

DOI:10.13203/j.whugis20200159



文章编号:1671-8860(2020)12-1989-08

# 结合自注意力机制和结伴行为特征的行人轨迹预测模型

周于涛<sup>1</sup> 吴华意<sup>1</sup> 成洪权<sup>1</sup> 郑 杰<sup>1</sup> 李学锡<sup>2</sup>

1 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉,430079

2 江苏易图地理信息科技股份有限公司,江苏 扬州,225009

**摘要:**理解并准确预测行人的移动轨迹,对提高自动驾驶技术的水平,减少交通事故的发生有重要的意义。针对现有轨迹预测方法预测精度不高,对行人交互信息利用不充分等问题,提出了一种结合自注意力机制和结伴行为特征的行人轨迹预测模型,该模型考虑了每个行人的运动信息及其与周围行人的交互作用,使用循环神经网络和图卷积网络分别对行人的行走状态和行人间的交互进行建模。在图卷积网络中,定义图的节点表示行人的运动信息,图的边表示行人之间的交互,使用自注意力机制计算行人间的交互程度。此外,为了增加模型捕捉结伴行走行为特征的能力,提高对该类轨迹预测的精度,提出了同伴损失函数的概念。在公共数据集上的实验表明,该模型在预测精度上相比其他方法有较大的提升。

**关键词:**行人轨迹预测;自注意力机制;图卷积网络;循环神经网络;深度学习

中图分类号:P208

文献标志码:A

近年来,随着智能交通系统的发展,对行人行为模式的研究不断深入。行人轨迹预测可以应用于诸多领域,如自动驾驶、机器人导航、智能追踪系统等<sup>[1-3]</sup>。准确的行人轨迹预测对自动驾驶系统至关重要,它可以帮助系统在车辆行驶过程中制定正确的决策,如刹车、减速、转弯或直行等动作。与车辆相比,行人的运动轨迹更加自由、灵活、复杂,对自动驾驶系统决策的影响更为显著,尤其是在人和车高度混合的场景,如住宅区、街道市场、停车场等。此外,准确的行人轨迹预测是智能交通管理的基础,如果能准确预测行人在十字路口的行为,就能提前识别潜在的风险,通过交通信号设施和车辆通信系统将警告信息发送到行人和车辆,使交通事故的概率大大降低。

行人的轨迹主要由行人计划行至的目的地决定,同时也会受到周围行人运动的影响。行人轨迹预测中一个关键的任务就是对同一场景下多个行人之间的相互作用进行建模。这些相互作用将导致一系列的运动行为,包括为了避免碰撞而加速、减速或转向,受他人阻挡而被迫跟随行走,以及其他复杂而微妙的运动行为。由于这

种相互作用与社会习俗、场景环境,甚至行人自身的个性有关,因此需要设计复杂的模型捕捉这种相互作用。轨迹预测的准确度很大程度上取决于对行人相互作用的理解程度。

行人轨迹预测问题近年来得到了广泛的研究。随着深度学习技术的发展,循环神经网络(recurrent neural network, RNN)模型被广泛用于捕获行人自身的移动模式。对于行人之间相互作用的捕捉,一些方法使用池化机制共享神经网络中的隐藏状态<sup>[4-5]</sup>,模拟行人之间的交互;另一些方法使用注意力机制来捕获行人之间的相对影响<sup>[6-7]</sup>,根据神经网络的隐藏状态所代表的运动模式来分配不同的权重。与基于池化机制的模型相比,基于注意力的模型能获得更丰富的交互信息,实现更准确的预测。

虽然现有方法在行人轨迹预测方面取得了一定的效果,但是对于行人交互作用的获取还不够充分,对行人交互信息的分析还不够全面。首先,基于池化机制与基于注意力机制的模型将行人两两之间的交互作用单独考虑。对于目标行人周围存在多个行人的情况,仅仅将周围行人的

收稿日期:2020-04-17

项目资助:国家重点研发计划(2017YFB0503800)。

第一作者:周于涛,硕士生,研究方向为轨迹数据挖掘与高性能地理信息计算。zhouyutao@whu.edu.cn

通讯作者:郑杰,博士,副研究员。zhengjie@whu.edu.cn

交互信息简单融合,忽略了行人整体之间的联系。本文使用图卷积神经网络结合自注意力机制,将行人间彼此的交互视作一个整体,共同捕捉行人间的交互信息。其次,结伴行走是人们行走时的常见现象,结伴而行的人可以是彼此认识的人,也可以是有着共同目的地的陌生人,他们在一段时间内将保持相似的运动状态,此前大多数研究的基本对象是场景中的独立行人,而没有考虑可能存在的同伴或群体。在行人群体的识别方法上,本文设计了一个特殊的损失函数来自动识别群体结伴行走行为,使模型在预测每个行人未来轨迹的基础上,同时顾及潜在群体轨迹的整体相似程度,帮助模型提高预测精度。在公共数据集上的实验表明,本文模型能够有效捕获行人间的交互特征,并预测行人的未来轨迹。

## 1 相关研究

传统的行人轨迹预测方法是通过建立行人行为模型,模拟每个行人的动作以及行人之间的交互,从而预测他们未来的轨迹。段炼等<sup>[8]</sup>使用贝叶斯模型预测时空位置,融入时空语义信息对行人移动特征进行建模。Helbing 等<sup>[9]</sup>提出了社会力模型,该模型通过建立吸引和排斥的运动模型来描述行人之间的相互作用。Trautman 等<sup>[10]</sup>开发了交互式高斯过程,这是一种基于高斯过程的非参数统计模型,用于估计人群交互作用。然而,这些方法大多是基于人工定义的特征或特定的规则。对于复杂的场景,需要进行手动调整以获得稳定的结果。并且由于模型复杂度高,难以用于规模较大、实时性要求较高的场景中。

近年来,基于深度学习的数据驱动的方法取得了更好的结果。RNN 及其变体结构,如长短时记忆网络(long short-term memory, LSTM)<sup>[11]</sup>和门控递归单元(gated recurrent unit, GRU)<sup>[12]</sup>被设计用于序列预测任务。基于 RNN 的网络可以捕获观测到的序列模式,并根据这些模式生成预测序列。在行人轨迹预测中,由于每个独立的 RNN 只获取一个行人的运动信息,无法利用行人之间的交互信息辅助预测,研究者引入运动特征和 RNN 内部的隐藏状态来整合附近行人的交互信息<sup>[4-5, 13]</sup>。Alahi 等<sup>[4]</sup>通过社交池化层在多个 LSTM 之间共享隐藏状态。Gupta 等<sup>[5]</sup>使用最大池化的神经网络结构,最大限度地汇集了行人之间的行为信息。这些方法比只使用独立的 RNN 而不考虑交互信息的方法取得了更好的效果。

但是,这些方法假设只有在一定距离内的行人才能影响目标行人的运动,并且对不同状态下的周围行人给予同等程度的关注,这在大多数情况下是不合理的。

注意力机制的主要思想是给模型输入的每个部分分配不同的权重,在解码过程中根据权重大小给予不同程度的关注。与简单的 RNN 相比,注意力机制可以使模型重视有助于预测的部分,而忽略不相关的部分,从而作出更准确的判断。在行人轨迹预测问题中,许多基于注意力的方法被用来衡量多个附近行人的相对重要性<sup>[6-7, 13-14]</sup>。Vemula 等<sup>[7]</sup>提出了一个社交注意力模型,在预测过程中为附近的每个行人分配不同的权重。Fernando 等<sup>[6]</sup>使用了一个软硬结合的注意力模型,将运动信息从附近的各个行人映射到目标行人的未来位置。这些工作表明,基于注意力的机制可以反映不同行人影响的差异。本文采用自注意力机制<sup>[15]</sup>来捕获行人间的影响程度。

图卷积网络(graph convolutional network, GCN)<sup>[16]</sup>是一种强大的神经网络结构,用于在具有图结构的数据上进行特征提取和信息融合。GCN 利用图的结构特征初始化各节点上的信息,通过构建神经网络从图中各个节点的局部邻域传递特征信息,捕获节点之间的依赖关系,在交通速度预测<sup>[17-18]</sup>和推荐系统<sup>[19-20]</sup>等任务上取得了显著的效果。对于轨迹预测问题,场景中的行人可以看作是图上的节点,而行人之间复杂的交互作用可以看作是图的边(如图 1 所示),行人的运动特征决定了这些边的权值。通过描述行人运动特征的张量在卷积层上的正向传播,行人的运动信息能够在行人之间共享和传递,由此形成的交互信息可用于辅助轨迹预测。

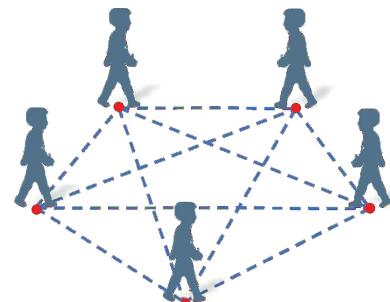


图 1 图模型描述的行人交互  
Fig.1 Pedestrian Interaction Described by Graph Model

结伴而行是行人行走时的普遍现象。它是指两个或多个互相认识的行人,或者虽不认识,但在一定范围内具有相同目的地的陌生人,保持

较为恒定的相对位置,共同行走一段路程的行为。Ge 等<sup>[21]</sup>使用相对速度特征的层次聚类来寻找这样的群体,Lu 等<sup>[22]</sup>使用扩展的元胞自动机模型来模拟行人的群体行走行为,并分析其对行人整体的影响。借助精心设计的物理模型,这些方法在识别行人群体方面较为准确,能获得可量化的指标。然而,由于这些模型的复杂性,轨迹预测的实时性要求难以满足,这些模型都是针对特定场景人工设计的特征,不能自动适应不同的场景。

针对以上问题,本文提出了一种将 RNN 编码器-解码器与 GCN 相结合的模型,用于在拥挤场景下对行人的轨迹进行预测,以更好地获取行人的运动特征和交互特征。此外,在绝对坐标的角度误差基础上,本文加入了额外的同伴损失,以提升模型预测结伴行走行为的能力。

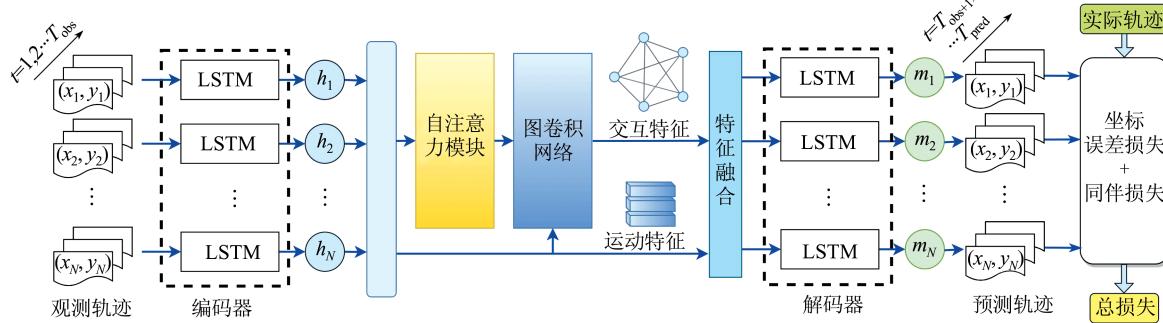


图 2 结合自注意力机制和结伴行为特征的行人轨迹预测模型架构图

Fig.2 Architecture of Pedestrian Trajectory Prediction Model Based on Self-Attention Mechanism and Group Behavior Characteristics

### 2.2.1 运动信息编码

每个行人都有自己的运动特征,运动特征是指行人在一段时间内的运动行为,包括绝对坐标、移动速度和移动方向。这些运动特征可以用于推断未来轨迹的大致方向,并可从中进一步提取行人之间的相互作用。为获取每个行人的运动信息,本文使用多层感知机(multi-layer perception, MLP)将每个行人的位置坐标嵌入到一个向量中,即:

$$\mathbf{v}_i^t = \phi(x_i^t, y_i^t; \mathbf{W}_e) \quad (1)$$

式中,  $\phi$  是多层感知机,  $\mathbf{W}_e$  是其权重。获得的各个时间步的嵌入向量将依次输入到 LSTM 中,获得编码后的运动特征向量  $\mathbf{h}$ :

$$\mathbf{h}_i^t = \text{LSTM}(\mathbf{h}_{i-1}^{t-1}, \mathbf{v}_i^t; \mathbf{W}_r) \quad (2)$$

式中,  $\mathbf{h}_i^t$  是行人  $i$  在时间步  $t$  下 LSTM 的隐藏状态,该隐藏状态反映了行人  $i$  在前  $t$  个时间步内的运动特征。

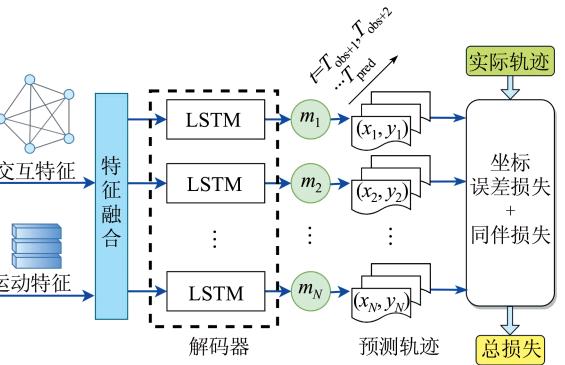
## 2 模型描述

### 2.1 问题定义

行人轨迹预测问题可以抽象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  $p_i^t = (x_i^t, y_i^t)$  表示, 已知行人  $i$  的位置  $p_i^t$  ( $t=1, 2 \dots T_{\text{obs}}$ ), 目标是预测它们从  $T_{\text{obs}+1}$  到  $T_{\text{pred}}$  的位置, 用  $\hat{p}_i^t = (\hat{x}_i^t, \hat{y}_i^t)$  表示预测值。

### 2.2 模型架构

模型的架构图如图 2 所示, 主要结构为编码器-解码器模型。本文使用一个 LSTM 作为编码器来获取行人的移动特征, 并使用 GCN 来提取行人之间的交互特征。另一个 LSTM 作为解码器, 它将运动特征和交互特征的组合作为输入, 依次输出未来的轨迹。



### 2.2.2 行人交互提取

在拥挤的场景中, 行人的运动会受到周围行人的影响, 从而导致转向、跟随、加速或减速。这种影响往往是相互的, 称之为行人交互, 它是影响行人未来轨迹的重要因素。模型使用 GCN 获得行人的交互信息。研究证明, GCN 可以根据图的结构聚合节点信息<sup>[23]</sup>。如图 1 中将场景中的行人视为一个有向完全图  $G(V, E, A)$ 。节点集合  $V$  表示场景中各个行人运动特征向量的集合, 边集  $E$  中的每条边表示一端行人对另一端行人存在交互影响,  $A$  表示邻接矩阵, 矩阵中的每个值表示对应节点每两个行人之间的交互程度, 即对应有向边的权重。GCN 以节点向量集合和邻接矩阵作为输入, 在该问题中, 节点向量是运动特征编码后的向量, 邻接矩阵则反映行人彼此之间的交互程度。本文设计了一个基于自注意力机制的模块, 使用数据驱动的方法, 利用行人彼此之间编码后的运动信息, 通过神经网络计算他们之间的

交互特征,生成邻接矩阵。

首先,将每个行人在编码器中的隐藏状态输入到多层感知机中,并将同一场景下所有行人的输出合并为张量  $F$ 。表示如下:

$$f_i = \phi(\mathbf{h}_i; \mathbf{W}_a) \quad (3)$$

$$F = [f_1, f_2, \dots, f_N] \quad (4)$$

式中,  $\phi$  是多层感知机;  $\mathbf{W}_a$  是其权重;  $N$  是一个场景下行人的总数;  $F \in \mathbb{R}^{N \times d_e}$ ,  $d_e$  是多层感知机的输出维度。然后将合并后的张量  $F$  分别与两个权重矩阵相乘:

$$\mathbf{K} = F\mathbf{W}_K \quad (5)$$

$$\mathbf{Q} = F\mathbf{W}_Q \quad (6)$$

式中,  $\mathbf{W}_K$  和  $\mathbf{W}_Q$  是权重矩阵, 维度是  $d_e \times d_k$ ,  $d_k$  是注意力向量的长度。最后, 将  $\mathbf{K}$  和  $\mathbf{Q}$  矩阵相乘并除以  $\sqrt{d_k}$ , 再经过 Softmax 层归一化, 即可得到邻接矩阵  $A$ :

$$A = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \quad (7)$$

需要注意的是, 矩阵  $A$  是实数矩阵, 且不是对称矩阵。这符合行人之间的相互作用连续性和不对称性。考虑到行人交互作用只发生在一定距离内, 我们使用一个阈值截断邻接矩阵。即如果两个行人之间的距离大于某个阈值, 则将其邻接矩阵的对应值设置为0。

获得邻接矩阵后, 使用 GCN 网络将行人的运动信息在所确定的图结构上进行信息传递和向量融合, 其定义如下:

$$\mathbf{R}_{(l+1)} = \sigma(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} \mathbf{R}_{(l)} \mathbf{W}_{(l)}) \quad (8)$$

式中,  $\hat{A} = A + I$ ,  $I$  是单位矩阵;  $\hat{D}$  是  $\hat{A}$  的度矩阵;  $\mathbf{R}_{(l)}$  是 GCN 第  $l$  层所有节点向量合并而成的矩阵, 特别地, 第 0 层是 GCN 的输入(即  $\mathbf{R}_{(0)}$ ), 表示每个行人对应的 LSTM 编码器在  $T_{\text{obs}}$  时间步的隐藏向量;  $\mathbf{W}_{(l)}$  对应第  $l$  层的 GCN 网络的权重。最后一层输出的矩阵  $\mathbf{R}$  即代表行人间交互的特征张量, 它将与行人自身的运动特征相结合, 用于生成未来的轨迹。

### 2.2.3 未来轨迹生成

针对未来轨迹的生成方法, Alahi 等<sup>[4]</sup>将其视为从某一分布中进行采样, 即利用处理后的神经网络隐藏状态确定一个二元高斯分布的参数, 并从其中随机采样得到未来的坐标。然而, 由于采样过程是不可微的<sup>[5]</sup>, 给训练过程中的反向传播带来了困难。本文通过直接从隐藏状态预测坐标来避免这种情况。将 GCN 输出的交互特征  $\mathbf{R}$

与 LSTM 编码器输出的运动特征  $H$  连接, 送入另一个 LSTM 解码器, 依次产生预测的位置坐标表示为:

$$\mathbf{m}_i^{t+1} = \text{LSTM}([\mathbf{r}_i, \mathbf{h}_i]; \mathbf{W}_d) \quad (9)$$

$$(\hat{x}_i^{t+1}, \hat{y}_i^{t+1}) = \phi(\mathbf{m}_i^{t+1}; \mathbf{W}_f) \quad (10)$$

式中,  $t = T_{\text{obs}}, T_{\text{obs}+1}, \dots, T_{\text{pred}}$ ;  $\mathbf{m}_i^t$  是行人  $i$  对应解码器的隐藏状态;  $\phi$  是多层感知机;  $\mathbf{W}_f$  是其权重。

### 2.2.4 结伴损失

结伴行走是行人一种常见的社会行为。靠得很近的行人可以被认为是同伴, 他们通常有相同的行走方向和目的地。同伴可以是一起行走的朋友, 也可以是在一定范围内有着相同运动目标的陌生人。在大多数情况下, 特别是拥挤的场景中, 他们的轨迹在短时间内非常相似, 且彼此间距离变化不大。为使模型能捕捉行人之间的结伴关系, 本文在绝对坐标的均方误差损失函数的基础上, 增加一个结伴损失函数  $L_s$ , 表达式为:

$$L_s = \frac{1}{|S|} \sum_{i,j \in S, i \neq j} \sum_{t=T_{\text{obs}+1}}^{T_{\text{pred}}} |D_t(i,j) - \hat{D}_t(i,j)| \quad (11)$$

$$D_t(i,j) = \sqrt{(x_i^t - x_j^t)^2 + (y_i^t - y_j^t)^2} \quad (12)$$

$$\hat{D}_t(i,j) = \sqrt{(\hat{x}_i^t - \hat{x}_j^t)^2 + (\hat{y}_i^t - \hat{y}_j^t)^2} \quad (13)$$

式中,  $S$  表示同伴对的集合, 如果  $i$  和  $j$  两个行人之间的距离在观测时段内均小于阈值  $d$ , 则  $i, j$  属于  $S$ , 即  $D_t(i,j) \leq d, t = 1, 2, \dots, T_{\text{obs}}$ ;  $|S|$  表示  $S$  中同伴对的数量。

若两个行人在前  $T_{\text{obs}}$  时间步内的距离小于阈值  $d$ , 则认为他们是一对同伴。结伴损失函数  $L_s$  与每对同伴之间的相对距离之和成正相关关系, 即相对距离越大, 则结伴损失函数值越大。

模型的总损失由坐标误差损失与结伴损失乘上系数后相加而得, 即:

$$L = L_c + \lambda L_s \quad (14)$$

式中, 坐标误差损失  $L_c$  是预测坐标与真实坐标的均方误差;  $\lambda$  是系数。

### 2.3 模型训练

本文提出的模型需要获取行人的交互特征, 为了便于计算, 在模型训练的过程中, 将同一时刻处于同一场景下的所有行人同时输入到模型中, 经过前向传播, 获得这些行人在未来  $T_{\text{obs}+1}, T_{\text{obs}+2}, \dots, T_{\text{pred}}$  时刻的预测位置, 结合他们在这些时刻的真实位置, 即可求出  $L_c$  和  $L_s$ , 完成前向传播过程。

在反向传播中, 根据反向传播算法, 分别计算总损失  $L$  相对于解码器、图卷积网络、编码器以

及多层感知机的模型参数的梯度值。这些模块的参数是共享的,与行人数量无关,且都能通过自动微分计算梯度。求出参数的梯度后,使用合适的优化器和学习率,更新模型的参数,如此迭代循环,直至收敛。

### 3 实验和分析

#### 3.1 实验数据和环境

在ETH<sup>[24]</sup>和UCY<sup>[25]</sup>这两个公共的数据集上评估本文模型。这些数据集由真实世界中具有丰富交互行为的行人轨迹组成。ETH数据集包含两个场景,分别称为UNIV和HOTEL。UCY数据集包含3个场景,分别称为ZARA01、ZARA02和UNIV。这5组数据一共包含1536位行人,所有轨迹数据均被转换为位置坐标,并按0.4 s的间隔插值,以获得连续的轨迹。所有的实验都是在相同的硬件环境下进行的,处理器为Intel Core i9-9900K CPU @ 3.60 GHz,显卡为NVIDIA 2080Ti GPU。

#### 3.2 评价指标

实验时,模型输入8个时间步长(3.2 s)的轨迹,输出预测的未来12个时间步长(4.8 s)的轨迹。与之前的研究相同<sup>[4-5]</sup>,本文使用两个指标来评估预测误差。

1)平均位置误差(average displacement error, ADE):每一时间步的预测坐标与真实坐标之间的均方误差,计算公式为:

$$ADE = \frac{1}{TN} \sum_{t=T_{obs}+1}^{T_{pred}} \sum_{i=1}^N \left\| p_i^t - \hat{p}_i^t \right\|_2 \quad (15)$$

2)最终位置误差(final displacement error, FDE):在预测的最后一个时间步 $T_{pred}$ ,预测坐标与真实坐标的误差。计算公式为:

$$FDE = \frac{1}{N} \sum_{i=1}^N \left\| p_i - \hat{p}_i \right\|_2 \quad (16)$$

#### 3.3 实验结果

模型在各个数据集上的实验结果如表1、表2所示,本文还与传统方法和当前主流的几种方法如线性回归、LSTM、社会长短时记忆网络(social LSTM, S-LSTM)<sup>[4]</sup>及社会生成对抗网络(social generative adversarial networks, SGAN)<sup>[5]</sup>进行了比较。

#### 3.4 结果分析

从表1、表2中平均结果来看,本文提出的模型优于其他对比模型。在UCY-ZARA02和UCY-UNIV场景下,线性回归模型和LSTM模

型的性能比本文提出的模型相差较大,原因是它们只捕获个体的运动,而不获取行人之间的交互信息。在这两个交互行为丰富的数据集中,交互信息缺失将导致预测结果与真实结果偏离较大。

表1 不同方法在各数据集场景下的 ADE/m

Tab.1 ADE of Several Models on Each Data Set /m

数据集场景	线性 回归	LSTM	S- LSTM	SGAN	本文 方法
ETH-UNIV	1.02	0.97	0.86	0.81	0.82
ETH-HOTEL	0.41	0.54	0.49	0.67	0.32
UCY-ZARA01	0.61	0.48	0.43	0.34	0.40
UCY-ZARA02	0.46	0.40	0.38	0.42	0.32
UCY-UNIV	0.74	0.69	0.61	0.60	0.50
平均值	0.65	0.62	0.55	0.58	0.47

表2 不同方法在各数据集场景下的 FDE/m

Tab.2 FDE of Several Models on Each Data Set /m

数据集场景	线性 回归	LSTM	S- LSTM	SGAN	本文模 型方法
ETH-UNIV	2.18	1.80	1.77	1.52	1.72
ETH-HOTEL	0.84	1.15	1.00	1.37	0.62
UCY-ZARA01	1.19	1.01	0.91	0.69	0.87
UCY-ZARA02	0.90	0.81	0.80	0.84	0.71
UCY-UNIV	1.43	1.36	1.24	1.26	1.10
平均值	1.31	1.23	1.14	1.18	1.01

本文提出的模型在ETH-HOTEL和UCY-UNIV场景下显著优于S-LSTM和SGAN。这些数据场景中同时存在静止的行人和朝着相同或相反方向移动的行人。对于这些场景,S-LSTM和SGAN只考虑目标行人附近区域的行人,且给予同等的关注。本文模型考虑了更大的范围,并根据相互作用的程度给予差异化的关注。与SGAN相比,本文方法在ADE和FDE两项指标上分别减少了19.0%和14.4%。

需要指出的是,根据文献[5],SGAN模型在预测每条轨迹时都进行了20次采样,并使用与真值最接近的一次结果作为最终的预测结果。而本文提出的模型对每条轨迹只进行了一次预测,不存在参考真值并挑选最优结果的情况。

#### 3.5 同伴损失验证

为了验证本文提出的同伴损失函数的效果,这里增加了一个模型,该模型是本文模型的一种变体,其损失函数是绝对坐标的均方误差损失,没有附加本文提出的同伴损失,除此之外,模型的其他组成部分和参数保持不变。评估结果如表3所示,可以看出,含有同伴损失的模型比不含

同伴损失的模型在 ADE 指标上提高了 7.80%，在 FDE 指标上提高了 8.10%。特别是在 ETH-UNIV 和 ETH-HOTEL 上，同伴损失的效果比较显著。主要原因是在这两个数据集场景中有较多的结伴行走的情况，并且大多数结伴群体中行人的轨迹是保持相对不变的。在这种情况下，同伴损失能够有效帮助模型在预测过程中保持同伴轨迹的相似性，有助于提高模型的预测精度。

表 3 同伴损失函数验证结果(ADE/FDE) /m

Tab.3 Verification Results of Peer Loss Function/m

数据集场景	含有同伴损失	不含同伴损失
ETH-UNIV	0.82/1.72	0.88/1.91
ETH-HOTEL	0.32/0.62	0.38/0.74
UCY-ZARA01	0.40/0.87	0.42/0.92
UCY-ZARA02	0.32/0.71	0.33/0.74
UCY-UNIV	0.50/1.10	0.53/1.16
平均值	0.47/1.01	0.51/1.10

### 3.6 定性评价

为了分析模型提取的行人交互信息与实际情况的吻合程度，本文将自注意力机制模块所生成的 GCN 邻接矩阵值进行可视化。选取了数据集中的几个场景，可视化结果如图 3 所示。图 3 中，轨迹以“▲”结尾的行人是本文关注的目标行人，其他以圆点结尾的是其附近行人，圆点周围的圆圈大小正比于对应的邻接矩阵的权值。

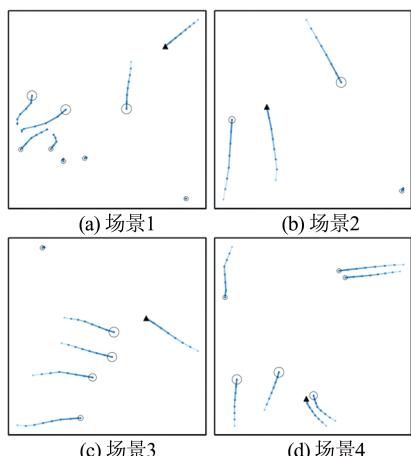


图 3 自注意力机制所生成的邻接矩阵权值可视化  
Fig.3 Visualization of Weights in Adjacency Matrix  
Obtained by Self-Attention Mechanism

可以看到，在图 3(a)中，在目标行人附近，移动的行人比静止的行人获得了更多的注意（即邻接矩阵的权重更大）。在图 3(b)中，目标行人对反方向走来的行人给予了更大的关注度，大于与其同方向而行的行人的关注度。如图 3(c)所示，

与远处行人相比，模型赋予了近处行人更高的关注，且权重大小随距离呈现反比关系。在图 3(d)中，与目标行人同向而行的 3 个行人获得的关注度基本相同，同时大于远处行走的行人。这些例子表明，本文提出的模型中的自注意力机制模块能够根据周围行人的运动状态为其分配合理的权重，从而帮助模型进行轨迹预测。

本文还将模型预测的轨迹与真实轨迹的示意图进行了可视化，以验证同伴损失函数提升预测群体行走行为的能力。图 4(a)、4(b)是两个场景下模型预测的示意图，其中，左侧子图对应的模型使用的是绝对坐标的均方误差作为损失函数，右侧子图则在左侧子图基础上增加了同伴损失。

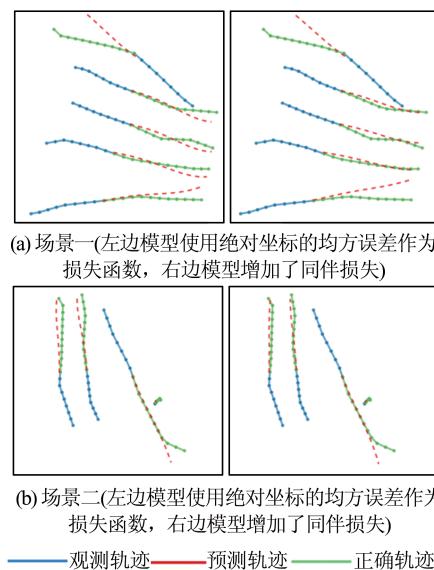


图 4 同伴损失提升预测效果示意图  
Fig.4 Visualization of Peer Loss Effects

从图 4 可以看出，这两个场景都存在明显的结伴行走的行为特征，有同伴损失的模型对群体预测的轨迹比没有同伴损失的模型预测的轨迹更加准确。这说明了含有同伴损失的模型能够准确识别场景中可能存在的结伴行走行为，并在预测时考虑群体内部两两之间的相对距离，使模型在预测每个行人个体轨迹的基础上，从全局的角度增加了对整个行人群体轨迹的约束，从而使预测的轨迹更加接近于真实轨迹。

## 4 结语

本文提出了一种结合自注意力机制和结伴行为特征的行人轨迹预测模型，将基于 LSTM 的编码器-解码器模型与图卷积网络相结合，把行人行走时的交互抽象成一个有向完全图。模型使

用基于LSTM的编码器获取行人的运动特征,使用GCN获取行人之间的交互特征,另一个基于LSTM的解码器通过融合运动特征和交互特征,生成预测的轨迹。为了捕捉行人结伴行走的行为特征,本文提出了同伴损失函数的概念,使神经网络对结伴而行的行人群体生成更为准确的预测。通过在两个公开数据集上的实验证明,证明本文方法相比已有方法预测精度显著提高,并且能够满足实时性的要求。今后可在此基础上引入更多的信息,如环境、行人个体特征等,进一步提高预测的精度。

## 参 考 文 献

- [1] Badue C, Guidolini R, Carneiro R V, et al. Self-Driving Cars: A Survey [J]. *Expert Systems with Applications*, 2020, 165:113 816
- [2] Zhao Qing, Chen Yong, Luo Bin, et al. A Local Path Planning Algorithm Based on Pedestrian Prediction Information [J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(5) : 667-675(赵青, 陈勇, 罗斌, 等. 一种融合行人预测信息的局部路径规划算法[J]. 武汉大学学报·信息科学版, 2020, 45(5):667-675)
- [3] Xiong Hanjiang, Guo Sheng, Zheng Xianwei, et al. Indoor Pedestrian Mobile Activity Recognition and Trajectory Tracking[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(11) : 1 696-1 703(熊汉江, 郭胜, 郑先伟, 等. 室内行人移动行为识别及轨迹追踪[J]. 武汉大学学报·信息科学版, 2018, 43(11):1 696-1 703)
- [4] Alahi A, Goel K, Ramanathan V, et al. Social LSTM: Human Trajectory Prediction in Crowded Spaces[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016
- [5] Gupta A , Johnson J , Fei F L, et al. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018
- [6] 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
- [7] Vemula A, Muelling K, Oh J. Social Attention: Modeling Attention in Human Crowds [C]. IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 2018
- [8] Duan Lian, Hu Tao, Zhu Xinyan, et al. Spatio-Temporal Prediction of Suspect Location by Spatio-Temporal Semantics [J]. *Geomatics and Information Science of Wuhan University*, 2019, 44 (5) : 765-770(段炼, 胡涛, 朱欣焰, 等. 顾及时空语义的疑犯位置时空预测[J]. 武汉大学学报·信息科学版, 2019, 44(5): 765-770)
- [9] Helbing D, Molnar P. Social Force Model for Pedestrian Dynamics[J]. *Physical Review E*, 1995, 51 (5):4 282
- [10] Trautman P, Krause A. Unfreezing the Robot: Navigation in Dense, Interacting Crowds [C]. IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, China, 2010
- [11] Hochreiter S, Schmidhuber J U R. Long Short-Term Memory[J]. *Neural Computation*, 1997, 9(8) : 1 735-1 780
- [12] Chung J, Gulcehre C, Cho K H, et al. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling [OL]. <http://arxiv.org/abs/1412.3555>, 2014
- [13] Sadeghian A, Kosaraju V, Sadeghian A, et al. Sophie: An Attentive Gan for Predicting Paths Compliant to Social and Physical Constraints[C]. IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019
- [14] Su H, Zhu J, Dong Y, et al. Forecast the Plausible Paths in Crowd Scenes[C]. 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017
- [15] Vaswani A, Shazeer N, Parmar N, et al. Attention is all You Need[C]. Advances in Neural Information Processing Systems, Long Beach, CA, USA, 2017
- [16] Kipf T N, Welling M. Semi-supervised Classification with Graph Convolutional Networks [OL]. <http://arxiv.org/abs/1609.02907>, 2016
- [17] Li Y, Yu R, Shahabi C, et al. Diffusion Convolutional Recurrent Neural Network: Data-Driven Traffic Forecasting [OL]. <http://arxiv.org/abs/1707.01926>, 2017
- [18] Yu B, Yin H, Zhu Z. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting [OL]. <http://arxiv.org/abs/1709.04875>, 2017
- [19] Ying R, He R, Chen K, et al. Graph Convolutional Neural Networks for Web-Scale Recommender Systems [C]. 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 2018
- [20] Ge Yao, Chen Songcan. Graph Convolutional Network for Recommender Systems [J]. *Journal of*

- Software*, 2020, 31(4): 1 101-1 112 (葛尧, 陈松灿. 面向推荐系统的图卷积网络[J]. 软件学报, 2020, 31(4):1 101-1 112)
- [21] Ge W, Collins R T, Ruback R B. Vision-Based Analysis of Small Groups in Pedestrian Crowds[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(5): 1 003-1 016
- [22] Lu L, Chan C Y, Wang J, et al. A Study of Pedestrian Group Behaviors in Crowd Evacuation Based on an Extended Floor Field Cellular Automaton Model[J]. *Transportation Research Part C: Emerging Technologies*, 2017, 81: 317-329
- [23] Li Y, Tarlow D, Brockschmidt M, et al. Gated Graph Sequence Neural Networks[OL]. <http://arxiv.org/abs/1511.05493>, 2015
- [24] Pellegrini S, Ess A, Schindler K, et al. You'll Never Walk Alone: Modeling Social Behavior for Multi-Target Tracking[C]. IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 2009
- [25] Lerner A, Chrysanthou Y, Lischinski D. Crowds by Example[M]//Computer Graphics Forum. Oxford, UK: Blackwell Publishing Ltd, 2007

## Pedestrian Trajectory Prediction Model Based on Self-Attention Mechanism and Group Behavior Characteristics

ZHOU Yutao<sup>1</sup> WU Huayi<sup>1</sup> CHENG Hongquan<sup>1</sup> ZHENG Jie<sup>1</sup> LI Xuexi<sup>2</sup>

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

2 Jiangsu Easymap Geographic Information Technology Co.Ltd, Yangzhou 225009, China

**Abstract:** Understanding and accurately predicting the trajectory of pedestrians is of great significance to improve the level of auto-driving technology and reduce the occurrence of traffic accidents. Aiming at the problems of low prediction accuracy and insufficient utilization of pedestrian interaction information, a pedestrian trajectory prediction model based on self-attention mechanism and group behavior characteristics is presented. The model considers the movement information of each pedestrian and its interaction with the surrounding pedestrians. The recurrent neural network and graph convolutional network are used to model the pedestrian's walking state and the pedestrian's interaction separately. In the graph convolutional network, nodes of the graph represent the movement information of pedestrians, and edges of the graph represent the interaction between pedestrians. Self-attention mechanism is used to calculate pedestrian interaction. In addition, in order to increase the ability of the network to capture the group walk behavior and decrease the prediction error, a peer loss function is proposed. Experiments on public datasets show that the proposed model in this paper improves prediction accuracy significantly.

**Key words:** pedestrian trajectory prediction; self-attention mechanism; graph convolutional network; recurrent neural network; deep learning

**First author:** ZHOU Yutao, postgraduate, majors in trajectory data mining and high-performance geographic information computing. E-mail: zhouyutao@whu.edu.cn

**Corresponding author:** ZHENG Jie, PhD, associate professor. E-mail: zhengjie@whu.edu.cn

**Foundation support:** The National Key Research and Development Program of China (2017YFB0503800).

**引文格式:** ZHOU Yutao, WU Huayi, CHENG Hongquan, et al. Pedestrian Trajectory Prediction Model Based on Self-Attention Mechanism and Group Behavior Characteristics[J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(12):1989-1996. DOI:10.13203/j.whugis20200159(周于涛,吴华意,成洪权,等.结合自注意力机制和结伴行为特征的行人轨迹预测模型[J].武汉大学学报·信息科学版,2020,45(12):1989-1996.DOI:10.13203/j.whugis20200159)