

DOI: 10.13382/j.jemi.B2105072

结合速度控制的时空图网络行人轨迹预测模型*

王海峰 桑海峰 王金玉 陈旺兴

(沈阳工业大学信息科学与工程学院 沈阳 110870)

摘要:行人轨迹预测中最重要的任务是建立行人轨迹交互模型,针对在模型中缺乏关于时间和速度等信息的建模,提出了一种结合速度控制的时空图网络算法来建立行人交互模型并对轨迹进行预测.整体模型采用条件生成对抗网络架构,其中采用速度预测模块预测行人未来速度并作为条件生成对抗网络的控制条件,显式地将速度信息引入行人轨迹预测,避免较大偏差速度对轨迹的影响.在生成器中设计了基于图卷积注意力机制的时空信息融合模块,在提取行人轨迹序列运动特征并关注其空间上相互作用关系的同时,显式地编码行人序列的时间相关性.最后,将结合时空信息和速度信息的轨迹交互特征解码,完成轨迹的预测.此外,考虑到现有评价方法的不足,采用平均碰撞次数作为轨迹合理性的评判.在公开数据集 ETH 和 UCY 上进行验证,实验结果表明,该文所提出的算法能更好地完成行人轨迹预测,平均位移误差为 0.40 m 和最终位移误差为 0.79 m.

关键词:行人轨迹预测;生成对抗网络;速度控制;时空图网络;平均碰撞次数

中图分类号: TP391;TN98 **文献标识码:** A **国家标准学科分类代码:** 520.20

Spatial-temporal graph network with speed control pedestrian trajectory prediction model

Wang Haifeng Sang Haifeng Wang Jinyu Chen Wangxing

(The School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870, China)

Abstract: The most important task in pedestrian trajectory prediction is to establish a pedestrian trajectory interaction model. Aiming at the lack of semantic information about time and speed in the model, a spatial-temporal graph network algorithm combined with speed control is proposed to establish pedestrian interaction model and predict trajectory. The overall model adopts the conditional generative adversarial networks architecture, in which the speed prediction module is used to predict the future speed of pedestrians, and the control condition of the conditional generative adversarial networks. The speed information is explicitly introduced into the pedestrian trajectory prediction to avoid the influence of large deviation speed on the trajectory. A spatial-temporal information fusion module is designed in the generator. While extracting the motion features of pedestrian trajectory sequence and paying attention to its spatial interaction, it explicitly encodes the temporal correlation of pedestrian sequence. Finally, the trajectory interactive features combined with space-time information and speed information are decoded to complete the trajectory prediction. In addition, considering the shortcomings of the existing evaluation methods, the average collision times is used as the evaluation of trajectory rationality. The model is verified on the public datasets ETH and UCY. The experimental results show that the proposed algorithm can better complete the pedestrian trajectory prediction, with an average displacement error of 0.40 m and a final displacement error of 0.79 m.

Keywords: pedestrian trajectory prediction; generative adversarial network; speed control; spatial-temporal graph network; average collision times

0 引言

在复杂的动态场景中建立行人交互模型并完成轨迹预测是近年来一个热门的研究方向,广泛应用于众多领域,例如自动驾驶^[1]、智能移动服务机器人^[2]和智慧城市交通监控系统^[3]等。行人轨迹的正确预测有助于智能导航系统预测其他目标位置以保持安全距离并避免相应的碰撞,规划出更加合理有效的路径来进行移动^[4,5]。

随着深度学习技术的兴起,行人轨迹预测的研究从传统的人工定义特征函数^[6-7]来对行人交互进行建模,转换成通过复杂数据驱动的深度神经网络结构对变量结果的潜在学习。长短时记忆网络(long short-term memory, LSTM)在行人轨迹预测时表现优异,但 LSTM 网络存在预测所有未来轨迹平均值的问题,这并不符合轨迹预测多模态性的要求。生成对抗网络^[8-10](generative adversarial networks, GAN)可以很好解决 LSTM 类模型的缺陷,完成轨迹的多模态预测。其中,社会生成对抗网络模型^[11]是应用最为广泛的 GAN 类模型。在 GAN 模型的基础上为更好地结合上下文信息,张志远等^[12]在模拟行人交互时采用注意力机制将行人间的距离及方向信息纳入模型。Sadeghian 等^[13]建立 SoPhie 模型将环境信息纳入考量,结合了场景语义分割模型和生成对抗网络的特征,完成行人的轨迹预测。

在模拟行人交互行为时,此前大多数研究多采用池化机制来完成,这导致了大量信息的丢失。而图卷积神经网络(graph neural network, GNN)可以获取图结构中的信息并进行学习,是一种建模行人间交互关系的有效方法。Huang 等^[14]基于图注意力网络(graph attention network, GAT)为临近每个行人节点匹配不同的注意力实现了节点之间有效的加权信息传递;Kosaraju 等^[15]模型用 GAT 对场景中行人间的交互行为进行编码并结合场景信息。

上述方法都能很好地模拟行人间的交互关系,但是还都存在一定的局限。首先,这些方法在提取行人之间的交互信息时所考虑的信息是片面的,忽略了显式速度信息等因素对目标行人未来轨迹的影响,仅仅只考虑了 LSTM 编码中速度的隐式信息;其次,行人在场景中移动时轨迹具有连续性,大多数现有的方法仅仅只考虑行人间的空间相关性,而忽略了与场景中涉及的其他行人互动的持续性。最后,目前大多数方法预测行人轨迹时仅仅只考虑轨迹预测的精度,而很少考虑到预测轨迹的合理性。

针对上述问题,本文提出一种结合速度控制的时空图卷积神经网络算法(spatial-temporal graph network with speed control generative adversarial networks, STGS-GAN)

对轨迹进行预测。首先,以条件生成对抗网络^[16](conditional generative adversarial networks, CGAN)模型为主体结构,将预测的速度信息作为 CGAN 的控制条件,更好地提升模型对速度因素影响的敏感性。其次,在 CGAN 模型的生成器中设计了时空信息融合模块,可以更好地提取行人间的时空交互信息。此外,将平均碰撞次数(average collision times, ACT)作为评价指标,作为轨迹合理性的考量。最后,经实验分析证明,本文所提出的模型在数据集上具有较高精度,可以很好的模拟在复杂场景下的行人交互行为,并生成多条符合日常认知的预测轨迹,同时避免发生碰撞。

1 理论分析

1.1 问题定义

行人轨迹预测问题可以定义为一个序列生成问题,对于场景中的每个行人 $i(i = 1, 2, \dots, N)$,其在时刻 t 的位置用坐标 $Y_i^t = (x_i^t, y_i^t)$ 表示,已知行人的位置 $Y_i^t(t = 1, 2, \dots, T_{obs})$,目标是预测它们从 T_{obs+1} 到 T_{pred} 的位置,用 $\hat{Y}_i^t = (\hat{x}_i^t, \hat{y}_i^t)$ 表示预测值。

本文研究可以显式地预测行人的运动速度,同时将其作为控制条件并以此预测多个行人的运动轨迹。给定 (x^t, y^t) 作为时间 t 时的坐标, S 代表行人运动速度,本文设定函数 f 来生成下一个时间段位置坐标 (x^{t+1}, y^{t+1}) ,如式(1)所示。

$$(x^{t+1}, y^{t+1}) = f(x^t, y^t | S) \quad (1)$$

其中,未来时间段的生成坐标基于速度 S ,会根据速度不同发生变化。在预测环境中,未来时间步长的速度通常是未知的。为了能够确定整个时间框架的速度,包括尚未生成的轨迹的未来速度,可以使用从数据中学习到估计速度 \hat{S} 。

1.2 模型框架

STGS-GAN 模型的架构图如图 1 所示。主体结构为条件生成对抗网络模型。本文在生成器中设计了时空信息融合模块,其中采用空间 LSTM(S-LSTM)网络提取行人轨迹序列的运动特征,通过 GAT 来关注行人间空间上的交互信息,同时设置时间 LSTM(T-LSTM)来编码行人轨迹序列的时间相关性。为更好地模拟行人运动速度对行人轨迹预测的影响,额外采用一个由 LSTM 组成的速度预测模块来从潜在空间预测行人未来速度,并据此作为 C-GAN 的控制条件。在解码器中采用 LSTM 网络,以运动特征,时间特征、交互特征和速度特征的组合为输入,输出为未来的轨迹。最后通过判别器中的 LSTM 网络输出“真/假”评价。

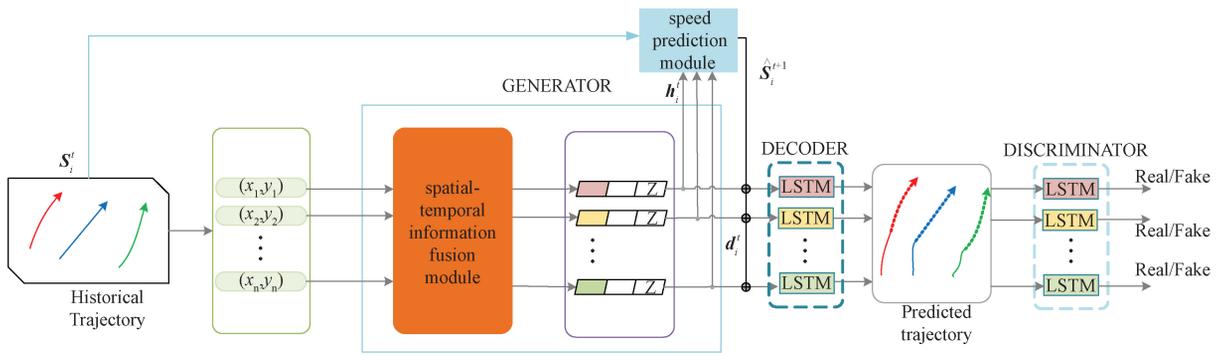


图 1 STGS-GAN 行人轨迹预测模型

Fig. 1 STGS-GAN Pedestrian trajectory prediction model in graph structure

1.3 预处理

行人的历史轨迹包含其位置和速度等运动特征信息,通过这些特征信息可以获取到行人间的相互作用关系并完成轨迹的预测。本文在编码行人轨迹信息时,不直接采用历史轨迹的行人坐标,而是采用相对位置坐标 $(\Delta x_i^t, \Delta y_i^t)$,这相当于同时获取行人位置信息和速度信息,如式(2)和(3)所示。

$$\Delta x_i^t = x_i^t - x_i^{t-1} \quad (2)$$

$$\Delta y_i^t = y_i^t - y_i^{t-1} \quad (3)$$

1.4 时空信息融合模块

在建模行人间交互作用时,简单地提取行人间空间信息不能很好地模拟人与人的互动行为,引入时间信息可以捕获每个时间步长上的时间交互,更好地避免可能出现的碰撞情况。考虑到 LSTM 无法充分地共享不同行人之间的信息和池化机制会丢失大量行人交互信息等问题,本文使用 GAT 作为交互信息的共享机制。如图 2 所示,将图定为数组 $G(\mathbf{V}, \mathbf{E}, \mathbf{A})$,其中 \mathbf{V} 定义了一组包含行人运动特征向量的节点, \mathbf{E} 是表示各个行人之间交互的边集, \mathbf{A} 是图的邻接矩阵。

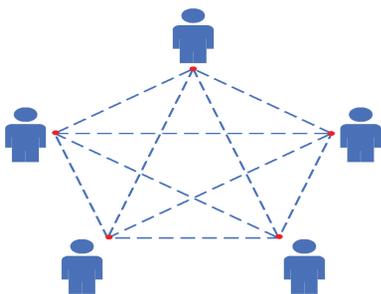


图 2 图结构中行人交互示意图

Fig. 2 Diagram of pedestrian interaction

本文将 GAT 应用于图上,用图上的边来表示行人间交互作用的存在,通过关注其周围行人的信息来计算并叠加每个图节点的特征。

时空信息融合模块如图 3 所示,由 S-LSTM、T-LSTM 和 GAT 组成。

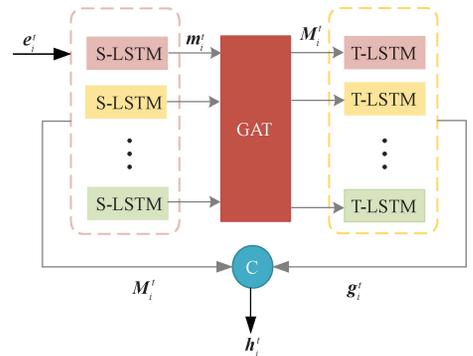


图 3 时空信息融合模块图

Fig. 3 Spatial-temporal information fusion module

在完成时空信息融合时,首先将相对位置坐标 $(\Delta x_i^t, \Delta y_i^t)$ 传入多层感知器和 S-LSTM 中进行运动信息编码,如式(4)和(5)所示。

$$\mathbf{e}_i^t = \phi_1(\Delta x_i^t, \Delta y_i^t; \mathbf{W}_{ec}) \quad (4)$$

$$\mathbf{m}_i^t = S - LSTM(\mathbf{m}_i^{t-1}, \mathbf{e}_i^t; \mathbf{W}_s) \quad (5)$$

式中:将多层感知器 ϕ_1 作为嵌入函数, \mathbf{W}_{ec} 是其权重, \mathbf{e}_i^t 是特征向量维度为 16。经 S-LSTM 编码后的运动特征向量 \mathbf{m}_i^t 维度为 32, \mathbf{W}_s 是 LSTM 的权重,在场景内所有行人间共享。

其次,将在观察行人运动历史轨迹期间运动特征向量 $\mathbf{m}_i^t (t = 1, \dots, T_{obs}) \in \mathbf{V}$ 作为节点信息输入到 GAT 以获得注意权重系数 α_{ij}^t ,计算方法如式(6)所示。

$$\alpha_{ij}^t = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{W}_{gat} \mathbf{m}_i^t \parallel \mathbf{W}_{gat} \mathbf{m}_j^t]))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(\mathbf{a}^T [\mathbf{W}_{gat} \mathbf{m}_i^t \parallel \mathbf{W}_{gat} \mathbf{m}_k^t]))} \quad (6)$$

其中, \parallel 表示数据的连接, N_i 表示与节点 i 临近的其他行人, \mathbf{a} 是单层前馈神经网络的权重向量, α_{ij}^t 是节点 j 到 i 在 t 时刻的注意系数, \mathbf{W}_{gat} 是每个节点的共享线性变

换的权重矩阵。在进行归一化处理时,增添 leakyReLU 作为激活函数以避免失去节点 i 的信息,导致实际聚合没有效果。在获得注意力系数后, $\hat{\mathbf{m}}_i^t$ 为在 t 时刻节点 i 的一个图注意力输出如式(7)所示。

$$\hat{\mathbf{m}}_i^t = \sigma \left(\sum_{j \in N_i} \alpha_{ij}^t \mathbf{W}_{gat} \mathbf{m}_j^t \right) \quad (7)$$

其中, σ 是一个非线性函数。GAT 网络存在过度平滑的问题,在两层模型中性能优异,而单纯叠加层数会导致降低模型性能的问题,因此本文仅叠加个两层注意力层。

原始 GAT 体系结构有一个缺陷,即在学习注意力系数时不使用任何可用的边集信息。为解决这一缺点,本文通过添加归一化层,增加节点特征向量的信息,提高了 GAT 层的学习能力,如式(8)所示。

$$\mathbf{M}_i^t = \hat{\mathbf{m}}_i^t \parallel \mathbf{l}_{ij} \quad (8)$$

其中, \mathbf{M}_i^t 是节点 i 的扩增特征向量, \mathbf{l}_{ij} 是连接节点 i 及其邻居 j 的边特征向量。

最后,采用一个 T-LSTM 来显式地多步长建模行人交互之间的时间相关性,输出如式(9)所示:

$$\mathbf{g}_i^t = T - \text{LSTM}(\mathbf{g}_i^{t-1}, \mathbf{M}_i^t; \mathbf{W}_g) \quad (9)$$

其中, \mathbf{g}_i^t 是 T-LSTM 的输出, \mathbf{W}_g 是 T-LSTM 权重,在所有行人轨迹序列中共享。

在搭建模型时将这两部分结合起来,以完成时空信息的融合。在对每个时间步长进行计算时,每个行人的两个 LSTM 存在 \mathbf{M}_i^t 和 \mathbf{g}_i^t 两个隐藏变量,分别将 \mathbf{M}_i^t 和 \mathbf{g}_i^t 输入到两个不同的多层感知器 (ϕ_2, ϕ_3) 中并在之后进行连接,在编码器特征信息中存储,最终 \mathbf{h}_i^t 如式(10)所示。

$$\mathbf{h}_i^t = \phi_2(\mathbf{M}_i^t) \parallel \phi_3(\mathbf{g}_i^t) \quad (10)$$

1.5 速度预测模块

此前研究中,大多数模型对速度的预测都采用的是 LSTM 特征向量的隐式信息,而几乎没有显式地提出速度对于模型预测结果准确性的影响。本文与此前研究不同,采用了一个由 LSTM 组成的模块显式地编码速度信息,并作为 C-GAN 的控制条件,使预测速度对模型精度的影响更加具体精确,出现较大偏差速度的轨迹不会生成。本文根据数据集每两个连续时间帧之间的欧氏距离计算速度标签,并在 (0,1) 范围内缩放。速度预测模块图如图 4 所示。

此前研究中,大多数模型对速度的预测都采用的是 LSTM 特征向量的隐式信息,而几乎没有显式地提出速度对于模型预测结果准确性的影响。本文与此前研究不同,采用了一个由 LSTM 组成的模块显式地编码速度信息,并作为 C-GAN 的控制条件,使预测速度对模型精度的影响更加具体精确,出现较大偏差速度的轨迹不会生

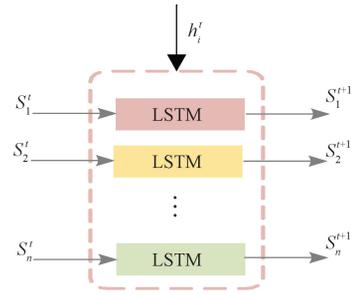


图 4 速度预测模块示意图

Fig. 4 Schematic diagram of speed prediction module

成。本文根据数据集每两个连续时间帧之间的欧氏距离计算速度标签,并在 (0,1) 范围内缩放。

速度预测模块用从编码器中获得的隐藏变量 \mathbf{h}_i^t 初始化速度预测模块的隐藏状态 \mathbf{v}_{si}^t 。输入是当前时间帧的速度 $S = (\Delta x_i^t, \Delta y_i^t)$, 通过具有信号激活的全连接层 (fully connected layers, FC) 传递隐藏状态, 计算未来速度估计 \hat{S}_i^{t+1} 如式(11)和(12)所示。

$$\mathbf{v}_{si}^t = \text{LSTM}_{sp}(S_i^t, \mathbf{v}_{si}^{t-1}; \mathbf{W}_{sp}) \quad (11)$$

$$\hat{S}_i^{t+1} = \gamma_{sp}(\mathbf{v}_{si}^{t-1}; \mathbf{W}_{\gamma_{sp}}) \quad (12)$$

其中, γ_{sp} 为 FC 层, \mathbf{W}_{sp} 和 $\mathbf{W}_{\gamma_{sp}}$ 为 LSTM 层和 FC 层的权重, 预测模块与其他组件同时训练, 使用地面真相 S_i^{t+1} 作为反馈信号。

1.6 解码器

在解码器中保持过去行人轨迹序列的特性, 本文用 \mathbf{h}_i^t 初始化其隐藏状态 \mathbf{h}_{di}^t , 并在训练期间根据控制条件输入相对位置的嵌入向量, 添加高斯分布噪声 z 获得解码输入特征向量 \mathbf{d}_i^t , 如式(13)所示。

$$\mathbf{d}_i^t = \alpha_d((\Delta x_i^t, \Delta y_i^t) \oplus S_i^{t+1}; \mathbf{W}_{\alpha_d}) \parallel z \quad (13)$$

在预测环境中, 本文采用预测模块的估计 \hat{S}_i^{t+1} 代替 S_i^{t+1} 。 \mathbf{h}_{di}^t 是 LSTM 的隐藏状态通过 FC 网络输入, 它输出每个代理的预测相对位置如式(15)所示。

$$\mathbf{h}_{di}^t = \text{LSTM}_{dec}(\mathbf{d}_i^t, \mathbf{h}_{di}^{t-1}; \mathbf{W}_{dec}) \quad (14)$$

$$(\hat{x}_i^{t+1}, \hat{y}_i^{t+1}) = \gamma_d(\mathbf{h}_{di}^t, \mathbf{W}_{\gamma_d}) \quad (15)$$

其中, \mathbf{W}_{dec} 是 LSTM 的权重, 而 \mathbf{W}_{γ_d} 是 FC 网络的权重。

1.7 判别器

本文在模型中采用一个 LSTM 编码器块作为判别器, 它约束生成器不仅要生成符合现实和社会可接受的轨迹, 而且要符合 CGAN 的控制条件。对真实样本与标签的配对需要接近于 1, 对于生成样本与标签需要接近于 0, 对于真实样本与不相符的标签, 判别器的输出应该接近于 0。对判别器的实际输入可表述如式(16)所示。

$$\mathbf{O}_i = \langle (\Delta x_i^t, \Delta y_i^t), S_i^t \mid t = 0, \dots, T \rangle \quad (16)$$

包括观察到的 ($t = 0, \dots, t_{obs}$) 和未来地面真相 ($t = t_{obs+1}, \dots, T$) 的相对位置。假输入如式(17)所示。

$$\hat{O}_i = \langle (\Delta x_i^t, \Delta y_i^t), S_i^t \mid t = 0, \dots, t_{obs} \rangle \quad (17)$$

$$\oplus \langle (\hat{x}_i^t, \hat{y}_i^t), S_i^t \mid t = t_{obs} + 1, \dots, T \rangle$$

结合观察到的和预测的相对位置,判别器输入如式(18)所示。

$$h_{dsi}^t = \text{LSTM}_{dsi}(\zeta_{di}(\mathbf{o}_i^t; \mathbf{W}_\zeta), h_{dsi}^{t-1}; \mathbf{W}_{dsi}) \quad (18)$$

其中, ζ_{di} 是具有相应权重 \mathbf{W}_ζ 的嵌入函数, \mathbf{W}_{dsi} 是 LSTM 权重。 \mathbf{o}_i^t 是来自真实 O_i 或假序列 \hat{O}_i 的输入数据。通过对 LSTM 的最终隐藏状态应用具有修正线性单元的多层 FC 网络来计算真实的或假的分类分数,如式(19)所示。

$$\hat{C}_i = \gamma_{di}(h_{dsi}^t; \mathbf{W}_{\gamma_{di}}) \quad (19)$$

式中: γ_{di} 是具有相应权重 $\mathbf{W}_{\gamma_{di}}$ 的嵌入函数。

1.8 损失函数

为更好地提升 C-GAN 训练结果,本文将 L_2 损失应用于生成器所生成的预测轨迹,并将 L_1 损失应用于速度预测模块。网络训练通过最小化损失函数,依次进行,判别器损失如式(20)所示。

$$\ell_D(\hat{C}_i, C_i) = -C_i \ln(\hat{C}_i) - (1 - C_i) \ln(1 - \hat{C}_i) \quad (20)$$

生成器整体损失 L_2 和 L_1 损失如式(21)所示。

$$\ell_G(\hat{O}_i) + \ell_2((x_i^t, y_i^t), (\hat{x}_i^t, \hat{y}_i^t)) + \ell_1(S_i^t, \hat{S}_i^t) \quad (21)$$

对于 $t = t_{obs} + 1, \dots, T$ 生成器损失是判别器将生成器生成的数据正确分类为“假”的能力,如式(22)所示。

$$\ell_G(\hat{O}_i) = -\ln(1 - \hat{C}_i) \quad (22)$$

其中, \hat{C}_i 是判别器的分类分数。

2 实验验证

2.1 数据集与评价指标

本文在 ETH^[17] 和 UCY^[18] 这两个公开的行人轨迹数据集上来评估所采用模型的性能。ETH 数据集有 ETH 和 HOTEL 两个场景,UCY 数据集有 ZARA1、ZARA2 和 UNIV 这 3 个场景。所有数据集都采用了监控下视角,并且还涵盖了具有挑战性的群体行为^[19]。ETH 和 UCY 数据集信息统计表如表 1 所示。实验采用留一法^[11],使用 4 个数据集进行训练和验证,在剩下一个数据集上进行测试。本文采用 3 种评价标准,其中两种同先前的工作一样^[11]:

1) 平均位移误差 (average displacement error, ADE)。预测轨迹和地面真实轨迹中所有估计位置上的均方误差,单位为 m。

$$ADE = \frac{1}{TN} \sum_{i=1}^{T_{pred}} \sum_{t=1}^N \| Y_i^t - \hat{Y}_i^t \|_2 \quad (23)$$

2) 最终位移误差 (final displacement error, FDE), 预测结束时的预测终点与真实轨迹终点间的误差,单位为 m。

$$FDE = \frac{1}{N} \sum_{i=1}^N \| Y_i^t - \hat{Y}_i^t \|_2 \quad (24)$$

考虑在行人运动时避免发生碰撞是一个基本特征,本文将 ACT 作为衡量轨迹合理性的一个评价标准。ACT 的计算公式如式(23)所示。

$$ACT = \frac{1}{N} \sum_{n=1}^N \sum_{t=t_{obs}+1}^{T_{obs}+T_{pred}} C_n^t \quad (25)$$

其中, C_n^t 表示在 t 时刻,第 n 个样本中发生的碰撞次数。在实验中,在 t 时刻两个行人间距离小于阈值 d 记为一次碰撞,考虑到 0.3 m 为社交中比较舒服的距离,其中阈值 d 设置为 0.3 m。

表 1 ETH 和 UCY 数据集信息统计

Table 1 Eth and UCY datasets information statistics

数据集	EHT	HOTEL	UNIV	ZARA1	ZARA2
数据量	5 492	6 543	21 813/17 953	5 153	9 722
截取帧数	876	1 168	444/451	872	1 050
行人总数	360	389	415/434	148	204
一帧最大行人数	27	18	75/52	20	20
一帧最小行人数	1	1	21/13	1	1
每帧平均行人数	6.27	5.60	49.13/33.18	5.91	9.24
单个人存在最大帧数	114	100	352/289	197	584
单个人存在最小帧数	2	2	4/5	10	7
行人存在平均帧数	15.26	16.82	52.56/41.37	34.82	47.66

2.2 超参数设置与实验环境

实验基于 Pytorch 建立网络环境,使用 Adam 优化器

进行参数优化,生成器学习率设置为 0.001,判别器学习率为 0.001,批处理大小为 64,训练数据集训练轮数为

800。所有实验都在相同的硬件环境下完成,所用的处理器为 Intel Core i7-8700k CPU@3.70 GHz,显卡为 NVIDIA 2080TI GPU。

2.3 算法对比

为评估 STGS-GAN 的性能,本文选取了 6 种算法进行 ADE 和 FDE 的对比,集合了生成对抗网络与图卷积网络相关模型。分别为 S-LSTM^[20]、SGAN^[11]、SoPhie^[13]、S-BIGAT^[15]、CGNS^[21]和 STGAT^[14]。所有算法所采用的

观测时间为 3.2 s,预测时间为 4.8 s,使用 ADE、FDE 和 ACT 进行评价,最终所有生成多样本的算法均生成 20 个样本。STGS-GAN 与其他 6 个算法在 ETH 和 UCY 数据集上进行对比,对比 ADE 和 FDE 的结果如表 2 所示,表 2 中黑体为最好的预测结果。

如表 2 所示,STGS-GAN 模型在整体 ADE 和 FDE 上优于其他方法,在 HOTEL、ZARA1 和 ZARA2 数据集上也比其他方法更加优秀。

表 2 本文算法与其他算法的比较结果

Table 2 The results of comparison between our algorithm and other algorithms

数据集	ADE/FDE/m					
	EHT	HOTEL	UNIV	ZARA1	ZARA2	平均值
S-LSTM	1.09/2.35	0.79/1.76	0.67/1.40	0.47/1.00	0.53/1.17	0.72/1.54
SGAN	0.81/1.52	0.72/1.61	0.76/1.52	0.34/0.69	0.42/0.84	0.58/1.18
SoPhie	0.70/1.43	0.76/1.67	0.54/1.24	0.30/0.63	0.38/0.78	0.54/1.15
S-BIGAT	0.69/1.29	0.49/1.01	0.55/1.32	0.30/0.62	0.36/0.75	0.48/1.00
STGAT	0.65/ 1.12	0.35/0.66	0.52/ 1.10	0.34/0.69	0.29/0.60	0.43/0.83
CGNS	0.62/1.40	0.70/0.93	0.48/1.22	0.32/ 0.59	0.35/0.71	0.49/0.97
STGS-GAN	0.60/1.19	0.30/0.51	0.51/ 1.10	0.30/0.59	0.28/0.57	0.40/0.79

其中,在 ETH 数据集中 ADE 和 FDE 指标在所有模型中表现都相对较差。由表 1 可知该数据集样本数量较大,通过分析数据集视频可知在此场景下行人运动大多数为非线性运动,停留等待时间较长,行人运动的选择具有多样性。在 UNIV 数据集中,行人总数和每帧最大行人人数均为所有数据集中最大,说明此数据集中行人密集,行人间相互作用明显,此数据集要求更好地行人交互模型,因此采用图卷积的 S-BIGAT,STCAT,CGNS 和本文模型 STGS-GAN 表现较好。在 ZARA1 数据集和 ZARA2 数据集中,测试场景为商店门口,观察测试集视频可以发现行人运动多为线性。尤其是在 ZARA2 数据集中,行人密度不大且其中行人多保持匀速运动或静止不动,所以很适合采用速度条件的生成对抗网络,因此 STGS-GAN 在这两个数据集中表现最为突出。

在 ACT 的对比实验中,选取 SGAN^[11]、SGAN-P^[11]、STGAT^[14]与本文模型 STGS-GAN 进行对比。在表 3 所示的结果中可以看出,STGS-GAN 的 ACT 结果较低,在所有数据集上优于其他方法,表明 STGS-GAN 可以很好地减少轨迹发生的碰撞,对预测轨迹合理性有明显的提升。

表 3 ACT 指标对比

Table 3 Comparison of ACT indicators

数据集	EHT	HOTEL	UNIV	ZARA1	ZARA2
SGAN	0.343 6	0.427 4	12.272 1	0.202 1	1.158 3
SGAN-P	0.391 4	0.465 0	16.679 6	0.439 0	3.839 0
STGAT	0.420 7	0.321 3	12.321 9	0.338 9	1.195 0
STGS-GAN	0.219 2	0.131 5	11.447 2	0.263 5	0.843 5

本文模型与 S-LSTM^[18]和 SGAN^[9]的推理时间对比如表 4 所示。

表 4 推理速度对比

Table 4 Comparison of speed

模型	S-LSTM	S-GAN-P	STGS-GAN
推理时间/s	97.12	8.57	18.721

结果显示 STGS-GAN 的推理速度显著高于 S-LSTM,同时低于 S-GAN-P。推理速度相比与 S-GAN-P 较慢的原因在于,本文模型交互方案上采用了 GAT,引入了更多的信息来模拟行人间的交互行为,在提高预测精度的同时会影响推理速度。

2.4 消融研究

在评估 STGS-GAN 中速度预测模块和时空信息融合模块对模型的影响时,采取调整算法模块的方法。本文算法与调整各项模块后的算法在 5 个数据集上进行预测精度比较的结果如表 5 所示,表中黑体为最好的预测结果。

1) 算法 1 在时空融合模块中去掉时间 LSTM 模块,在编码器中仅采用 LSTM 结合 GAT 的运动编码。研究目的是确定时间 LSTM 模块是否影响行人轨迹预测的精度;

2) 算法 2 去掉速度预测模块,主要是为比较采用速度控制的条件 GAN 与普通的 GAN 对模型性能的不同影响,确定速度预测模块对行人轨迹预测精度的影响;

3) 算法 3 训练样本数为 1,测试样本数为 1;

4) 算法 4 训练样本数为 1,测试样本数为 20。

算法 3 和 4 设定的主要目的是确定训练样本与测试样本多样化对模型预测行人轨迹精度的影响。

表 5 本文算法在调整指定模块情况下的比较结果

Table 5 Comparison results of our algorithm with different control setting

序号	(ADE/FDE)/m					
	EHT	HOTEL	UNIV	ZARA1	ZARA2	平均值
算法 1	0.70/1.24	0.34/0.53	0.52/1.15	0.35/0.71	0.29/0.61	0.44/0.85
算法 2	0.72/1.26	0.31/0.53	0.52/1.12	0.35/0.71	0.30/0.60	0.43/0.84
算法 3	0.88/1.76	0.39/0.69	0.55/1.29	0.41/0.81	0.35/0.69	0.52/1.05
算法 4	0.78/1.44	0.36/0.63	0.53/1.21	0.38/0.76	0.32/0.66	0.47/0.94
STGS-GAN	0.60/1.19	0.30/0.51	0.51/1.10	0.30/0.59	0.28/0.57	0.40/0.79

由表 5 可分析得知,与算法 1 作对比,STGS-GAN 由于结合行人序列间的时序关系,使行人在交互时会考虑到其他行人历史轨迹对其行为的影响,STGS-GAN 的预测结果全面优于对比算法 1;与算法 2 作对比,STGS-GAN 采用了速度控制条件的 GAN,对需要速度因素影响的数据集表现更为优异,因此 ADE 和 FDE 都较算法 2 下降,表明部分改进来自速度预测模块;STGS-GAN 与算法 3 和算法 4 作对比,这两种算法采用的训练样本和测试样本较少,偶然性增大,影响行人轨迹预测精度,而本文算法考虑了轨迹的多样性和不确定性,在预测误差上明显小于算法 3 和算法 4,说明 STGS-GAN 满足了轨迹预测的多模态性,可以生成多条符合日常认知的预测轨迹。

2.5 定性评价

行人轨迹预测研究的最终任务是为了设计出更好的行人交互模型以提高预测精度,为了更好地分析所设计的模型效果,本文将模型进行可视化,通过可视化的效果展示所采用的算法实现结果是否与实际情况相符。选择 ZARA2 数据集进行可视化,行人轨迹预测可视化如图 5 所示,图中实线代表行人在指定场景内运动的真实轨迹,虚线代表模型生成的预测轨迹。

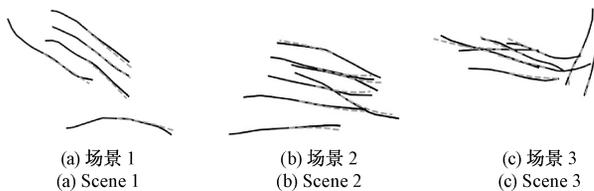


图 5 行人轨迹预测可视化图

Fig. 5 Visualization of pedestrian trajectory prediction

图 5(a) 中所示各个行人的历史轨迹间距离基本未发生变化各个行人均保持匀速运动,且远端的行人不影响这个集体的行动轨迹。图中所示与实际相符,说明本文算法可以很好地模拟行人在保持匀速运动时各个行人间的交互关系,同时远处的行人对算法的预测没有影响。图 5(b) 与 (c) 中皆存在行人逆向行走,逆向行走的行人对运动轨迹有较大影响,相距越近的行人受到的影响越大,所以他们的轨迹都发生了一定的偏转,这一点也符合

实际。同时在这两幅图中也存在同向行人交叉运动,当行人轨迹交叉时,行人间相互作用关系较强,影响轨迹预测效果,使预测轨迹发生偏转。虽然预测轨迹与真实轨迹稍有偏差,但这也说明了所建立的行人轨迹交互模型是有效果的。通过分析可视化预测结果,证明本文所建立的行人轨迹交互模型基本能与实际相符,同时精度较高。

3 结论

针对在复杂场景下行人轨迹预测模型缺乏关于时间和速度等信息的问题,本文提出一种结合速度控制的时空图卷积神经网络算法对轨迹进行预测。首先,本文所采用的模型通过速度预测模块显式地将预测速度作为 CGAN 的控制条件,提升速度信息对模型准确性的影响。其次,通过时空信息融合模块将行人交互存在的空间信息和时间信息相结合,更好地模拟了行人间相互作用关系。最后,将 ACT 作为行人轨迹合理性的评价标准,对轨迹的合理进行分析。在 ETH 和 UCY 数据集上的实验验证,证明本文的方法对比于同类模型能够更好地模拟行人间的交互作用,可以有效地提高预测的精度和合理性。

参考文献

- [1] RASOULI A, TSOTSOS J K. Autonomous vehicles that interact with pedestrians: A Survey of theory and practice[J]. IEEE Transactions on Intelligent Transportation Systems, 2019: 1-19.
- [2] CHEN C, LIU Y, KREISS S, et al. Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning [C]. 2019 International Conference on Robotics and Automation (ICRA), 2019: 6015-6022.
- [3] KOOIJ J F P, FLOHR F, POOL E A I, et al. Context-based path prediction for targets with switching dynamics [J]. International Journal of Computer Vision, 2019, 127(3): 239-262.
- [4] HUDNELL M, PRICE T, FRAHMF J M. Robust

- aleatoric modeling for future vehicle localization [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020.
- [5] RUDENKO A, PALMIERI L, HERMAN M, et al. Human motion trajectory prediction: A survey[J]. The International Journal of Robotics Research, 2020, 39(8): 895-935.
- [6] TRAUTMAN P, KRAUSE A. Unfreezing the robot: Navigation in dense, interacting crowds[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems, 2010:797-803.
- [7] CHERNOVA S, THOMAZ A L. Robot learning from human teachers [J]. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2014, 8(3): 1-121.
- [8] 罗佳,黄晋英.生成式对抗网络研究综述[J].仪器仪表学报,2019,40(3):74-84.
LUO J, HUANG J Y. Generative adversarial network: An overview [J]. Chinese Journal of Scientific Instrument,2019,40(3):74-84.
- [9] 陈亮,吴攀,刘韵婷,等.生成对抗网络 GAN 的发展与最新应用[J].电子测量与仪器学报,2020,34(6):70-78.
CHEN L, WU P, LIU Y T, et al. Development and application of the latest generation against the network of GAN[J]. Electronic Measurement Technology, 2020, 34(6):70-78.
- [10] 王延年,李文婷,任劼.基于生成对抗网络的单帧图像超分辨率算法[J].国外电子测量技术,2020,39(1):26-32.
WANG Y N, LI W T, REN J. Single image super-resolution algorithm based on generation adversarial networks [J]. Foreign Electronic Measurement Technology,2020,39(1):26-32.
- [11] GUPTA A, JOHNSON J, LI F F, 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.
- [12] 张志远,刁英华.结合社会特征和注意力的行人轨迹预测模型[J].西安电子科技大学学报,2020,47(1):10.
ZHANG ZH Y, DIAO Y H. Pedestrian trajectory prediction model with social features and attention[J]. Journal of Xidian University, 2020, 47(1):10.
- [13] SADEGHIAN A, KOSARAJU V, SADEGHIAN A, et al. SoPhie: An Attentive GAN for predicting paths compliant to social and physical constraints [C]. The IEEE Conference on Computer Vision and Pattern Recognition, 2019: 1349-1358.
- [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), 2019: 6272-6281.
- [15] KOSARAJU V, SADEGHIAN A, MARTÍN-MARTÍN R, et al. Social bigat: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks [C]. Proceedings of Annual Conference on Neural Information Processing Systems, 2019: 1-10.
- [16] MIRZA M, OSINDERO S. Conditional Generative Adversarial Nets [J]. Computer Science, 2014: 2672-2680.
- [17] PELLEGRINI S, ESS A, GOOL L V. Improving data association by joint modeling of pedestrian trajectories and groupings [C]. European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2010: 452-465.
- [18] LEALTAIXE L, FENZI M, KUZNETSOVA A, et al. Learning an image-based motion context for multiple people tracking[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA,2014: 3542-3549.
- [19] 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, Kyoto, Japan, 2009: 261-268.
- [20] 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. Las Vegas, USA: IEEE CS, 2016: 961-97.
- [21] 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). IEEE, Macau, China, 2019: 6150-6156.

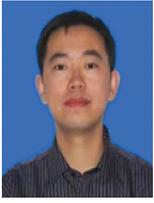
作者简介



王海峰,2018年于燕山大学获得学士学位,现为沈阳工业大学信息科学与工程学院硕士研究生,主要研究方向为行人轨迹预测。

E-mail: 798466420@qq.com

Wang Haifeng received his B. Sc. degree from Yanshan University in 2018. Now he is M. Sc. candidate at the School of Information Science and Engineering, Shenyang University of Technology. His main research interest includes pedestrian trajectory prediction.



桑海峰 (通信作者), 沈阳工业大学信息科学与工程学院教授, 主要研究方向为机器视觉检测、生物特征识别、深度学习、行人轨迹预测。

E-mail: sanghaif@163.com

Sang Haifeng (Corresponding author) is a professor at the School of Information Science and Engineering, Shenyang University of Technology. His main research interests include machine vision, biometrics identification, deep learning, pedestrian trajectory prediction.



王金玉, 2018 年于沈阳工业大学获得学士学位, 现为沈阳工业大学信息科学与工程学院博士研究生, 主要研究方向为行人轨迹预测。

E-mail: 191113198@qq.com

Wang Jinyu received her B. Sc. degree

from Shenyang University of Technology in 2018. Now she is a Ph. D. candidate at the School of Information Science and Engineering, Shenyang University of Technology. Her main research interest includes pedestrian trajectory prediction.



陈旺兴, 2020 年于沈阳工业大学获得学士学位, 现为沈阳工业大学信息科学与工程学院硕士研究生, 主要研究方向为行人轨迹预测。

E-mail: 1909703861@qq.com

Chen Wangxing received his B. Sc. degree from Shenyang University of Technology in 2020. Now he is a M. Sc. candidate at the School of Information Science and Engineering, Shenyang University of Technology. His main research interest includes pedestrian trajectory prediction.