基于全局自适应有向图的行人轨迹预测

孔 玮,刘 云,李 辉,崔雪红,杨浩冉
(青岛科技大学信息科学技术学院,山东青岛 266061)

摘 要: 由于行人交互的复杂性和周围环境的多变性,行人轨迹预测仍是一项具有挑战性的任务.然而,基于图结构的方法建模行人之间的交互时,存在着网络感受野小、成对行人间的相互交互对称、固定的图结构不能适应场景变化的问题,导致预测轨迹与真实轨迹偏差较大.为了解决这些问题,本文提出一种基于全局自适应有向图的行人轨迹预测方法(pedestrian trajectory prediction method based on Global Adaptive Directed Graph,GADG).设计全局特征更新(Global Feature Updating,GFU)和全局特征选择(Global Feature Selection,GFS)分别提升空间域和时间域的网络感受范围,以获取全局交互特征.构建有向特征图,定义行人间的不对称交互,提高网络建模的方向性.建立自适应图模型,灵活调整行人间的交互关系,减少冗余连接,增强图模型的自适应能力.在ETH和UCY数据集上的实验结果表明,与最优值相比,平均位移误差降低14%,最终位移误差降低3%.

关键词: 轨迹预测; 自适应图; 有向图; 感受野; 行人轨迹; 图卷积
中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112(2022)08-1905-12
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20211613

Pedestrian Trajectory Prediction Based on Global Adaptive Directed Graph

KONG Wei, LIU Yun, LI Hui, CUI Xue-hong, YANG Hao-ran
(*School of Information Science and Technology*, *Qingdao University of Science and Technology*, *Qingdao*, *Shandong* 266061, *China*)

Abstract: Due to the complexity of pedestrian interaction and the variability of the surrounding environment, pedestrian trajectory prediction is still a challenging task. However, when modeling pedestrian interaction based on graph structure, there are some problems, such as small sensing field of the network, symmetrical interaction between pedestrians, and fixed graph structure that can not adapt to scene changes, which lead to a large deviation of the predicted trajectory from the real trajectory. To solve these problems, a pedestrian trajectory prediction method based on global adaptive directed graph is proposed. Global feature updating(GFU) and global feature selection(GFS) are designed to improve the perception range in spatial and temporal domain respectively and get global interaction features. A directed feature graph is constructed to define the asymmetric interaction between pedestrians and improve the directionality of network modeling. An adaptive graph model is established to flexibly adjust the relationship between pedestrians, reduce redundant connections and enhance the adaptive ability of the graph. The experimental results on ETH and UCY datasets show that comparing with the optimal value, the average displacement error is reduced by 14% and the final displacement error is reduced by 3%.

Key words: trajectory prediction; adaptive graph; directed graph; sensing field; pedestrian trajectory; graph convolution

1 引言

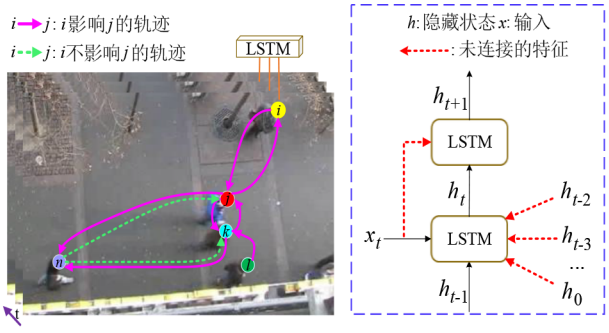
行人轨迹预测旨在利用观察到的行人轨迹,预测行人未来的运动轨迹[1].行人轨迹预测在自动驾驶[2,3]、视觉识别[4]、目标跟踪[5]和视频监控[6]等领域得到了广泛的应用.但受客观环境的影响,人与人之间、人与环境的交互变得复杂抽象,准确预测行人的轨迹仍然具有复杂性和挑战性.

随着深度学习[7]的发展,神经网络为行人轨迹预测提供了必要条件.尤其是,用于序列学习的递归神经网络(Recurrent Neural Network,RNN)、生成对抗网络

收稿日期:2021-12-03;修回日期:2022-02-07;责任编辑:朱梅玉
基金项目:国家自然科学基金(No.61702295);山东省高等学校优秀青年创新团队计划(No.2019KJN047)

(Generative Adversarial Networks, GAN)及图卷积网络(Graph Convolutional Network, GCN)成为行人轨迹预测建模的主要网络. Social LSTM[8]是循环神经网络在行人轨迹预测领域的典型应用,它通过池化层建模行人之间的相互作用. 基于GANs的方法[9~12]预测未来轨迹的分布时,模型的生成器也是使用递归神经网络设计. 这些方法的局限性在于递归架构的使用,使得网络模型的参数多,训练成本高. 基于RNN的轨迹预测方法在建模行人之间的交互时,不能单独处理空间上下文,而是需要借助额外的结构对相邻行人的信息进行编码,既不直观也不直接.

图卷积网络是另一种被广泛应用于行人轨迹预测的模型. 很多研究者将时空图[13~16]应用于行人轨迹预测,并实现了不错的预测性能. 时空图包含丰富的特征信息,比聚集的方法(例如池化)[17]直观有效. 由于行人在轨迹预测中的重要性不同,注意力机制更有助于编码行人之间的相对影响和潜在互动. 基于图注意力网络的轨迹预测方法[14~16,18~21]打破了RNN网络的顺序依赖性,利用注意力机制实现了行人特征信息的加权融合. 然而,在建立时空图模拟行人交互的过程中仍存在许多问题,如图1所示.



(a) 行人的空间交互 (b) 时域模型LSTM

图1 行人交互的时空场景分析

首先,网络的时空感受野小,无法获得行人的全局信息. 图1(a)表示行人的空间交互,当融合行人j的交互特征时,往往根据距离只关注行人k和行人l的信息,而忽略远距离行人i的特征,这使得网络的输入范围变小. 在时域中,基于长短期记忆网络(Long Short-Term Memory, LSTM)的行人轨迹预测只依赖前一时刻的隐藏状态,不能像卷积神经网络(Convolutional Neural Network, CNN)那样实现并行处理,如图1(b)表示的时域模型LSTM中缺失的连接所示. 这导致模型运行时间长,感知范围狭窄. 其次,以往的研究在空间域构造图模型时,不同的行人在同一时间通常定义为全连通图,默认行人之间的相互影响是对等的,忽视了行人间的不对称交互关系,方向性不强,导致网络模型不能准确模拟行人之间的真实互动. 例如在图1(a)中,行走在后面的行人n的运动轨迹不会影响前面的行人j和k(绿色虚线所示),而这两个行人的运动轨迹却对行人n的未来轨迹产生了重要的作用(红色实线所示). 最后,全连通图不能随着行人运动状态的变化及时调整图结构,行人间的交互冗余,自适应能力差. 为此,本文提出了基于全局自适应有向图的行人轨迹预测方法(pedestrian trajectory prediction method based on Global Adaptive Directed Graph, GADG). 针对以上问题,本文的研究贡献总结如下:

(1)设计全局特征更新GFU(Global Feature Updating)和全局特征选择GFS(Global Feature Selection),关联相互交互的行人的全局特征,扩展网络感受野,强化网络学习时空特征的能力.

(2)构建有向特征图模型,有效提取成对行人之间的非对称社交互动,增强网络的方向性,提高网络模拟真实场景的能力.

(3)建模自适应交互图,定义行人之间的自适应交互关系,减少不必要的交互连接,增强图模型适应场景变化的能力.

2 相关工作

2.1 行人之间的交互

人与人之间的交互建模经历了社会力模型、多模型方法、混合估计方法和基于模式的方法. 人与人的交互不仅包括成对行人间的交互,还涉及复杂的群组行为[22]. 而基于模式的方法从数据中拟合不同的函数(如神经网络)来学习行人之间的交互关系,提高了模型的灵活性. 例如,RNN和CNN联合建模空间关系[23]以捕获行人之间的交互. Social LSTM[8]利用LSTM计算隐藏状态,聚集一定范围内的行人交互影响. Social GAN[9]建立新的池化机制确定行人间的交互关系. 然而,这些基于RNN的模型在长序列训练中容易出现梯度消失和爆炸. 基于图结构的模型表现出基于图数据的依赖关系进行建模的强大功能,可以更好地模拟场景中人与人之间的交互. STGAT[14]通过图注意力网络(Graph Attention network, GAT)学习行人间的影响权重. Social-STGCNN[15]将轨迹直接建模为图形,根据相对距离确定行人之间的相互关系. GraphTCN[16]以输入感知的方式捕获时空交互. 然而,这些方法忽略了行人交互建模的方向性,认为两个行人之间的相互交互是对等的. 在行人运动的过程中,后面的行人总是会注意前面的行人,而前面的行人通常对后面的行人不关注. 所以,行人之间的相互交互具有不对称性. 为了体现这种不对称关系,本文把行人之间的互动建模为有向图,不仅能捕捉对目标行人产生重要影响的交互对象,还能提取他们之间的方向信息.

2.2 基于图架构的行人轨迹预测

递归神经网络虽然具备显著的序列建模能力,但

缺乏直观的高层时空结构. 在行人运动过程中,行人的运动轨迹不确定[24],行人之间的交互没有规律,图结构是表示行人交互行为的自然方法. 时空图[14~16,19,20]是比较流行的工具,可以同时捕获空间和时间关系. 这些方法通常将行人表示为节点,将他们的交互表示为连接. 但这些方法在每一个时间步都会引入一个固定结构的图,图结构不能随着场景的变化而改变. 与上述方法不同的是,本文提出的自适应图模型在不同的时间点是动态变化的,可以自适应调整行人之间的连接. 有些方法把图模型与LSTM等深层序列模型结合建模,并在此基础上进行拓展. 例如,Zhang等人[25]在位置和运动方向上构建图模型,并使用层次LSTM逐步解码. 递归社交行为图[26]递归更新交互范围内的个体特征来强化社交互动. 这些方法只建模了局部交互,不能体现深层交互关系,网络的空间感受野小. 为此,本文设计全局特征更新GFU,打破行人地理位置的限制,捕获网络全局空间特征.

2.3　基于注意力机制的行人轨迹预测

由于相邻行人对轨迹预测的重要性不同,注意力机制更有助于编码行人之间的相对影响和潜在交互. Su等人[27]根据速度计算邻居的相关性. SoPhie[12]与CNN结合,为行人添加双向注意力. Vemula等人[28]利用隐藏状态计算注意力分数. 图注意力网络利用软注意力或转移机制来区分邻居的重要性,实现了节点之间的加权消息传递和更好的群体理解. STGAT[14]和Social-STGCNN[15]通过引入灵活的图注意力机制来改善行人之间的交互关系. GraphTCN[16]使用边缘图注意力网络捕获行人间的空间交互. Social-BiGAT[18]通过图注意力网络学习网络中可靠的特征表示. 然而,这些方法只根据距离来确定行人之间的相互影响,忽略了时域注意力,导致注意力分配不符合行人行走的客观规律. 本文构建空间注意力(Spatial Attention,SA),融合行人轨迹中隐含的距离、速度和方向信息,克服仅使用位置特征的不足. 设计时域注意力模块(Temporal Attention Module,TAM),激励网络调整在时间维度上的权值比重. 这使模型具备了更好的时空建模能力.

2.4　基于CNN的行人轨迹预测

递归神经网络及其变体在行人轨迹预测领域广泛应用,表现出了良好的预测性能. ST-RNN[29]使用时空转换矩阵建模每个层的时空上下文. Social GAN[9]在Social LSTM的基础上增加对抗性训练,提高了预测性能. SR-LSTM[30]激活邻居的当前意图,迭代细化了行人的当前状态. 但基于RNN的轨迹预测模型只依赖前一时刻的输出,忽略了其他时刻对轨迹预测的影响,时域感知范围小. 而CNN可以实现并行处理并能提取丰富的上下文信息,一些方法证实了基于CNN的模型在轨迹预测方面具有竞争性. 例如,Yi等人[31]使用一个大的感受野来模拟行人的行为;Yagi等人[32]开发了一种深度神经网络来预测行人位置. 但是,仅利用CNN来集中附近行人的特征会丢失一些运动信息,这限制了预测精度. 为了提升时域的感知范围,本文将CNN与LSTM进行组合,在利用LSTM进行轨迹预测之前,设计了全局特征选择GFS,并在LSTM上增加残差连接. 消融实验表明,此设计进一步提高了网络的预测性能.

3　算法描述

本文提出的模型GADG是一种编解码结构,总体框架如图2所示. 编码器包括图注意力网络和自适应有向图学习(Adaptive Learning,APL),解码器包括全局特征选择GFS和轨迹预测. 其中,编码器中的全局特征更新GFU、自适应有向图学习APL和解码器中的全局特征选择GFS是本文的主要创新点.

行人轨迹预测的定义表示为:在时间$t=1,\cdots,T_m$期间,假设行人i在t时刻的二维坐标为$\left(x_i^t, y_i^t\right)$,$N$个行人的坐标$p_i^t=\left\{\left(x_i^t, y_i^t\right)|t=1,2,\cdots,T_m, i=1,2,\cdots,N\right\}$表示为历史轨迹. 根据历史轨迹,当$t=T_{m+1},\cdots,T_{\text{end}}$时,$\hat{p}_i^t=\left\{\left(\hat{x}_i^t, \hat{y}_i^t\right)|t=T_{m+1},\cdots,T_{\text{end}}, i=1,2,\cdots,N\right\}$表示预测的行人轨迹.

3.1　图注意力网络

3.1.1　单人运动特征编码

每个行人在运动过程中有不同的运动状态,而LSTM已被证明能从行人轨迹中提取可以描述或预测行人运动模式的隐藏特征. 行人下一时刻的运动趋势受到当前时刻运动状态的较大影响,为了强化行人当前的运动意图,增强当前特征信息的传输,本文在LSTM中添加残差连接,形成TS-LSTM,使得行人获取更丰富的特征信息,增强运动决策的合理性和准确性. 增加残差连接前后的对比情况见4.2节中的消融实验,具体实现如式(1)和式(2)所示.

$$\begin{aligned}(\Delta x_i^t, \Delta y_i^t)&=\left(x_i^{t+1}-x_i^t, y_i^{t+1}-y_i^t\right)\\ v_i^t&=\varphi\left(\Delta x_i^t, \Delta y_i^t, \boldsymbol{W}_v\right)\end{aligned}\tag{1}$$

$$h_i^t=\text{LSTM}\left(h_i^{t-1}, v_i^t, \boldsymbol{W}_h\right)+v_i^t\tag{2}$$

其中,φ是嵌入函数,$\boldsymbol{W}$是权重,v_i^t作为TS-LSTM的输入来计算隐藏特征h_i^t.

3.1.2　全局特征更新(GFU)

$\boldsymbol{H}=\left\{h_i^t|t=1,2,\cdots,T_m, \forall i=\{1,2,\cdots,N\}\right\}$作为图3的输入. GFU通过卷积运算$\theta$和$\beta$计算图中所有行人之间的特征关联程度(亲密度),来获得目标行人的全局更新特征.

在实验过程中,式(3)中的亲密度函数$d(h_i,h_j)$有4种定义,4.2节中的消融实验验证了它们的有效

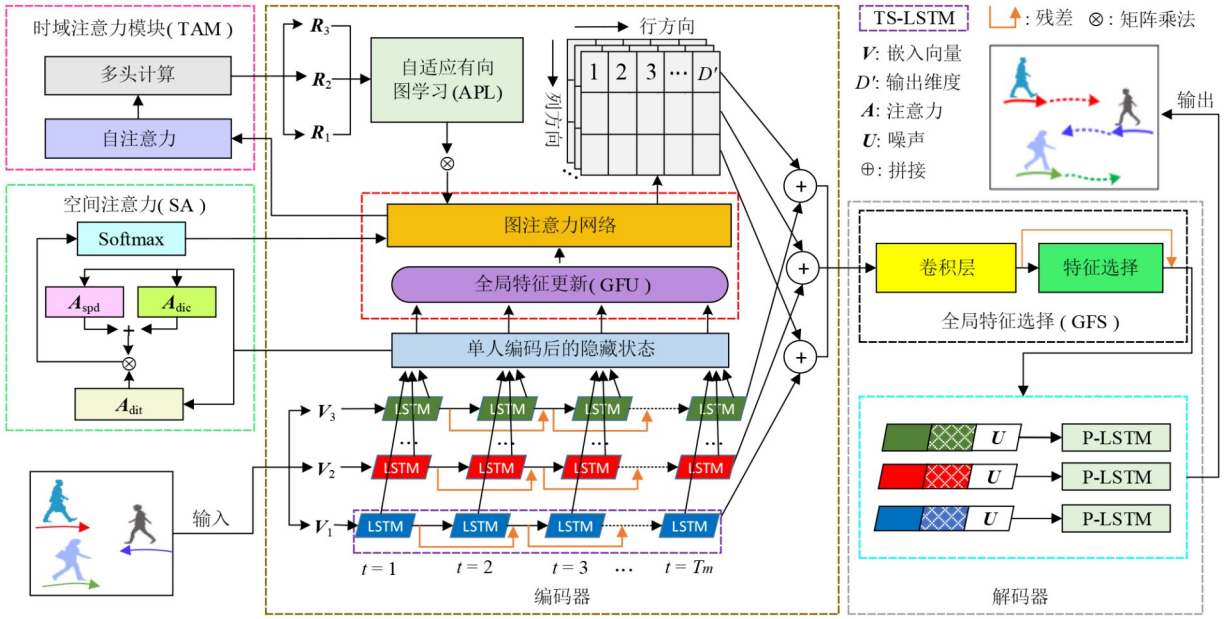


图2 模型的技术路线图

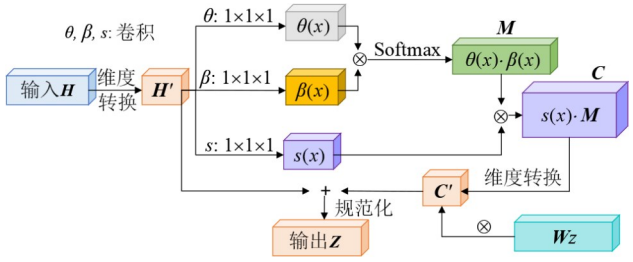


图3 全局特征更新GFU的流程图

性. $h_i^{'t}$ 和 $c_i^{'t}$ 分别是 h_i^t 和 c_i^t 的维度转换结果,T表示转置. 式(4)中的 $s(\cdot)$ 是一个显示函数,用于计算相邻行人的特征. GFU不再局限于近距离的行人特征,所以提升了网络在空间域的感受野. 经过GFU后,$\boldsymbol{H}$ 被扩展为 $\boldsymbol{Z}=\left\{\boldsymbol{Z}_i^t|t=1,2,\cdots,T_m,\forall i=\{1,2,\cdots,N\}\right\}$,$\boldsymbol{Z}$ 表示全局更新特征.

$$\begin{aligned}&\text{(高斯函数)}\quad d\left(\boldsymbol{h}_i,\boldsymbol{h}_j\right)=\mathrm{e}^{\left(\boldsymbol{h}_i\right)^{\mathrm{T}}\boldsymbol{h}_j}\\&\text{(嵌入高斯)}\quad d\left(\boldsymbol{h}_i,\boldsymbol{h}_j\right)=\mathrm{e}^{\theta\left(\boldsymbol{h}_i\right)^{\mathrm{T}}\phi\left(\boldsymbol{h}_j\right)}\\&\text{(点乘函数)}\quad d\left(\boldsymbol{h}_i,\boldsymbol{h}_j\right)=\theta\left(\boldsymbol{h}_i\right)^{\mathrm{T}}\varphi\left(\boldsymbol{h}_j\right)\\&\text{(拼接函数)}\quad d\left(\boldsymbol{h}_i,\boldsymbol{h}_j\right)=\mathrm{ReLU}\left(\boldsymbol{w}_d^{\mathrm{T}}\left[\theta\left(\boldsymbol{h}_i\right)^{\mathrm{T}}\varphi\left(\boldsymbol{h}_j\right)\right]\right)\end{aligned}\tag{3}$$

$$\begin{aligned}&M_{ij}^t=d\left(h_i^{'t},h_j^{'t}\right)\Big/\sum_{\forall j}d\left(h_i^{'t},h_j^{'t}\right),\ s\left(h_j^{'t}\right)=w_s h_j^{'t}\\&c_i^t=M_{ij}^t s\left(h_j^{'t}\right),\ \boldsymbol{Z}_i^t=w_z c_i^{'t}+h_i^{'t}\end{aligned}\tag{4}$$

3.1.3 时空注意力

(1)空间注意力(SA)

空间注意力综合了行人间的距离、速度和方向信息. 因为数据集的采样时间是0.4 s,输入为相对距离,所以相对速度等于相对距离除以采样时间. 相对方向是计算行人间的余弦相似性. 当融合距离 $\boldsymbol{A}_{dit}$、速度 $\boldsymbol{A}_{spd}$ 和方向 $\boldsymbol{A}_{dic}$ 信息后,空间注意力 $\boldsymbol{A}_e$ 的计算如式(5)所示. 距离 $\boldsymbol{A}_{dit}$ 构造了图的邻接矩阵,建立了行人间的连接关系. 为了分别突出速度和方向对图上行人交互的不同影响,$\boldsymbol{A}_{dit}$ 分别与速度 $\boldsymbol{A}_{spd}$ 和方向 $\boldsymbol{A}_{dic}$ 相乘后,再通过加法进行特征融合,即 $\boldsymbol{A}_{dit}\boldsymbol{A}_{spd}+\boldsymbol{A}_{dit}\boldsymbol{A}_{dic}=\boldsymbol{A}_{dit}(\boldsymbol{A}_{spd}+\boldsymbol{A}_{dic})$,距离、速度和方向对预测性能影响的消融实验见4.2节.

$$\begin{aligned}&\alpha_{ij}^t=\mathrm{Softmax}\left(\mathrm{ReLU}\left(\boldsymbol{a}^{\mathrm{T}}\left[\boldsymbol{W}^t h_i^t\|\ \boldsymbol{W}^t h_j^t\right]\right)\right)\\&\boldsymbol{A}_{\mathrm{dit}}=\left\{\alpha_{ij}^t|t=1,2,\cdots,T_m,\ \forall ij\in\{1,2,\cdots,N\}\right\}\\&\boldsymbol{A}_e=\mathrm{Softmax}\left(\boldsymbol{A}_{\mathrm{dit}}\otimes\left(\boldsymbol{A}_{\mathrm{spd}}+\boldsymbol{A}_{\mathrm{dic}}\right)\right)\end{aligned}\tag{5}$$

其中,$\boldsymbol{a}\in\boldsymbol{R}^{2D'}$ 是单层感知机的权值向量,$\boldsymbol{W}^t\in\boldsymbol{R}^{D'\times D}$ 是实现线性变换的共享权重,D 和 D' 是输入输出维度,‖是拼接操作,j 表示行人 i 的邻居,⊗表示矩阵的乘法.

(2)图卷积

结合注意力 $\boldsymbol{A}_e$ 和全局特征 $\boldsymbol{Z}$,图卷积的输出如式(6)所示.

$$\boldsymbol{Z}^{(l+1)}=\sigma\left(\boldsymbol{A}_e\boldsymbol{Z}^{(l)}\boldsymbol{W}_e\right)\tag{6}$$

其中,$\boldsymbol{Z}^{(l)}\in\boldsymbol{R}^{N\times D_l}$ 是 N 个行人在第 l 层上的特征矩阵;D_l 是特征维度,在本文中,l=2;$\boldsymbol{W}_e\in\boldsymbol{R}^{D_l\times D_{l+1}}$ 是可学习的参数矩阵;$\boldsymbol{Z}^{(l+1)}$ 由多头注意力连接而成,注意力头数是4,消融实验见4.2节. 多头图注意力的卷积运算如图4所示.

(3)时域注意力模块(TAM)

由于行人在不同时刻的运动状态不同,且不同历史时刻的运动特征对行人未来轨迹的影响力度也不同,因此,时域注意力TAM可以定义行人在不同时刻的

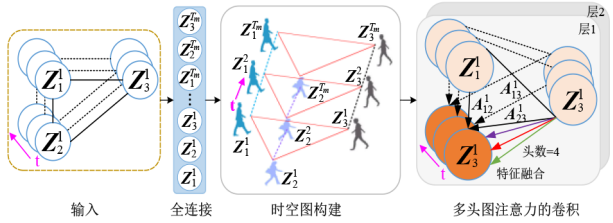


图4　多头图注意力网络

运动状态的重要程度,激励网络调整在时间维度上的权值比重,以进一步模拟真实场景,提高网络的预测性能. 给定来自式(6)的输入$\boldsymbol{Z}$,通过TAM进行时间关联后,输出变成$\boldsymbol{R}$. 首先,$\boldsymbol{Z}$被共享的线性变换函数$f=\boldsymbol{x}\boldsymbol{w}$($\boldsymbol{x}$是输入,$\boldsymbol{w}$是可学习的权值参数)转换维度,经过3次不同的权值参数$\boldsymbol{w}$的转换,变成式(7)中的3个不同的张量$\boldsymbol{Q}_i$、$\boldsymbol{K}_i$和$\boldsymbol{V}_i$;其次,用$\boldsymbol{Q}_i\boldsymbol{K}_i^{\mathrm{T}}$计算不同时间步之间的关联程度,也就是时间注意力;再次,通过$\boldsymbol{V}_i$转换维度;最后,把时间注意力加权到$\boldsymbol{V}_i$中得到式(8)的单头注意力head_j. TAM的计算过程如图5所示.

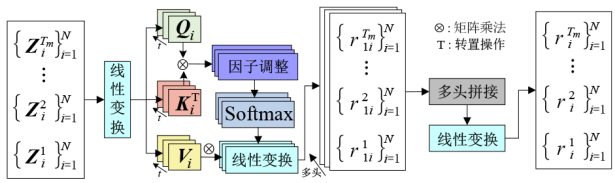


图5　TAM的计算过程

s_f是将输出调整到合理范围的比例因子,$0<s_f<1$. 根据实验结果,当$s_f=0.5$时,预测性能最优. 为使网络获取更丰富的特征信息,用式(9)计算多头注意力. 其中,时域注意力头的数量h_num=8,消融实验见4.2节.

$$\boldsymbol{Q}_i=f_q\left\{\boldsymbol{Z}_i^t\right\}_{t=1}^{T_m},\ \boldsymbol{K}_i=f_k\left\{\boldsymbol{Z}_i^t\right\}_{t=1}^{T_m},\ \boldsymbol{V}_i=f_v\left\{\boldsymbol{Z}_i^t\right\}_{t=1}^{T_m} \tag{7}$$

$$\mathrm{head}_j=\left(\mathrm{Softmax}\left(\boldsymbol{Q}_i\boldsymbol{K}_i^{\mathrm{T}}/s_f\right)\right)\boldsymbol{V}_i \tag{8}$$

$$\boldsymbol{R}_i=f_c\left(\left\|\left\{\mathrm{head}_j\right\}_{j=1}^{\mathrm{h\_num}}\right.\right) \tag{9}$$

3.2　自适应有向图学习

式(9)的$\boldsymbol{R}\in\boldsymbol{R}^{T_m\times N\times N}$堆叠了场景中的行人在每个时间步的交互作用,但存在行人间的交互方向性不强和图结构固定的问题. 图的自适应学习过程解决了这些问题,学习流程如图6所示.

3.2.1　建立有向特征图

为了体现行人交互的方向性和不对称性,本文设计了行与列的级联卷积,交叉融合行人i对行人j的影响和行人j对行人i的影响. 在实现过程中,首先把$\boldsymbol{R}$表示的图结构利用1×1的卷积进行时空融合,产生时空密集交互$\boldsymbol{R}'\in\boldsymbol{R}^{T_m\times N\times N}$,然后,对$\boldsymbol{R}'$分别实现行卷积和列卷积,最后把两种卷积结果融合,如式(10)所示. $\boldsymbol{E}^{(0)}=\boldsymbol{R}'$,

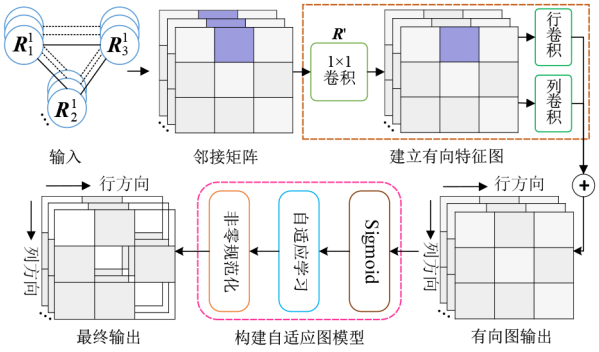


图6　自适应有向图APL的学习过程

$\mathcal{K}$是卷积核. 本文设置7层卷积,最终获得的高级交互特征表示为$\boldsymbol{E}$.

$$\begin{aligned}\boldsymbol{E}_{\mathrm{row}}^{(l+1)}&=\mathrm{Conv}\left(\boldsymbol{E}^{(l)},\ \mathcal{K}_{(1\times 3)}\right)\\ \boldsymbol{E}_{\mathrm{col}}^{(l+1)}&=\mathrm{Conv}\left(\boldsymbol{E}^{(l)},\ \mathcal{K}_{(3\times 1)}\right)\\ \boldsymbol{E}^{(l+1)}&=\mathrm{PReLU}\left(\boldsymbol{E}_{\mathrm{row}}^{(l+1)}+\boldsymbol{E}_{\mathrm{col}}^{(l+1)}\right)\end{aligned} \tag{10}$$

3.2.2　构建自适应图模型

(1)自适应学习

级联卷积使行人间的交互具有了方向,但图结构不能随着场景的变化而改变,存在很多冗余连接. 比如在图结构中,后面的行人仍会对前面的行人轨迹产生影响. 为此,本文学习阈值$\xi\in[0,1]$来消除不必要的交互. 通过实验,当$\xi=0.5$时,网络的预测性能最好. 在式(11)中,$\mathbb{I}(\cdot)$是指示函数,如果不等式成立输出1,否则输出0.

$$\boldsymbol{F}=\mathbb{I}\left\{\mathrm{Sigmoid}\left(\boldsymbol{E}\right)\geq\xi\right\} \tag{11}$$

(2)非零规范化

为了增加自连接,在$\boldsymbol{F}$中需增加大小相等的单位矩阵$\boldsymbol{I}$. 然后通过元素相乘形成特征矩阵$\boldsymbol{G}_{\mathrm{sp}}$,如式(12)所示,⊙代表元素相乘. 本文对编码结果归一化时发现,零输入值经过Softmax后变成非零值,使得没有交互连接的行人被重新影响,冗余连接再次产生. 为了避免这个问题,本文设计了调整因子ϵ,来保持特征矩阵的稀疏性.

设$\boldsymbol{G}_{\mathrm{sp}}^t=\left[x_1,x_2,\cdots,x_N\right]$,$\boldsymbol{G}_{\mathrm{sp}}=\left\{\boldsymbol{G}_{\mathrm{sp}}^t|t=1,2,\cdots,T_m\right\}$,经过式(13)的操作之后,特征矩阵$\boldsymbol{G}_{\mathrm{sp}}$变成了规范化的自适应有向图$\boldsymbol{G}'_{\mathrm{sp}}$.

$$\boldsymbol{G}_{\mathrm{sp}}=(\boldsymbol{F}+\boldsymbol{I})\odot\boldsymbol{R}' \tag{12}$$

$$\epsilon-\mathrm{Softmax}\left(x_i\right)=\frac{\left(\exp\left(x_i\right)-1\right)^2}{\sum_j^N\left(\exp\left(x_j\right)-1\right)^2+\epsilon} \tag{13}$$

(3)编码输出

首先,把自适应有向图$\boldsymbol{G}'_{\mathrm{sp}}$输入图注意力网络,输出为$\boldsymbol{G}$,表达式如式(14),$\boldsymbol{Z}$来自式(6). 其次,在行人运动过程中,目标行人的轨迹变化不仅来自周围行人的相互作用,还取决于目标行人自身的影响.

设 $\boldsymbol{G}=\left\{g_i^t|t=1,2,\cdots,T_m,\forall i=\{1,2,\cdots,N\}\right\}$，本文将来自式(2)的$h_i^t$和式(14)的$g_i^t$拼接起来以完成编码．编码器的最终输出如式(15)所示，δ是多层感知机．

$$\boldsymbol{G}=\boldsymbol{G}'_{\mathrm{sp}}\otimes\boldsymbol{Z} \tag{14}$$

$$c_{0i}^t=\delta_h\left(h_i^t\right)\|\ \delta_g\left(g_i^t\right) \tag{15}$$

3.3 解码器

3.3.1 全局特征选择(GFS)

在使用LSTM预测轨迹之前，为了提高时域的感知范围，选择重要的行人特征并控制特征信息的流动，本文设计GFS.

GFS由卷积层和特征选择组成，具体结构如图7所示．输入来自式(15)，由$\boldsymbol{C}_0$表示，具体的表达式为$\boldsymbol{C}_0=\left\{c_{0i}^t|t=1,\cdots,T_m,\forall i\in\{1,\cdots,N\}\right\}$.

(1)卷积层

在图7左侧中，GFS有3个卷积层，卷积核是3×3.为了确保输入和输出的长度相同，需要使用填充操作来保持卷积前后的特征映射不变．观察图中红线的变化可以发现，随着卷积层的加深，感受野变得越来越大．例如，假设把图中的省略号表示的多个时间步看成一个时间步，那么经过3层卷积，输出的一个时间步特征$\boldsymbol{C}_3^{T_m}$能感知输入的7个时间步的特征，这便提高了网络在时域的接收范围．经过每个时间步特征的相互叠加，网络便获取了全局时域特征．

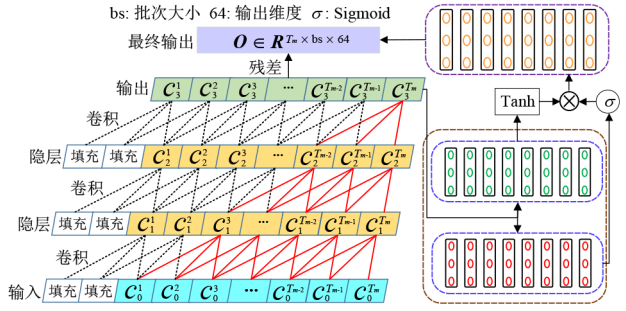


图7 全局特征选择GFS的架构图

(2)特征选择

为了从卷积层中选择重要的行人特征并控制特征信息的流动，图7右侧设计了由两个激活函数组成的选通机制．当两个激活函数分别为Tanh和Sigmoid时，模型表现最好．图中的一个圆可以代表许多行人，方框表示不同的时间步．GFS之后，最终输出如式(16)所示．

$$\boldsymbol{O}=\mathrm{Tanh}\left(\boldsymbol{W}_a\boldsymbol{C}_3+b_a\right)\odot\sigma\left(\boldsymbol{W}_\sigma\boldsymbol{C}_3+b_\sigma\right)+\boldsymbol{C}_3 \tag{16}$$

其中，$\boldsymbol{W}_a$和$\boldsymbol{W}_\sigma$是两个激活函数的权重，b是偏差，$\boldsymbol{C}_3$是最后卷积层的输出．

3.3.2 轨迹预测

图2中的解码部分是在LSTM上增加残差连接形成P-LSTM来预测轨迹．P-LSTM的结构类似于TS-LSTM.为了模拟真实场景，在训练过程中，对服从标准正态分布$N(0,1)$的随机噪声$\boldsymbol{U}$进行采样，并与$\boldsymbol{O}$连接作为P-LSTM的输入，如式(17)所示．$e_i^{T_m}$是初始隐藏状态，来自式(1)的$v_i^{T_m}$表示初始输入，$\boldsymbol{W}_e$是P-LSTM的可更新权重．式(18)的$\left(\Delta x_i^{T_{m+1}},\Delta y_i^{T_{m+1}}\right)$是最终预测的行人相对位置．通过后续输入，相对位置可以转换为绝对位置．

$$\begin{aligned}e_i^{T_m}&=\mathrm{o}_i^{T_m}\ \|\ u\\ e_i^{T_{m+1}}&=\mathrm{LSTM}\left(e_i^{T_m},v_i^{T_m},\boldsymbol{W}_e\right)+e_i^{T_m}\end{aligned} \tag{17}$$

$$\left(\Delta x_i^{T_{m+1}},\Delta y_i^{T_{m+1}}\right)=\delta_e\left(e_i^{T_{m+1}}\right) \tag{18}$$

为了模拟行人运动的不确定性，本文使用多样性损失策略．受随机噪声$\boldsymbol{U}$的影响，k个结果可在一次训练中生成．这些结果分别计算$L2$距离，并将最小值作为损失，如式(19)所示．

$$\mathrm{Loss}=\min_k\left\|\boldsymbol{Y}_i-\hat{Y}_i^k\right\|_2 \tag{19}$$

其中，$\boldsymbol{Y}_i$是真实轨迹，$\hat{Y}_i^k$是预测轨迹，k是超参数，在本文中，k=20.

4 实验及分析

4.1 实验设置及运行细节

(1)数据集

实验在2个开放数据集ETH和UCY上进行了验证．这2个数据集包括5个室外拍摄的鸟瞰场景，共2 206条行人轨迹，详细介绍见表1. 本文参考了Social GAN[9]的数据预处理策略，所有数据都转换为世界坐标．

表1 ETH/UCY数据集

数据集	场景	帧数	人数	分组数	障碍物数
ETH	ETH	1 448	360	243	44
	HOTEL	1 168	390	623	25
UCY	UNIV	541	434	297	16
	ZARA1	866	148	91	34
	ZARA2	1 052	204	140	34

(2)评估指标

式(20)为平均位移误差(Average Displacement Error，ADE)和最终位移误差(Final Displacement Error，FDE)的计算方式，主要用于计算预测轨迹和真实轨迹之间的差异．指标值越小，网络性能越好．

$$\begin{aligned}\mathrm{ADE}&=\frac{\sum_{i=1}^{N}\sum_{t=1}^{T_{\mathrm{end}}}\left\|\hat{p}_i^t-p_i^t\right\|_2}{N\times T_{\mathrm{end}}}\\ \mathrm{FDE}&=\frac{\sum_{i=1}^{N}\left\|\hat{p}_i^t-p_i^t\right\|_2}{N},\ t=T_{\mathrm{end}}\end{aligned} \tag{20}$$

(3)实验细节

实验在Pytorch=1.2的环境中运行. 训练过程使用两个NVIDIA GeForce GTX-1080 GPU. 行人的相对坐标是模型的输入. TS-LSTM的隐藏状态和图卷积的输出为32维向量,随机噪声 $\boldsymbol{U}$ 为16维. 模型使用Adam进行优化,批量大小为64. 观测的历史轨迹为3.2秒(8个时间步),预测轨迹为4.8秒(12个时间步).

4.2 消融实验

消融实验在ZARA2数据集上进行. 由于基线模型的预测长度为12个时间步,所以在验证各个模块对网络性能的影响时,预测长度设置为12个时间步. 其余消融实验的预测长度设为8.

4.2.1 模块内的消融实验

表2是超参数的设置实验,由于这些超参数是基线模型自带的参数,所以表2的消融实验以基线为基础,用黑色粗体突出最好的结果. 当图卷积层数 l=2、多头图注意力 h=4和预测次数 k=20的时候,模型取得了较好的性能. 这说明,图卷积网络具有浅层特征,多头图注意力可以强化模型的学习能力以及 k 表示的多样性轨迹能体现行人运动的不确定性.

表3用黑色粗体突出的是最好结果,可以看出,与基线相比,当亲密度函数是嵌入高斯函数时,模型的表现最好. 在LSTM上增加残差连接后,ADE和FDE分别比基线降低10%和7.5%,这证明了残差连接对于预测性能的提升是有效的.

表2　图卷积层数 l、图注意力头数 h 和预测次数 k 的消融实验

评估指标	图卷积层数 l(h=4,k=20)				图注意力头数 h(l=2,k=20)				预测次数 k(l=2,h=4)				
	1	2	3	5	2	4	8	16	5	10	15	20	25
ADE↓	0.21	**0.20**	0.21	0.22	0.22	**0.20**	0.21	0.21	0.21	0.22	0.21	**0.20**	0.23
FDE↓	0.42	**0.40**	0.41	0.43	0.42	**0.40**	0.41	0.42	0.42	0.42	0.42	**0.40**	0.44

表3　亲密度函数与LSTM上残差连接的消融实验

评估指标	基线	亲密度函数(增残差前)				增残差后(嵌入高斯)
		高斯函数	嵌入高斯	点乘函数	拼接函数	
ADE↓	0.20	0.22	0.20	0.21	0.22	**0.18**
FDE↓	0.40	0.45	0.40	0.44	0.44	**0.37**

表4和表5中用黑色粗体突出最好的结果. 表4显示,融合了行人的距离、速度和方向的空间注意力,能使网络获得详细的行人交互,多特征融合能提升网络的预测性能. 表5中的数据不仅体现了多头注意力的有效性,还确定了最佳时域注意力头数是8. 时域注意力体现的是目标行人在不同时刻的历史运动状态对其未来轨迹的影响,而多头注意力能从多个角度关联历史运动信息.

表4　行人间的距离、速度和方向对预测性能的影响

评估指标	特征融合的空间注意力			
	距离	距离+速度	距离+方向	距离+速度+方向
ADE↓	0.180 5	0.179 9	0.178 7	**0.178 0**
FDE↓	0.371 3	0.370 4	0.369 3	**0.367 7**

表5　时域注意力头数的设置实验

评估指标	时域注意力TAM			
	2	4	8	16
ADE↓	0.21	0.19	**0.17**	0.20
FDE↓	0.42	0.37	**0.36**	0.37

4.2.2 模块间的消融实验

基线STGAT[14]的图注意力网络根据距离获得行人间的空间交互,使用两个LSTM分别对时域的个人运动状态和行人交互进行编码. 在预测行人轨迹时,也使用了LSTM,预测长度为12个时间步. 本节主要是验证全局特征更新GFU、自适应学习APL和全局特征选择GFS对模型性能的影响,实验结果如表6所示,用黑色粗体突出最好的结果. Res是在LSTM上添加的残差连接. 表6中的数据证明了在GADG中设计的各个模块可以进一步提高预测性能. 尤其是同时增加GFU,APL和GFS后,模型的性能达到最优,这也证明了本文提出的模型GADG的有效性. 在基线上增加全局特征更新GFU,并在LSTM上增加残差连接的网络,本文称之为扩展图注意力网络(Extended Graph Attention Network, EGAT),以便于后面的轨迹比较.

表6　各个模块的消融实验

基线	GFU	Res	SA	TAM	APL	GFS	ADE↓	FDE↓
√							0.291 7	0.604 3
√	√	√					0.264 9	0.574 3
√	√	√	√				0.264 5	0.574 1
√	√	√	√	√			0.264 2	0.573 9
√	√	√	√	√		√	0.254 6	0.564 8
√	√	√	√	√	√	√	**0.251 5**	**0.551 3**

4.3 实验结果比较

4.3.1 与先进技术的比较

在表7中,排在前三位的预测指标值分别用红、绿、蓝三种颜色表示. 表中标有*的模型生成确定的轨迹,未标记的模型生成多种轨迹,并选择最佳轨迹进行对比. 实验结果表明,与其他模型相比,本文提出的模型GADG在所有场景数据集中都优于基线STGAT,ADE和FDE的平均值分别比STGAT降低14%和12%. 与最优值相比,ADE和FDE的平均值分别降低14%和3%. ETH的ADE/FDE,HOTEL的ADE/FDE,ZARA2的ADE以及ADE和FDE的均值都达到最优. 在UNIV中,高密度人群涉及更多的行人交互,迫使目标行人在转弯、穿越人群等不同选项中做出决策,这使得预测更具有挑战性. 在ZARA1中,行人的轨迹经常受到周围行人和障碍物的影响,这可能会改变或限制人类活动,导致模型无法捕捉更多的社交互动.

表7 在ETH/UCY数据集上的实验结果比较

数据集	ETH		HOTEL		UNIV		ZARA1		ZARA2		平均值	
评估指标	ADE ↓	FDE ↓	ADE ↓	FDE ↓	ADE ↓	FDE ↓	ADE ↓	FDE ↓	ADE ↓	FDE ↓	ADE ↓	FDE ↓
Linear*[8]	1.33	2.94	0.39	0.72	0.82	1.59	0.62	1.21	0.77	1.48	0.79	1.59
SR-LSTM*[30]	0.63	1.25	0.37	0.74	0.51	1.10	0.41	0.90	0.32	0.70	0.45	0.94
Social LSTM[8]	1.09	2.35	0.79	1.76	0.67	1.40	0.47	1.00	0.56	1.17	0.72	1.54
Social GAN[9]	0.87	1.62	0.67	1.37	0.76	1.52	0.35	0.68	0.42	0.84	0.61	1.21
SoPhie[12]	0.70	1.43	0.76	1.67	0.54	1.24	0.30	0.63	0.38	0.78	0.54	1.15
CGNS[33]	0.62	1.40	0.70	0.93	0.48	1.22	0.32	0.59	0.35	0.71	0.49	0.97
PIF[6]	0.73	1.65	0.30	0.59	0.60	1.27	0.38	0.81	0.31	0.68	0.46	1.00
STSGN[25]	0.75	1.63	0.63	1.01	0.48	1.08	0.30	0.65	0.26	0.57	0.48	0.99
GAT[18]	0.68	1.29	0.68	1.40	0.57	1.29	0.29	0.60	0.37	0.75	0.52	1.07
Social-BiGAT[18]	0.69	1.29	0.49	1.01	0.55	1.32	0.30	0.62	0.36	0.75	0.48	1.00
Social-STGCNN[15]	0.64	1.11	0.49	0.85	0.44	0.79	0.34	0.53	0.30	0.48	0.44	0.75
STGAT[14]	0.65	1.12	0.35	0.66	0.52	1.10	0.34	0.69	0.29	0.60	0.43	0.83
GADG	0.56	0.99	0.28	0.55	0.45	0.96	0.30	0.61	0.25	0.55	0.37	0.73

4.3.2 推断时间

表8比较了不同方法的推理时间,通过比较可以发现,GADG在推理过程中具有较高的计算效率. 这归因于其计算过程只使用视觉信息,不需要在场景中检测和跟踪行人. 但由于GADG使用了递归网络LSTM进行部分时态推理,因此,本模型的推理速度略慢于Social-STGCNN. 但与STGAT相比,GADG的推理速度依然很快. 这是因为GADG不仅增加了感受野,提高了数据并行处理的效率,还能利用图的自适应学习精简模型结构.

表8 推断时间比较

方法	推断时间/秒
Social-LSTM[8]	1.473 6
SR-LSTM[30]	0.197 3
Social GAN[9]	0.121 0
PIF[6]	0.143 1
Social-STGCNN[15]	0.002 5
STGAT[14]	0.031 0
Introvert[34]	0.120 0
GADG	0.012 7

4.4 实验分析

4.4.1 训练过程对比

在相同的实验环境下,GADG和STGAT的训练过程在图8中进行了比较. 图中ADE和FDE的变化趋势存在几个特点. 首先,GADG随着训练进度的推进更加稳定,比STGAT更快地拟合. 其次,拟合后,GADG的ADE和FDE均优于STGAT,且都超过了最优值. 最后,STGAT在ADE上的变化先降后升,说明更多的迭代使得STGAT的性能没有提高反而下降. 也就是,尽管STGAT能够适应样本,但对样本的拟合能力不强.

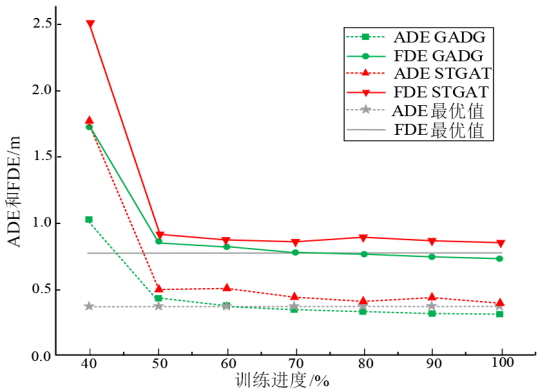


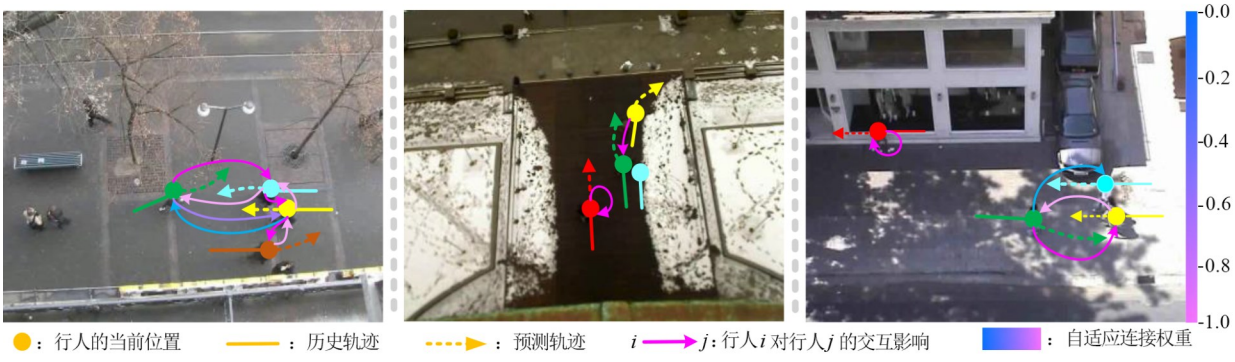
图8 训练过程分析

4.4.2　自适应有向图的可视化

图9不仅展示了模型在不同场景中行人之间的交互影响,而且还能捕捉到行人具体的交互对象.图中由实线带箭头表示的交互连接具有不同的方向和颜色,说明了行人间的交互具有方向性和不对称性.连接颜色越深,行人间的影响越大,且影响程度从蓝色、紫色到红色依次递增.例如,在图9(a)中,由于绿色节点到蓝色节点的连接颜色(深红色)比蓝色节点到绿色节点的连接颜色(淡红色)深,所以绿色节点对蓝色节点的影响大于蓝色节点对绿色节点的影响,这与现实场景是一致的.在图9(b)和图9(c)中,通过交互连接的方向可以发现,红色节点的轨迹仅受自身历史轨迹的影响.此外,根据交互连接的指示方向,模型还可以动态捕获目标行人的交互对象.例如,图9(a)中的蓝色节点与绿色和黄色节点交互,与棕色节点无交互关系;在图9(c)中,除红色节点外,绿色节点与所有节点交互,但蓝色节点的交互节点只有黄色节点.

4.4.3　轨迹可视化

图10比较了行人在同向或异向行走、多人并行、相遇、群组行走的轨迹变化,黄色虚线(预测轨迹)和蓝色实线(真实轨迹)的重合度越高,预测精度越高.对于群体运动,行人交互是复杂的,观察重合度可以看出GADG预测的轨迹比EGAT和STGAT更准确.STGAT擅长预测线性轨迹,而GADG可以推断行人轨迹的变化,如图10(c)(e)(f)所示.当行人直行时,STGAT可以预测符合现实的轨迹,但精度比EGAT差.这是因为EGAT在融合运动特征时利用全局特征更新GFU捕获了行人的全局交互.但是与GADG相比,EGAT的预测精度较差.其原因是GADG能在自适应学习过程中建立合理的自适应有向图,并能利用全局特征选择GFS提升时域的感知范围并获取行人在运动过程中的显著特征.当行人非线性移动(如转弯、曲折行走)时,如图10(a)(b)(d)(e),STGAT不能准确预测行人的未来轨迹,但GADG却可以合理地预测贴近真实的轨迹.在图10(e)中,当一名身穿黑色T恤衫的女士穿过人群时,STGAT预测的黄色虚线较短,与蓝色实线表示的真实轨迹相差很大.也就是,STGAT预测该女士将在原地等待.但EGAT和GADG却推断出她即将穿过人群,这主要得益于GFU实现的全局特征关联.但是,GADG的预测精度更好,这就证明APL和GFS对预测性能的提升是有效的.在图10(b)中,EGAT和GADG能判断静止行人(轨迹由点表示)并预测其未来的静止状态,而STGAT将静止行人视为移动行人.这些可视化结果直观地表明,与STGAT生成的轨迹相比,本文提出的模型GADG能够更好地捕捉全局交互和显著的运动特征,并能生成更可靠的行人轨迹.



(a)　绿蓝节点间的不对称交互　　(b)　红节点只影响自身　　(c)　各节点的不同交互对象

图9　自适应有向图的可视化

在UNIV数据集中,行人的数量不多但密集度很高,建立的图模型比较复杂,行人之间存在着更加复杂的交互.图11展示了在密集行人的场景中预测的未来轨迹.根据真实轨迹和预测轨迹的重合度可以发现,本文提出的模型能取得较好的预测效果.由于观测轨迹是8个时间步,预测轨迹是12个时间步,在建立图模型的过程中,模型会忽略当前场景中达不到要求的行人.所以,图11显示的是达到上述要求的部分行人的预测轨迹,而不满足要求的行人多为刚进入或即将走出场景以及正在行走但未达到时间步数量的人.

4.4.4　存在的问题及研究方向

当场景中同时有大量行人出现时,由于行人比较密集,因此行人之间的特征差异减小,导致空间注意力均匀分布,如图12所示.在图12中,周围行人上的圆圈越大,说明此行人对目标行人的影响越大.而图中却显示了大小差不多的圆圈,即模型产生了均匀分布的注意力.因此,未来的研究重点将是为模型添加额外的辅助信息,例如场景信息、行人的社会属性信息等.只有对这些信息进行整合,才能把握行人的运动意图,模拟行人的最终行为决策.另外,面对异常复杂的人群数据集,还需要提升模型的泛化性能.

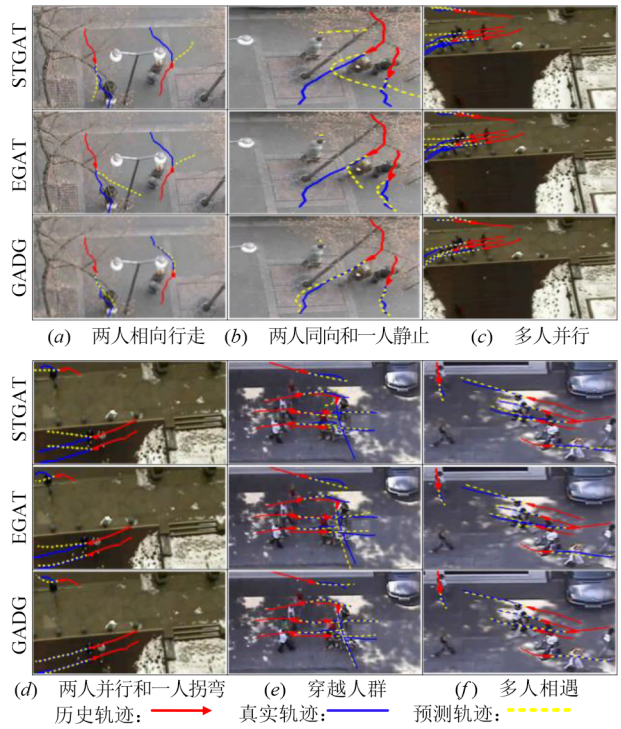


图 10 预测轨迹的可视化

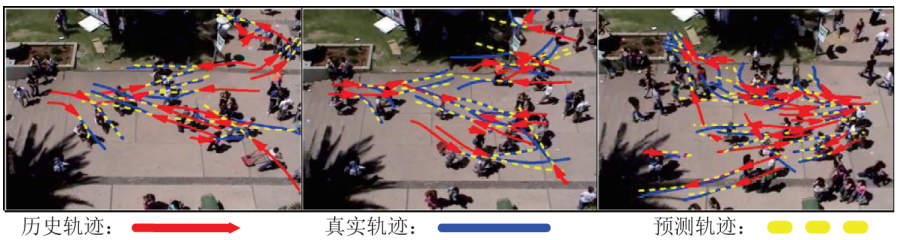


图 11 密集行人的预测轨迹

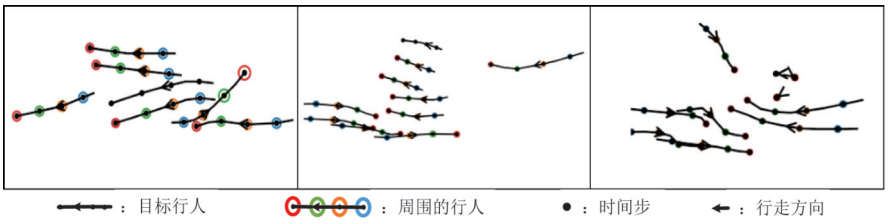


图 12 空间注意力均匀分布

5 总结

本文提出了一种基于全局自适应有向图的行人轨迹预测方法GADG，旨在解决行人轨迹预测过程中存在的时空感知范围小、行人之间的交互对称和图结构固定不随场景变化的问题. 模型在5个开放的场景数据集上取得了优异的实验性能. 实验结果表明，GADG能提高模型的时空感知范围，根据行人之间的不对称交互强化方向感知，自适应调整图结构，并能预测更可靠的行人运动轨迹. 然而，当场景中突然出现许多行人时，行人之间的特征差异随着行人数量的增加而减小，导致注意力均匀分布. 所以，结合场景、行人社会属性等信息，及时判断行人的运动意图，为将来的研究指明了方向.

参考文献

[1] 孔玮, 刘云, 李辉, 等. 基于深度学习的行人轨迹预测方法综述[J]. 控制与决策, 2021, 36(12): 2841-2850.
KONG W, LIU Y, LI H, et al. Survey of pedestrian trajectory prediction methods based on deep learning[J]. Control and Decision, 2021, 36(12): 2841-2850. (in Chinese)

[2] WU P X, CHEN S H, METAXAS D N. MotionNet: Joint perception and motion prediction for autonomous driving

based on bird's eye view maps[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 11382-11392.

[3] LUO Y F, CAI P P, BERA A, et al. PORCA: Modeling and planning for autonomous driving among many pedestrians[J]. IEEE Robotics and Automation Letters, 2018, 3 (4): 3418-3425.

[4] HU T, LONG C J, XIAO C X. A novel visual representation on text using diverse conditional GAN for visual recognition[J]. IEEE Transactions on Image Processing, 2021, 30: 3499-3512.

[5] SALEH F, ALIAKBARIAN S, SALZMANN M, et al. ArTIST: Autoregressive trajectory inpainting and scoring for tracking[EB/OL]. (2020-04-16)[2021-12]. https://arxiv.org/abs/2004.07482.

[6] LIANG J W, JIANG L, NIEBLES J C, et al. Peeking into the future: Predicting future person activities and locations in videos[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach, CA, USA: IEEE, 2019: 5718-5727.

[7] 张顺, 龚怡宏, 王进军. 深度卷积神经网络的发展及其在计算机视觉领域的应用[J]. 计算机学报, 2019, 42(3): 453-482.
ZHANG S, GONG Y H, WANG J J. The development of deep convolution neural network and its applications on computer vision[J]. Chinese Journal of Computers, 2019, 42(3): 453-482. (in Chinese)

[8] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: Human trajectory prediction in crowded spaces [C]//2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Las Vegas, NV, USA: IEEE, 2016: 961-971.

[9] 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(CVPR). Salt Lake City, UT, USA: IEEE, 2018: 2255-2264.

[10] SUN H, ZHAO Z Q, HE Z H. Reciprocal learning networks for human trajectory prediction[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle: IEEE, 2020: 7414-7423.

[11] YANG B, YAN G C, WANG P, et al. TPPO: A novel trajectory predictor with pseudo oracle[EB/OL]. (2020-02-04)[2021-12]. https://arxiv.org/abs/2002.01852.

[12] 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). Long Beach, CA, USA: IEEE, 2019: 1349-1358.

[13] IVANOVIC B, PAVONE M. The trajectron: probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs[C]//2019 IEEE/CVF International Conference on Computer Vision(ICCV). Seoul, Korea: IEEE, 2019: 2375-2384.

[14] 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). Seoul, Korea: IEEE, 2019: 6271-6280.

[15] 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). Seattle: IEEE, 2020: 14412-14420.

[16] WANG C X, CAI S F, TAN G. GraphTCN: Spatio-temporal interaction modeling for human trajectory prediction [C]//2021 IEEE Winter Conference on Applications of Computer Vision(WACV). Waikoloa: IEEE, 2021: 3449-3458.

[17] 毛琳, 巩欣飞, 杨大伟, 等. 空时社交关系池化行人轨迹预测模型[J]. 计算机辅助设计与图形学学报, 2020, 32 (12): 1918-1925.
MAO L, GONG X F, YANG D W, et al. Space-time social relationship pooling pedestrian trajectory prediction model[J]. Journal of Computer-Aided Design & Computer Graphics, 2020, 32(12): 1918-1925. (in Chinese)

[18] KOSARAJU V, SADEGHIAN A, MARTÍN-MARTÍN R, et al. Social-BiGAT: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks[C]//33rd Annual Conference on Neural Information Processing Systems(NIPS). Vancouver, BC, Canada: NIPS, 2019: 1-10.

[19] HADDAD S, WU M Q, WEI H, et al. Situation-aware pedestrian trajectory prediction with spatio-temporal attention model[EB/OL]. (2019-02-13)[2021-12]. https://arxiv.org/abs/1902.05437.

[20] YU C J, MA X, REN J W, et al. Spatio-Temporal graph transformer networks for pedestrian trajectory Prediction [C]//European Conference on Computer Vision(ECCV). Glasgow, UK: Springer, 2020: 507-523.

[21] LIANG J W, JIANG L, MURPHY K, et al. The garden of forking paths: Towards multi-future trajectory prediction [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Seattle: IEEE, 2020: 10505-10515.

[22] 丰艳, 张甜甜, 王传旭. 基于伪3D残差网络与交互关系建模的群组行为识别方法[J]. 电子学报, 2020, 48(7): 1269-1275.

FENG Y, ZHANG T T, WANG C X. Group activity recognition method based on pseudo 3D residual network and interaction modeling[J]. Acta Electronica Sinica, 2020, 48(7): 1269-1275. (in Chinese)

[23] ZHAO T Y, XU Y F, MONFORT M, et al. Multi-agent tensor fusion for contextual trajectory prediction[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach, CA, USA: IEEE, 2019: 12118-12126.

[24] 程媛, 迟荣华, 黄少滨, 等. 基于非参数密度估计的不确定轨迹预测方法[J]. 自动化学报, 2019, 45(4): 787-798.
CHENG Y, CHI R H, HUANG S B, et al. Uncertain trajectory prediction method using non-parametric density estimation[J]. Acta Automatica Sinica, 2019, 45(4): 787-798. (in Chinese)

[25] ZHANG L D, SHE Q, GUO P. Stochastic trajectory prediction with social graph network[EB/OL]. (2019-07-24) [2021-12]. https://arxiv.org/abs/1907.10233.

[26] SUN J H, JIANG Q H, LU C W. Recursive social behavior graph for trajectory prediction[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 657-666.

[27] SU H, DONG Y P, ZHU J, et al. Crowd scene understanding with coherent recurrent neural networks[C]//Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI). New York, NY, USA: IJCAI, 2016: 3469-3476.

[28] VEMULA A, MUELLING K, OH J. Social attention: Modeling attention in human crowds[C]//2018 IEEE International Conference on Robotics and Automation(ICRA). Brisbane, QLD, Australia: IEEE, 2018: 4601-4607.

[29] LIU Q, WU S, WANG L, et al. Predicting the next location: A recurrent model with spatial and temporal contexts [C]//30th AAAI Conference on Artificial Intelligence (AAAI). Phoenix, AZ, USA: AAAI, 2016: 194-200.

[30] ZHANG P, OUYANG W L, ZHANG P F, et al. SR-LSTM: State refinement for LSTM towards pedestrian trajectory prediction[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Long Beach, CA, USA: IEEE, 2019: 12077-12086.

[31] YI S, LI H S, WANG X G. Pedestrian behavior understanding and prediction with deep neural networks[C]// 14th European Conference on Computer Vision(ECCV). Amsterdam, Netherlands: Springer, 2016: 263-279.

[32] YAGI T, MANGALAM K, YONETANI R, et al. Future person localization in first-person videos[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Salt Lake City, UT, USA: IEEE, 2018: 7593-7602.

[33] LI J C, MA H B, TOMIZUKA M. Conditional generative neural system for probabilistic trajectory prediction[C]// 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS). Macau, China: IEEE, 2019: 6150-6156.

[34] SHAFIEE N, PADIR T, ELHAMIFAR E. Introvert: Human trajectory prediction via conditional 3D attention[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Nashville: IEEE, 2021: 16810-16820.

作者简介

孔 玮 女,1986年生,山东济南人. 青岛科技大学信息科学技术学院博士研究生. 主要研究方向为计算机视觉、轨迹预测.

刘 云 男,1962年生,山西太原人. 青岛科技大学信息科学技术学院教授、博士生导师. 主要研究方向为计算机视觉、轨迹预测等.

李 辉(通讯作者) 男,1984年生,河南平顶山人. 青岛科技大学信息科学技术学院副教授、硕士生导师. 主要研究方向为计算机视觉、多目标跟踪与轨迹预测等.
E-mail: lihui@qust.edu.cn

崔雪红 女,1978年生,山东菏泽人. 青岛科技大学信息科学技术学院高级实验师. 主要研究方向为计算机视觉、多目标检测与跟踪.

杨浩冉 女,1997年生,河北邯郸人. 青岛科技大学信息科学技术学院硕士研究生. 主要研究方向为3D点云目标跟踪.