的能力;同时利用图嵌入和卷积神经网络,对路线特征中的空间关系进行有效提取,充分学习特征关联关系和时空关系对配送用时的影响。在真实数据集上的实验表明,本文提出的模型在预测精度上比其他模型有一定提升,进一步提高了即时配送时间预测的准确性。

由于本文是在配送路线已知的情况下对即时配送用时进行预测,考虑的影响因素相对固定,是对现实即时配送场景的简化,未来可研究配送路线未知场景下的预测模型,更好地模拟现实情况进而提高预测精度。

### [参 考 文 献]

- [1] 田丹,唐加福,任悦. O2O 模式下即时配送服务系统弹性的提升策略优化[J]. 系统工程理论与实践, 2021, 41(2): 310-318.
- [2] KHALED A, ELSIR A M T, SHEN Y M. GSTA: gated spatial-temporal attention approach for travel time prediction[J]. Neural Computing and Applications, 2022, 34: 2307-2322.
- [3] HONG H T, LIN Y C, YANG X Y, et al. HetETA: heterogeneous information network embedding for estimating time of arrival[C]//The 26th ACM SIGKDD International Conference on

Knowledge Discovery & Data Mining, New York: Association for Computing Machinery, 2020: 2444-2454.

- [4] DHIVYABHARATHI B, KUMAR B A, VANAJAKSHI L. Real time bus arrival time prediction system under Indian traffic condition[C]//2016 IEEE International Conference on Intelligent Transportation Engineering (ICITE). Singapore: IEEE, 2016: 18-22.
- [5] ACHAR A, BHARATHI D, KUMAR B A, et al. Bus arrival time prediction: a spatial Kalman filter approach[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 21(3): 1298-1307.
- [6] MASIERO L, CASANOVA M, CARVALHO M. Travel time prediction using machine learning[C]//The 4th ACM SIGSPATIAL International Workshop on Computational Transportation Science (CTS'11). Chicago: Association for Computing Machinery, 2011: 34-38.
- [7] HILDEBRANDT F D, ULMER M W. Supervised learning for arrival time estimations in restaurant meal delivery[J]. Transportation Science, 2022, 56(4): 1058-1084.
- [8] ZHANG Y R, HAGHANI A. Gradient boosting method to improve travel time prediction[J]. Transportation Research Part C(Emerging Technologies), 2015, 58: 308-324.
- [9] KANKANAMGEK D, WITHARANAGEY R, WITHANAGEC S, et al. Taxi trip travel time prediction with isolated XGBoostregression[C]//2019 Moratuwa Engineering Research Conference (MERCCon). Moratuwa: [s. n.], 2019: 54-59.
- [10] GUNDUZ G, ACARMAN A. Vehicle travel time estimation using sequence prediction [J]. Promet-Traffic & Transportation, 2020, 32(1): 1-12.
- [11] LEE E H, KHO S Y, KIM D K, et al. Travel time prediction using gated recurrent unit and spatio-temporal algorithm [J]. Proceedings of the Institution of Civil Engineers-Municipal Engineer, 2021, 174(2): 88-96.
- [12] CHEN K Q, CHU G W, YANG X X, et al. HSETA: a heterogeneous and sparse data learning hybrid framework for estimating time of arrival [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(11): 21873-21884.
- [13] XU S J, ZHANG R Q, CHENG W J, et al. MTLM: a multi-task learning model for travel time estimation[J]. Geoinformatica, 2022, 26: 379-395.
- [14] WANG Z, FU K, YE J P. Learning to estimate the travel time[C]//The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'18). New York: Association for Computing Machinery, 2018: 858-866.
- [15] GAO C L, ZHANG F, WU G Q, et al. A deep learning method for route and time prediction in food delivery service[C]//The 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD'21). New York: Association for Computing Machinery, 2021: 2879-2889.

(下转第 1274 页)

出结论如下。

1) 随迎流角从  $0^\circ$  增大到  $45^\circ$ , 下游桩周围河床的床面切应力分布模式由屏蔽状态向分离流冲击状态转变。

2) 间距比为 1.0 时上游桩的剪切层对下游桩的屏蔽角在  $5^\circ$  左右; 随间距比的增大, 上游桩的剪切层对下游桩的屏蔽效应减小。

3) 间距比为 1.5 或 2.0 工况下, 迎流角位于  $20^\circ \sim 30^\circ$  之间时, 各桩阻力系数值均处于 1.0 左右, 屏蔽效应最小。

4) 当迎流角为  $0^\circ$  时, 对比典型间距下的间隙流流速, 当间距比大于 1.5 后桩间区域的间隙流受圆形桩几何特征的影响越来越小; 实际工程中可通过调整迎流角、间距比的布置, 优化设计方案, 降低水流的动力学作用, 保障工程结构安全。

### 参 考 文 献

- [1] 殷长山. 方型布置四圆柱绕流的试验及数值研究[D]. 杭州: 浙江大学, 2017.
- [2] 涂佳黄, 唐彪, 曹波, 等. 正方形排列四圆柱体绕流特性及互扰效应研究[J]. 应用力学学报, 2020, 37(1): 32-40, 471.
- [3] 李聪洲, 张新曙, 胡晓峰, 等. 高雷诺数下多柱绕流特性研究[J]. 力学学报, 2018, 50(2): 233-243.
- [4] 庄茜, 李绍武, 祁泽鹏. 基于有限体积法的群桩绕流数值模拟[J]. 天津大学学报(自然科学与工程技术版), 2015, 48(5): 445-454.
- [5] 杨熠琳. 桥梁群桩基础局部冲刷机理研究[D]. 北京: 北京

交通大学, 2020.

- [6] 喻鹏. 改进数值模型与群桩基础桥墩局部冲刷的 CFD 模拟[D]. 长沙: 湖南大学, 2017.
- [7] 金亚昆. 桥墩局部冲刷坑发展中的三维紊流场研究[D]. 北京: 北京交通大学, 2014.
- [8] MENTER F R. Two-equation eddy-viscosity turbulence models for engineering applications [J]. AIAA Journal, 1994, 32(8): 269-289.
- [9] PATANKAR S V, SPALDING D B. A calculation procedure for heat, mass and momentum transfer in three-dimensional parabolic flows[J]. Journal of Heat Mass Transfer, 1972, 15(10): 1787-1806.
- [10] ONG L, WALLACE J. The velocity field of the turbulent very near wake of a circular cylinder[J]. Experiments in Fluids, 1996, 20(6): 441-453.
- [11] KRAVCHENKO A G, MOIN P. Numerical studies of flow over a circular cylinder at  $Re=3\ 900$  [J]. Physics of Fluids, 2000, 12(2): 403-417.
- [12] 端木玉, 万德成. 雷诺数为 3 900 时三维圆柱绕流的大涡模拟[J]. 海洋工程, 2016, 34(6): 11-20.
- [13] LAM K, LI J Y, SO R M C. Force coefficients and Strouhal numbers of four cylinders in cross flow[J]. Journal of Fluids and Structures, 2003, 18(3/4): 305-324.
- [14] WANG X K, GONG K, LIU H, et al. Flow around four cylinders arranged in a square configuration[J]. Journal of Fluids & Structures, 2013, 43: 179-199.
- [15] 陈小莉. 局部绕流冲刷机理及数值模拟研究[D]. 北京: 清华大学, 2008.

(责任编辑 吴 亮)

### (上接第 1254 页)

- [16] ZHU L, YU W, ZHOU K R, et al. Order fulfillment cycle time estimation for on-demand food delivery [C]//The 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'20). New York: Association for Computing Machinery, 2020: 2571-2580.
- [17] 王强, 林友芳, 万怀宇. 基于多任务学习的快件送达时间预测方法[J]. 计算机工程, 2022, 48(4): 324-330.
- [18] ADITYA G, JURE L. Node2vec: scalable feature learning for networks[C]//The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16). New York: Association for Computing Machinery, 2016: 855-864.
- [19] 陈颖婷, 林耿, 陈梦, 等. 融合多头自注意力的问答社区专家推荐算法[J]. 计算机应用研究, 2023, 40(5): 1452-1458.

- [20] SONG W P, SHI C C, XIAO Z P, et al. AutoInt: automatic feature interaction learning via self-attentive neural networks[C]//The 28th ACM International Conference on Information and Knowledge Management (CIKM'19). New York: Association for Computing Machinery, 2019: 1161-1170.
- [21] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [22] SUN Y W, WANG Y L, FU K, et al. FMA-ETA: estimating travel time entirely based on FNN with attention [C]//ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto: IEEE, 2021: 3355-3359.

(责任编辑 张 镅)