计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(12):284−291 [doi: 10.15888/j.cnki.csa.009335]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于时空图的行人轨迹预测①

朱鹏飞, 张德平

(南京航空航天大学 计算机科学与技术学院, 南京 211106)
通信作者: 张德平, E-mail: depingzhang@163.com

摘　要: 在蓬勃发展的自动驾驶技术中, 行人轨迹预测的结果往往会影响到自动驾驶的安全性. 行人轨迹预测技术目前面临着在实际场景中应用时与他人的交互问题, 需要在预测轨迹的同时考虑社会交互性与逻辑自洽. 因此, 提出了一种基于时空图的行人轨迹预测方法, 该方法采用图注意力网络对场景中的行人交互进行建模, 并使用一种自动生成正负样本的方法来通过对比学习降低输出轨迹的碰撞率, 达到了提高输出轨迹的安全性以及逻辑自洽的效果. 在 ETH 和 UCY 数据集上进行模型训练与测试, 结果分析表明, 本文提出的方法有效降低了碰撞率, 且预测准确度优于主流算法.

关键词: 轨迹预测; 注意力机制; 图神经网络; 对比学习

引用格式: 朱鹏飞,张德平.基于时空图的行人轨迹预测.计算机系统应用,2023,32(12):284–291. http://www.c-s-a.org.cn/1003-3254/9335.html

Pedestrian Trajectory Prediction Based on Spatio-temporal Graph

ZHU Peng-Fei, ZHANG De-Ping

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: In the booming autonomous driving technology, the results of pedestrian trajectory prediction often affect autonomous driving safety. Pedestrian trajectory prediction technology currently faces the problem of interaction with others when applied to practical scenarios, requiring consideration of social interaction and logical consistency during predicting trajectories. Therefore, this study proposes a pedestrian trajectory prediction method based on spatio-temporal graphs. This method employs graph attention networks to model pedestrian interactions in the scenarios and adopts a method of automatically generating positive and negative samples to reduce the collision rate of the output trajectory through contrastive learning, thus improving the safety and logical consistency of the output trajectory. Model training and testing are conducted on ETH and UCY datasets, and the results show that the proposed method reduces the collision rate and has better prediction accuracy than mainstream algorithms.

Key words: trajectory prediction; attention mechanism; graph neural network (GNN); contrastive learning

1　引言

根据观察到的行人轨迹, 行人轨迹预测旨在预测行人未来的位置坐标序列, 在自动驾驶、视频监控和视觉识别等各种应用中发挥着关键作用[1]. 然而, 行人轨迹预测是一项极具挑战性的任务, 其难点主要在于: (1) 人与人之间的交互是复杂的且难以捕捉的. 对于行人而言, 为了遵守一定的社会规范甚至规避危险, 在移动过程中也需要预测场景中其他行人的轨迹的能力, 从而动态的调整自己的路线, 但这种能力是难以建模的. (2) 行人轨迹预测是一个多模态问题. 基于行人过

① 基金项目: 国防基础科研项目 (JCKY2022605C006)
收稿时间: 2023-06-11; 修改时间: 2023-07-12; 采用时间: 2023-07-21; csa 在线出版时间: 2023-10-27
CNKI 网络首发时间: 2023-10-30

于的轨迹, 应当产生多条可能的未来轨迹, 因此在进行轨迹输出时, 需要在考虑合理性的同时也需要考虑多样性. (3) 轨迹输出的自洽性. 对于轨迹预测而言, 如果输出的轨迹会产生碰撞, 那么不但无法逻辑自洽, 而且在应用层面可能带来巨大的风险, 这通常是不能接受的.

目前, 国内外已经有多名学者对轨迹预测的行人交互问题而言, 进行了深入研究. 经典模型通过手工制作的能量函数来捕获人与人之间的交互[2], 成功实现了真实社会中的交互问题, 但这种方法往往无法在拥挤的空间中建立人群交互. 而随着深度神经网络研究的进展, 已经有越来越多的循环神经网络模型应用到了轨迹预测任务之中并展现出了广阔的前景与强大的能力. 基于 RNN 的方法通过其潜在状态来捕获行人的运动, 并通过合并行人的潜在状态来模拟人与人之间的交互作用[3,4]. 近年来, Transformer[5]网络强大的能力使其变得流行, 它抛弃了语言序列的顺序性质, 只用强大的自我注意机制来建模时间依赖关系. 与 RNN 相比, Transformer 架构的主要好处是使用自我注意显著改善了时间建模, 可以更好地捕捉行人之间潜在的互动[6–8]. 然而, 无论是利用 RNN 模型或是自注意力机制进行人与人交互建模, 往往都会忽略空间结构信息. 而图神经网络[9]的特殊形式, 让行人交互的空间结构的信息得到了直观且有效的利用, 在行人轨迹预测领域做出很多应用[10–12], 效果得到了良好的提升.

行人轨迹预测结果的碰撞问题一直是此领域挥之不去的难题. 对于轨迹预测结果而言, 若产生碰撞, 则会带来模型自洽问题甚至安全问题. 有的避免碰撞方法是对预测结果强行加上一个手工的约束函数[13], 这种物理约束的方法可能导致预测轨迹的生硬, 在合理性上稍显欠缺. 更多的深度学习模型通常通过让模型学习行人交互问题的建模[14,15]来解决碰撞问题, 这种方法通常通过数据驱动数据集的学习, 可以产生合理轨迹, 但当产生碰撞情景时却无法修复——因为模型一直在学习正确的社会交互, 不知道何为“错误情况”. 因此减少碰撞的难度在于碰撞样本的缺失和训练方法的选择.

基于以上难点, 本文提出一种融合安全性的轨迹预测模型 (safe spatio-temporal graph Transformer network for pedestrian trajectory prediction, S-STGTN), 如图 1. S-STGTN 是基于图注意力机制的神经网络模型, 通过图注意力机制来合理提取行人交互特征, 并提出一种无需手工标注的负样本生成方法, 通过对比表示学习来降低碰撞率以提高安全性.

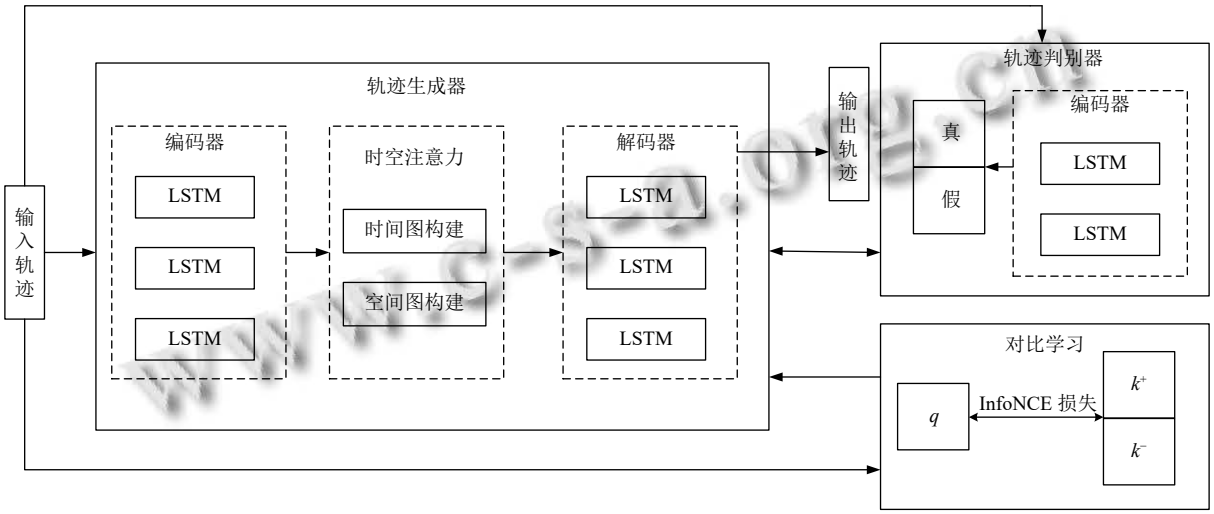

图 1 S-STGTN 算法整体框架

2 融合对比学习的时空图网络

本文的主要结构使用了编解码结构. 首先是使用图注意力网络来提取行人特征, 通过此方法可以在进行轨迹预测时更好的考虑到他人带来的影响. 而针对轨迹预测中的安全问题, 即尽量地避免碰撞, 本文使用了对比表示学习的方法, 来在无需手动标注的情况下生成负样本, 解决了实际生活中碰撞样本少的问题, 来告知模型何种情况是不能接受的. 此外, 由于轨迹预测是一个多模态问题, 本文使用对抗生成网络来生成多种轨迹.

2.1 问题定义

行人轨迹预测, 即根据过去一段时间内的轨迹来推断出未来一段时间的可能轨迹. 假设场景中有 N 个行人, 且行人过去轨迹定义为已 p_t^i, 行人未来预测轨迹定义为 y_t^i, 行人未来真实轨迹定义为 $\hat{y}_t^i$. 则本问题可以描述为已知行人 i 以及场景中其他行人 j 的历史时间段 1 到 T_{obs} 的轨迹 p_t^i, p_t^j 目标为预测他在未来时间段 $T_{\text{obs}}+1$ 到 T_{pre} 时间段内的轨迹 y_t^i. 其中:

$$p_t^i = \{(x_t^i, y_t^i) \mid t = 1,2,\cdots,T_{\text{obs}}\} \tag{1}$$

$$p_t^j = \{(x_t^j, y_t^j) \mid t = 1,2,\cdots,T_{\text{obs}}, j \in N, j \neq i\} \tag{2}$$

$$y_t^i = \{(x_t^i, y_t^i) \mid t = T_{\text{obs}}, T_{\text{obs}+1},\cdots,T_{\text{pre}}\} \tag{3}$$

$$\hat{y}_t^i = \{(\hat{x}_t^i, \hat{y}_t^i) \mid t = T_{\text{obs}}, T_{\text{obs}+1},\cdots,T_{\text{pre}}\} \tag{4}$$

2.2 行人时空图网络

行人自身的轨迹是预测其轨迹最重要的输入之一, 而 LSTM 已经被证实了能从轨迹中提取带有描述甚至预测行人未来轨迹的隐藏特征, 因此本文采用 LSTM 对单人的历史轨迹进行编码.

图注意力网络 (graph attention network, GAT)[9]是一种基于注意力机制的图神经网络, 应用非常广泛, 它可以用于许多不同的任务, 包括图分类、图分割、图推理、图生成等, 在计算机视觉、自然语言处理、生物信息学、社交网络分析等解决了诸多实际问题[16]. 本文利用图注意力网络的结构来建模场景中的行人网络, 并利用其注意力机制建模行人之间的交互, 有效模拟了真实社交场景下的未来轨迹输出. 与使用图卷积网络的 SGCN[12]等方法相比, 本文使用注意力机制对节点进行加权, 能够更准确地捕捉节点之间的依赖关系, 从而提供更精确的节点表示和预测能力.

行人特征提取模块利用轨迹数据作为输入, 分为两个支部分提取图特征向量提取, 分别为时间特征提取和交互特征提取. 图结构可以完好地保存空间信息, 而自注意力机制则能提取人与人之间的交互问题. 然后将两个解耦的时空向量进行融合, 最后进行预测轨迹的输出. 图 2 展示了本文行人交互特征的提取过程, 行人交互特征的提取分为上下两个模块, 分别是时间模块与空间模块.

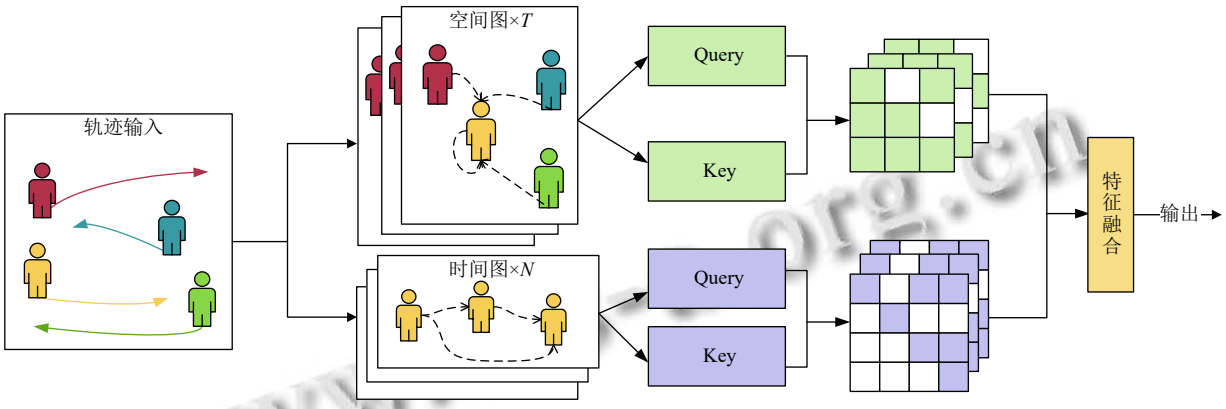

图 2 行人时空图构建

时间模块考虑到行人 i 自己本身的过去轨迹, 是一个典型的时间序列问题. 对于轨迹数据而言, 其中包含了在时间 T_{obs} 内场景中所有 N 个人的轨迹数据, 将这 N 个人的轨迹数据分别提取出来, 形成 N 个时间图 $G_{\text{tmp}}(i) = (V_i, U_i)$, $i \in 1,2,\cdots,N$, 由 V_i 和 U_i 构成, 其中 $V_i = \{v_i^t | i = 1,2,\cdots,T_{\text{obs}}\}$ 是 G_{tmp} 的顶点, 代表了行人 i 在观察时间 T_{obs} 内每秒的所有轨迹点的集合, $U_i = \{u_i^{m,n} | m, n = 1,2,\cdots,T_{\text{obs}}\}$ 是 G_{tmp} 的边, 代表了行人 i 在时间 m 和时间 n 的轨迹点的关系. 有了此图之后, 就可以对行人 i 的轨迹点做出图注意力机制, 具体公式为:

$$ATT\left(Q^i, K^i, V^i\right) = \frac{Softmax\left(Q^i K^{i\mathrm{T}}\right)}{\sqrt{d_k}} V^{i\mathrm{T}} \tag{5}$$

其中,

$$Q^i = f_Q\left(\left\{h_j^i\right\}_{j=1}^{T_{\text{obs}}}\right) \tag{6}$$

$$K^i = f_K\left(\left\{h_j^i\right\}_{j=1}^{T_{\text{obs}}}\right) \tag{7}$$

$$V^i = f_V\left(\left\{h_j^i\right\}_{j=1}^{T_{\text{obs}}}\right) \tag{8}$$

而空间图则保存了在某一时刻 t 场景中人与人的交互信息. 对于轨迹数据而言, 在每个时间点 t 场景中都会有 N 个人, 这 N 个人之间的空间信息就形成了 T_{obs} 个空间图 $G_{\text{spa}}(t)=(V^t,U^t), t\in 1,2,\cdots,T_{\text{obs}}$, 其中, $V^t=\{v_n^t|n=1,2,\cdots,N\}$, 为 t 时刻场景中的所有人, 代表了 $G_{\text{spa}}(t)$ 的顶点, $U^t=\{u_{i,j}^t|i,j=1,2,\cdots,N\}$ 为 t 时刻这些人之间的连接关系, 代表了 $G_{\text{spa}}(t)$ 的边. 有了此图之后, 就可以对行人 i 和行人 j 在 t 时刻的轨迹点做出图注意力机制, 具体公式为:

$$ATT\left(q^t,k^t,v^t\right)=\frac{Softmax\left(\left(q_i^{t\mathrm{T}}k_j^t\right)_{i,j=1:n}\right)}{\sqrt{d_k}}[v_i]_{i=1}^n \tag{9}$$

其中, $q_i^t=f_q\left(h_i^t\right)$, $k_i^t=f_k\left(h_i^t\right)$, $v_i^t=f_v\left(h_i^t\right)$.

经过时间图与空间图的自注意力机制后, 就可以对两个图中的节点分别进行更新, 具体公式为:

$$h_i'=ATT(i)+h_i \tag{10}$$

在得到更新后的时间图与空间图节点信息之后, 将他们连接起来通过全连接层, 得到融合了行人时空交互的新特征.

2.3 轨迹对比学习

长期以来的轨迹预测任务的指标大多聚焦在准确率即预测轨迹与真实轨迹的吻合率上, 但会忽视轨迹的碰撞率. 在基于数据驱动的轨迹训练任务中, 碰撞率总体来看保持在一个较低水准, 因为数据集中的轨迹点几乎不会产生碰撞, 模型自然也不会去“模仿”碰撞事件. 然而碰撞率是一个很重要的安全指标, 因此即使较小概率的碰撞率也是不可接受的, 因此有必要专门开拓一个模块来减少碰撞概率[17]. 为此, 本系统加入了避免碰撞模块.

避免碰撞模块的主要思想利用了对比学习. 对于一个模型而言, 想让它知道什么是最佳答案, 则在告知其正确答案之外, 还需要告诉它错误答案, 这就是对比学习的主要思想. 对比学习可以在无需额外人工标注的基础之上获得更多的“正样本”与“负样本”. 其中, 如何对一个普通数据集进行采样而得到正负样本是对比学习的核心任务之一. 在轨迹预测的避免碰撞模块中, 正样本即普通的数据集中的轨迹 (无碰撞样本), 而负样本则是碰撞样本, 设置负样本可以让模型在进行轨迹输出的时候, 尽量避免出现行人 i 与行人 j 在某一时间点处于相同的位置 (即碰撞), 希望通过此举降低轨迹输出的碰撞率, 从而达到逻辑自洽的效果. 但由于负样本很少出现, 且人工标注开销过大, 所以本文采用了一种自动生成正负样本的方法, 如图 3, 具体为以下公式:

$$\text{正样本抽取}: k_{t+t_0}^{i,+}=p_{t+t_0}^i+\epsilon \tag{11}$$

$$\text{负样本抽取}: k_{t+t_0}^{i,-}=p_{t+t_0}^j+\Delta p+\epsilon \tag{12}$$

其中, $k_{t+t_0}^{i,+}$ 为行人 i 在 $t+t_0$ 时刻的正样本, 只有一个. $p_{t+t_0}^i$ 为行人 i 在 $t+t_0$ 时刻的真实位置坐标, ϵ 为一个很小的常数, 作为随机噪声, 以防止过拟合. 而 $k_{t+t_0}^{i,-}$ 为行人 i 在 $t+t_0$ 时刻的负样本, 有多个. $p_{t+t_0}^j$ 为行人 i 的邻居 j 在 $t+t_0$ 时刻的真实位置坐标, 同样 ϵ 为随机噪声. 而 Δp 则是负样本采样范围参数, 可以确定为一定的数值. 例如, 如果仅考虑两个行人不会发生碰撞, 则可以将 Δp 设定为 0.3 m , 即形成了以位置为圆心, 半径为 0.3 的圆, 即图 3 中的碰撞区.

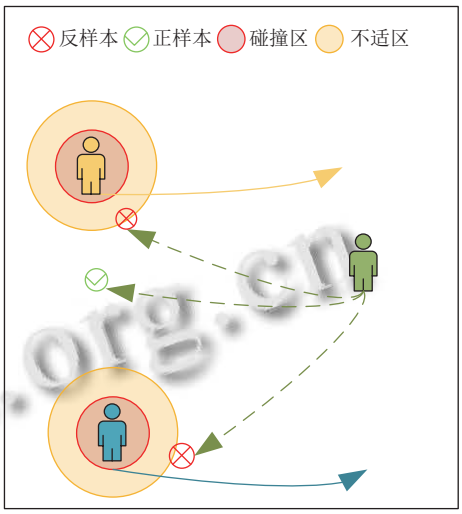

图 3 正负样本生成

同理, 如考虑到社交礼仪, 即正常情况下人并不会与其他路人贴着行走, 将 Δp 设置为 0.5 m, 则不但保证了不会发生碰撞, 也保证了社交距离. 因此形成了以行人位置为圆心, 半径为 0.5 的圆, 即图 3 中的不适区, 在此不适区内进行负样本采样, 可以让输出轨迹尽量远离此区域.

在得到了正负样本之后, 就可以引入对比学习的损失函数. 对比学习的损失函数通常包含一个度量函数与一个实际损失函数, 度量函数用于衡量数据样本与正负样本之间的距离, 实际损失函数用于对度量函

数计算出的进行距离约束. 普通对比学习[18]的的损失函数L_{pair}为:

$$L_{\text{pair}}=\begin{cases} D(q,k)^2, k\sim p^+(|q) \\ \max\left(0,m-D(q,k)^2\right), k\sim p^-(|q) \end{cases} \tag{13}$$

其中, $D(q,k)=\|q-k\|_2$为欧几里得距离. 这种最初版本的对比损失可以理解为, 让正样本之间的距离尽可能的近、且让负样本之间的距离尽量地大于指定边界 m. 但这种损失收敛的速度较为缓慢, 样本之间的距离也较远. 因此在这里, 使用噪声对比估计损失函数L_{NCE}[19], 具体的公式为:

$$L_{\text{NCE}}=-\log\frac{\exp(s(q,k^+))}{\exp(s(q,k^+))+\exp(s(q,k^-))} \tag{14}$$

其中, q 代表查询样本, k^+代表正样本, k^-代表负样本, 在轨迹预测任务中分别为预测行人 i 的过去的轨迹向量、预测行人 i 的未来的真实轨迹向量、行人 i 的邻居 j 的未来真实轨迹向量, 而 $s()$ 为度量函数, 这里可以设为余弦相似度. L_{NCE}的核心思想是学习数据样本与噪声样本之间的数据分布的区别, 可以理解为一个二分类问题, 训练一个分类器以学习数据样本与噪声样本之间的区别.

2.4 轨迹生成

在轨迹预测任务中, 由于人的轨迹的不确定性, 因此需要生成多条备选轨迹. 本系统的轨迹生成模块主要采用 InfoGAN 结构[20], 如图 4.

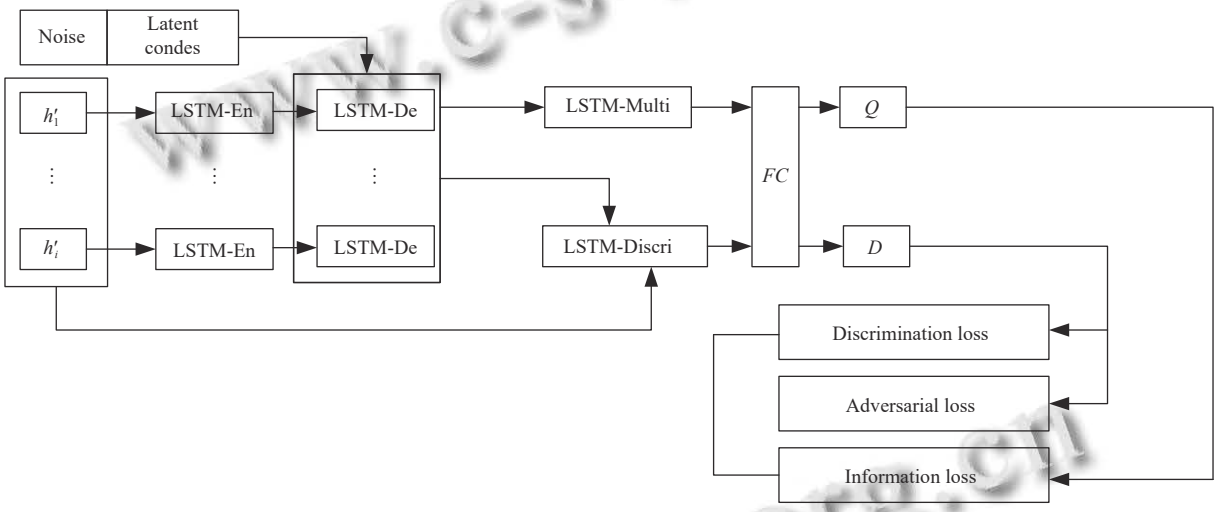

图 4 InfoGAN 框图

不同于普通的对抗生成网络, InfoGAN 加入了互信息损失来避免模式崩溃的问题. 本文中采取该结构可以有效地避免在生成不同的轨迹预测模块时, 生成轨迹过分相似的问题, 实现预测轨迹的多样化. 其中, 损失函数可以写为:

$$\min_G\max_D V(D,G)=E_{x\sim p_{\text{data}}(x)}[\log D(x)]+E_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{15}$$

其中, D 表示生成器, G 表示判别器. x 为来自真实样本的数据, 而 z 则是随机噪声. 通过这样的损失函数, 可以生成和真实样本很相近的数据, 即在轨迹预测任务中就是与真实未来轨迹相似的轨迹. 然而, GAN 在生成多模态轨迹时, 很容易引起模式崩溃.

InfoGAN 与 GAN 的区别主要在于输入与损失函数. 相比 GAN 的输入, InfoGAN 首先将生成器输入噪声 z 分为了 (z, c) 两个部分, 其中 z 和 GAN 的 z 并无不同, 均为随机噪声, 而 c 则可以理解为可解释的隐变量. 而关于损失函数, 则分为了两部分. 对于噪声 z 的损失与 GAN 相同, 而对于 c 则引入了新的互信息损失 $I(c;G(z,c))$, 因此, InfoGAN 的损失函数可以写为:

$$\min_G\max_D V_I(D,G)=V(D,G)+\lambda I(c;G(z,c)) \tag{16}$$

其中, $V(D,G)$为图 4 的生成器损失和对抗损失, $I(c;G(z,c))$为互信息损失. 互信息量用字母 I 表示, $I(X;Y)$ 就代表了 X 和 Y 之间的互信息量的大小.

根据图 4 可知, 生成器模块主要是由编码器-解码器结构构成, 解码器的输入为编码器的输出与随机噪声和可接受的隐变量, 输出为预测的轨迹. 而判别器模

块分为两个部分, 分别是 LSTM-Multi 和 LSTM-Discri. LSTM-Multi 的输入为生成器中解码器的输出, 用于生成更多不同的样本, LSTM-Discri 的输入为生成器中解码器的输出与真实轨迹向量, 用于鉴别收到样本的真假.

3 实验结果与分析

3.1 数据集

在本节中, 为了验证本文算法的有效性, 首先在行人轨迹预测的基准数据集 ETH[21]和 UCY[22]上进行测试. 这两个数据集是包含了大量的社会交互, 这两个数据集中加起来共有 1 536 名行人, 上千条真实轨迹, 包含行人绕开障碍物、单个行人与人群的相向而行、路口行人转弯等多种真实场景. 是行人轨迹预测常用的基准数据集. 其中, 又分别包含了 5 个人群数据集, ETH 包含 ETH 和 HOTEL 数据集, UCY 包含 ZARA1, ZARA2 和 UNIV 这 3 个数据集. 这在训练或评估阶段, 输入的是过去 3.2 s 的轨迹, 输出的是未来 4.8 s 的轨迹.

对于验证轨迹预测的准确性, 本文采用最终位移误差 (final displacement error, *FDE*) 和平均位移误差 (average displacement error, *ADE*), 出于安全因素的考量还引入了碰撞率 (collision rate, *CR*) 行人之间的预测轨迹之间发生碰撞率. 公式分别如下:

$$ADE = \frac{\sum_{p=1}^{N}\sum_{t=1}^{T_f} y_p^t - \hat{y}_p^t}{N \times T_f} \tag{17}$$

$$FDE = \frac{\sum_{p=1}^{N} y_p^t - \hat{y}_p^t}{N} \tag{18}$$

$$CR = \frac{\sum_{p=1}^{N}\sum_{t=1}^{T_f} y_p^t - \hat{y}_p^t}{N \times T_f} \left| \left(y_p^t - \hat{y}_p^t \right) < dis \right. \tag{19}$$

3.2 实验结果

定量评估是通过运用数学模型对所要进行分析的对象在各项关键性能指标上的评估分析. 表 1 中列出了 ETH 和 UCY 数据集上以 *ADE*、*FDE* 和 *CR* 为指标模型的评估结果, *ADE* 与 *FDE* 用斜线隔开.

表 1 各模型在 ETH/UCY 测试结果对比

模型	ETH		HOTEL		UNIV		ZARA1		ZARA2		平均值	
	ADE/FDE	*CR*	*ADE/FDE*	*CR*	*ADE/FDE*	*CR*	*ADE/FDE*	*CR*	*ADE/FDE*	*CR*	*ADE/FDE*	*CR*
Social LSTM	1.09/2.35	3.27	0.79/1.76	5.22	0.67/1.40	10.2	0.47/1.00	3.65	0.56/1. 17	10.3	0.72/1.54	6.52
Social GAN	0.87/1.62	2.24	0.67/1.37	3.98	0.76/1.52	10.54	0.35/0.68	3.12	0.42/0.84	9.04	0.61/1.21	5.78
GAT	0.68/1.29	2.11	0.68/1.40	4.02	0.57/1.29	9.88	0.29/0.60	2.81	0.37/0.75	8.11	0.52/1.07	5.38
Social-STGCNN	0.64/1. 11	1.33	0.49/0.85	3.82	0.44/0.79	9.11	0.34/0.53	2.27	0.30/0.48	6.86	0.44/0.75	4.69
S-STGTN	0.55/1.09	0.51	0.27/0.52	2.98	0.42/1.01	6.23	0.32/0.61	1.01	0.44/1.22	3.33	0.40/0.87	2.98

表 1 是本文的定量分析结果, 使用了多个经典模型或前沿模型进行对比. Social LSTM 是最先提出“社交池”的模型, 利用池化层来通过邻居节点的特征捕捉社会交互, 轨迹更加合理合理, 也带来了更好的结果, 从而表现出了“社会交互”在行人轨迹预测领域的重要性. Social GAN[23]是本文的轨迹生成器参考的模型, 考虑到行人轨迹的多模态性, 只输出一条轨迹无法满足真实场景的需求, 因此 Social GAN 输出了多条轨迹, 结果证明这更有利于得到准确轨迹. GAT 是本文在提取时空特征时使用的模型, 通过图神经网络与自注意力机制结合, 有效地得到了社会交互特征, 得到良好结果. 而本文的 S-STGTN 模型的结果显然处于第 1 梯队, 值得注意的是, 虽然 STGCNN[7]最终的结果与本文模型相差不大, 但 S-STGTN 预测的结果的碰撞率得到大幅度的降低, 使轨迹更加的合理, 在将预测的轨迹应用到实际场景中也更加的安全.

在表 2 中, S-STGTN 为本文的模型, S-STGTN-CL 为消融对比表示学习模块之后的模型, 揭示了对比学习模块的重要作用. 可以看出, 对比表示学习模块在降低碰撞率的功能上十分有效, 在 5 个行人数据集上分别提升了 58.1%, 20.9%, 31.3%, 54.2%, 50.4%, 大幅度提升了安全性. 同时, 对于 *ADE* 和 *FDE* 两个指标也有提升作用.

表 2 对比学习模块消融对比

模型	ETH		HOTEL		UNIV		ZARA1		ZARA2	
	ADE/FDE	*CR*	*ADE/FDE*	*CR*	*ADE/FDE*	*CR*	*ADE/FDE*	*CR*	*ADE/FDE*	*CR*
S-STGTN	0.66/1.13	0.51	0.41/0.76	2.98	0.47/1.01	6.23	0.34/0.61	1.01	0.31/0.69	3.33
S-STGTN-CL	0.67/1.13	1.22	0.44/0.78	3.77	0.46/1.01	9.07	0.41/0.63	2.21	0.34/0.71	6.71

而考虑到算法的稳定性与可靠性, 当所需的输出时长增加时, 算法的 *ADE* 与 *FDE* 误差应该控制在一个合理的范围内.

图 5 和图 6 分别表现出了当预测时长从 0.8 s 到 4.2 s 时, 各个算法的误差趋势. 可以看出, 本文提出的 S-STGTN 算法在输出时间增加时, *ADE* 与 *FDE* 的斜率相比其他流行的算法增加幅度较小, 说明了本算法的稳定性. 本文提出了一种融合安全性考虑的轨迹预测模型 S-STGTN. 本模型基于图注意力网络, 并结合对比表示学习降低碰撞率, 使轨迹更加合理和安全. 经过在公开数据集 ETH 和 UCY 上的测试, S-STGTN 模型取得了较好的预测效果, 并且与其他轨迹预测模型相比, 本模型有效降低了碰撞率, 且算法稳定性得到提升.

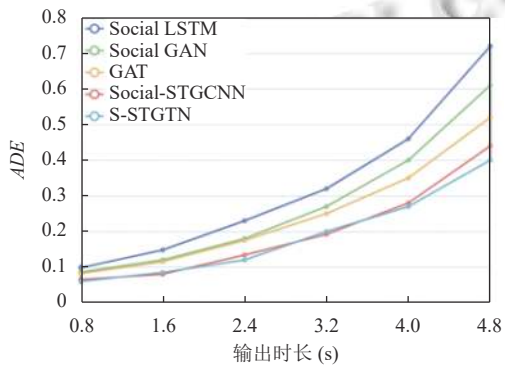

图 5 *ADE* 与输出时长关系

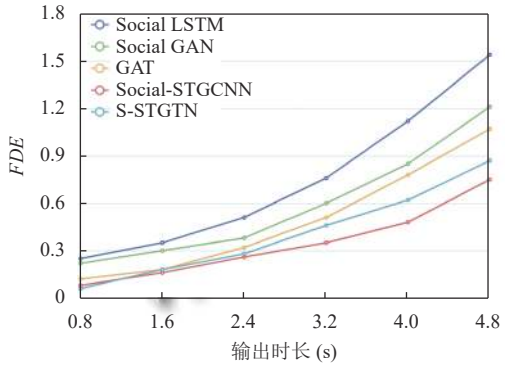

图 6 *FDE* 与输出时长关系

参考文献

1 Dai MM, Wang JX, Yin GD, *et al*. Dynamic output-feedback robust control for vehicle path tracking considering different human drivers' characteristics. Proceedings of the 36th Chinese Control Conference (CCC). Dalian: IEEE, 2017. 9407–9412.

2 Yi S, Li HS, Wang XG. Understanding pedestrian behaviors from stationary crowd groups. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 3488–3496.

3 Alahi A, Goel K, Ramanathan V, *et al*. Social LSTM: Human trajectory prediction in crowded spaces. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 961–971.

4 Yang ZF, Liu D, Ma L. Vehicle trajectory prediction based on LSTM network. Proceedings of the 2022 International Conference on Artificial Intelligence and Computer Information Technology (AICIT). Yichang: IEEE, 2022. 1–4.

5 Vaswani A, Shazeer N, Parmar N, *et al*. Attention is all you need. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6000–6010.

6 Giuliari F, Hasan I, Cristani M, *et al*. Transformer networks for trajectory forecasting. Proceedings of the 25th International Conference on Pattern Recognition (ICPR). Milan: IEEE, 2021. 10335–10342.

7 Mohamed A, Qian K, Elhoseiny M, *et al*. Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction. Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 14412–14420.

8 Liu YC, Zhang JH, Fang LJ, *et al*. Multimodal motion prediction with stacked transformers. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 7573–7582.

9 Velickovic P, Cucurull G, Casanova A, *et al*. Graph attention networks. Proceedings of the 6th International Conference on Learning Representations. Vancouver: ICLR, 2018.

10 Yu CJ, Ma X, Ren JW, *et al*. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 507–523.

11 Gilles T, Sabatini S, Tsishkou D, *et al*. Thomas: Trajectory heatmap output with learned multi-agent sampling. Proceedings of the 10th International Conference on Learning Representations. ICLR, 2022.

12 Shi LS, Wang L, Long CJ, *et al*. SGCN: Sparse graph convolution network for pedestrian trajectory prediction. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021. 8994–9003.

13 Tay MKC, Laugier C. Modelling smooth paths using Gaussian processes. In: Laugier C, Siegwart R, eds. Field and Service Robotics: Results of the 6th International Conference. Berlin: Springer, 2008. 381–390.

14 Sadeghian A, Kosaraju V, Sadeghian A, *et al*. SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 1349–1358.

15 Mangalam K, Girase H, Agarwal S, *et al*. It is not the journey but the destination: Endpoint conditioned trajectory prediction. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 759–776.

16 Xia F, Sun K, Yu S, *et al*. Graph learning: A survey. IEEE Transactions on Artificial Intelligence, 2021, 2(2): 109–127. [doi: 10.1109/TAI.2021.3076021]

17 Liu YJ, Yan Q, Alahi A. Social NCE: Contrastive learning of socially-aware motion representations. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 15098–15109.

18 Chopra S, Hadsell R, LeCun Y. Learning a similarity metric discriminatively, with application to face verification. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005). San Diego: IEEE, 2005. 539–546.

19 Jozefowicz R, Vinyals O, Schuster M, *et al*. Exploring the limits of language modeling. arXiv:1602.02410, 2016.

20 Chen X, Duan Y, Houthooft R, *et al*. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets. Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 2180–2188.

21 Pellegrini S, Ess A, Schindler K, *et al*. You'll never walk alone: Modeling social behavior for multi-target tracking. Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto: IEEE, 2009. 261–268.

22 Lerner A, Chrysanthou Y, Lischinski D. Crowds by example. Computer Graphics Forum, 2007, 26(3): 655–664. [doi: 10.1111/j.1467-8659.2007.01089.x]

23 Gupta A, Johnson J, Fei-Fei L, *et al*. Social GAN: Socially acceptable trajectories with generative adversarial networks. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 2255–2264.

(校对责编: 孙君艳)