计算机科学
COMPUTER SCIENCE

基于图注意力交互的行人轨迹预测方法

刘宏鉴, 邹丹平, 李萍

引用本文

刘宏鉴, 邹丹平, 李萍. 基于图注意力交互的行人轨迹预测方法[J]. 计算机科学, 2026, 53(1): 97-103.

LIU Hongjian, ZOU Danping, LI Ping. Pedestrian Trajectory Prediction Method Based on Graph Attention Interaction [J]. Computer Science, 2026, 53(1): 97-103.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.250300132

基于图注意力交互的行人轨迹预测方法

刘宏鉴　邹丹平　李　萍
上海交通大学电子信息与电气工程学院　上海 200240
(liuhongjian@sjtu.edu.cn)

摘　要　行人轨迹预测在自动驾驶领域和智慧交通领域均取得了显著的研究进展。由于行人的行为受到自身和环境因素的双重影响,其轨迹具有不确定性和复杂性,因此准确利用轨迹数据的交互特征生成多模态轨迹仍存在较大挑战。目前,该领域中的主要挑战是准确建模行人之间的时空交互。面对复杂的行人时空交互,提出了一种基于图注意力的时空图神经网络,其量化表示行人之间的空间交互并重点关注关键交互,从而将行人轨迹信息表示为有向时空图,利用图注意力机制提取空间位置特征和交互特征,同时结合自注意力机制在时间维度提取时间特征并融合时空特征信息,最后生成结合历史轨迹和交互信息的多模态未来轨迹。在ETH-UCY数据集上的实验表明,与最佳基线模型相比,所提出的方法在平均位移误差(ADE)和最终位移误差(FDE)方面分别降低3.4%和2.1%,并具有较短的推理时间,确保实现实时推理响应。可视化的结果表明,所提出的方法能够生成具有可接受性的未来行人轨迹,展现了良好的工程应用前景。

关键词:轨迹预测;时空图;图神经网络;图注意力;时空交互
中图分类号　TP391

Pedestrian Trajectory Prediction Method Based on Graph Attention Interaction

LIU Hongjian,ZOU Danping and LI Ping
School of Electronics,Information and Electrical Engineering,Shanghai Jiao Tong University,Shanghai 200240,China

Abstract　Pedestrian trajectory prediction has made significant research progress in the fields of autonomous driving and intelligent transportation. Due to the dual influence of individual and environmental factors,pedestrian trajectories exhibit uncertainty and complexity. Accurately generating multimodal trajectories by leveraging the interactive features of trajectory data remains a challenge. One of the primary challenges in this field is the accurate modeling of spatial temporal interactions among pedestrians. To address the complexity of pedestrian spatial temporal interactions,this paper proposes a spatial temporal graph neural network based on graph attention. The proposed method quantitatively represents the spatial interactions between pedestrians,focusing on key interactions,and represents pedestrian trajectory information as a directed spatial temporal graph. The spatial position features and interaction features are extracted using a graph attention mechanism,while the temporal features are obtained using a self-attention mechanism. By integrating spatial temporal feature information,the model generates multimodal future trajectories based on historical trajectory data and interaction information. Experiments conducted on the publicly available ETH-UCY dataset demonstrate that the proposed method outperforms the baseline models,achieving improvements of 3.4% and 2.1% in ADE and FDE,respectively. Additionally,the proposed model has a shorter inference time,ensuring real-time inference responses. Visualization results further indicate that the predicted pedestrian trajectories are plausible and socially acceptable,demonstrating promising prospects for engineering applications.

Keywords　Trajectory prediction,Spatial temporal graph,Graph neural networks,Graph attention,Spatial temporal interaction

1　引言

轨迹预测已成为自动驾驶[1]和交通控制[2]领域的研究热点。自动驾驶系统需要通过环境感知和行为规划来确保自身和乘客的安全,因此,准确预测车辆的未来轨迹对于智能汽车高效、安全的规划未来路径起至关重要的作用[3]。根据Daamen等的研究[4],行人个体行走行为可能受到交通状况、目的方位和群体移动等因素的影响,这些因素可通过客观的观察和分析进行研究。鉴于行人交通行为复杂且难以预测,近年来,基于学习的轨迹预测方法不断涌现。相关研究[5-7]充分利

到稿日期:2025-03-24　返修日期:2025-05-22
基金项目:国家重点研发计划(2022YFB3903802)
This work was supported by the National Key R&D Program of China(2022YFB3903802).
通信作者:邹丹平(dpzou@sjtu.edu.cn)

用了历史轨迹数据，并考虑了邻近个体轨迹的相关性，以建模行人个体之间的社会交互，从而实现了合理的行人轨迹预测。

大量具有相互关系的数据可以表示为图的形式。图结构数据可被用来高效建模节点之间的关系，并保留学习节点与边之间的拓扑信息[8]。例如 STGCN[2] 的方法采用了时空图的结构表示交通流量的时空信息，完成了在智慧交通领域的时序预测任务。基于图神经网络的轨迹预测方法在推理效率和可解释性方面取得了显著的提升[7]，因此图神经网络在轨迹预测领域得到迅速推广，并逐步成为该领域内的主流技术手段之一，对相关研究与发展产生了深远影响[9]。许多现有方法使用时空图来表示多行人的轨迹，可以通过图的结构表达时空交互信息。例如，Social-STGCNN[7] 在时空图中建模行人场景交互；SGCN[10] 充分利用了图数据的特性，通过注意力机制对行人交互信息进行建模；Social-TAG[11] 还添加了行人群组的表示，从而突出表示伴随和避让这一空间交互问题。这些方法都建立在图结构的基础上，以充分利用轨迹数据的多种表示形式，从而使模型能够学习更多的特征。

然而，现有方法在行人社会交互建模方面存在局限性，空间交互建模不够准确，且大多数方法忽视了时间维度的交互信息，导致难以捕捉行人在行走时产生的行为模式变化，如速度或者方向的变化。这是因为传统的行人社交建模方法通常采用人工定义的方式在节点之间建立空间交互，这种交互建立方式缺乏灵活性和可调整性。另外，过去的方法在节点之间建立和处理的连接是无向的，这限制了它们在实际场景中提供交互的定向表示能力。实际上，行人代理之间的交互可以有向表示，这意味着两个行人代理之间的空间影响不一定是相互的。

为了解决行人交互建模存在的问题，本文提出了根据现实影响构建有向时空图的方法，这种方法可以量化行人之间的非对称影响。时空有向图使用图注意力(Graph Attention，GAT)网络[12]来处理有向的空间交互，并通过注意力机制来处理时间交互。Zhao 等的实验证明了多个维度特征融合在时间序列特征化方面是有效的[13]。因此，文中提出了一种时空图神经网络来融合轨迹节点在空间和时间两个维度的交互，并基于学习到的交互生成行人未来轨迹，从而提高行人轨迹预测的准确性。

本文的主要工作可以总结为：构建有向图，以定量表示行人之间的交互；提出了时空图神经网络，通过图注意力和自注意力处理和学习空间和时间维度的交互信息，并将不同维度交互融合；随后，利用时间卷积网络(Temporal Convolution Network，TCN)来预测行人轨迹的分布，减少了资源消耗；最后，通过公开数据集对所提方法进行系统性评估，并对生成的轨迹进行可视化定性分析，以比较和分析不同轨迹生成方法的结果。

2 相关工作

2.1 行人交互建模

研究者们对行人轨迹预测这一任务已经进行了广泛的研究，目前的研究大多数基于鸟瞰视角进行行人轨迹预测。行人轨迹预测任务的主要挑战包括行人之间的社会交互以及行人运动模式的多样性。目前大多数现有方法主要关注行人之间的交互建模问题，而针对行人运动的不确定性提出的解决方案较少[6]。

现有的轨迹预测方法主要基于深度学习完成行人轨迹与社会交互建模。Social LSTM[5] 是其中一个典型的例子，它是早期基于深度学习的行人轨迹预测方法之一。Social LSTM 利用循环神经网络(Recurrent Neural Network，RNN)对行人运动进行建模，并通过池化机制聚合 RNN 的输出以提取特征。另一种常用的方法是基于图神经网络(Graph Neural Network，GNN)的方法[9]。在这类方法中，行人被表示为图中的节点，行人之间的交互关系被定义为节点之间的边。例如，Social-STGCNN[7] 利用图卷积网络(Graph Convolution Network，GCN)建模行人之间的社会交互。SGCN[10] 则采用注意力机制处理行人之间的交互关系，同时利用稀疏矩阵进行缩放，以缓解行人交互带来的挑战。后续的研究提出了其他方法，以进一步提高行人交互建模的准确性和泛化能力。

然而，基于 SGCN 和 Social-STGCNN 的方法主要针对无向图交互数据，这在解决行人社会交互中的非对称性问题时存在局限性。为了解决行人社会交互有向性的问题，本文提出了一种专注于行人实际交互的方法，通过构建更精确的有向交互矩阵，有效捕捉行人之间复杂的社会交互关系，从而更加准确地建模和预测行人的行为。

2.2 自注意力机制

自注意力机制最初由 Transformer[14] 提出，并在更复杂的领域中得到推广。它在深度学习领域中具有重要的应用价值，能够有效建模数据的远程依赖关系和基于上下文的交互问题。与传统的序列模型和卷积神经网络相比，自注意力机制能够在序列中的所有成对交互之间动态分配权重。传统模型通常按顺序处理数据，或使用固定的感受野；而自注意力机制的权重具有高度的可解释性，并且能够按需处理和分配注意力。

由于自注意力机制对数据输入结构具有更强的承载能力和出色的可扩展性，因此它被广泛应用于多种架构中，以处理数据本身的结构特性。例如，在自然语言处理(Natural Language Processing，NLP)领域，基于自注意力机制的 BERT 和 GPT 等模型取得了显著成果；在计算机视觉领域，视觉 Transformer(Vision Transformer，ViT)[15] 利用自注意力机制进行图像特征提取，表现出优异的性能。

在本研究中，所提出的方法利用自注意力机制来计算时间序列之间的潜在关系，从而更全面地捕捉序列数据中的时序特征和依赖关系。

2.3 图注意力

图注意力网络[12]是一种处理图数据的方法。与 GCN 相比，GAT 的优势主要体现在注意力机制的应用使其能够处理并捕捉节点之间的细微关系；同时，它还具备处理异构数据或有向图的能力。通过在聚合过程中动态地为邻居节点分配权重，GAT 利用注意力机制增强了图数据结构的表示能力。它计算节点之间的成对注意力系数，使模型能够专注于局部相

关的邻居节点。这一创新不仅提升了模型的可解释性，还在空间域中提供了强大的特征提取能力。后续的研究，如GATv2[16]，引入了改进的注意力机制，以解决GAT中的静态注意力问题，从而确保了更强的拟合能力。GAT已广泛应用于多个领域，例如社交网络分析[17]、生物信息学[18]和推荐系统[19]，并在某些场景下展现出了卓越的性能。

3 行人轨迹预测方法

3.1 建立有向时空图

多个行人轨迹的序列可以构建为时空图。在给定的时间 t 下，N 条不同行人代理的轨迹 $\{p_1,p_2,\cdots,p_N\}$ 可以构建为包含相对位置的图 G_t。该图可以表示为 $G_t=(V_t,E_t)$，其中 $V_t=\{v_t^i \mid \forall i\in\{1,\cdots,N\}\}$ 表示包含行人代理空间位置的节点，$E_t\subseteq\mathbb{R}^{N\times N}$ 表示图 G_t 的边，$E_t=\{e_t^{i,j} \mid \forall i,j\in\{1,\cdots,N\}\}$。对于图的边 E，若节点 i 存在到节点 j 之间的有向图连接，则 $e_{i,j}>0$；否则 $e_{i,j}=0$。在对连接的节点边进行加权后，将其作为空间交互的初始化注意力，该方法通过这些初始化的注意力权重学习节点之间的影响力。

在相同场景中，并非所有行人个体之间都存在社会交互，因此不能仅将两个节点之间的欧几里得距离作为行人代理之间的边权重。例如，基于行人视线原因，位于某行人后方的行人不会对其行为产生影响，在建模时需要考虑此类情况。为了准确模拟现实场景中的社会交互，应综合考虑与行人社会交互相关的多种因素，例如两个行人的速度、方向以及行人之间的距离。这些因素有助于创建一个有向图。有向图的边权重表示方法如下：

$$e_t^{i,j}=\frac{\boldsymbol{u}_t^{i,j}\cdot\boldsymbol{l}_t^{i,j}+\boldsymbol{u}_t^{j,i}\cdot\boldsymbol{l}_t^{j,i}}{|\boldsymbol{l}_t^{i,j}|^2} \tag{1}$$

$$\hat{e}_t^{i,j}=\sigma(0,e_t^{i,j}) \tag{2}$$

其中，$\boldsymbol{u}^i$ 表示 i 的速度向量；$\boldsymbol{l}^{i,j}$ 表示从 i 到 j 的位移向量；σ 表示一个非线性函数，由Softmax与ReLU组合而成，目的是使计算后的权重数值可以保持在(0,1]的范围内。

3.2 时间交互注意力

该方法引入了一个时间注意力机制模块来提取时空图中的时间信息。然而，注意力机制对向量的位置信息不敏感，它会忽略每个轨迹点的时间关系[14]。在此之前，将时间位置编码数据输入图节点信息的向量嵌入中，其位置编码如下所示：

$$Penc_{(pos,2i)}=\sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \tag{3}$$

$$Penc_{(pos,2i+1)}=\cos\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \tag{4}$$

其中，pos 表示时空图序列中节点元素的位置编号，i 表示特征维度的索引，d 是注意力层嵌入向量的维度。

将这些组合后的数据纳入时间注意力机制层，使该方法能够学习不同时间节点之间的相互关系。不同的时间点对整个时间过程具有不同程度的影响，注意力机制可以很好地区分这种影响。多头注意力(Multi-head Attention)通过多个查询(queries)、键(keys)和值(values)集合同时进行学习。本方法利用多头自注意力构建该模块，使其能够以灵活的方式处理时间维度的连接。在这个过程中，查询、键和值矩阵是通过向量嵌入将带有位置编码的图节点信息 $\boldsymbol{G}_{\text{enc}}$ 转换到时间序列的维度得到的，其中 $\boldsymbol{Q}=\boldsymbol{G}_{\text{enc}}\times\boldsymbol{W}_q$，$\boldsymbol{K}=\boldsymbol{G}_{\text{enc}}\times\boldsymbol{W}_k$，$\boldsymbol{V}=\boldsymbol{G}_{\text{enc}}\times\boldsymbol{W}_v$；$\boldsymbol{W}_q$，$\boldsymbol{W}_k$ 和 $\boldsymbol{W}_v$ 为参数。每个头 i 的注意力计算式为：

$$Att_i(\boldsymbol{Q}_i,\boldsymbol{K}_i,\boldsymbol{V}_i)=\text{Softmax}\left(\frac{\boldsymbol{Q}_i\boldsymbol{K}_i^{\mathrm{T}}}{\sqrt{d_k}}\right)\cdot\boldsymbol{V}_i \tag{5}$$

其中，$\boldsymbol{Q},\boldsymbol{K}$ 具有相同的维度，记为 d_k。不同头的注意力表示将被拼接输出。这个过程可以获得与时间维度相关的注意力权重，用于获取时间维度的时空图特征。

3.3 空间图注意力

本方法采用图注意力网络来处理有向时空图在空间维度上的交互特征。图注意力网络由图注意力层组成。其中，输入到图注意力层的特征为 $\boldsymbol{h}^{(l)}=\{h_1^{(l)},h_2^{(l)},\cdots,h_N^{(l)}\}$，$h_i^{(l)}\in\mathbb{R}^F$，其中 N 表示图中行人节点的数量，F 表示节点的特征数量。此图注意力层生成一组新的特征 $\boldsymbol{h}^{(l+1)}=\{h_1^{(l+1)},h_2^{(l+1)},\cdots,h_N^{(l+1)}\}$，$h_i^{(l+1)}\in\mathbb{R}^{F'}$。其中，$F$ 和 F' 可能不相等。输入到注意力层的初始特征是行人的二维相对位置，通常可以认为输出特征数量为2。输入时空图在空间维度上的初始权重，通过两种不同版本的图注意力网络处理计算空间维度特征，并将其转化为更高维度特征。GAT[12]的计算式如下：

$$e'_{i,j}=\text{LeakyReLU}(\boldsymbol{a}^{\mathrm{T}}[\boldsymbol{W}h_i\parallel\boldsymbol{W}h_j]) \tag{6}$$

$$\alpha_{i,j}=\sigma(e'_{i,j})=\frac{\exp(\text{LeakyReLU}(\boldsymbol{a}^{\mathrm{T}}\cdot[\boldsymbol{W}h_i\parallel\boldsymbol{W}h_j]))}{\sum_{k=1}^{N}\exp(\text{LeakyReLU}(a^{T}\cdot[\boldsymbol{W}h_i\parallel\boldsymbol{W}h_k]))} \tag{7}$$

为了解决可能由初始化引起的注意力单调收敛问题，Brody等提出了GATv2[16]，其计算式如下：

$$e'_{i,j}=\boldsymbol{a}^{\mathrm{T}}\text{LeakyReLU}(\boldsymbol{W}[h_i\parallel h_j]) \tag{8}$$

$$\alpha_{i,j}=\sigma(e'_{i,j})=\frac{\exp(\boldsymbol{a}^{\mathrm{T}}\text{LeakyReLU}(\boldsymbol{W}\cdot[h_i\parallel h_j]))}{\sum_{k=1}^{N}\exp(\boldsymbol{a}^{\mathrm{T}}\text{LeakyReLU}(\boldsymbol{W}\cdot[h_i\parallel h_k]))} \tag{9}$$

其中，$e'_{i,j}$ 表示节点 j 的特征对节点 i 的重要性；σ 采用Softmax函数调整计算的注意力，以便归一化计算得到的注意力；$\boldsymbol{W}\in\mathbb{R}^{F*F'}$ 为特征变换矩阵，“$\parallel$”表示连接或取平均操作。LeakyReLU的负斜率设置为0.2。图注意力 l 层中节点 i 的输出可以表示为：

$$h_i^{(l+1)}=\sigma\Big(\sum_{j\in N_i}\overset{\wedge}{\alpha}_{i,j}^{(l)}\boldsymbol{W}^{(l)}h_j^{(l)}\Big) \tag{10}$$

其中，σ 为激活函数；$\boldsymbol{W}^{(l)}$ 为图注意力 l 层的参数；$\overset{\wedge}{\alpha}_{i,j}^{(l)}$ 表示注意力矩阵根据有向图边的关系经过乘法调整后的结果，调整方法定义如下：

$$\overset{\wedge}{\alpha}_{i,j}^{(l)}=f_\sigma(\alpha_{i,j}^{(l)}\odot\hat{e}_{i,j}) \tag{11}$$

其中，f_σ 表示ReLU函数；$\odot$ 表示相同位置相乘的操作，在这里充当掩码操作。空间图注意力采用多头注意力机制计算。

3.4 时间卷积网络

最后，时空图提取融合的特征信息，通过时间卷积神经网络回归生成轨迹的二维高斯分布，从而预测行人的未来轨迹。提取的时空图嵌入，经过图注意力和时间注意力步骤后，输入到该网络中。然后，原始的时空图根据需要预测的时间扩展为新的长度，从而得到预测的输出。时间

卷积层的计算式可以表示为：

$$\boldsymbol{X}_{(l+1)}=\sigma\cdot Conv(\boldsymbol{W}_{(l)},\boldsymbol{X}_{(l)}) \tag{12}$$

其中，$\boldsymbol{W}_{(l)}$ 表示 l 层二维卷积核的参数，$Conv$ 表示卷积操作，σ 表示采用 ReLU 作为卷积操作的非线性激活函数，卷积层之间一般存在残差连接。

3.5 方法定义和损失函数

行人的轨迹通常用坐标表示，在本方法中行人的轨迹表示为二维坐标，对一个包含 N 名行人的场景 $\boldsymbol{X}_t(t\in T_i, T_i$ 表示输入的时间长度)，行人轨迹可以表示为 $\boldsymbol{p}_t^n=(x_t^n,y_t^n)$，$n\in N$。类似地，输出的生成轨迹可以表示为 $\boldsymbol{X}_o, t\in T_o$，其中每个行人的轨迹可以表示为 $\hat{p}_t^n$。(x_t^n,y_t^n) 是描述行人 n 在时间 t 的二维空间位置的随机变量，其遵循二维高斯分布 $(x_t^n,y_t^n)\sim N(\mu_t^n,\sigma_t^n,\rho_t^n)$，则 $\hat{p}_t^n$ 同样遵循二维高斯分布。该方法的优化过程使用负对数似然损失，损失函数定义如下：

$$\mathcal{L}^n(W)=-\sum_t^{T_o}\log_e(\mathbb{P}(\hat{p}_t^n|\mu_t^n,\sigma_t^n,\rho_t^n)) \tag{13}$$

图 1 展示了基于空间图注意力和时间注意力的行人轨迹预测方法的流程图。该方法首先接收行人历史轨迹，并将其表示为有向时空图，随后分别通过在时间和空间维度上应用自注意力和图注意力来提取特征，从而获得高维向量。将两个部分获得的高维向量连接在一起，而后通过多层感知机(Multilayer Perceptron，MLP)整合特征。此步骤下，提取到的时空特征可以被同时融合，这进一步增强了行人历史行为场景的表示。在空间维度上，其特征可以有效捕捉人群避让、跟随等社交行为；而时间维度特征可以识别行人速度、方向变化等演变规律。最后，采用时间卷积网络生成未来轨迹的二维高斯分布，并采用随机采样器生成多模态的行人未来轨迹。

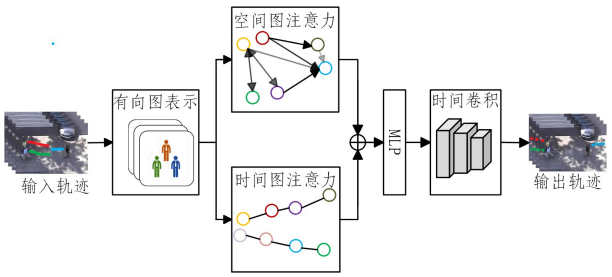

图 1 基于空间图注意力和时间注意力的行人轨迹预测方法的框架

Fig. 1 Structure of the pedestrian trajectory prediction model based on spatial GAT and temporal attention

4 实验及分析

4.1 数据集

行人轨迹模型基于公开行人轨迹数据集 ETH[20] 和 UCY[21]，其中包含了不同场景的大量轨迹数据。在 ETH-UCY 数据集下的训练策略与 Social-LSTM[5] 相同，去掉数据集的一部分用于测试，而其余部分均用于训练。仍然沿用先前工作的设置，在这些数据集下的实验都观察了 8 帧(3.2 s)的轨迹，并预测了接下来的 12 帧(4.8 s)，因此在训练和定量评估的过程中需要得到某个行人 20 帧(8 s)的完整轨迹。

4.2 评估方法和实验参数设置

使用两个指标来评估模型：平均位移误差(Average Displacement Error，ADE)和最终位移误差(Final Displacement Error，FDE)。最终的预测结果呈二维高斯分布。比较的过程均采用 Best-of-20 评估方法来进行评估，即生成多模态轨迹数量为 20，并在其中选择最接近真实轨迹的作为误差计算的依据。评估指标 ADE 和 FDE 可以定义为：

$$ADE=\frac{\sum_{n\in N}\sum_{t\in T_p}\|\widehat{p_t^n-p_t^n}\|_2}{N\times T_p} \tag{14}$$

$$FDE=\frac{\sum_{n\in N}\|\widehat{p_t^n-p_t^n}\|_2}{N},t=T_p \tag{15}$$

在实验中，模型使用了一个时空注意力层和 5 个时间卷积层进行生成。多头注意力头的个数均设置为 4，时间注意力的隐藏层向量嵌入维度设置为 64。训练的批次大小设置为 128，模型经过 200 个轮次训练，使用验证数据集来决定是否更新权重。初始学习率设置为 0.001，并且每 100 个训练轮次衰减一个因子 0.1。该方法在 Ubuntu 20.04.6 和 PyTorch 1.11.0 上实现。所有训练和实验都在 NVIDIA RTX 3090 和 Xeon Platinum 8255C 上进行。

4.3 消融实验

本节设计了一系列消融实验来验证网络设计的有效性。在单独的分析中，将有向交互图的输入转化为无向交互图的输入，类似于 Social-STGCNN[7]，并去除了时间维度的注意力，使用不同版本的图注意力网络进行计算。在消融实验中，分别采用去除时间注意力机制("w/o T-Attn")、去除有向社会交互而使用欧几里得距离影响的无向交互("w/o Directed Edge")、并分别比较不同版本的 GAT，在 ETH-UCY 数据集上训练，记录位移误差。实验结果如表 1 所列。

表 1 在 ETH-UCY 公开数据集上的消融实验 ADE/FDE 结果

Table 1 ADE/FDE results of ablation study on ETH-UCY dataset

(m)

	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
w/o T-Attn	0.55/0.89	0.28/0.50	0.28/0.48	0.22/0.40	0.20/0.39	0.306/0.532
w/o Directed Edge	0.54/0.88	0.30/0.53	0.27/0.44	0.20/0.35	0.19/0.29	0.300/0.498
gat v1	0.47/0.78	0.25/0.44	0.28/0.47	0.22/0.38	0.20/0.35	0.284/0.484
gat v2	0.47/0.75	0.26/0.49	0.26/0.44	0.20/0.36	0.19/0.31	0.276/0.470

结果表明，当去除时间维度的注意力时，模型在轨迹生成中的性能受到不利影响，尤其是在生成较长序列的数据时，因此可以认为时间维度的注意力可以学习到行人速度和方向趋势这一特征，并融合这一特征来生成未来的轨迹模式。建立有向的行人间交互使模型能够准确建立和学习行人之间可能产生的社会交互活动，从而判断其他行人位置对自身行为

的影响，这对于轨迹生成是必要的。结果还表明，在行人行走比较快速而道路狭窄的场景中，模型具备更小的位移误差。所提方法建立的有向交互能够针对性地调整行人之间的交互关系，去除实际场景中很难存在的社会交互影响，从而强化可能影响较大的行人交互。此外，不同版本的图注意力网络在行人轨迹预测任务中对预测结果的影响较小，这是由轨迹数据组成图的结构导致的结果。定量对比表明，GATv2 的位移误差比 GATv1 更小。

4.4 定量对比

4.4.1 位移误差对比

选择了一些最先进的方法和一些最新的基于图神经网络的方法作为基准，以比较验证所提方法的性能。表 2 分析并比较了所提方法与其他模型在 ADE/FDE 指标上的性能结果，所有方法均采用根据过去 8 帧生成未来 12 帧的方式进行准确性对比，其中带有“*”的方法均采用图的结构表示行人的轨迹，其余方法均采用序列数据的方式表示行人的轨迹。

表 2 与过去方法 ADE/FDE 对比结果

Table 2 Comparison results of ADE/FDE with the previous methods

(m)

Model	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
Social LSTM[5]	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[6]	0.87/1.62	0.67/1.37	0.76/1.52	0.35/0.68	0.42/0.84	0.61/1.21
Social-BiGAT*[22]	0.69/1.29	0.49/1.01	0.55/1.32	0.30/0.62	0.36/0.75	0.48/1.00
PECNet[23]	0.54/0.87	0.18/0.24	0.35/0.60	0.22/0.39	0.17/0.30	0.29/0.48
Social-STGCNN*[7]	0.64/1.11	0.49/0.85	0.44/0.79	0.34/0.53	0.30/0.48	0.44/0.75
SGCN*[10]	0.63/1.03	0.32/0.55	0.37/0.70	0.29/0.53	0.25/0.45	0.37/0.65
Social TAG*[11]	0.61/1.00	0.37/0.56	0.51/0.87	0.33/0.50	0.30/0.49	0.42/0.68
DSTCNN[24]	0.53/1.08	0.19/0.34	0.29/0.53	0.23/0.43	0.23/0.43	0.29/0.53
Our Method	0.47/0.75	0.26/0.49	0.26/0.44	0.20/0.36	0.19/0.31	0.28/0.47

总体而言，所提方法在 ADE 和 FDE 上都得到了改进。相较于 Social-STGCNN[7]，所提方法在 ADE 上降低了 36.4%，这是由于其能够建立有向的行人交互，因此能够学习到准确的空间交互表征。相比于 SGCN[10]，所提方法在 ADE 指标的准确性上提高了 24.3%，这是因为其采用了构建有向图的形式来去除行人之间的冗余交互，能够更好地学习行人之间的交互而避免不必要的影响。相比于 SGCN[10]，所提方法在 FDE 的准确性上提升了 27%，这得益于模型关注时间维度的数据，可以提供更精确的速度和方向概率。与一些过去的生成模型相比，所提方法在 ADE 和 FDE 指标上更优。相比于准确性较高的 PECNet[23]，所提方法在 ADE 性能方面提升了 3.4%，在 FDE 性能方面提升了 2.1%。

4.4.2 推理时间和参数大小对比

表 3 对所提方法与其他方法的推理时间和参数量进行了比较。

表 3 各方法的推理时间和参数大小对比

Table 3 Inference time and parameter size comparison of each method

Model	Generator	Inference Speed/s	Parameters
Social GAN[6]	GAN	0.0968	4.63×10^4
PECNet[23]	VAE	0.4357	2.1×10^6
Social-STGCNN[7]	TCN	0.0020	7.6×10^3
SGCN[10]	TCN	0.0042	2.53×10^4
DSTCNN[24]	TCN	0.0013	4.1×10^3
Our Method	TCN	0.0058	3.32×10^4

从表中可以明显看出，采用基于 TCN 生成的方法在参数量和推理时间上明显较低，大大减少了资源消耗。这对于有速度要求的任务和具有适度计算能力的机器是有利的。模型参数主要由时间维度的注意力机制和空间维度的图注意力机制表示。具体来说，所提方法和 PECNet[23] 在准确性指标上接近，但前者在模型参数大小和推理时间方面远远更优。在参数数量上，所提方法与 PECNet[23] 相比减少了 98.4%，与 Social-GAN[6] 方法相比减少了 28.3%，这是因为所提方法在生成结构上简单。然而，相比于一些模型，如 Social-STGCNN[7] 和 DSTCNN[24]，所提方法的参数数量稍大，因为时间注意力模块通常需要比 CNN 更多的参数。总体而言，所提方法在准确性和模型大小之间实现了一定的平衡性。

4.5 定性分析

准确性并不是评估模型性能的唯一指标。通过可视化的定性分析，通常有助于更好地评估模型的实用性，即能否生成真实可用的轨迹。图 2 展示了几个不同场景下的行人轨迹分布图（用于观测的过去轨迹用虚线表示，未来真实轨迹用实线表示），采用生成多模态轨迹的方法显示生成轨迹的核密度估计（Kernel Density Estimation，KDE）图，颜色越深表示行人未来有更大的可能性会移动到这个位置。将所提模型与常用于定性分析的两种先进方法 Social-STGCNN[7] 和 SGCN[10] 进行了对比，这两种方法均以图结构表示行人轨迹数据。前两个场景展示了行人平行行走的情况。在第一个场景中，两名行人并排走，而第二个场景中则是行人前后走的情况。可以观察到，所提方法通过有意减弱前后行人之间的关系，展示了更准确的行人移动趋势，这是因为建立的有向图可以去除多余的交互，从而使行人的影响更加符合现实情况。最后两个场景展示了行人相对行走的情况。根据常识，模型应该预测两名行人会避让对方。尽管第三个场景中行人路径并不十分接近，但两人的未来轨迹都趋向于避开对方，所提方法很好地实现了这一点，证明了其合理性。相比之下，第四个场景中有两组行人朝相反方向行走。所提方法不仅相对准确地预测了每个行人的运动趋势，还能使并排行走的个体未来轨迹预测方向趋于一致，展示了更好的准确性，这也是因为本方法能够准确考虑到必要的空间交互，并且结合时间交互在复杂的场景下生成准确的轨迹。同样，所提方法能够根据历史轨迹辨识行人的行走趋势。这也证明了深度学习模型在一定程度上能够适应这一任务。

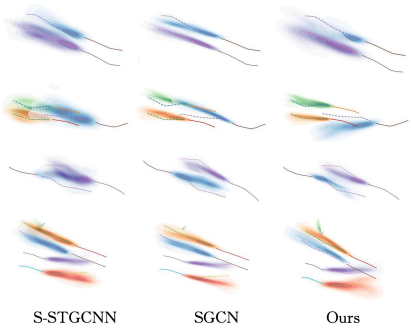

图 2 模型的定性分析对比

Fig. 2 Qualitative analysis of our method

图 3 展示了模型经过图注意力后的空间注意力。我们选择了两个时空场景作为基准进行分析比较，由于节点自身注意力在模型中会设置为常数值，因此不考虑将行人节点对自身的注意力加入比较。该图表示在某一时刻，其他行人对某个行人的影响。白色表示模型在调整之后不会考虑两个节点之间的注意力机制，而颜色越深，表示另一个节点对该节点的影响越大。模型可以基于图注意力处理节点之间的交互，并利用已建立的有向交互进一步影响节点的决策。结果表明，所提方法能够建立有向的空间交互，从而使模型在提取空间特征时更准确地得到必要的社会交互，避免无用交互可能引发的问题。

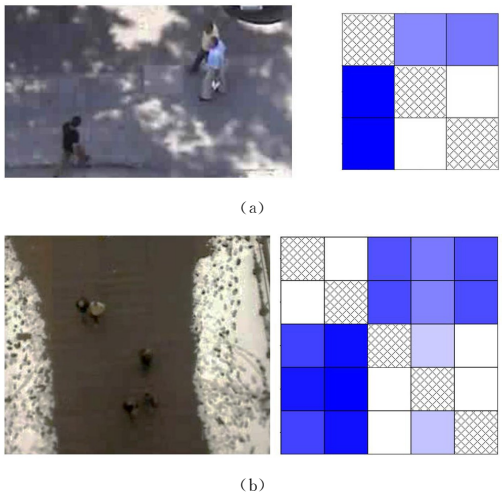

图 3 空间注意力处理之后的交互强弱可视化示意图

Fig. 3 Visualization of the interaction after spatial attention processing

结束语 本文提出了一种基于时空图的行人轨迹预测方法。该方法通过构建有向图并结合图注意力和时间注意力机制来准确提取行人轨迹特征和社会交互信息，从而提升模型的性能。通过对比分析不同图注意力机制变体在行人轨迹预测任务上的性能得出，该方法能够自适应地分配行人个体之间的空间交互关系并考虑时空图时间维度的交互，从而取得出色的轨迹预测结果。该方法在公共数据集上表现出较强的性能，同时具有较低的计算成本。未来工作将探索和丰富更复杂的图结构设计，利用图神经网络在更广泛且更复杂的场景中识别、分类和预测行人或其他交通参与者。此外，未来还可以将该方法集成到无人系统中，并结合感知结果，用于路径规划。

参 考 文 献

[1] GAO J, SUN C, ZHAO H, et al. Vectornet: Encoding hd maps and agent dynamics from vectorized representation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11525-11533.

[2] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[J]. arXiv: 1709.04875, 2017.

[3] 易虹宇，杨智宇，杜力. 基于变分自动编码器的车辆轨迹预测研究[J]. 重庆工商大学学报(自然科学版)，2024，41(2)：60-65.

[4] DAAMEN W, HOOGENDOORN S P. Experimental research of pedestrian walking behavior[J]. Transportation Research Record, 2003, 1828(1): 20-30.

[5] ALAHI A, GOEL K, RAMANATHAN V, et al. Social lstm: Human trajectory prediction in crowded spaces[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 961-971.

[6] GUPTA A, JOHNSON J, FEI-FEI L, et al. Social gan: Socially acceptable trajectories with generative adversarial networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2255-2264.

[7] MOHAMED A, QIAN K, ELHOSEINY M, et al. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 14424-14432.

[8] HOU L, LIU J, YU X, et al. Review of graph neural network [J]. Computer Science, 2024, 51(6): 282-298.

[9] CAO J, CHEN Y, LI H, et al. A survey of pedestrian trajectory prediction based on graph neural network[J]. Computer Engineering & Science, 2023, 45(6): 1040-1053.

[10] SHI L, WANG L, LONG C, et al. SGCN: Sparse graph convolution network for pedestrian trajectory prediction[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 8994-9003.

[11] ZHANG X, ANGELOUDIS P, DEMIRIS Y. Dual-branch spatio-temporal graph neural networks for pedestrian trajectory prediction[J]. Pattern Recognition, 2023, 142: 109633.

[12] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. arXiv: 1710.10903, 2017.

[13] ZHAO H, WANG Y, DUAN J, et al. Multivariate time-series anomaly detection via graph attention network[C]// 2020 IEEE International Conference on Data Mining (ICDM). IEEE, 2020: 841-850.

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv: 1706.03762, 2017.

[15] DOSOVITSKIY A. An image is worth 16×16 words: Transformers for image recognition at scale[J]. arXiv: 2010.11929,

2020.

[16] BRODY S,ALON U,YAHAV E. How attentive are graph attention networks? [J]. arXiv:2105.14491,2021.

[17] BUSBRIDGE D,SHERBURN D,CAVALLO P,et al. Relational graph attention networks[J]. arXiv:1904.05811,2019.

[18] SHAO K,ZHANG Y,WEN Y,et al. DTI-HETA:prediction of drug-target interactions based on GCN and GAT on heterogeneous graph[J]. Briefings in Bioinformatics,2022,23(3):bbac109.

[19] WANG X,HE X,CAO Y,et al. Kgat:Knowledge graph attention networkfor recommendation[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019:950-958.

[20] LERNER A,CHRYSANTHOU Y,LISCHINSKI D. Crowds by example[C]// Computer graphics forum. Oxford,UK:Blackwell Publishing Ltd,2007:655-664.

[21] 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. IEEE,2009:261-268.

[22] KOSARAJU V,SADEGHIAN A,MARTÍN-MARTÍN R,et al. Social-bigat:Multimodal trajectory forecasting using bicycle-gan and graph attention networks[C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems,2019:137-146.

[23] MANGALAM K,GIRASE H,AGARWAL S,et al. It is not the journey but the destination:Endpoint conditioned trajectory prediction[C]// Computer Vision-ECCV 2020:16th European Conference,Glasgow,UK,August 23-28,2020,Proceedings,Part II 16. Springer International Publishing,2020:759-776.

[24] CHEN W,SANG H,WANG J,et al. DSTCNN:Deformable spatial-temporal convolutional neural network for pedestrian trajectory prediction [J]. Information Sciences, 2024, 666: 120455.

LIU Hongjian, born in 1999, postgraduate. His main research interest is perception, prediction and planning control of autonomous systems.

ZOU Danping, born in 1982, Ph.D, professor, Ph.D supervisor. His main research interests include synchronous positioning and map construction, 3D visual perception and autonomous systems.

(责任编辑:柯颖)