

基于分级包络域适应的行人轨迹预测模型

李勇明, 李文正, 张小恒, 王 品, 胡 杰

(重庆大学微电子与通信工程学院, 重庆 400044)

摘要: 复杂环境下行人轨迹短时预测在自动驾驶、社交机器人控制、智能安防及智慧城市等领域有着广泛用途。行人与行人、行人与环境之间的交互具有多尺度复杂性和不确定性, 具有挑战。现有深度学习模型虽然有助于挖掘行人的复杂交互关系, 但都假设行人轨迹在不同场景遵循相同运动模式, 未考虑场景间存在的潜在分布差异; 域适应模型虽然考虑了这一点, 但仍未考虑行人间和行人环境间的多层次特性。为了解决上述问题, 本文提出了一种基于分级包络域适应的行人轨迹预测模型。通过构造局部层次行人邻接关系设计局部层次包络样本, 通过个体层次行人关系设计个体层次包络样本, 并将两者融合形成双级包络样本。基于双级包络样本构造模块, 求得行人轨迹的时空特征分布, 从而构造全局层次包络样本。基于注意力机制和跨域分布对齐, 分别设计了局部层次包络域适应模块和全局层次包络域适应模块, 构建加权预测损失函数将两者融合一体, 并联合优化。实验部分选取了2个有代表性的公共数据集, 并与5个相关代表性算法模型进行对比。通过消融实验、参数分析、方法对比和轨迹可视化等来进行综合验证。在ETH和UCY的实验结果表明, 相比于T-GNN, 本文方法的平均位移误差降低了22.7%, 终点位移误差降低了19.8%。文章完整版参见链接: <https://github.com/LWZ9910/MESC-HEDA.git>。

关键词: 行人轨迹预测; 分级包络; 域适应; 多层次; 包络样本

基金项目: 国家自然科学基金(No.U21A20448, No.61771080)

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112(2025)04-1308-14

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20240562

The Pedestrian Trajectory Prediction Model Based on Hierarchical Envelope Domain Adaptation

LI Yong-ming, LI Wen-zheng, ZHANG Xiao-heng, WANG Pin, HU Jie

(School of Microelectronics and Communication Engineering, Chongqing University, Chongqing 400044, China)

Abstract: In complex environments, short-term pedestrian trajectory prediction finds extensive applications in autonomous driving, social robotics, intelligent security, and smart city infrastructures. Interactions among pedestrians and between pedestrians and their environment exhibit multi-scale complexities and uncertainties, posing substantial challenges. Although current deep learning models are effective in uncovering complex pedestrian interactions, they typically assume uniform motion patterns across various scenes, thereby neglecting potential distributional discrepancies. While domain adaptation models partially address this issue, they often overlook the multi-level characteristics of pedestrian interactions and environmental influences. To address these challenges, this study proposes a pedestrian trajectory prediction model founded on hierarchical envelope domain adaptation. We design a local-level envelope sample construction module by establishing local-level pedestrian adjacency relationships. An individual-level envelope sample construction module is devised based on individual pedestrian relationships. These two modules are subsequently integrated to form a bi-level envelope sample construction module. Leveraging the bi-level envelope sample construction module, we compute the spatio-temporal feature distribution of all pedestrian trajectories to construct global-level envelope samples. Employing the attention mechanism and cross-domain distribution alignment, we respectively design the local-level envelope domain adaptation and global-level envelope domain adaptation modules. These modules are then integrated into a unified framework using a weighted prediction loss function, which is jointly optimized. The experimental section utilizes two representative public datasets and compares them with five representative algorithm models. Comprehensive validation is conducted through ablation studies, parameter

analysis, method comparison, and trajectory visualization. The experimental results in the ETH and UCY datasets show that compared with T-GNN, the average displacement error is reduced by 22.7% and the final displacement error is reduced by 19.8%. For the full version of the article, please refer to the link: <https://github.com/LWZ9910/MESC-HEDA.git>.

Key words: pedestrian trajectory prediction; hierarchical envelopes; domain adaptation; multilevel; envelope samples
Foundation Item(s): National Natural Science Foundation of China (No.U21A20448, No.61771080)

1 引言

复杂环境下行人轨迹预测极具挑战性,受诸多复杂因素影响^[1],包括行人运动状态、目的地、年龄和性别等本身属性^[2],以及多个行人之间的互动^[3].早期研究成果大多是基于社会力模型、等速模型等知识模型^[4,5],但行人交互的非线性和多种复杂因素导致上述方法普适性差,存在较大局限性.近年来,基于机器学习尤其是深度学习的行人轨迹预测模型成为研究热点,具有较高预测精度^[6-9].但是,大多数基于深度学习模型方法往往假定训练集和测试集轨迹边缘概率分布一致,侧重于学习统一的行人间及行人与地理环境间的交互模式,而忽略训练集和测试集的域偏移问题,从而导致模型泛化性能差^[10,11].例如,在游乐园或商场等人流密集场所,行人之间交互会非常频繁,通常需要不断避让、绕行或等待以避免碰撞和阻塞,导致轨迹曲折和复杂,且不断突变;而街道上行人间碰撞交互关系较少,行人更多遵循道路规划设计方向前进,运动轨迹多为直线.忽略上述场景间分布差异会导致从一个域构建的模型应用到另一个域时,不可避免地影响行人轨迹预测的准确性.

行人轨迹预测模型性能分析大多基于 ETH、HOTEL、UNIV、ZARA1、ZARA2 这 5 个轨迹场景进行. Xu 等人^[10]通过对样本数量、行人数量、行走速度、加速度等指标量化评估,发现领域间存在显著差异,不同场景同一指标大小差异十分明显,这通常被称为行人轨迹场景域偏移问题.其通常解决思路是:将训练场景看作源域,将测试场景看作目标域,并通过域适应相关方法实现分布对齐.尽管近期出现了基于域适应的行人轨迹预测模型,但其仅考虑局部社交层次样本且仅考虑局部层次域适应,在面对行人之间、行人与环境之间交互的多层次复杂性和不确定性时具有局限性.事实上,行人轨迹走向受个体行人自身属性、邻近行人社交特性以及全局地理环境特性 3 个方面因素的共同影响,因此考虑从这 3 个层次去描述样本特性并构建对应的分层域适应是有必要的.

基于上述分析,本文在现有基于域适应的行人轨迹预测模型基础上,做如下创新工作:(1)构造个体层次包络样本和局部层次包络样本,融合构建双级包络样本;(2)基于双级包络样本的时空特征分布,进一步生成全局层次包络样本,从而构造出多级包络样本,以获取面向行人轨迹数据在不同层面更为丰富的时空特

征表征;(3)针对各级包络样本,设计源域和目标域之间的局部层次和全局层次域适应模块,从而实现分级包络域适应模块;(4)综合上述创新,形成一种基于分级包络域适应的行人轨迹预测模型.文末通过 20 种场景组合实验验证了本文方法的有效性.本文工作将现有“基于单一层面域适应的行人轨迹预测模型”转化为“基于多个层面域适应的行人轨迹预测模型”,对相关研究具有一定的参考价值.

2 相关工作综述

2.1 非域适应行人轨迹预测

现有行人轨迹预测方法大多是非域适应的.传统方法有社会力模型^[12,13]、多模型方法和混合估计法^[5,14]等,但面临复杂拥挤环境时预测精度难以提升.近年来出现的深度学习相关方法在有效提取高级特征、处理非线性数据上具有显著优势,因此成为复杂环境中行人轨迹预测的研究热点之一.

基于深度学习模型的行人轨迹预测主要采用以下 3 种架构及其变体:长短期记忆网络(Long Short-Term Memory, LSTM)、生成对抗网络(Generative Adversarial Network, GAN)、卷积神经网络(Convolutional Neural Network, CNN).在 LSTM 方面,Alahi 等人^[15]考虑行人之间存在相互影响,并将单个行人作为网络节点创建一个社交信息共享系统,即 Social-LSTM 行人轨迹预测模型.Xue 等人^[16]提出一种基于 LSTM 的分层网络.Hug 等人^[17]提出一种混合密度层结合粒子滤波的 LSTM 多模型行人轨迹预测模型.

在 GAN 方面,Gupta 等人^[18]提出基于 GAN 和 LSTM 的混合模型,基于噪声和改进的多样性损失函数生成多样性轨迹,并使用最大池化方法对全局行人进行交互分析,但该模型的 GAN 架构网络训练不太稳定.Amirian 等人^[19]通过引入 Info-GAN 架构解决了 Social-GAN 模型训练不稳定的问题.Sadeghian 等人^[20]提出了一种基于 GAN 的可解释轨迹预测模型 SoPhie,通过结合社交和物理注意力机制,聚合了不同智能体的交互信息,并利用 GAN 生成更真实样本预测针对行人的社交和物理上的合理路径.基于生成式网络架构预测方法在进行行人轨迹预测时能保证较高的精度,提升模型预测速度,使复杂模型建立更为合理有效.

在 CNN 方面,Nikhil 等人^[21]提出基于 CNN 的轨迹

预测算法,以提高计算效率,实现并行化,并且不依赖前一个时间步,预测速度可以更快,而且预测误差不会像RNN(Recurrent Neural Network)算法一样累积. Yan等人^[22]提出基于图卷积神经网络(Graph Convolutional Network, GCN)的轨迹预测算法,在提取行人与其他对象之间的空时交互信息上具有优势. Mohamed等人^[23]提出了Social-STGCNN模型,通过使用时间卷积网络(Temporal Convolutional Network, TCN)从空间和社交交互特征中提取时空特征,并利用CNN作为时间外推器. Zhang等人^[24]提出了Social-IWSTCNN,通过池化机制和注意力权重聚合来模拟行人与其他人之间相互作用,并使用CNN和TCN进行行人轨迹预测. Lv等人^[25]提出了社交软注意力图卷积网络SSAGCN,旨在同时处理行人之间的社交互动以及行人与环境之间的场景互动.

2.2 基于域适应的行人轨迹预测

目前主流的行人轨迹预测模型大多是非域适应的,其假设训练场景和测试场景具有相同的数据分布. 但是这个假设在现实中往往难以成立. 例如,在ETH场景中人流稀少,行人轨迹模式相对稳定,而UNIV场景人流密集,行人轨迹模式复杂多变. 因此,不同场景的轨迹域之间存在显著的分布差异. 针对这一问题,近年来有学者提出了一些基于域适应的行人轨迹预测方法,取得了一定的进展,但相关研究仍较为有限,目前主要包括以下4篇文献: Xu等人^[10]提出了一种可迁移图神经网络T-GNN框架,采用域不变GNN减少领域特定知识结构,该框架同时进行轨迹预测和域对齐; Zhi等人^[11]首次结合域适应和持续学习预测轨迹,其中域适应模块分析源域和目标域的数据分布差异,持续学习模块通过限制模型参数变化应对灾难性遗忘,从而提高了模型泛化性; Huang等人^[26]提出跨域轨迹预测网络CTP-Net,借助编码器-解码器结构以及域适应方法,构建将目标数据映射到源轨迹特征域的网络; Geng等人^[27]提出了一种可分层孪生网络T-HSTN,主要探究数据不足或异常的城市无信号交叉口的异构多智能体轨迹预测任务,填补了异构智能体轨迹预测中多源迁移学习任务的解决方案和基准的空白.

类似上述涉及域适应的行人轨迹预测模型研究较少,但总的来说,已有相关文章开始探讨训练场景和测试场景间的潜在分布差异,一定程度上缓解了行人轨迹场景域偏移问题,但也存在局限性. 首先,关于域适应,现有模型往往未能充分应对不同场景间的频繁切换,导致在跨域预测时精度下降. 部分研究虽然尝试引入域适应机制,但大多局限于局部层次的域适应,忽略了全局层次域适应的重要性,这限制了模型的泛化能力和对域偏移问题的全面处理能力. 相比之下,本文模型通过引入分级包络域适应模块,实现了局部层次与

全局层次域适应的有机结合,显著增强了模型在不同场景中的多层次跨域时空特征的迁移学习能力,有效提升了泛化能力,并减少了交互时冗余信息对预测结果的影响. 其次,在多层次交互特征捕捉方面,现有模型往往难以同时兼顾行人的空间层次和时间层次关联关系. 即便有模型考虑了这些关联,也往往未能全面覆盖全局层次和局部层次的空-时关联. 本文模型则通过构建表征行人轨迹多尺度时空特征的多级包络样本,巧妙地利用图卷积网络和注意力机制,深入挖掘了行人轨迹序列中多层次、多维度的时空特征表征. 图卷积网络有效捕捉了行人之间的空间关联关系,而注意力机制则进一步强化了时间层次上的关联建模,使得模型能够更有效地挖掘行人轨迹序列中多层次的时空特征表征.

3 基于分级包络域适应的行人轨迹预测模型

3.1 问题定义

给定场景中 N 个行人在时间段 $T_1 \sim T_{\text{obs}}$ 的观察轨迹 $P_{\text{obs}}^N = \{\mathbf{p}_1^i, \mathbf{p}_2^i, \dots, \mathbf{p}_{\text{obs}}^i | \forall i \in \{1, 2, \dots, N\}\}$, 通过可学习参数为 \mathbf{W}^* 的模型 $f(\cdot)$ 预测未来较短时间段 $T_{\text{obs}+1} \sim T_{\text{pre}}$ 的轨迹 $P_{\text{pre}}^N = \{\mathbf{p}_{\text{obs}+1}^i, \mathbf{p}_{\text{obs}+2}^i, \dots, \mathbf{p}_{\text{pre}}^i | \forall i \in \{1, 2, \dots, N\}\}$, $\mathbf{p}_t^i = [x_t^i, y_t^i] \in \mathbb{R}^2$ 表示第 i 个行人在时间帧 t 时刻的二维空间坐标,表示如下:

$$P_{\text{pre}}^N = \{f(P_{\text{obs}}^N) \mathbf{W}^*\} \quad (1)$$

3.2 模型描述

3.2.1 包络样本定义

现有深度学习思想主要基于原始数据样本建模,自动提取出反映原始数据样本的深层特征,而本文提出的包络样本学习机制旨在挖掘原始样本间关联信息,将其转化为新样本,从而可将原始样本映射为多层包络样本,能从多个层面去表征目标对象. 研究表明,相较于现有直接基于原样本建模法,基于分级包络样本建模法有助于提高准确率^[28,29]. 包络样本学习机制简述如下:

给定数据集 $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, 对于任一样本 $\mathbf{x}_i \in X$, 通过算子 η 找到其近邻样本并组成近邻样本集为 $N(\mathbf{x}_i), N(\mathbf{x}_i) \subseteq X$; 然后将 \mathbf{x}_i 与 $N(\mathbf{x}_i)$ 合并成一个新样本集 $X_c = \{\mathbf{x}_i, N(\mathbf{x}_i)\}$, 同时构建邻接矩阵 $\mathbf{A}(X_c)$ 表达其邻接关系,得到样本集 $\{X_c, \mathbf{A}(X_c)\}$ 的过程为样本包络化的过程; 其次,通过变换算子 $\Gamma(\cdot)$,对样本集 $X_c' = \{X_c, \mathbf{A}(X_c)\}$ 进行变换,即 $X_e = \Gamma(X_c')$,则称生成的新样本集 X_e 为包络样本. 从 X 变换到 X_e 的过程为包络样本变换. 多层次的包络样本集 $\{X_e^l\}_{l=1}^L$ 和现有样本集 X 构成分层包络样本空间 $X_e^L = \{X, \{X_e^l\}_{l=1}^L\}$, 其中 $X_e^l = \Gamma^l(X_c')$. 据

此,基于原样本的模型 $H(X)$ 转变为基于多级包络样本的模型 $H(X_e^L)$,即实现了多级包络样本构造.

3.2.2 模型总体介绍

本文提出的基于分级包络域适应的行人轨迹预测模型如图 1 所示. 给定源域和目标域的轨迹, 构建双级包络样本构造模块, 将原始轨迹数据按照个体层次和局部层次分别构造出个体层次包络样本和局部层次包络样本, 而后通过特征提取与特征融合得到双级包络

样本. 基于双级包络样本的时空特征分布, 设计一种时空信息协同整合机制, 对时间层面和空间层面信息进行捕捉与整合, 构建全局层次包络样本. 结合双级包络样本, 形成表征行人轨迹多尺度时空特征的多级包络样本. 在构造多级包络样本的基础上, 构建分级包络域适应模块进行迁移学习, 通过加权融合局部层次和全局层次特征表示的域分布差异, 进一步增强模型的多层次跨域时空特征的迁移学习能力, 实现域分布对齐下的行人轨迹预测.

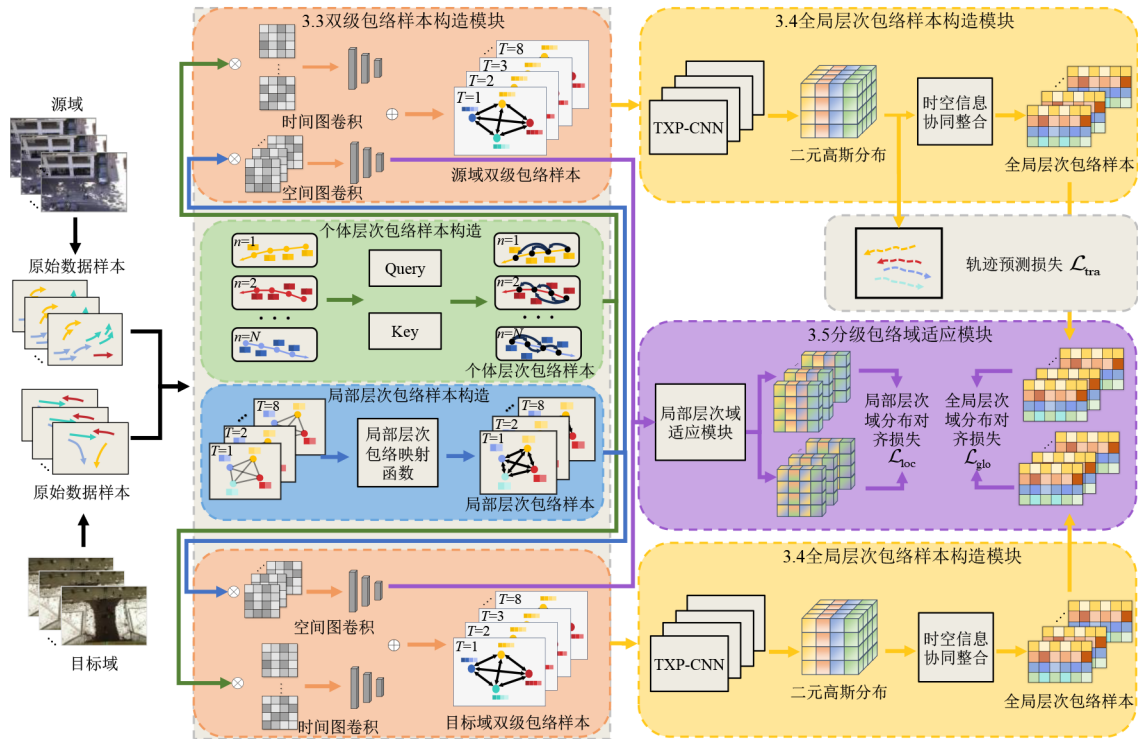


图 1 MES-C-HEDA 模型框图

3.3 双级包络样本构造模块

3.3.1 局部层次包络样本构造模块

局部层次包络样本构造模块如图 2 所示. 基于场景中原始行人轨迹序列样本, 利用各个行人在空间维度上的关系, 通过局部层次包络映射函数 $F_{loc}(\cdot)$ 挖掘原始轨迹样本在空间层次的邻接关系, 从而构造一系列带有邻接矩阵的空间图结构; 而后通过合适的交互力度筛选和归一化操作得到局部层次包络样本 $[F_{loc}, A'_{loc}]$, 实现包络样本变换. 具体过程如下:

(1) 局部层次包络样本的行人邻接关系构造

首先, 针对同一场景相同时刻的各个行人之间局部关系, 构造一系列空间图结构表示行人轨迹的局部层次样本, 即将时间轴 t 时刻 N 个行人的坐标 $P_t =$

$\{p_i^t | \forall i \in \{1, 2, \dots, N\}\}$ 处理为图表示 $G_{loc} = (V_{loc}, E_{loc})$, 其中 $V_{loc} = \{p_{T_1}^i, p_{T_2}^i, \dots, p_{T_{obs}}^i | \forall i \in \{1, 2, \dots, N\}\}$ 为图表示 G_{loc} 的顶点集, $E_{loc} = \{e_{T_1}^{ij}, e_{T_2}^{ij}, \dots, e_{T_{obs}}^{ij} | \forall i, j \in \{1, 2, \dots, N\}\}$ 是 G_{loc} 的边集, E_{loc} 中各条边是用局部层次行人邻接关系 R_{loc} 和邻接矩阵 A_{loc} 表示. $F_{loc} = \{v_{T_1}^i, v_{T_2}^i, \dots, v_{T_{obs}}^i | \forall i \in \{1, 2, \dots, N\}\}$ 是各个行人对应的特征向量, 其中 $v_t^i = (x_t^i - x_{t-1}^i, y_t^i - y_{t-1}^i)$ 代表时间轴 t 时刻的第 i 个行人的速度矢量.

通过综合考虑行人群体的局部层次交互关系可构造出局部层次包络的邻接关系 R_{loc} . 首先分别定义 2 个半径分别为 r_1 和 r_2 的局部层次行人群体包络圈. 当目标行人与其他行人之间的距离处于 2 个圆圈的范围时, 可将这些行人视为局部层次包络样本. L_t^j 表示 t 时刻包络圈内的行人 i 与行人 j 的欧式距离的倒数.

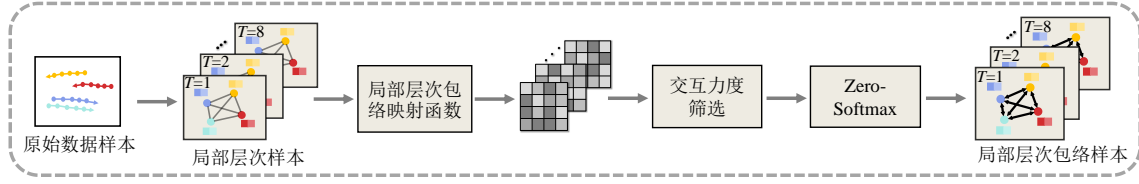


图2 局部层次包络样本构造模块

$$L_t^{ij} = \begin{cases} \frac{1}{\sqrt{(x_t^i - x_t^j)^2 + (y_t^i - y_t^j)^2}}, & r_1 < \sqrt{(x_t^i - x_t^j)^2 + (y_t^i - y_t^j)^2} < r_2 \\ 0, & \text{Otherwise} \end{cases} \quad (2)$$

α_t^{ij} 表示 t 时刻的行人 i 的速度矢量 \mathbf{v}_t^i 与行人 j 的当前位置连线之间的夹角,对夹角 α_t^{ij} 取余弦函数得 $c_t^{ij} = \cos \alpha_t^{ij}$. H_t^{ij} 表示 t 时刻行人 i 与行人 j 的速度夹角综合度量.

$$H_t^{ij} = \max(0, |\mathbf{v}_t^i|c_t^{ij} + |\mathbf{v}_t^j|c_t^{ji}) \quad (3)$$

此外,定义 $\tau_t^{ij} = \min\left(1, \frac{|\mathbf{v}_t^i|}{|\mathbf{v}_t^j|}\right)$. 综合来看,局部层次包络映射函数 $F_{\text{loc}}(\cdot)$ 包含了群体中行人速度与周边行人之间的距离,以及相对运动方向上的夹角等多种因素,则局部层次行人邻接关系 \mathbf{R}_{loc} 构造如下:

$$\mathbf{R}_{\text{loc}} = F_{\text{loc}}(\mathbf{v}_t^i, \mathbf{v}_t^j, \mathbf{p}_t^i, \mathbf{p}_t^j) = \begin{cases} L_t^{ij} H_t^{ij} \tau_t^{ij}, & i \neq j \\ 0, & \text{Otherwise} \end{cases} \quad (4)$$

(2)局部层次包络样本交互力度筛选

接下来,对 \mathbf{R}_{loc} 进行局部层次包络样本交互力度筛选,通过引入阈值超参数 $\eta \in [0, 1]$,生成局部层次掩码 \mathbf{M}_{loc} .

$$\mathbf{M}_{\text{loc}} = \mathbb{I}(\mathbf{R}_{\text{loc}} \geq \eta) \quad (5)$$

其中, $\mathbb{I}(\cdot)$ 是指示函数,当 $\mathbf{R}_{\text{loc}} \geq \eta$ 时,则输出1,否则输出0. 然后通过逐元素乘法运算将实现对 \mathbf{R}_{loc} 筛选,同时叠加单位阵 \mathbf{I} 得到局部层次邻接矩阵 \mathbf{A}_{loc} ,其中 \odot 表示逐元素乘法.

$$\mathbf{A}_{\text{loc}} = \mathbf{M}_{\text{loc}} \odot \mathbf{R}_{\text{loc}} + \mathbf{I} \quad (6)$$

而后采用“Zero-Softmax”方法^[30]对 \mathbf{A}_{loc} 内的各个元素 a_t^{ij} 进行归一化操作,如下所示:

$$\text{Zero-Softmax}(a_t^{ij}) = \frac{\exp(a_t^{ij}) - 1}{\sum_{k=1}^N (\exp(a_t^{ik}) - 1) + \epsilon} \quad (7)$$

其中, ϵ 是极小常数,以确保数值稳定性. LESC的伪代

码描述如算法1所示.

算法1 局部层次包络样本构造模块-LESC

输入: 行人轨迹坐标 $\mathbf{X}_t^i = [x_t^i, y_t^i]$, $i = 1, 2, \dots, N$, $t = T_1, T_2, \dots, T_{\text{obs}}$

输出: 局部层次包络样本 $[\mathbf{F}_{\text{Loc}}, \mathbf{A}_{\text{Loc}}]$

步骤:

1. 对行人轨迹坐标 \mathbf{X}_t^i 按照空间近邻关系构造局部层次特征 \mathbf{F}_{loc}
2. For $t = T_1$ to T_{obs}
3. For $i = 1$ to N
4. For $j = 1$ to N
5. 通过式(2)求得各个行人局部层次行人群体包络圈内的欧氏距离的倒数 L_t^{ij}
6. 求行人 i 的速度矢量 \mathbf{v}_t^i 与节点 j 的当前位置连线之间的夹角 α_t^{ij}
7. 通过式(3)、式(4)求得局部层次行人邻接关系 \mathbf{R}_{loc}
8. End for
9. End for
10. End for
11. 通过式(5)、式(6)得到局部层次邻接矩阵 \mathbf{A}_{loc}
12. 通过式(7)进行归一化操作,构造局部层次包络样本 $[\mathbf{F}_{\text{loc}}, \mathbf{A}'_{\text{loc}}]$
13. 堆叠得 $\mathbf{A}_{\text{Loc}} = [\mathbf{A}'_1, \mathbf{A}'_2, \dots, \mathbf{A}'_{\text{obs}}]$, $\mathbf{F}_{\text{Loc}} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_{\text{obs}}]$

3.3.2 个体层次包络样本构造模块

个体层次包络样本构造模块如图3所示. 基于场景中原始行人轨迹序列样本,利用同一行人在时间维度上的关联关系,挖掘原始轨迹样本在时间层次上的邻接关系,从而构造一系列带有邻接矩阵的时间图结构;而后通过合适的交互力度筛选和归一化操作得到个体层次包络样本 $[\mathbf{F}_{\text{ind}}, \mathbf{A}'_{\text{ind}}]$,实现包络样本变换. 具体过程如下:

(1)个体层次包络样本的行人邻接关系构造

首先,针对同一场景不同时刻的同一行人个体层

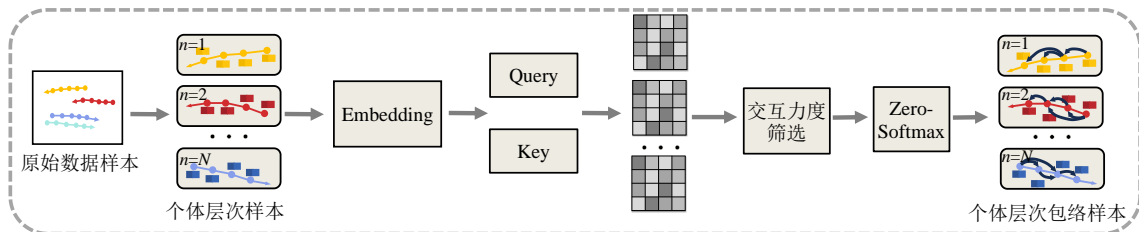


图3 个体层次包络样本构造模块

次轨迹关系,通过构建一系列时间图结构表示行人轨迹的个体层次样本;将时间段 $T_1 \sim T_{\text{obs}}$ 内观察的同一个行人的坐标 $P^i = \{p_t^i | \forall t \in \{T_1, T_2, \dots, T_{\text{obs}}\}\}$ 处理为图表示 $G_{\text{ind}} = (V_{\text{ind}}, E_{\text{ind}})$, 其中 $V_{\text{ind}} = \{p_t^1, p_t^2, \dots, p_t^N | \forall t \in \{T_1, T_2, \dots, T_{\text{obs}}\}\}$ 为 G_{ind} 的顶点集, $E_{\text{ind}} = \{e_{p,q}^1, e_{p,q}^2, \dots, e_{p,q}^N | \forall p, q \in \{T_1, T_2, \dots, T_{\text{obs}}\}\}$ 是 G_{ind} 的边集, E_{ind} 中各条边由行人邻接关系 R_{ind} 和邻接矩阵 A_{ind} 表示. $F_{\text{ind}} = \{v_t^1, v_t^2, \dots, v_t^N | \forall t \in \{T_1, T_2, \dots, T_{\text{obs}}\}\}$ 是行人在时间段 $T_1 \sim T_{\text{obs}}$ 内对应的特征向量, $v_t^i = (x_t^i - x_{t-1}^i, y_t^i - y_{t-1}^i)$ 代表时间轴 t 时刻的第 i 个行人的速度矢量.

个体层次包络样本的行人邻接关系 R_{ind} 可由注意力机制得到, Fra_t 表示 t 时刻的时间帧数, $E = [\text{Fra}_{T_1}, \text{Fra}_{T_2}, \dots, \text{Fra}_{T_{\text{obs}}}]$ 对应单个行人轨迹 p_t^i 在时间段 $T_1 \sim T_{\text{obs}}$ 内的位置编码张量. 在原始的个体层次特征向量 F_{ind} 基础上加入位置编码张量 E , 通过可学习参数 W_{ind} 进行线性变换, 得到带有位置编码信息的高维张量 F'_{ind} .

$$F'_{\text{ind}} = (F_{\text{ind}} \oplus E) W_{\text{ind}} \quad (8)$$

然后, 通过可学习参数 W_Q 和 W_K 分别得到自注意力机制的 query 和 key. 最后, 对 query 和 key 的乘积采用 D 进行放缩保持稳定性, 并用 Softmax 函数进行归一化, 可得个体层次包络样本行人邻接关系 R_{ind} .

$$\begin{aligned} Q_{\text{ind}} &= F'_{\text{ind}} W_Q \\ K_{\text{ind}} &= F'_{\text{ind}} W_K \\ R_{\text{ind}} &= \text{softmax} \left(\frac{Q_{\text{ind}} K_{\text{ind}}^T}{D} \right) \end{aligned} \quad (9)$$

(2) 个体层次包络样本交互力度筛选

类似于局部层次包络样本交互力度筛选, 通过设定阈值超参数 $\psi \in [0, 1]$ 对 R_{ind} 进行个体层次包络样本交互力度的筛选, 生成个体层次掩码 M_{ind} .

$$M_{\text{ind}} = \mathbb{I}(R_{\text{ind}} \geq \psi) \quad (10)$$

将 M_{ind} 与个体层次行人邻接关系 R_{ind} 融合后, 得到个体层次邻接矩阵 A_{ind} .

$$A_{\text{ind}} = M_{\text{ind}} \odot R_{\text{ind}} \quad (11)$$

通过式(7)对个体层次包络样本邻接矩阵进行归一化操作得 $A'_{\text{ind}} = \text{Zero-Softmax}(A_{\text{ind}})$. IESC 的伪代码描述如算法 2 所示.

3.3.3 双级包络样本特征融合模块

基于局部层次包络样本 $[F_{\text{loc}}, A'_{\text{loc}}]$, 将时间轴 T_1 时刻到 T_{obs} 时刻的 A'_{loc} 堆叠为 $A_{\text{Loc}} = [A'_1, A'_2, \dots, A'_{\text{obs}}]$; 将时间轴 T_1 时刻到 T_{obs} 时刻的各个顶点特征矩阵 F_{loc} 堆叠为 $F_{\text{Loc}} = [F_1, F_2, \dots, F_{\text{obs}}]$; 再对局部层次样本的图结构执行图卷积运算, W_{Loc} 是可学习参数, 可得局部层次包络样本图神经网络输出.

$$F'_{\text{Loc}} = \sigma(A_{\text{Loc}} F_{\text{Loc}} W_{\text{Loc}}) \quad (12)$$

再基于个体层次包络样本 $[F_{\text{ind}}, A'_{\text{ind}}]$, 将时间轴 T_1

算法 2 个体层次包络样本构造模块-IESC

输入: 行人轨迹坐标 $X_t^i = [x_t^i, y_t^i], i = 1, 2, \dots, N, t = T_1, T_2, \dots, T_{\text{obs}}$

输出: 个体层次包络样本 $[F_{\text{ind}}, A_{\text{ind}}]$

步骤:

1. 对行人轨迹坐标 X_t^i 按照时间近邻的关系构造个体层次特征 F_{ind} ;
2. For $i = 1$ to N
3. 通过式(8)、式(9)求得个体层次包络样本行人邻接关系 R_{ind} ;
4. End for
5. 通过式(10)、式(11)得到个体层次邻接矩阵 A_{ind} ;
6. 通过式(7)进行归一化操作, 构建个体层次包络样本 $[F_{\text{ind}}, A'_{\text{ind}}]$;
7. 堆叠得 $A_{\text{Ind}} = [A'_1, A'_2, \dots, A'_N], F_{\text{Ind}} = [F_1, F_2, \dots, F_N]$

时刻到 T_{obs} 时刻的所有行人的邻接矩阵 A'_{ind} 堆叠为 $A_{\text{Ind}} = [A'_1, A'_2, \dots, A'_N]$, 各个节点的特征矩阵 F_{ind} 堆叠为 $F_{\text{Ind}} = [F_1, F_2, \dots, F_N]$, 并对构建的个体层次样本图结构执行图卷积运算, W_{Ind} 是可学习参数, 可得个体层次包络样本图神经网络输出如下式所示:

$$F'_{\text{Ind}} = \sigma(A_{\text{Ind}} F_{\text{Ind}} W_{\text{Ind}}) \quad (13)$$

为表征行人轨迹数据在不同层面的时空特征, 后将多层图卷积网络得到的局部层次特征表示 F'_{Loc} 和个体层次特征表示 F'_{Ind} 进行拼接融合, 得到双级包络样本的特征表示 F_{tef} :

$$F_{\text{tef}} = F'_{\text{Ind}} \oplus F'_{\text{Loc}} \in \mathbb{R}^{N \times L_{\text{obs}} \times 2D} \quad (14)$$

3.4 全局层次包络样本构造模块

将融合的双级包络样本的特征 F_{tef} 输入多层 TXP-CNN, 可得到预测轨迹. 形式上, 对于第 l 层 TXP-CNN:

$$F_{\text{tef}}^{(l+1)} = \text{TXP-CNN}(F_{\text{tef}}^{(l)}; W^{(l)}) \quad (15)$$

其中, $W^{(l)}$ 为第 l 层可学习参数, $F_{\text{tef}}^{(l+1)} \in \mathbb{R}^{N \times L_{\text{pre}} \times 2D}$ 表示预测输出. 将第 l 层输出 $F_{\text{tef}}^{(l+1)}$ 通过一个线性层确定预测轨迹坐标二元高斯分布的 5 个参量^[23], 其中 W_l 为该线性层的可学习参数.

$$F_{\text{final}} = \text{Linear}(F_{\text{tef}}^{(l+1)}; W_l) = (\hat{\mu}_t^i, \hat{\sigma}_t^i, \hat{\rho}_t^i) \quad (16)$$

基于二元高斯分布参量中的轨迹均值参量 $\hat{\mu}_t^i = (\hat{\mu}_x^i, \hat{\mu}_y^i)_t^i$, 引入时空信息协同整合机制, 构造全局层次包络样本. 计算每个行人在每个时间帧的线速度 u_t^i , 再将其在时间段 $T_{\text{obs}+1} \sim T_{\text{pre}}$ 的线速度形成向量 U^i , 并拼接场景中 N 个行人的 U^i 得到全局样本 $F_{\text{glo}} \in \mathbb{R}^{N \times L_{\text{pre}}}$, 其中 L_{pre} 表示预测的轨迹长度.

$$u_t^i = \sqrt{(\hat{\mu}_x^i)_t^i + (\hat{\mu}_y^i)_t^i}$$

$$U^i = [u_{\text{obs}+1}^i, u_{\text{obs}+2}^i, \dots, u_{\text{pre}}^i]^T \in \mathbb{R}^{L_{\text{pre}} \times 1} \quad (17)$$

$$F_{\text{glo}} = [U^1, U^2, \dots, U^N]^T \in \mathbb{R}^{N \times L_{\text{pre}}}$$

基于融合的双级包络样本, 从全局视角出发对时间层面和空间层面信息进行捕捉与整合, 构造全局层次包络样本特征表示, 包含时间层面和空间层面全局关系、供源域和目标域的全局层次域适应.

3.5 分级包络域适应模块

为解决源域和目标域轨迹特征分布不一致问题,本文构建了多层次的分级包络域适应模块.与现有域适应方法显著不同的是,行人轨迹预测中的单个“样本”由不同数量行人轨迹组合而成,即存在不同尺度、不同维度的问题.对不同尺度样本提取统一特征会导致无法精细化表征样本特征,产生信息丢失.一方面,本文采用基于注意力机制的样本特征提取方式,可以更有效地适应不同样本尺度,避免了特征尺度表示误差导致的局部层次以及全局层次域适应不精准的问题.另一方面,局部层次和全局层次包络样本相结合,包含了所有行人运动信息,能够很好地解决域偏移问题.

(1) 局部层次包络域适应模块

源域和目标域局部层次包络样本特征表示分别为 $F'_{\text{Loc(s)}} \in \mathbb{R}^{N_s \times L_{\text{obs}} \times D}$ 和 $F'_{\text{Loc(t)}} \in \mathbb{R}^{N_t \times L_{\text{obs}} \times D}$, 其中 N_s 和 N_t 分别表示源域和目标域局部层次样本的行人个数, L_{obs} 为观测轨迹长度.为关注局部层次多尺度特征表示,源域和目标域局部层次特征表示 $F'_{\text{Loc(s)}}$ 和 $F'_{\text{Loc(t)}}$ 可按行人进行分解:

$$\begin{aligned} F'_{\text{Loc(s)}} &= [f_{(s)}^1, f_{(s)}^2, \dots, f_{(s)}^{N_s}] \\ F'_{\text{Loc(t)}} &= [f_{(t)}^1, f_{(t)}^2, \dots, f_{(t)}^{N_t}] \end{aligned} \quad (18)$$

$f_{(s)}^i$ 和 $f_{(t)}^i$ 分别对应源域和目标域某个局部层次样本中的某个行人特征, $f_{(s)}^i \in \mathbb{R}^{L_{\text{obs}} \times D}$, $f_{(t)}^i \in \mathbb{R}^{L_{\text{obs}} \times D}$, 将 $f_{(s)}^i$ 和 $f_{(t)}^i$ 特征展平为 D_c 大小的特征向量, 其中 $D_c = L_{\text{obs}} \times D$. 特征向量 f^i 含有局部层次样本中某个行人的时空信息, 但源域和目标域局部层次特征表示维度并不相同, 按照经典域适应方法无法进行特征对齐操作, 且局部层级域适应需要挖掘行人之间的社交关系, 因此采用注意力机制学习目标行人与其对应的局部层次群体中行人之间的相对相关性, 并使得不同域的特征表示保持相同维度. 注意力分数计算方法如下所示:

$$\begin{aligned} Z_{(s)}^{i,j} &= \frac{(f_{(s)}^i f_{(s)}^j)^T / \sqrt{D_c}}{\sum_{j=1}^{N_s} (f_{(s)}^i f_{(s)}^j)^T / \sqrt{D_c}} \\ Z_{(t)}^{i,j} &= \frac{(f_{(t)}^i f_{(t)}^j)^T / \sqrt{D_c}}{\sum_{j=1}^{N_t} (f_{(t)}^i f_{(t)}^j)^T / \sqrt{D_c}} \end{aligned} \quad (19)$$

通过属于同一个局部层次包络样本中不同行人特征 $f_{(s)}^i$ 和 $f_{(s)}^j$ 的乘积表达交互关系权重, 并计算所有两两行人交互影响的累计和, 注意力分数为特定 2 个行人在其中所占的比例, 显然比例值越大, 交互影响越强, 否则越弱. 特征数 D_c 起归一化作用.

$$\begin{aligned} J_{(s)} &= \frac{\sum_{i=1}^{N_s} \left(\sum_{j=1}^{N_s} (Z_{(s)}^{i,j} f_{(s)}^i) \right)}{N_s} \\ J_{(t)} &= \frac{\sum_{i=1}^{N_t} \left(\sum_{j=1}^{N_t} (Z_{(t)}^{i,j} f_{(t)}^i) \right)}{N_t} \end{aligned} \quad (20)$$

$J_{(s)}$ 和 $J_{(t)}$ 是源域和目标域局部层次样本行人间关系交互影响的总和表达. 假设源域轨迹和目标域轨迹分别含有 $K_{(s)}$ 个和 $K_{(t)}$ 个局部层次样本, 则整个源域和目标域局部行人注意力特征可表示为 $M_{(s)}$ 和 $M_{(t)}$:

$$\begin{aligned} M_{(s)} &= [J_{(s)}^1, J_{(s)}^2, \dots, J_{(s)}^{K_{(s)}}] \\ M_{(t)} &= [J_{(t)}^1, J_{(t)}^2, \dots, J_{(t)}^{K_{(t)}}] \end{aligned} \quad (21)$$

基于上述局部层次行人关系注意力特征表示, 可引入局部层次域分布对齐损失 \mathcal{L}_{loc} :

$$\mathcal{L}_{\text{loc}} = \text{Dist}_1(M_{(s)}, M_{(t)}) \quad (22)$$

(2) 全局层次包络域适应模块

在引入局部层次包络域适应模块基础上, 构建全局层次包络域适应模块. 首先求出对源域轨迹和目标域轨迹在观测轨迹长度 L_{obs} 上的线速度差异倍率 κ , 而后针对源域和目标域全局层次包络样本特征 $F_{\text{glo(s)}}$ 和 $F_{\text{glo(t)}}$, 引入全局层次域分布对齐损失 \mathcal{L}_{glo} :

$$\mathcal{L}_{\text{glo}} = \text{Dist}_2(F_{\text{glo(s)}}, \kappa \times F_{\text{glo(t)}}) \quad (23)$$

距离函数 $\text{Dist}_1()$ 和 $\text{Dist}_2()$ 有多种选择, 如 MMD 损失、CORAL 损失、GFK 损失和 L2 距离等. 本文通过联合计算目标域和源域的局部层次特征和全局层次特征的分布差异, 进一步增强了网络的多层次跨域时空特征的迁移学习能力.

3.6 轨迹预测加权融合机制

假设认为行人坐标 (x_t^i, y_t^i) 遵循二元高斯分布 $(x_t^i, y_t^i) \sim \mathcal{N}(\hat{\mu}_t^i, \hat{\sigma}_t^i, \hat{\rho}_t^i)^{[15]}$, 其中 $\hat{\mu}_t^i = (\hat{\mu}_x^i, \hat{\mu}_y^i)^T$ 为均值, $\hat{\sigma}_t^i = (\hat{\sigma}_x^i, \hat{\sigma}_y^i)^T$ 为标准差, $\hat{\rho}_t^i$ 为相关系数. 因此, 预测轨迹的预测损失 \mathcal{L}_{tra} 可转化为最小化负对数似然损失.

$$\mathcal{L}_{\text{tra}} = - \sum_{i=1}^N \sum_{t=T_{\text{obs}}+1}^{T_{\text{rec}}} \log(P((x_t^i, y_t^i) | \hat{\mu}_t^i, \hat{\sigma}_t^i, \hat{\rho}_t^i)) \quad (24)$$

因此, 总体目标函数由 3 项损失组成. 模型通过联合最小化轨迹的预测损失 \mathcal{L}_{tra} 、源域和目标域的局部层次域分布对齐损失 \mathcal{L}_{loc} 和全局层次域分布对齐损失 \mathcal{L}_{glo} 进行训练.

$$\mathcal{L} = \mathcal{L}_{\text{tra}} + \sigma \mathcal{L}_{\text{loc}} + \lambda \mathcal{L}_{\text{glo}} \quad (25)$$

预测损失 \mathcal{L}_{tra} 只与源域样本相关, 与目标域样本无关. σ 和 λ 是平衡这 3 项损失的超参数.

4 实验结果与分析

4.1 实验环境

(1)数据集. 实验选择2个现实世界代表性公共数据集:ETH^[31]和UCY^[32]. 这2个数据集已经被大量行人轨迹预测方法研究广泛使用. 数据集含有行人与周边行人间社交互动以及行人与环境之间的交互,例如人群的聚集停留、人群的躲避碰撞和群体移动等. ETH数据集由ETH和HOTEL这2个场景组成,UCY数据集由UNIV、ZARA1、ZARA2这3个场景组成. 其中,UCY数据集场景偏向人流密集的场景,往往蕴含更复杂行人社交互动,而ETH数据集场景偏向人流稀疏场景,意味着行人间社交互动较少.

(2)实验设置. 参照现有基于域适应的轨迹预测研究公认范式,将每个场景视为一个轨迹域,模型首先在一个域上训练,然后在其他4个域的验证集上进行验证,最终分别在其他4个域上进行测试,通过比较其预测准确率来验证模型的有效性. 本文用A、B、C、D、E分别代表ETH、HOTEL、UNIV、ZARA1、ZARA2这5个场景的行人轨迹域,则总共有20个轨迹预测任务:A→B/C/D/E, B→A/C/D/E, C→A/B/D/E, D→A/B/C/E, E→A/B/C/D. 在训练阶段,本文模型只能访问验证集中观察到的轨迹,即验证集仅作为域适应用途. 实验完整版参见<https://github.com/LWZ9910/MESC-HEDA.git>.

(3)评估指标. 模型性能评估指标采用平均位移误差(ADE)和最终位移误差(FDE). ADE用于评估预测轨迹的平均预测精度,FDE用于评估最后一帧轨迹端点的预测精度. 目前由于Social-LSTM^[15]、Social-STGCNN^[23]和Social-GAN^[18]都采用模型输出服从双变

量高斯分布的评估准则,本文也遵循这一统一准则用于性能比较.

(4)参数设置. 局部层次包络样本的GCN层数为1,局部层次包络样本交互力度 η 设置为0.05,个体层次包络样本的GCN层数为2,个体层次包络样本交互力度 ψ 设置为0.7,双级包络样本特征融合后的TXP-CNN层数为5,特征维度设置为64. 在训练阶段,批量大小设置为16,局部层次域适应损失权重超参数 σ 设置为9,全局层次域适应损失权重超参数 λ 设置为5. 模型采用随机梯度下降法进行了200次训练,初始学习率为0.01,经过100次后学习率改为0.005. 实验软件平台为PyTorch=2.3,硬件平台采用CPU i9-13900H,GPU采用RTXA3000,内存64 GB. 观测行人历史轨迹为3.2 s(8个时间帧),预测轨迹为4.8 s(12个时间帧).

4.2 消融实验

为验证各个模块的有效性,对本文提出的MESC-HEDA模型进行消融实验. 主要分为4种情况:

(1)LESC(Local Envelope Sample Construction):基于局部层次包络样本的行人轨迹预测模型.

(2)BESC(Bilevel Envelope Sample Construction):基于双级包络样本的行人轨迹预测模型.

(3)BESC-LEDA(Bilevel Envelope Sample Construction - Local Envelope Domain Adaptation):基于双级包络样本和局部层次域适应的行人轨迹预测模型.

(4)MESC-HEDA(Multilevel Envelope Sample Construction-Hierarchical Envelope Domain Adaptation):本文提出的完整模型,基于多级包络样本和分级包络域适应的行人轨迹预测模型.

消融实验的结果如表1所示.

表1 消融实验

Variants	模块组件			AVE
	双级包络样本构造模块	局部域适应模块	分级包络域适应模块	ADE/FDE
LESC	×	×	×	0.65/1.13
BESC	√	×	×	0.62/1.06
BESC-LEDA	√	√	×	0.56/0.96
MESC-HEDA	√	√	√	0.51/0.89

LESC和BESC的对比结果表明,“双级包络样本构造模块”结合了个体层次包络样本和局部层次包络样本构造,相比于仅使用局部层次包络样本的情况,能更有效地捕获行人轨迹的个体层次和局部层次的特征信息,使得预测结果更准确. 其中,ADE降低了4.6%,FDE降低了6.2%. BESC和BESC-LEDA的对比结果表明,引入“局部域适应模块”后,实现了源域和目标域的局部层次时空特征分布对齐. 相比于无域适应模块的情况,ADE降低了9.7%,FDE降低了9.4%. BESC-LEDA和

MESC-HEDA的对比结果表明,实现多级包络样本构造,并引入“分级包络域适应模块”后,通过不同域的局部层次特征和全局层次特征的分布对齐,进一步增强了模型在跨域多层次时空特征的迁移学习能力. 相比于单层次域适应算法,ADE降低了8.9%,FDE降低了7.3%. 以上实验结果说明,引入相关模块,性能得到了相应提升.

4.3 参数影响分析

局部层次包络样本映射核函数比较分析,以及不

同核函数交互力度比较分析(缩写“w/”则代表“with”,即表示加入某一模块).

如表2所示,不同核函数构造的局部层次包络样本对 ADE/FDE 指标有较大影响.式(7)所示局部层次行人邻接矩阵 $\mathbf{M}_{(s)}$ 构造,综合考虑了行人群体的速度、距离、夹角和交互力度等因素,并与其他核函数进行比较分析.表2中 LESC 指仅构造局部层次包络样本,而不考虑双级包络样本的构建和域适应模块.此外,本文还评估了核函数的6种变体, $a_{L_2,t}^{ij}$ 采用 L2 范数以及 $a_{sim,t}^{ij}$ 采用 L2 范数的倒数模拟行人之间的相互影响^[23], $a_{att,t}^{ij}$ 采用图注意力网络对行人之间的交互关系进行更新^[10], $a_{vl,t}^{ij}$ 对综合速度和距离进行定义如式(26)所示, $a_{cos,t}^{ij}$ 对综合角度和距离进行定义如式(27)所示, $a_{com,t}^{ij}$ 对综合速度、角度和距离进行定义^[25].

表2 不同核函数的比较

Variants	ADE	FDE
LESC w/ $a_{L_2,t}^{ij}$	1.13	1.85
LESC w/ $a_{sim,t}^{ij}$	1.22	1.98
LESC w/ $a_{att,t}^{ij}$	0.69	1.16
LESC w/ $a_{vl,t}^{ij}$	0.72	1.25
LESC w/ $a_{cos,t}^{ij}$	0.83	1.46
LESC w/ $a_{com,t}^{ij}$	0.72	1.23
LESC-V1	0.68	1.17
LESC-V2	0.65	1.13
LESC-V3	0.68	1.20
LESC-V4	0.67	1.14

$$a_{vl,t}^{ij} = \max\left(0, \frac{|\mathbf{v}_i^i| + |\mathbf{v}_i^j|}{l_{ij}^i}\right) \quad (26)$$

$$a_{cos,t}^{ij} = \max\left(0, \frac{\cos c_t^{ij} + \cos c_t^{ji}}{l_{ij}^i}\right) \quad (27)$$

此外, LESC-V 指通过式(7)构造局部层次包络样本的邻接矩阵,并设置不同的 η 值进行交互力度筛选的实验.其中, LESC-V1 指设置 $\eta=0$, 即不筛选局部层次包络样本的交互力度, 此时局部层次行人邻接关系非常密集. LESC-V2 指设置 $\eta=0.05$, 即适当筛选局部层次包络样本的交互力度, 此时局部层次行人邻接关系合理且适宜. LESC-V3 指设置 $\eta=0.1$, 即进一步筛选局部层次包络样本的交互力度, 此时局部层次行人邻接关系较为稀疏. LESC-V4 指设置 $\eta=0.2$, 即严格筛选局部层次包络样本的交互力度, 此时局部层次行人邻接关系非常稀疏. LESC-V2 对应本文方法的实验设置, 如表2所示. 相较于其他核函数和其他交互力度的选择, 本文局部层次包络样本能够取得相对更优的整体平均性能, 这意味着对局部层次包络样本的交互力度进行一

定程度筛选是必要且有效的.

超参数 σ 用于平衡损失方程中的轨迹损失和局部域对齐损失, 当模型只引入局部层次域适应时, 对 σ 取不同数值进行性能比较, 如表3所示, 参数若设置太大会导致域对齐比重过大, 相反则会导致域对齐无法发挥有效作用. 通过设置6组数值 {0.1, 1, 3, 6, 9, 10} 进行实验可找出较优 σ 值. 实验结果表明, 当设置 σ 为9时, 模型可获得最佳性能, ADE 指标为0.56, FDE 指标为0.96. 模型引入分级包络域适应后, 即同时考虑局部层次域适应和全局层次域适应, 固定 σ 取值为9, 并对 λ 取不同数值进行性能比较, 如表4所示, 当设置 λ 为5时, 模型可以获得最佳性能, ADE 指标为0.51, FDE 指标为0.89. 当 λ 为10时, 会导致域分布对齐过多, 反而影响模型预测精度, 参数若设置太小会导致域分布对齐无法有效发挥作用.

表3 不同 σ 值平均性能对比

σ	0.1	1	3	6	9	10
ADE	0.69	0.66	0.62	0.61	0.56	0.60
FDE	1.19	1.13	1.09	1.07	0.96	1.03

表4 不同 λ 值平均性能对比

λ	0.1	1	5	10
ADE	0.59	0.63	0.51	0.71
FDE	1.03	1.06	0.89	1.04

4.4 方法对比

在前述同样的实验设置下, 将5种相关模型 STGAT^[33]、Transformer^[34]、Social-STGCNN^[23]、SGCN^[30]、T-GNN^[10] 与本文提出的 MESC-HEDA 模型进行比较. 表5和表6分别统计了该模型在20个场景任务组合中的 ADE 和 FDE 度量结果.

如表5和表6所示, 本文所提出的 MESC-HEDA 模型的 ADE 和 FDE 指标平均值始终优于其他5个参考模型. 总体而言, MESC-HEDA 模型的平均位移误差相比 SGCN 和 T-GNN 模型降低了 25.0% 和 22.7%, MESC-HEDA 模型的终点位移误差相比 SGCN 和 T-GNN 模型降低了 23.3% 和 19.8%. MESC-HEDA 模型在20项任务中有14项任务的 ADE 指标表现最优, 在20项任务中有12项任务的 FDE 指标表现最优. 具体表现如下:

(1) 在人流稀疏场景下的表现: 例如, 在以场景 Hotel 作为目标域的情况下, MESC-HEDA 模型在对应的4项任务中的3项任务的 ADE 和 FDE 指标均表现最优, 平均 ADE 为0.51, 平均 FDE 为0.86. 相较于 Transformer 模型, 平均 ADE 为0.69, 平均 FDE 为1.17, MESC-HEDA 模型在 ADE 和 FDE 上分别降低了 26.1% 和 26.5%.

(2) 在人流密集场景下的表现: 例如, 在以场景 ZARA1 作为目标域的情况下, MESC-HEDA 模型在对应

的4项任务中的3项任务的ADE和FDE指标均表现最优,平均ADE为0.39,平均FDE为0.67.相较于次优的T-GNN模型,平均ADE为0.66,平均FDE为1.10,MESC-HEDA模型在ADE和FDE上分别降低了40.9%和39.1%.

从上述结果可以看出,无论是在人流稀疏场景还是人流密集场景的不同跨域任务下,MESC-HEDA模型均表现出了良好的性能,模型泛化能力强.可能的原因是,分级包络域适应模型通过构建多级包络样本,获取了时空序列数据在不同层次更为丰富的时空特征表

征,同时通过学习从源域到目标域可迁移的多层次时空特征信息实现域分布对齐,进一步增强了模型多层次跨域时空特征迁移学习能力.从标准差的角度来看,MESC-HEDA模型的ADE和FDE指标在20项任务中的平均标准差均为0.003,是所有模型的20项任务中平均标准差最小的.这一低标准差表明本文提出的模型在多次实验中的表现具有较高的稳定性,其标准差显著低于其他模型,说明其预测性能相比其他模型更为一致和可靠.

表5 MESC-HEDA模型与参考模型的ADE比较

Method		STGAT ^[33]	Transformer ^[34]	Social-STGCNN ^[23]	SGCN ^[30]	T-GNN ^[10]	MESC-HEDA
Source	Target						
A	B	1.19±0.005	1.04±0.009	1.13±0.014	0.84±0.009	0.92±0.005	0.90±0.007
	C	0.82±0.002	0.84±0.008	0.72±0.001	0.57±0.001	0.99±0.001	0.53±0.001
	D	0.54±0.001	0.65±0.010	0.67±0.004	0.45±0.002	0.77±0.004	0.44±0.002
	E	0.52±0.005	0.49±0.004	0.62±0.003	0.35±0.001	0.79±0.003	0.40±0.001
B	A	1.42±0.010	1.41±0.016	1.86±0.028	2.17±0.034	1.39±0.025	1.12±0.006
	C	0.84±0.001	1.04±0.004	0.97±0.001	0.81±0.001	0.83±0.001	0.56±0.001
	D	1.47±0.003	2.14±0.004	1.64±0.006	1.76±0.006	1.17±0.008	0.53±0.004
	E	0.95±0.003	1.29±0.009	1.10±0.002	1.06±0.001	0.88±0.003	0.79±0.002

表6 MESC-HEDA模型与参考模型的FDE比较

Method		STGAT ^[33]	Transformer ^[34]	Social-STGCNN ^[23]	SGCN ^[30]	T-GNN ^[10]	MESC-HEDA
Source	Target						
A	B	2.22±0.007	1.77±0.018	1.71±0.022	1.47±0.012	1.38±0.005	1.43±0.007
	C	1.61±0.003	1.16±0.013	0.85±0.003	0.93±0.002	1.56±0.001	0.78±0.001
	D	1.08±0.003	0.93±0.027	0.96±0.018	0.58±0.003	1.14±0.004	0.66±0.002
	E	1.05±0.011	0.64±0.011	0.78±0.008	0.56±0.001	1.13±0.003	0.56±0.001
B	A	2.18±0.020	2.25±0.029	2.61±0.109	3.01±0.072	2.40±0.025	1.89±0.006
	C	1.65±0.002	1.84±0.005	1.59±0.005	1.49±0.001	1.48±0.001	1.02±0.001
	D	2.76±0.006	3.91±0.009	2.93±0.012	3.31±0.007	2.02±0.008	0.98±0.004
	E	1.75±0.005	2.32±0.014	1.97±0.011	2.00±0.003	1.63±0.003	1.45±0.002

4.5 预测轨迹对比

本节定性分析本文模型、T-GCN和Social-STGCNN这3种模型的轨迹预测准确性.图4可视化3种模型对行人在并行、相遇、静止和密集场景下的预测轨迹变化,从而分析不同模型的性能差异.红色实线表示历史轨迹,蓝色实线表示真实轨迹,黄色虚线表示预测轨迹,蓝色实线和黄色虚线的重合度越高意味着预测精度越高.

图4(a)和图4(d)面向的是行人并行和静止的情况,分别以ETH和ZARA2场景作为目标域,以ZARA1场景作为源域.源域和目标域之间存在域差异,特别是ETH场景中的行人密度较少且行人运动速度较快.从图4(a)可以看出,T-GCN可以预测基本符合现实的轨迹,但精度要差于本文的MESC-HEDA模型,而Social-STGCNN模型表现较差,预测的轨迹较为曲折,重合度

较低.其中可能的原因是本文的MESC-HEDA模型引入了分级包络域适应模块,从而充分考虑了局部层次和全局层次的分布对齐.

图4(b)、图4(c)和图4(e)面向的是多个行人相遇和静止的情况,分别以UNIV和ZARA2场景作为目标域,以ZARA1场景作为源域,此时源域和目标域之间存在的域差异相对较小,但场景中行人密度较多,导致行人间的交互关系更为复杂.从图4(b)可以看出,MESC-HEDA模型能够准确预测行人结伴而行和有效避让了迎面相遇行人的轨迹,而T-GCN和Social-STGCNN模型表现稍差,轨迹重合度较低.从图4(c)和图4(e)可以看出,3个模型都捕获了静止行人不受其他行人影响的事实,但T-GCN和Social-STGCNN模型在避让静止行人时产生了多余的交互,预测精度较差,而MESC-HEDA模型能够准确预测行人有效避让静止行人,且能够合理

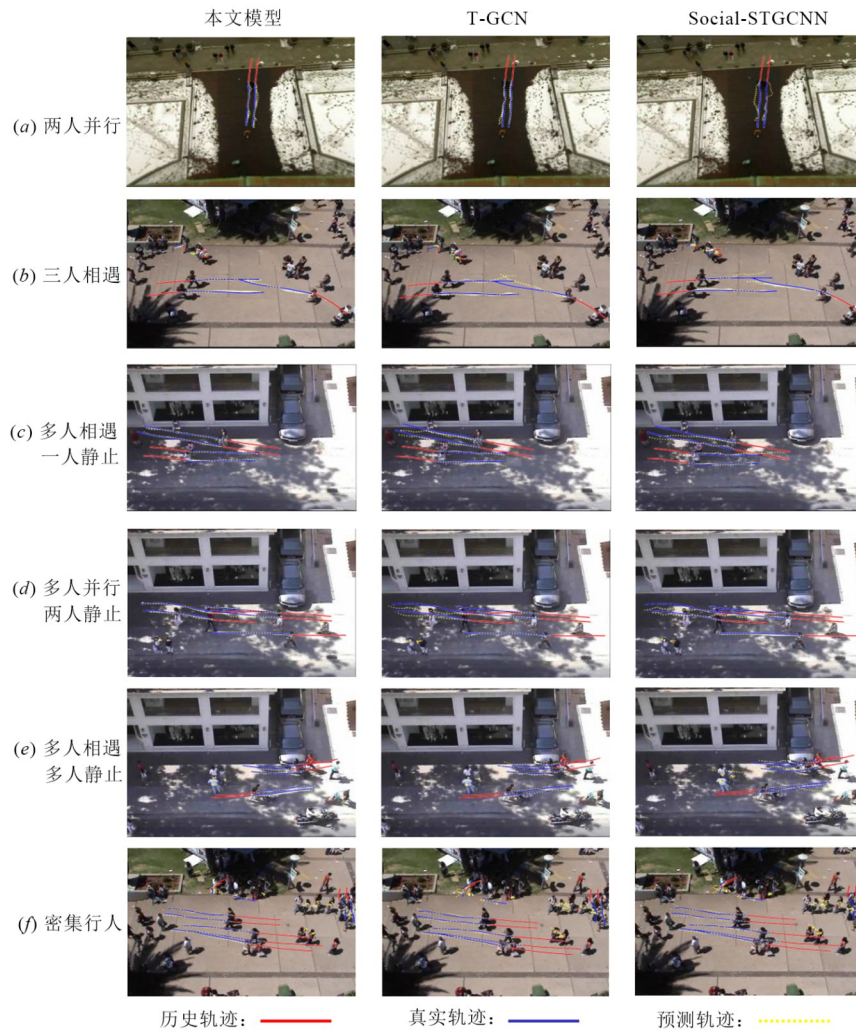


图4 预测轨迹的可视化

地预测更贴近真实的轨迹。

图4(f)面向的是密集场景中的行人并行、相遇、静止和行走中止的情况,以UNIV场景作为目标域,以ZARA1场景作为源域,其中目标域场景中的行人密度很高,行人之间存在着更加复杂的交互,建立的图结构较为复杂。从图4(f)可以看出,T-GCN和Social-STGCNN模型虽然能够预测静止行人的状态,但是在预测不同方向的行人结伴而行时精度较差,且并未判断出行人行走中止的情况。相比之下,本文的模型能够准确预测基本符合现实的轨迹,包括不同方向的行人结伴而行、行人静止和行人行走中止的情况。

4.6 时间开销对比

在前述同样的实验设置下,本节比较各个模型的时间开销。表7和表8分别统计了各模型在20个场景任务组合中的训练运行时间和测试运行时间的结果。

如表7和表8所示,本文所提出的MESC-HEDA模型的训练运行时间远低于Transformer模型的训练运

行时间,且本文提出的MESC-HEDA模型在20项任务中的平均测试运行时间为15.5 s,处于运行效率最高的模型行列之中。这一结果充分证明了该模型在执行效率上的优势,其开销完全契合实际应用场景的需求。进一步结合表5和表6中模型的预测准确率进行综合考量,本文模型兼具适宜的时间开销和较高的预测精度,相较于其他被比较的模型,具有显著的性能优势。

5 结论

针对不同的行人轨迹场景之间轨迹数据分布差异大的问题,本文提出了一种基于分级包络域适应的行人轨迹预测模型。通过构造多级包络样本,并基于此构建分级包络域适应轨迹预测模型,从而充分构建了行人轨迹数据在不同层面的时空特征表征,并有效地应对了不同场景轨迹域之间分布差异大的问题。实验结果表明,本文所提出的多级包络样本和分级包络域适应相对其他代表性算法具有更优性能,能够有效应对行人之间、行人与环境之间交互的多层次复杂性的问

表 7 MESC-HEDA 模型与参考模型的训练运行时间比较

单位: s

Method		STGAT ^[33]	Transformer ^[34]	Social-STGCNN ^[23]	SGCN ^[30]	T-GNN ^[10]	MESC-HEDA
Source	Target						
A	B	878.64	2 676.78	394.23	1 185.42	641.98	709.74
	C	875.06	52 766.17	545.24	1 634.92	967.22	733.09
	D	885.71	5 267.70	436.21	1 378.27	764.21	711.60
	E	870.05	11 889.72	552.47	1 622.37	924.98	751.36
B	A	949.12	1 895.08	628.73	1 993.40	1 636.80	2 848.63
	C	966.00	53 896.88	936.75	2 655.61	2 077.10	2 973.50
	D	975.86	6 357.88	804.09	2 388.25	1 888.71	2 980.04
	E	950.33	12 624.26	922.89	2 797.22	2 056.27	3 008.40

表 8 MESC-HEDA 模型与参考模型的测试运行时间比较

单位: s

Method		STGAT ^[33]	Transformer ^[34]	Social-STGCNN ^[23]	SGCN ^[30]	T-GNN ^[10]	MESC-HEDA
Source	Target						
A	B	11.40	31.00	7.26	9.60	6.76	6.13
	C	26.59	944.75	47.56	49.95	47.67	49.43
	D	18.16	79.22	7.43	10.83	7.64	7.15
	E	24.30	217.13	16.19	19.21	16.05	14.66
B	A	6.29	11.86	0.84	3.18	0.84	0.87
	C	26.90	1 109.46	48.41	50.68	47.20	49.81
	D	18.16	86.09	7.22	11.05	7.88	8.18
	E	24.08	240.16	15.07	18.88	16.06	15.24

题. 此外, 本文模型可与其他模型相结合, 对于相关研究具有较好的参考价值.

行人轨迹预测可以用于自动驾驶中行人的行为预判, 以繁忙的十字路口为例, 行人的运动受到红绿灯、其他行人和车辆的多重影响, 其轨迹呈现出高度的复杂性和不确定性. MESC-HEDA 模型能够综合考虑不同场景轨迹域的多种因素, 预测行人的行为轨迹变化. 这样的预测结果对于自动驾驶车辆来说至关重要, 有助于车辆提前做出合理的行驶规划, 避免潜在的碰撞风险. 下一步将考虑针对不同类型的场景, 在追求模型普适性的同时, 还需根据具体应用场景的特点进行更加精细化的模型调优与适配. 针对本文模型的潜在局限性的分析, 拟通过统计分析或关联机制进行最优的超参数设置和其可能的搜索范围, 或引入自适应机制以动态调整模型的相关阈值参数和权重, 以适配不同场景, 克服当前限制, 从而进一步提升模型的性能和适用性.

参考文献

- [1] 孔玮, 刘云, 李辉, 等. 基于全局自适应有向图的行人轨迹预测[J]. 电子学报, 2022, 50(8): 1905-1916.
KONG W, LIU Y, LI H, et al. Pedestrian trajectory prediction based on global adaptive directed graph[J]. Acta Electronica Sinica, 2022, 50(8): 1905-1916. (in Chinese)
- [2] ZHANG H J, LIU Y J, WANG C, et al. Research on a pedestrian crossing intention recognition model based on natural observation data[J]. Sensors, 2020, 20(6): 1776.
- [3] MOUSSAÏD M, PEROZO N, GARNIER S, et al. The walking behaviour of pedestrian social groups and its impact on crowd dynamics[J]. PLoS One, 2010, 5(4): e10047.
- [4] HELBING D, MOLNÁR P. Social force model for pedestrian dynamics[J]. Physical Review E, 1995, 51(5): 4282-4286.
- [5] SCHNEIDER N, GAVRILA D M. Pedestrian path prediction with recursive bayesian filters: A comparative study[M]// Pattern Recognition. Berlin: Springer Berlin Heidelberg, 2013: 174-183.
- [6] 裴焰, 邱文涛, 王森, 等. 基于Transformer动态场景信息生成对抗网络的行人轨迹预测方法[J]. 电子学报, 2022, 50(7): 1537-1547.
PEI Z, QIU W T, WANG M, et al. Pedestrian trajectory prediction method using dynamic scene information based transformer generative adversarial network[J]. Acta Electronica Sinica, 2022, 50(7): 1537-1547. (in Chinese)
- [7] ZHU W J, LIU Y H, ZHANG M Y, et al. Reciprocal consistency prediction network for multi-step human trajectory prediction[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(6): 6042-6052.

- [8] LI Y M, LIANG R Q, WEI W, et al. Temporal pyramid network with spatial-temporal attention for pedestrian trajectory prediction[J]. *IEEE Transactions on Network Science and Engineering*, 2022, 9(3): 1006-1019.
- [9] 桑海峰, 陈旺兴, 王海峰, 等. 基于多模式时空交互的行人轨迹预测模型[J]. *电子学报*, 2022, 50(11): 2806-2812. SANG H F, CHEN W X, WANG H F, et al. Pedestrian trajectory prediction model based on multi-model SpaceTime interaction[J]. *Acta Electronica Sinica*, 2022, 50(11): 2806-2812. (in Chinese)
- [10] XU Y, WANG L C, WANG Y Z, et al. Adaptive trajectory prediction via transferable GNN[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 6510-6521.
- [11] ZHI C Y, SUN H J, XU T. Adaptive trajectory prediction without catastrophic forgetting[J]. *The Journal of Supercomputing*, 2023, 79(14): 15579-15596.
- [12] YAN X, KAKADIARIS I A, SHAH S K. Modeling local behavior for predicting social interactions towards human tracking[J]. *Pattern Recognition*, 2014, 47(4): 1626-1641.
- [13] FERRER G, SANFELIU A. Behavior estimation for a complete framework for human motion prediction in crowded environments[C]//2014 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE, 2014: 5940-5945.
- [14] SCHULZ A T, STIEFELHAGEN R. A controlled interactive multiple model filter for combined pedestrian intention recognition and path prediction[C]//2015 IEEE 18th International Conference on Intelligent Transportation Systems. Piscataway: IEEE, 2015: 173-178.
- [15] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: Human trajectory prediction in crowded spaces[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 961-971.
- [16] XUE H, HUYNH D Q, REYNOLDS M. SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway: IEEE, 2018: 1186-1194.
- [17] HUG R, BECKER S, HÜBNER W, et al. Particle-based pedestrian path prediction using LSTM-MDL models[C]//2018 21st International Conference on Intelligent Transportation Systems (ITSC). Piscataway: IEEE, 2018: 2684-2691.
- [18] GUPTA A, JOHNSON J, LI F F, et al. Social GAN: Socially acceptable trajectories with generative adversarial networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 2255-2264.
- [19] AMIRIAN J, HAYET J B, PETTRÉ J. Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE, 2019: 2964-2972.
- [20] SADEGHIAN A, KOSARAJU V, SADEGHIAN A, et al. SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 1349-1358.
- [21] NIKHIL N, MORRIS B T. Convolutional neural network for trajectory prediction[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019: 186-196.
- [22] YAN S J, XIONG Y J, LIN D H. Spatial temporal graph convolutional networks for skeleton-based action recognition[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, 32(1): 77-82.
- [23] MOHAMED A, QIAN K, ELHOSEINY M, et al. Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 14424-14432.
- [24] ZHANG C, BERGER C, DOZZA M. Social-IWSTCNN: A social interaction-weighted spatio-temporal convolutional neural network for pedestrian trajectory prediction in urban traffic scenarios[C]//2021 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE, 2021: 1515-1522.
- [25] LV P, WANG W T, WANG Y X, et al. SSAGCN: Social soft attention graph convolution network for pedestrian trajectory prediction[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(9): 11989-12003.
- [26] HUANG P X, CUI Z H, LI J, et al. Cross-domain trajectory prediction with CTP-Net[M]//Artificial Intelligence. Cham: Springer Nature, 2022: 80-92.
- [27] GENG M S, LI J Y, LI C J, et al. Adaptive and simultaneous trajectory prediction for heterogeneous agents via transferable hierarchical transformer network[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(10): 11479-11492.
- [28] LI F, ZHANG X H, WANG P, et al. Deep instance enve-

lope network-based imbalance learning algorithm with multilayer fuzzy C-means clustering and minimum inter-layer discrepancy[J]. Applied Soft Computing, 2022, 123: 108846.

- [29] LI F, LI Y M, SHEN Y H, et al. Deep fuzzy envelope sample generation mechanism for imbalanced ensemble classification[J]. IEEE Transactions on Fuzzy Systems, 2024, 32(3): 1248-1262.
- [30] SHI L S, WANG L, LONG C J, et al. SGCN: Sparse graph convolution network for pedestrian trajectory prediction[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2021: 8994-9003.
- [31] PELLEGRINI S, ESS A, SCHINDLER K, et al. You'll never

walk alone: Modeling social behavior for multi-target tracking[C]//2009 IEEE 12th International Conference on Computer Vision. Piscataway: IEEE, 2009: 261-268.

- [32] LERNER A, CHRYSANTHOU Y, LISCHINSKI D. Crowds by example[J]. Computer Graphics Forum, 2007, 26(3): 655-664.
- [33] HUANG Y F, BI H K, LI Z X, et al. STGAT: Modeling spatial-temporal interactions for human trajectory prediction[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 6272-6281.
- [34] GIULIARI F, HASAN I, CRISTANI M, et al. Transformer networks for trajectory forecasting[C]//2020 25th International Conference on Pattern Recognition (ICPR). Piscataway: IEEE, 2021: 10335-10342.

作者简介



李勇明 男,1976年生,四川绵阳人.重庆大学微电子与通信工程学院教授,博士生导师.主要研究方向为医学信号处理、机器学习.
E-mail: yongmingli@cqu.edu.cn



王品 女,1979年生,江苏盐城人.重庆大学微电子与通信工程学院副教授,硕士生导师.主要研究方向为图像处理与识别.
E-mail: wangpin@cqu.edu.cn



李文正 男,1999年生,福建漳州人.重庆大学微电子与通信工程学院硕士研究生.主要研究方向为图神经网络、行人轨迹预测.
E-mail: fxxmlwz9910@163.com



胡杰 男,2001年生,四川遂宁人.重庆大学微电子与通信工程学院硕士研究生.主要研究方向为图神经网络、行人轨迹预测.
E-mail: 2357590375@qq.com



张小恒 男,1980年生,四川达州人.重庆大学微电子与通信工程学院博士研究生,副教授.主要研究方向为医学信号处理、机器学习、行人轨迹预测.
E-mail: 7818320@qq.com