

# 基于多模式时空交互的行人轨迹预测模型

桑海峰, 陈旺兴, 王海峰, 王金玉

(沈阳工业大学信息科学与工程学院, 辽宁沈阳 110870)

**摘要:** 在正确地规划合理路径方面, 行人轨迹预测具有重要的意义. 大多数现有轨迹预测方法在考虑周围行人的影响时, 都是简单地将周围行人全部考虑在内, 这必然带来的冗余信息. 本文提出了一种基于多模式时空交互的行人轨迹预测模型, 该模型通过多模式行人空间交互模块对不同行人在不同情况下给予不同的权重, 使得模型可以有效减小冗余信息带来的影响. 并且本文的模型针对于输入轨迹信息不同重要程度, 设计了加权信息融合模块在原轨迹信息的基础上融合了赋予不同权重的历史轨迹信息, 使得模型的轨迹信息更加有效. 此外, 该模型采用了时间卷积网络模块来捕获行人的时间交互. 实验结果表明, 在公开数据集 ETH 和 UCY 上, 相比于 Social-STGCNN 平均位移误差 (Average Displacement Error, ADE) 和终点位移误差 (Final Displacement Error, FDE) 分别降低了 15% 和 14%.

**关键词:** 行人轨迹预测; 多模式时空交互; 行人空间交互; 加权信息融合; 时间卷积网络; 时间交互

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 0372-2112(2022)11-2806-07

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

**DOI:** 10.12263/DZXB.20210752

## Pedestrian Trajectory Prediction Model Based on Multi-Model Space-Time Interaction

SANG Hai-feng, CHEN Wang-xing, WANG Hai-feng, WANG Jin-yu

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

**Abstract:** Pedestrian trajectory prediction plays an important role in correctly planning reasonable paths. Most of the existing trajectory prediction methods simply take all the pedestrians into account when considering the influence of the surrounding pedestrians, which inevitably brings redundant information. A pedestrian trajectory prediction model based on multi-mode space-time interaction is proposed. This model gives different weights to different pedestrians in different situations through multi-mode pedestrian space interaction module, which makes the model effectively reduce the impact of redundant information. Aiming at the different importance of the input trajectory information, the weighted information fusion module is designed to integrate the historical trajectory information with different weights on the basis of the original trajectory information, so as to make the trajectory information of the model more effective. In addition, the model uses time convolution network module to capture pedestrian time interaction. The experimental results show that compared with social-stgcnn, average displacement error(ADE) and final displacement error(FDE) is reduced by 15% and 14% respectively on the open data sets ETH and UCY.

**Key words:** pedestrian trajectory prediction; multi-mode space-time interaction; pedestrian space interaction; weighted information fusion; time convolution network; time interaction

### 1 引言

行人未来轨迹预测一直是近些年来一项基础, 但具有挑战性的项目. 由于行人的运动不仅受到周围行人和环境的影响而且取决于个体的社交习惯<sup>[1]</sup>, 这使得对于行人轨迹预测的建模非常困难.

最近几年, 深度学习<sup>[2]</sup>凭借着其完备的理论体系以

及丰富的网络模型在行人轨迹预测方面大展拳脚. 正是因为长短时记忆网络 (Long Short-Term Memory, LSTM) 在解决时间序列问题上的巨大优势以及行人轨迹的时序性, Social LSTM<sup>[3]</sup>、Stgat<sup>[4]</sup>、Sophie<sup>[5]</sup>等相关工作<sup>[6-11]</sup>都利用了 LSTM 来对行人之间复杂的交互进行建模. 基于生成对抗网络 (Generative Adversarial Network,

GAN)的行人轨迹预测模型<sup>[12-14]</sup>,首先通过编码器来提取行人特征再使用解码器对行人特征进行处理生成多条行人预测轨迹,解决了之前模型只能预测一条轨迹的问题.随着时空图快速发展,越来越多的行人轨迹<sup>[15-17]</sup>预测模型采用了时空图的方式即从空间和时间两个方向对行人交互进行建模.

尽管目前行人轨迹预测已经有了一定的成果,但是依然存在部分局限性.大部分的行人轨迹预测方法在提取特征时都是简单地对行人轨迹坐标信息进行同等考虑.并且对于行人之间交互信息的处理都是一次性整合,产生的冗余信息对预测轨迹准确性造成的影响也没有进行处理.

针对于上述问题,本文提出了一种基于多模式时空交互的行人轨迹预测模型.在原轨迹基础上,对坐标信息分别融合赋予不同权重的行人历史轨迹信息来提高轨迹数据的有效性.其次,将周围行人之间的空间交互分成多种模式,分别进行处理,由此减少行人空间交互时产生的冗余信息对轨迹预测造成的影响.

## 2 基于多模式时空交互的行人轨迹预测模型

行人轨迹预测任务可表示为,假设场景中有  $N$  个行人,通过观察一段时间中的行人轨迹,来预测行人未来轨迹.设观察第  $i$  个行人在时间  $t$  的位置表示为  $p_t^i = (x_t^i, y_t^i)$ ,其中  $i \in N$ .已知  $N$  个行人在一段时

间  $t_{gc} \in \{1, 2, \dots, T_{gc}\}$  上的位置为  $p_t^i = (x_t^i, y_t^i)$ ,其中  $t \in t_{gc}$ ,  $i \in N$ .根据已知的上一段时间的行人轨迹来预测该  $N$  个行人下一段时间  $t_{pred} \in \{T_{gc+1}, T_{gc+2}, \dots, T_{pred}\}$  的位置  $\hat{p}_t^i = (x_t^i, y_t^i)$ ,其中  $t \in t_{pred}$ ,  $i \in N$ .

本文沿用了 Social-STGCNN<sup>[15]</sup> 的假设,假设行人轨迹  $(x_t^i, y_t^i)$  服从二维高斯分布 (bivariate Gaussian distribution),使得  $p_t^i \sim N(\mu_t^i, \sigma_t^i, \rho_t^i)$ .其中  $\mu_t^i$  为均值,  $\sigma_t^i$  为标准差,  $\rho_t^i$  为相关系数.本文中的预测轨迹也遵循预测的二维高斯分布  $\hat{p}_t^i \sim N(\hat{\mu}_t^i, \hat{\sigma}_t^i, \hat{\rho}_t^i)$  其中  $\hat{\mu}_t^i$  为预测均值,  $\hat{\sigma}_t^i$  为预测标准差,  $\hat{\rho}_t^i$  为预测相关系数,  $\hat{p}_t^i$  为预测轨迹.本文模型通过最小化负对数似然损失函数来进行训练,损失函数如式(1)所示:

$$L^i(W) = - \sum_{t=T_{gc}+1}^{T_{pred}} \log \left( P \left( p_t^i | \hat{\mu}_t^i, \hat{\sigma}_t^i, \hat{\rho}_t^i \right) \right) \quad (1)$$

其中  $W$  为模型中所有可以训练的参数.

在本文中,提出了一种基于多模式时空交互行人轨迹预测模型,如图1所示.该模型由加权信息融合模块,多模式行人空间交互模块,时间卷积网络 (Temporal Convolutional Network, TCN) 模块组成.加权信息融合模块对输入的轨迹信息进行处理,使轨迹信息更加有效.多模式行人空间交互模块捕获不同模式中行人之间的交互,时间卷积网络模块捕获时间步长上的时间交互,利用时间卷积网络模块的输出来预测二维高斯分布的参数,从而生成预测的轨迹.

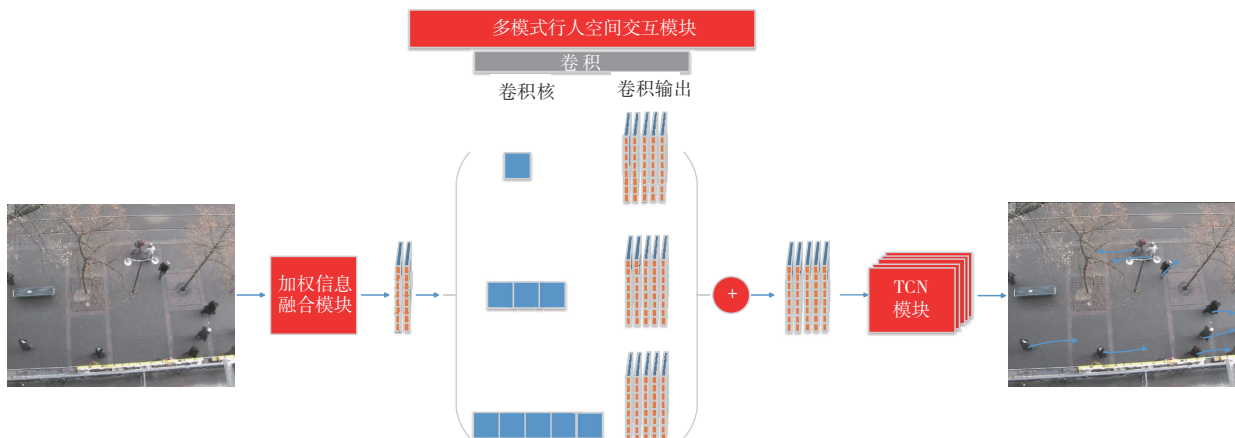


图1 基于多模式时空交互的行人轨迹预测模型框图

### 2.1 加权信息融合模块

简单地对于行人轨迹坐标  $x, y$  给与相同权重,势必会造成在某个坐标关注度不够,而在另外一个坐标又过度关注.本文提出的加权信息融合模块对于行人轨迹的坐标  $x, y$  坐标信息不再是之前的方式,而是在之前的基础上对  $x, y$  坐标信息上分别融合了赋予不同权重的该行人历史轨迹信息.

加权信息融合模块首先将行人轨迹整合成矩阵形式,如图2中(a)所示.然后将行人轨迹信息输入到卷积网络中,只对行人历史轨迹进行操作,通过卷积来对观察时间段内轨迹信息赋予不同权重,再与行人轨迹信息进行叠加,得到模块处理后的行人轨迹信息.加权信息融合模块具体结构如图2中(b)所示.

本文的模型对行人  $i$  在时间  $t$  上的位置信息  $p_t^i =$

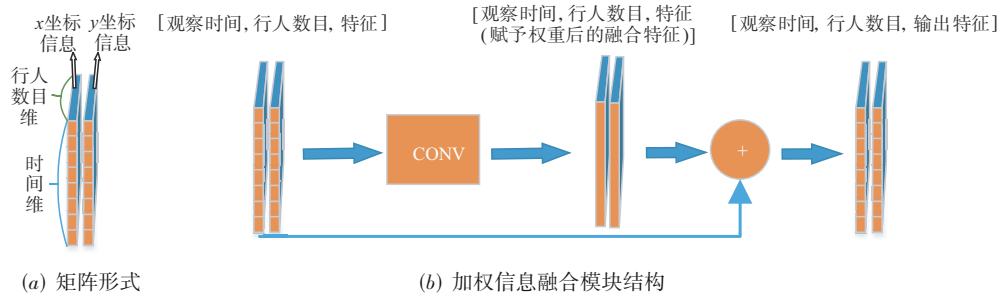


图2 加权信息融合模块图

$(x_i^t, y_i^t)$ 进行加权信息融合处理得到输出结果 $z_i^t$ . 加权信息融合的步骤是首先对行人 $i$ 在时间段 $t_{gc}$ 上的轨迹信息赋予不同的权重,然后再与对应时刻上的坐标信息 $x, y$ 对应相加得到了输出结果,如式(2)所示:

$$z_i^t = \left( \sum_{j=1}^{n=T_{gc}} w_j^x x_j^t + w_j^y y_j^t \right) + x_i^t, \left( \sum_{j=1}^{n=T_{gc}} w_j^y x_j^t + w_j^x y_j^t \right) + y_i^t \quad (2)$$

其中 $w_j^x$ 和 $w_j^y$ 是赋予观察阶段 $t_{gc}$ 上每时刻轨迹信息的权重, $z_i^t$ 是行人 $i$ 在时间 $t$ 上的加权信息融合的输出.

## 2.2 多模式行人空间交互模块

虽然在现有的模型中整合行人空间交互信息相当成熟,但是一次性整合所有行人之间的空间交互信息,必定会导致冗余信息的产生. 本文通过提出多模式行人空间交互模块来解决这个问题. 本文将行人空间交互情况分成三种模式,如图3所示. 第一种模式,每个行人只考虑自身的影响. 第二种模式,每次将三人分成一批,只考虑这三人之间的交互,用于考虑小规模交互情况. 第三种模式,将五人分成一批,只考虑这五个行人之间的交互情况,用于考虑大规模的交互情况.

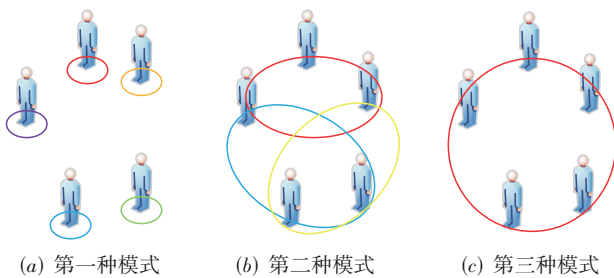


图3 多模式行人空间交互说明图

利用这种多模式行人空间交互模块,本文模型可以有效地获取多种不同模式下行人之间的交互信息. 通过对多种不同模式下行人之间的交互信息的叠加,来减少冗余信息对未来轨迹预测带来的影响. 本论文中的实现方法是通过将行人特征信息经过不同的卷积得到相同维度的信息再进行信息的叠加.

该结构采用了多个大小不同的卷积核,在卷积中实现提取行人之间交互信息. 根据卷积的公式,在行人数目维上的输出如式(3)所示:

$$\text{Nums} = \frac{\text{Num} - \text{kernelsize} + 2 \times \text{padding}}{\text{stride}} + 1 \quad (3)$$

公式中Num代表的是行人数目,kernelsize代表的是在卷积核在行人数目维上的大小,padding代表的是填充,stride代表的是步长,Nums代表的是卷积后的输出. 为了保证卷积前后行人数目的一致性,kernelsize只能取1,3,5,7...经过消融实验,本文取最好的情况即同时提取单个行人,三个行人和五个行人之间的交互信息.

如图4所示,本文的多模式行人空间交互模块是一个三级模块,采用的是二维卷积,为保证不产生时间上的交互,将卷积核的大小分别设为 $1 \times 1$ 、 $1 \times 3$ 和 $1 \times 5$ .

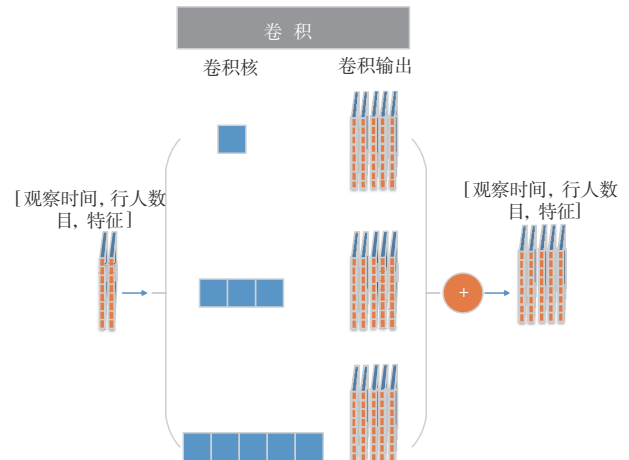


图4 多模式行人空间交互模块框图

并且通过对行人数目维上的填充,保证了卷积的正常进行以及卷积输出的结果的一致性. 这样保证了有效提取对应时刻下的行人空间交互信息,并且不会产生时间上的交互,具体实现见式(4)~(7):

$$g_1 = \text{CONV}(Z^i, W_1) \quad (4)$$

$$g_2 = \text{CONV}(Z^{i-1}, Z^i, Z^{i+1}, W_2) \quad (5)$$

$$g_3 = \text{CONV}(Z^{i-2}, Z^{i-1}, Z^i, Z^{i+1}, Z^{i+2}, W_3) \quad (6)$$

$$\alpha_i = g_1 + g_2 + g_3 \quad (7)$$

其中 $Z^{i-2}, Z^{i-1}, Z^i, Z^{i+1}, Z^{i+2}$ 分别为 $z_i^{i-2}, z_i^{i-1}, z_i^i, z_i^{i+1}, z_i^{i+2}$ 在时间上的堆叠, $W_1$ 是 $1 \times 1$ 卷积的参数, $W_2$ 是 $1 \times 3$

卷积的参数,  $W_2$  是  $1 \times 5$  卷积的参数,  $g_1$  是单人交互的输出,  $g_2$  是三人交互的输出,  $g_3$  是五人交互的输出.  $\alpha_i$  为多模式行人  $i$  空间交互的输出.

### 2.3 时间卷积网络模块

时间卷积网络(TCN)模块可以单独关注时间维度上的行人交互信息,从而真实地预测行人未来轨迹. 本文模型所用的TCN层结构如图5所示:

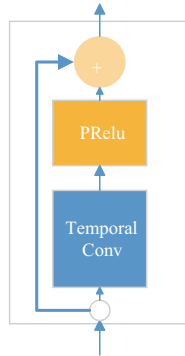


图5 TCN层框图

由于之前的多模式行人空间交互模块已经充分考虑了行人之间的交互情况,所以本文模型的时间卷积网络模块相比于之前模型中的TCN模块,通过改变了卷积核的大小从而避免了不必要的行人之间的信息融合. TCN模块公式如式(8)所示:

$$\beta_i = f_{\text{TCN}}(\alpha_i, W_{\text{TCN}}) \quad (8)$$

其中  $W_{\text{TCN}}$  为TCN模块的参数,  $\alpha_i$  是多模式行人  $i$  空间交互的输出,  $\beta_i$  是TCN模块输出的行人  $i$  的二维高斯分布参数,用于预测行人轨迹.

## 3 实验结果与分析

### 3.1 数据集和评价指标

该模型在两个行人轨迹预测数据集上进行训练: ETH<sup>[18]</sup>和UCY<sup>[19]</sup>. 其中ETH含有ETH和HOTEL两个场景,UCY含有ZARA1、ZARA2和UNIV三个场景. 这些数据集包含了上千条行人真实轨迹,蕴含了多种多样的行人交互. 在数据集的实验中,本文使用4个数据集来训练模型,然后在剩下的1个数据集上进行测试. 评估阶段,模型通过观察前3.2 s行人轨迹,来预测后来的4.8 s行人轨迹.

与先前的工作一样,使用了两种测量标准:

平均位移误差(Average Displacement Error, ADE): 在预测的所有时间步长内,地面真实和本文的预测之间的平均L2距离,如式(9)所示:

$$\text{ADE} = \frac{\sum_{n \in N} \sum_{t \in T_{\text{pred}}} \|\hat{p}_t^n - p_t^n\|^2}{N \times T_p}, T_p = T_{\text{pred}} - T_{\text{gc}} \quad (9)$$

最终位移误差(Final Displacement Error, FDE): 在预测时间结束时,预测的最终目的地与真正的最终目的地之间的距离,如式(10)所示:

$$\text{FDE} = \frac{\sum_{n \in N} \|\hat{p}_t^n - p_t^n\|^2}{N}, t = T_{\text{pred}} \quad (10)$$

本文模型将与以下模型进行对比,以下为各模型的简单介绍:

S-LSTM<sup>[3]</sup>: 每个行人都通过一个单独的LSTM进行建模,在每个时间步长上使用社交池化层(S-Pooling)聚集隐藏状态.

S-GAN<sup>[6]</sup>: 生成对抗网络应用于S-LSTM的预测模型.

SoPhie<sup>[5]</sup>: 基于生成式对抗网络框架的行人历史轨迹与场景上下文信息结合的模型.

Social-BiGAT<sup>[20]</sup>: 基于图卷积的生成对抗网络.

SR-LSTM<sup>[8]</sup>: 改进的SR模块的LSTM网络.

Social-STGCNN<sup>[15]</sup>: 由图卷积层和改进的TCN层组成的模型.

### 3.2 模型构造和训练设置

本文的模型包括一个加权信息融合模块,一个多模式行人空间交互模块,以及由五个TCN层组成的时间卷积网络模块. 通过模块堆叠数目实验,验证了一个多模式行人空间交互模块和五个TCN层堆叠对行人轨迹预测的效果最佳. 训练过程中,训练批次大小设为128. 本文的模型使用的激活函数是PReLU,采用了随机梯度下降法(Stochastic Gradient Descent, SGD)对模型进行250次训练. 初始学习率为0.01,再经过150次后学习效率转变成0.002. 本文中的实验均在相同硬件条件下进行的,实验计算机采用的处理器为Intel Core i7-10700k CPU @3.8 GHz, 8核16线程,显卡为NVIDIA 3090 GPU, 24 GB显存.

### 3.3 定量分析

#### 3.3.1 模型性能对比

在ADE/FDE指标上比较了本文模型的性能,具体结果见表1,表中的评价指标是ADE/FDE. 表1中的模型中除了SR-LSTM模型均是选取20条预测轨迹中最好的轨迹进行评估. 本文的模型在ADE/FDE指标上优于下面所提的模型. 本文的模型比Social-STGCNN模型在ADE指标上减低了15%,在FDE指标上降低了14%.

#### 3.3.2 模型参数和推理速度对比

本文模型的参数量为2828,比Social-STGCNN模型的参数量还要少. 本文模型与可供公众使用的模型之间模型参数和推理时间对比如下表所示,其中推理时间为是一些单一推理步骤的平均值,推理时间越低越

表1 几种经典模型与本文模型性能对比表

数据集场景	对比的模型						本文模型方法
	S-LSTM	S-GAN	SoPhie	Social-BiGAT	SR-LSTM	Social-STGCNN	
ETH	1.09/2.35	0.81/1.52	0.70/1.43	0.69/1.29	0.63/1.25	0.64/1.11	<b>0.62/1.08</b>
HOTEL	0.79/1.76	0.72/1.61	0.76/1.67	0.49/1.01	0.37/0.74	0.49/0.85	<b>0.28/0.44</b>
UNIV	0.67/1.40	0.60/1.26	0.54/1.24	0.55/1.32	0.51/1.10	0.44/0.79	<b>0.39/0.74</b>
ZARA1	0.47/1.00	0.34/0.69	0.30/0.63	0.30/0.62	0.41/0.90	0.34/0.53	<b>0.28/0.50</b>
ZARA2	0.56/1.17	0.42/0.84	0.38/0.78	0.36/0.75	0.32/0.70	0.30/0.48	<b>0.25/0.43</b>
AVG	0.72/1.54	0.58/1.18	0.54/1.15	0.48/1.00	0.45/0.94	0.44/0.75	<b>0.37/0.64</b>

好. 本文模型中之所以有这些成果,是因为该模型仅使用卷积神经网络,克服了递归体系架构和聚合机制这两大限制. 表2为各模型参数和推理速度对比表,其中模型参数的k代表千,推理时间的单位为s.

表2 模型参数和推理速度对比表

	模型参数	推理时间/s
S-LSTM	264k	1.1789
SR-LSTM	64.9k	0.1578
S-GAN-P <sup>[6]</sup>	46.3k	0.0968
Social-STGCNN	7.6k	0.0020
本文模型	2828	0.0008

### 3.3.3 加权信息融合模块对比

对比无加权信息融合模块,有加权信息融合模块可以在 ADE 和 FDE 指标下分别减低了 9% 和 8%. 表3为有无加权信息融合模块对比表,其中表中的有无代表的是加权信息融合模块的有无,评价指标是 ADE/FDE.

表3 有无加权信息融合模块对比表

	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
无	0.68/1.17	0.36/0.57	0.41/0.76	0.32/0.55	0.27/0.45	0.41/0.70
有	0.62/1.08	0.28/0.44	0.39/0.74	0.28/0.50	0.25/0.43	0.37/0.64

### 3.3.4 多模式消融实验对比

保证本文模型的多模式行人空间交互中选取的这三种模式的最优性,对此做了对应的消融实验,考虑多种情况中在 HOTEL 数据集上的效果. 表4为各种情况下行人交互预测效果对比表,其中,“√”符号表示了采用了该交互方式,多个“√”表示的是多种交互方式共同使用. 实验表明了单人交互加上三人交互再加上五人交互的情况预测准确度表现最好.

### 3.3.5 模型堆叠数目实验

通过考虑不同数目的多模式行人空间交互模块和 TCN 层堆叠后在 ETH 和 UCY 数据集上的平均预测效果来确定模块的具体选取数目. 在进行多次模块堆叠数目实验后,实验结果表明当多模式行人空间交互模块数目为 1 以及 TCN 层数目为 5 时达到最优效果. 表5为模块堆叠数目效果对比表,其中表中的第一行对应的

表4 各种情况下行人交互预测效果对比表

	单人交互	三人交互	五人交互	七人交互	ADE/FDE
情况1	√				0.33/0.57
情况2	√	√			0.34/0.61
情况3	√		√		0.30/0.47
情况4		√	√		0.31/0.51
情况5	√	√		√	0.29/0.48
情况6	√		√	√	0.33/0.55
情况7		√	√	√	0.30/0.50
情况8	√	√	√	√	0.33/0.58
情况9	√	√	√		0.28/0.44

是 TCN 层选取数目,从左边开始的第一列对应的是多模式行人空间交互模块选取数目,另外表中的评价指标是 ADE/FDE.

表5 模块堆叠数目效果对比表

	1	2	3	4	5	6
1	0.41/0.75	0.39/0.67	0.40/0.70	0.39/0.73	<b>0.37/0.64</b>	0.39/0.70
2	0.42/0.74	0.40/0.69	0.43/0.80	0.40/0.72	0.39/0.68	0.41/0.73
3	0.42/0.75	0.43/0.79	0.41/0.72	0.41/0.80	0.38/0.68	0.41/0.72
4	0.45/0.81	0.44/0.80	0.41/0.73	0.42/0.76	0.44/0.84	0.42/0.75

### 3.3.6 模型预测轨迹对比

本文的模型与 Social-STGCNN 的预测轨迹对比图,如图6所示. 上面的一行是 Social-STGCNN 的预测轨迹图,下面的一行是本文模型的轨迹预测图. 图中红线代表的是历史轨迹,黄线是预测轨迹,蓝线是真实轨迹. 根据预测轨迹对比图,可以明显看出本文的模型预测的轨迹要 Social-STGCNN 的轨迹更加接近真实轨迹.

## 3.4 定性分析

在定量分析部分表明了,本文提出的模型在 ADE/FDE 指标上要优于之前的先进水平. 现在定性的分析下,本文的模型的加权信息融合模块为什么会提高轨迹信息有效性,多模式行人空间交互模块为何能够减小冗余信息造成的影响.

在行人水平行走时,通常只在某个方向一直行走



图6 本文的模型与 Social-STGCNN 的预测轨迹对比图

下去,很少出现突然改变方向,简单地给予各坐标信息相同权重必定会影响信息的有效性. 本文的模型中的加权信息融合模块通过在原本坐标的基础上,对  $x$  和  $y$  坐标信息分别融合了赋予不同权重的行人观察阶段的轨迹信息,通过后续的训练使权重改变,这使得该模型可以重点去关注某个方向的轨迹信息. 在轨迹对比图 6 中,第一列的行人直线行走方式下,本文的模型通过考虑观察阶段的历史轨迹信息从而给予其中重要的轨迹坐标信息更大权重,预测出的行人轨迹也更加符合真实轨迹.

在轨迹对比图 6 中,第二列和第三列小规模行人交互情况中,正是因为本文模型中充分考虑了小规模行人交互情况,本文的模型预测的行人轨迹相比于 Social-STGCNN 模型预测的行人轨迹更加符合真实轨迹. 在轨迹对比的第四列大规模行人交互的情况中,本文的模型考虑了多种情况下的行人交互情况,减少了大规模交互时冗余信息带来的影响,预测的行人轨迹也更加准确.

#### 4 结论

本文提出一种基于多模式时空交互行人轨迹预测模型,在 ETH 和 UCY 数据集上预测行人轨迹的准确度高. 本文提出的模型首先通过对行人轨迹进行加权信息融合成功提高了输入信息的有效性. 然后通过模型中的多模式行人空间交互有效地降低了行人交互时产生的冗余信息对行人预测轨迹的影响. 最后通过时间卷积网络模块得到行人预测轨迹. 本文中模型中的空间行人交互部分和时间交互部分完全可分,可以准确预测行人未来轨迹. 此外该模型的模型参数少,推理速度快. 对于将来行人轨迹预测的研究,本文的工作为多模式时空交互的建模以及轨迹信息的有效利用提供了新的思路.

#### 参考文献

- [1] 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.
- [2] ZHAO D, CHEN Y, LE L. Deep reinforcement learning with visual attention for vehicle classification[J]. IEEE Transactions on Cognitive & Developmental Systems, 2017, 9(4): 356-367.
- [3] 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: IEEE CS, 2016: 961-971.
- [4] HUANG Y F, BI H K, LI Z X, et al. Stgat: Modeling spatial-temporal interactions for human trajectory prediction [C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul: CV/IEEE, 2019: 6271-6280.
- [5] SADEGHIAN A, KOSARAJU V, SADEGHIAN A, et al. Sophie: An attentive GAN for predicting paths compliant to social and physical constraints[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE CS, 2019: 1349-1358.
- [6] GUPTA A, JOHNSON J, FEI-FEI LI, et al. Social GAN: Socially acceptable trajectories with generative adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE CS, 2018: 2255-2264.
- [7] IVANOVIC B, PAVONE M. The trajectorytron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul: CV/IEEE, 2019: 2375-2384.

- [8] ZHANG P, OUYANG W, ZHANG P, et al. SR-LSTM: State refinement for LSTM towards pedestrian trajectory prediction[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE CS, 2019: 12077-12086.
- [9] XUE H, HUYNH D Q, REYNOLDS M. Bi-prediction: pedestrian trajectory prediction based on bidirectional LSTM classification[C]//Proceedings of the International Conference on Digital Image Computing: Techniques and Applications. Sydney: IEEE, 2017: 1-8.
- [10] XUE H, HUYNH D Q, REYNOLDS M. SS-LSTM: A hierarchical LSTM model for pedestrian trajectory prediction[C]//Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Lake Tahoe: IEEE, 2018: 1186-1194.
- [11] XUE H, HUYNH D Q, REYNOLDS M. A location-velocity-temporal attention LSTM model for pedestrian trajectory prediction[J]. IEEE Access, 2020, 8: 44576-44589.
- [12] LEE N, CHOI W, VERNAZA P, et al. Desire: Distant future prediction in dynamic scenes with interacting agents [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2165-2174.
- [13] FERNANDO T, DENMAN S, SRIDHARAN S, et al. Soft+ hardwired attention: An LSTM framework for human trajectory prediction and abnormal event detection [J]. Neural Networks, 2018, 108: 466-478.
- [14] SUN L, YAN Z, MELLADO S M, et al. 3DOF pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data[C]//Proceedings of the IEEE International Conference on Robotics and Automation. Brisbane: IEEE, 2018: 5942-5948.
- [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. Seattle: IEEE, 2020: 14412-14420.
- [16] ZHAO X, CHEN Y, GUO J, et al. A spatial-temporal attention model for human trajectory prediction[J]. IEEE/CAA Journal of Automatica Sinica, 2020, 7(4): 965-974.
- [17] WANG C X, CAI S F, TAN G. Graphtcn: Spatio-temporal interaction modeling for human trajectory prediction [C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021: 3450-3459.
- [18] PELLEGRINI S, ESS A, SCHINDLER K, VAN GOOL L. You'll never walk alone: Modeling social behavior for multi-target tracking[C]//2009 IEEE 12th International Conference on Computer Vision. Kyoto: IEEE, 2009: 261-268.
- [19] LERNER A, CHRYSANTHOU Y, LISCHINSKI D. Crowds by example[J]. Computer Graphics Forum, 2007, 26(3): 655-664.
- [20] 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. Vancouver: NeurIPS, 2019: 1-10.

### 作者简介



**桑海峰** 男, 1978年1月出生于辽宁省沈阳市. 现为沈阳工业大学教授、博士生导师. 主要研究方向为机器视觉检测技术和智能视频分析技术.

E-mail: sanghaif@163.com



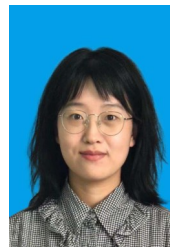
**陈旺兴(通讯作者)** 男, 1998年3月出生于江西省抚州市. 目前在沈阳工业大学攻读硕士学位, 主要研究方向为行人轨迹预测.

E-mail: 1909703861@qq.com



**王海峰** 男, 1995年3月出生于吉林省吉林市. 目前在沈阳工业大学攻读硕士学位, 主要研究方向为行人轨迹预测.

E-mail: 798466420@qq.com



**王金玉** 女, 1996年5月出生于辽宁省盖州市. 目前在沈阳工业大学攻读博士学位, 主要研究方向为行人轨迹预测.

E-mail: 1911131982@qq.com