基于 GAN 的社会和场景感知行人轨迹预测

李 兰 张 洁 刘 杰 胡克勇
(青岛理工大学信息与控制工程学院 山东 青岛 266520)

摘 要 针对状态精细化长短期记忆网络(SR-LSTM)未考虑周围物理场景对行人轨迹预测的影响,且无法生成多种可能性样本的问题,提出一种基于生成对抗网络(Generative Adversarial Networks,GAN)的社会和场景感知行人轨迹预测模型。此模型引入社会注意力及语义池机制,社会注意力建模邻人当前重要意图,以从相邻行人中选择重要的信息,语义池定义物理场景语义并学习与行人轨迹相关性。由于 GAN 易发生模式崩溃和下降,采用 Info-GAN 进行训练生成更真实的样本。在 ETH 和 UYC 两个数据集上进行实验,结果表明该方法较于 SR-LSTM,ADE 降低 8.9 百分点,FDE 降低 12.8 百分点,且可生成更多合理的样本。

关键词 行人轨迹预测 生成对抗网络 注意力机制 语义池机制 长短期记忆网络

中图分类号 TP3 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2024.06.011

SOCIAL AND SCENE AWARENESS PEDESTRIAN TRAJECTORY PREDICTION BASED ON GAN

Li Lan Zhang Jie Liu Jie Hu Keyong
(*School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266520, Shandong, China*)

Abstract In order to solve the problem that state refinement for LSTM (SR-LSTM) does not consider the influence of surrounding physical scenes on pedestrian trajectory prediction, and can not generate a variety of possible samples, a social and scene awareness pedestrian trajectory prediction model based on GAN is proposed. This model introduced social attention and semantic pool mechanism, and social attention mechanism was used to model the current important intention of adjacent pedestrians in order to select important information from adjacent pedestrians. Semantic pools defined the semantics of physical scenes and learn their correlation with pedestrian tracks. Because GAN was prone to mode collapse and decline, Info-GAN was used for antagonistic training to generate more real samples. The experiments on ETH and UYC data sets show that compared with SR-LSTM, ADE of this method is 8.9 percentage points lower and FDE is 12.8 percentage points lower, and more reasonable samples can be generated.

Keywords Pedestrian trajectory prediction Generative adversarial networks Attention mechanism Semantic pool mechanism Long and short-term memory

0 引 言

现阶段,许多应用程序都在大量使用有关行人运动的数据分析,某些情况下,它们可以通过在线方式预测下一步行人的行动,并推断出其短期或中期意图,以保证具有实时决策功能的监控系统在关键决策时触发早期警报或采取预防措施。例如,自动驾驶情况下,推断周围行人的意图,汽车才能合理规划其路径。因此,行人轨迹预测对避免相互碰撞至关重要。

然而,这个问题的解决是极其复杂的。行人的运动轨迹会受到多种因素的影响,如场景拓扑结构、行人行为以及人与人之间的交互作用等。尽管使用循环神经网络进行时间序列预测取得了有效的结果,但仍存

收稿日期:2020-12-31。国家自然科学基金项目(61902205)。**李兰**,教授,主研领域:数据挖掘,模式识别。**张洁**,硕士生。**刘杰**,硕士生。**胡克勇**,副教授。

在许多问题。例如,这种情况下,数据驱动的方法通常不知道周围物理元素的影响,而这些元素是行人方向变化的主要因素。

本文提出一种基于 GAN[1] 的社会和场景感知行人轨迹预测模型。此模型引入注意力机制对行人社会关系进行建模,并结合语义池机制实现与周围空间的交互,最后通过 Info-GAN[2] 训练得到行人预测轨迹。结果表明,相较于传统方法,模型在准确性上得到了有效提高,并可生成多条合理轨迹。

1　相关工作

此前,在仿真和群组动画等领域,已经引入了许多解释人体运动的闭合形式的数学模型。基于计算几何的方法[3] 通常在碰撞极限处产生最佳运动,而不是类似人类的真实运动。基于优化的方法[4] 通过优化目标函数的参数,以涵盖与运动相关的信息。

由于行人运动的复杂性,手工制定的数学模型无法适应广泛的环境,而基于机器学习的技术则得益于大量的人体运动数据集。随着基于神经网络的机器学习的出现,人们开始使用循环神经网络或更有效的变体(如 LSTM[5]、GAN 等)来进行预测任务。现有的方法中,对行人轨迹的预测,主要考虑以下两个方面的因素:

1) 行人间的社会交互。基于 LSTM 网络的研究主要依赖于 Alahi 等[6] 提出的 S-LSTM 模型。该模型使用一个社会汇聚池(Social-Pooling)合并附近行人的隐藏状态,以对行人社会关系进行建模。2019 年,Pu 等[7] 在 S-LSTM 行人注意的基础上提出一个运动门来有效地提取社会影响。继生成对抗网络(GAN)在其他领域的广泛应用,Gupta 等[8] 提出了通过将 GAN 输入的随机向量与其他行人轨迹的隐藏特征相结合来处理行人之间交互的 Social-GAN 模型。2020 年,Mangalam 等[9] 提出了预测端点条件网络(PECNet),通过推断远程轨迹端点来辅助远程多模态轨迹预测。

2) 物理场景约束。基于行人交互的模型并没有考虑附近的因素(如长凳或树木等)对轨迹的影响,有时这些物理环境因素才是行人改变方向的主要原因。Sadeghian 等[10] 通过考虑过去的交叉区域和语义上下文,使用 GAN 来预测的社交和上下文感知位置。2019 年,Haddad 等[11] 将图注意力机制(Graph Attention, GAT)[12] 引入轨迹预测,利用 LSTM 处理人-人、人-物理场景之间的时空交互特征,并在上层利用注意力机制[13] 进行整合。Lisotto 等[14] 基于对先前交叉区域的观察,添加导航张量,利用场景的先前信息来区分同样可能的预测位置。2020 年,Mohamed 等[15] 提出社会时空图卷积神经网络(Social-STGCNN),它通过将交互建模为一个图来代替聚合的方法预测行人轨迹。Liang 等[16] 基于真实世界的轨迹在一个 3D 模拟器创建了一个新的数据集,预测行人在多个可能的未来路径上的分布。

2　模型结构

2.1　问题定义

行人轨迹预测可以表述为给定一组静态物体 O 和一组行人 V 及其轨迹 X,在时间步长 $t=1,2,\cdots,T_{\text{obs}}$ 时,预测时间步长 $t=T_{\text{obs}}+1,\cdots,T_{\text{pred}}$ 的轨迹 $\hat{Y}$。

其中,第 i 个行人在时间步长 $t=1,2,\cdots,T_{\text{obs}}$ 的历史轨迹 X_i^t 表示为:

$$X_i^t=\{(x_i^t,y_i^t)\mid t=1,2,\cdots,T_{\text{obs}}\}$$

时间步长 $t+1=T_{\text{obs}}+1,\cdots,T_{\text{pred}}$ 预测的未来轨迹 Y_i^{t+1} 表示为:

$$\hat{Y}_i^{t+1}=\{(\hat{x}_i^{t+1},\hat{y}_i^{t+1})\mid t+1=T_{\text{obs}}+1,\cdots,T_{\text{pred}}\}$$

本文中行人的真实轨迹使用 Y 表示,第 i 个行人在时间步 $t=T_{\text{obs}}+1,\cdots,T_{\text{pred}}$ 的真实轨迹 Y_i^{t+1} 表示为:

$$Y_i^{t+1}=\{(x_i^{t+1},y_i^{t+1})\mid t+1=T_{\text{obs}}+1,\cdots,T_{\text{pred}}\}$$

下面用 $i,j\in\{1,2,\cdots,N\}$ 来表示行人,其中 N 是行人总数。对于行人 i,其邻域用 $N(i)$ 表示,相邻行人 $j\in N(i)$。

2.2　整体架构

如图 1 所示,本文模型由以下三个关键部分组成:

1) 社会注意力模块。本模块对行人进行特征提取,并突出显示这些特征中对行人轨迹产生影响的最重要信息。首先使用编码器(一个 LSTM 层,表示为 LSTM-E)提取行人特征,之后通过注意力模块学习行人之间的交互作用及其对行人未来路径的影响。

2) 语义池。对于场景中的物理约束,我们定义语义池机制,使用场景语义分割来识别不可跨越的区域。

3) 基于 LSTM 的 Info-GAN 网络。此网络包含一个生成器和一个判别器。生成器学习行人间的社会交互及物理场景对行人的影响,并将结果作为解码器(一个 LSTM 层,记为 LSTM-D)的输入,来为每个行人生成一系列合理的未来路径。判别器被用于提高生成器的性能,迫使其生成更真实的样本(轨迹)。

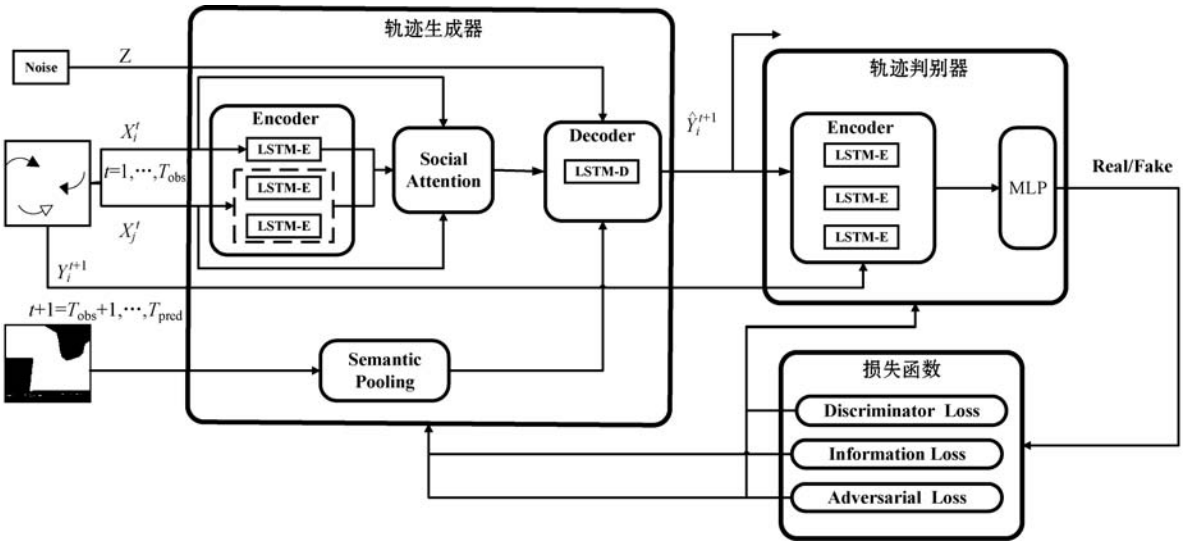


图 1 基于 GAN 的社会和场景感知轨迹预测模型

2.3 社会注意力模块

2.3.1 特征提取

该模块对每个行人的特征进行提取,并学习他们之间的社交规则。特征提取公式如式(1)、式(2)所示。使用单个 MLP 层嵌入行人位置,获得固定长度的嵌入向量 $\boldsymbol{e}_i^t$,并作为时间 t 时编码器 LSTM-E 的输入,来学习行人沿轨迹的时间特征。

$$\boldsymbol{e}_i^t = \boldsymbol{\Phi}_e(x_i^t, y_i^t; W_e) \tag{1}$$

$$h_{ei}^t = \mathrm{LSTM}_{ei}(\boldsymbol{e}_i^t, h_{ei}^{t-1}; W_{\mathrm{lstme}}) \tag{2}$$

式中:$\boldsymbol{\Phi}(\cdot)$是 ReLU 非线性嵌入函数;W_e是嵌入权重,编码器 LSTM-E 的权重 W_{lstme}在所有行人之间共享。

2.3.2 社会注意力

提取到的行人特征均负责不同的运动模式,包括行走的方向和速度。以往主要通过分析相邻行人的特征来预测行人轨迹,然而,相邻行人的特征(运动模式)对于预测行人的轨迹并不同等重要。如图 2 所示,左边行人在 t 时刻的轨迹意图由不同颜色深度的实线表示(神经元捕获这些运动特征),但右边的行显然会更注意与其产生碰撞情况的特征,这种潜在的注意力取决于行人间的相对位置和成对运动。为了自适应地关注最有用的邻人信息并进行状态传递,我们通过行人间的相对位置计算注意力系数,并通过行人的成对运动进行特征选择,两者结合形成具有社会意识的信息选择机制。

图 2 行人潜在注意力示意图

行人 i 对 j 的注意力系数 $\alpha_{i,j}^t$利用行人相对位置和隐藏状态进行计算,作用是为了强调重要的邻居,并控制邻里信息的数量。公式如下:

$$r_{i,j}^t = \boldsymbol{\Phi}_r(x_i^t - x_j^t, y_i^t - y_j^t; W_r) \tag{3}$$

$$U_{i,j}^t = W_u[r_{i,j}^t; h_{ei}^t, h_{ej}^t] \tag{4}$$

$$\alpha_{i,j}^t = \frac{e^{U_{i,j}^t}}{\sum_{k \in N(i)} e^{U_{i,k}^t}} \tag{5}$$

式中:$r_{i,j}^t$为相对空间位置,是进行特征选择的重要因素,通过嵌入函数 $\boldsymbol{\Phi}_r$ 嵌入;W_r表示嵌入函数 $\boldsymbol{\Phi}_r$的参数。

特征选择项 $g_{i,j}^t$综合考虑行人 i 和 j 的运动及其相对空间位置,根据 $r_{i,j}^t$、h_{ei}^t、h_{ej}^t的组合计算,计算公式如下:

$$g_{i,j}^t = \sigma(W_g[r_{i,j}^t; h_{ei}^t, h_{ej}^t] + b^g) \tag{6}$$

式中:δ 表示 sigmoid 函数;W^g、b^g表示参数。通过 $g_{i,j}^t$作用于行人 j 的隐藏状态 h_{ej}^t,使用元素乘积 $g_{i,j}^t \odot h_{ej}^t$选择特征。

则对于行人 i,在其邻域 $N(i)$ 内的社会注意力 a_i^t如式(7)所示。a_i^t结合注意力系数和特征选择项,共同从相邻行人中选择重要信息进行信息传递。

$$a_i^t = \sum_{j \in N(i)} W_a \alpha_{i,j}^t (g_{i,j}^t \odot h_{ej}^t) \tag{7}$$

2.4 语义池

由于多种原因,人们有可能会表现出方向的改变。例如,他们可能接近一个固定障碍物,但该障碍物必须绕行。为此,我们定义语义张量捕捉行人特殊的动作出现与周围空间的语义相关性。由于我们的数据集不提供任何的语义注释建模人与空间的交互,故定义语义类 C = {grass, building, barrier, bench, car, road, sidewalk},并采用 one-hot 编码来表示图像的语义。例如一个像素代表草,根据语义类 C,它的位置 j 用 $V_j \in R^7$ =

{10…0}表示。为行人 i 定义 $N(i)\times N(i)\times L$ 的张量 $\boldsymbol{S}_i^t$,并嵌入语义向量 $\boldsymbol{s}_i^t$ 中,如下所示:

$$\boldsymbol{S}_i^t(m,n) = \frac{1}{|S_{mn}|}\sum_{j\in S_{mn}} V_j \tag{8}$$

$$\boldsymbol{s}_i^t = \boldsymbol{\Phi}_s(\boldsymbol{S}_i^t;W^s) \tag{9}$$

式中:$\boldsymbol{V}_j$表示位置 j 的语义向量,S_{mn}表示第 i 个行人在 (m,n)单元格内的位置。$|S_{mn}|$是(m,n)单元格内的位置数。即对于每个单元,我们提取出每个语义类的出现频率。式(9)中,$\boldsymbol{\Phi}_s$表示 ReLU 激活函数,W_s表示权重。

2.5　网络结构

本文的 Info-GAN 网络观察 i 邻域 $N(i)$内所有行人历史时间步 $t=1,2,\cdots,T_{\text{obs}}$的轨迹及物理场景,学习训练数据中轨迹的分布和物理约束,生成独立的随机轨迹样本。如图 1 所示,它将 N 个行人的历史轨迹 X_i^t 和从固定分布 P_z中采样的随机向量 $\boldsymbol{z}$ 作为输入,在下一个时间步 $t+1=T_{\text{obs}}+1,\cdots,T_{\text{pred}}$中为行人 i 生成预测轨迹 $\hat{Y}_i^{t+1}$的合理分布。

一个 Info-GAN 网络包含在训练阶段两个相互对立的成分——轨迹生成器 G 和轨迹判别器 D。轨迹判别器 D 被训练成从真实样本中检测出假样本,而轨迹生成器 G 生成新的假样本来欺骗 D 并混淆其预测。对于行人 i,GAN 对其历史轨迹 $\boldsymbol{X}_i^t$及相邻行人的历史轨迹 $\boldsymbol{X}_j^t$进行编码,生成器使用噪声向量 $\boldsymbol{z}$ 生成预测轨迹。

2.5.1　生成器

与现有的轨迹生成器[10-11]类似,如图 1 轨迹生成器所示,此网络的生成器包含编码器、社会注意力模块、语义池模块和解码器四部分。编码器 LSTM-E 学习行人的轨迹并将结果传入社会注意力模块进行特征选择,语义池学习场景对行人的物理约束,最后将行人社会注意力 a_i^t、语义向量 $\boldsymbol{s}_i^t$及随机噪声 $\boldsymbol{z}$ 连接作为解码器 LSTM-D 的输入,生成 t 时刻的隐藏状态 h_{di}^t。

$$C_i^t = [a_i^t, s_i^t, z] \tag{10}$$

$$h_{di}^t = LSTM_{di}(c_i^t, h_{di}^{t-1}; W_{\text{lstmd}}) \tag{11}$$

对于轨迹样本的预测,通过将 LSTM-D 的输出编码成为二维高斯分布的参数$(\mu_i^{t+1},\sigma_i^{t+1},\rho_i^{t+1})$,预测得到的新坐标通过此高斯分布给出。

$$[\mu_i^{t+1},\sigma_i^{t+1},\rho_i^{t+1}] = W_l h_{di}^t \tag{12}$$

$$(\hat{x}_i^{t+1},\hat{y}_i^{t+1}) \sim N((x,y);\mu_i^{t+1},\sigma_i^{t+1},\rho_i^{t+1}) \tag{13}$$

2.5.2　判别器

判别器由一个单独的编码器和一个 MLP 层组成。编码器 LSTM-D 将行人的真实轨迹或生成器生成的预测轨迹作为输入,生成隐藏状态 $h_{dis,i}^t$,编码器最后一个隐藏状态使用 MLP 层来获得分类分数,即判别轨迹为真/假。

$$h_{dis,i}^t = LSTM_{d\_dis}(\hat{Y}_i^{t+1}, h_{dis,i}^{t-1}; W_{\text{lstm\_dis}}) \tag{14}$$

$$\text{score} = \text{MLP}(h_{dis,i}^t; W^m) \tag{15}$$

式中:score 为分类结果,$W_{\text{lstm\_dis}}$、W_m分别为 LSTM-D 和 MLP 层的权重。

2.5.3　损失函数

已知 GAN 训练是困难的,因为它可能不收敛,当生成器和判别器之间存在不平衡时,会出现消失梯度,或者可能受到模式崩溃的影响。在预测行人运动时,避免模式崩溃是至关重要的,因为它可能导致灾难性的决策。

不同于文献[10-11]的随机预测方法,我们的 GAN 训练有两大特点:1)由于 L2 损失项对生成样本多样性的不良影响,我们的损失函数不使用该项。2)使用 Info-GAN 体系结构,正如实验结果所展示的,与其他 GAN 相比,它在避免模式崩溃问题上具有很大改进。

Info-GAN 利用生成器学习数据分离和可解释的表示。其训练是通过添加新的隐变量 c,使 c 与生成的样本具有较高的互信息来完成的。这就需要在原始 GAN 损失函数 $V(D,G)$的基础上,再训练另一个子网络 $Q(c\mid\hat{Y}_i^{t+1})$(具有参数 θ_Q)来评估生成预测轨迹 $\hat{Y}_i^{t+1}$的可能性 $P(c\mid\hat{Y}_i^{t+1})$,其损失函数被定义为如下形式,λ 为超参数。

$$\begin{aligned}\min_{G,Q}\max_{D} V_I(D,G,Q) &= V(D,G) - \lambda L_I(G,Q) = \\ &E_{P_{\text{data}}(X_i^t,Y_i^{t+1})}[\log_2 D(Y_i^{t+1}\mid X_i^t;\theta_D)] + \\ &E_{P_z(z)}[\log_2(1-D(G(z\mid X_i^t,X_j^t;\theta_D);\theta_D))] - \\ &\lambda E_{P(c),P_z(Z)}[\log_2 Q(c\mid G(z\mid X_i^t,X_j^t;\theta_G);\theta_Q)]\end{aligned} \tag{16}$$

3　实验及结果分析

实验采用 Ubuntu 16.04 操作系统,NVIDIA GTX 1080Ti GPU,以及 Python 3.5、PyTorch 0.4、Cuda 9.0 的深度学习框架。

3.1　数据集及评价指标

使用两个公开的数据集 ETH[17]和 UCY[18]进行实验。这些数据集由真实世界的人类轨迹组成,以 2.5 帧每秒的速率手动标记。ETH 包含 ETH、Hotel 两个实验数据集,UCY 包含 ZARA1、ZARA2 和 Univ 三个实验数据集。采用交叉验证方法,每个数据集分成 5 个子集,在 4 个集合上训练和验证,并在剩余集合上进行测试。

本文使用两组代表性模型作为对比基准：

1）确定性预测模型，每次观测生成一条轨迹。

Linear：线性回归器，可以采用最小化 MSE 来优化回归器的参数。

LSTM：采用 vanilla LSTM 编解码器模型来预测每个行人的顺序。

S-LSTM：将每个行人关联到一个 LSTM 单元，并通过社会汇集机制收集相邻行人的隐藏状态来进行预测[6]。

SR-LSTM：在 S-LSTM 模型基础上引入社会感知信息选择机制，自适应地从邻域中提取重要的社会信息[7]。

2）随机预测模型，从预测分布中生成一组样本。

Social GAN：基于 GAN 的预测[10]。

SoPhie：在 GAN 预测中实现了社会和物理注意机制[11]。

S-Ways：基于 Info-GAN 的训练及预测[19]。

Social-BiGAT：基于图的生成对抗网络[20]。

RSBG：通过递归社会行为图以及引入 GCNs 进行轨迹预测[21]。

与之前的工作[8,11]类似，使用以下指标评估所提出的模型：

1）平均位移误差（Average Displacement Error，ADE），即在所有时间步长内，真实轨迹位置与预测位置之间的 L_2 距离：

$$A_{\mathrm{DE}} = \frac{1}{T}\sum_{t=1}^{T} \left\| Y_i^{t+1} - \hat{Y}_i^{t+1} \right\| \tag{17}$$

2）最终位移误差（Final Displacement Error，FDE），即真实轨迹位置与预测最终位置之间的 L_2 距离：

$$F_{\mathrm{DE}} = \left\| Y_i^{t+1} - \hat{Y}_i^{t+1} \right\| \tag{18}$$

3.2 模型精度分析

为评估随机模型，使用 KV-K 的方法，即训练和测试阶段对单个行人进行 K 次迭代，生成 K 个样本，取最接近真实轨迹的样本进行评估。K 表示模型对行人生成的轨迹条数，K 值越大，模型预测的情况越多，准确性越高，但同时会影响模型速率，为此，需选择一个合理的 K 值，使模型准确性和速率达到最优平衡。本文参考文献[10]中的工作，在 1V-K 及 KV-K 方法下，选取一些显著性 K 值，在 ETH、Hotel 和 ZARA1 三个数据集中进行实验，比较多个 K 值下单个行人的 ADE 值，得到合理的 K 值。设置 K 分别为 1、10、15、20、60、100，根据图 3 的比较结果可知，$K=20$ 时，两种性能达到最佳平衡。其中，1V-K 表示训练阶段对单个行人进行 1 次迭代，生成 1 个样本，测试阶段进行 K 次迭代，生成 K 个样本。

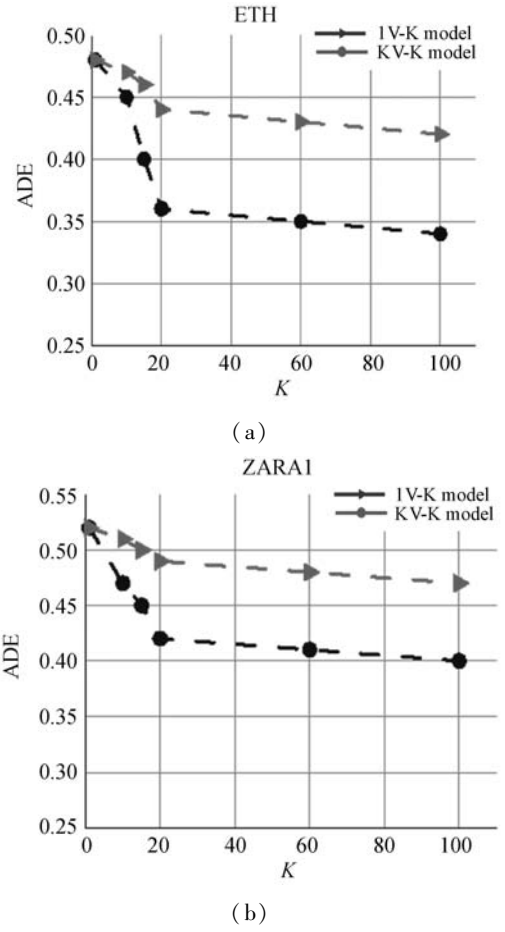


图 3 不同 K 值下的 ADE 结果比较

实验观察 $T_{\mathrm{obs}}=8$ 帧（3.2 秒）的轨迹，并预测 $T_{\mathrm{pred}}=12$ 帧（4.8 秒）时 ADE 和 FDE 的期望值，各模型对比结果如表 1 所示。据表 1 可知，本文提出的模型在 ETH 和 Hotel 数据集上预测指标最好，且总体预测效果优于表中其他模型。线性模型由于无法对不同行人之间复杂的社会交互进行有效建模，在整个轨迹预测任务中表现最差。S-LSTM 仅简单汇集相邻行人的隐藏状态，对于 LSTM 有相对提高，但效果仍较差。相对于 SR-LSTM，本文方法考虑物理场景的限制，且从生成的角度来处理问题，从而对其进行了改进。在 Hotel 数据集下，酒店场景包含更多的物理静态场景，FDE 实现了显著改进。S-way 从欧氏距离和方位角等方面考虑行人注意力，但未结合物理场景建模，总体效果较弱一些。Social-BIGAT 只将循环 GAN 用在了输入噪声 $\boldsymbol{z}$ 与生成轨迹之间，性能有所降低。结果表明，使用本文模型可以显著降低 ETH 和 Hotel 实验的预测误差，但在 ZARA 数据集中却没有，这可能是由于 ZARA 数据集中行人的道路宽度明显小于酒店和 ETH 场景，因此轨迹变化较小。本文模型本质上倾向于生成多个合理的样本，因此在更复杂的场景和非线性轨迹下可以获得良好的性能。

表 1　基于 ETH 和 UCY 数据集的模型对比

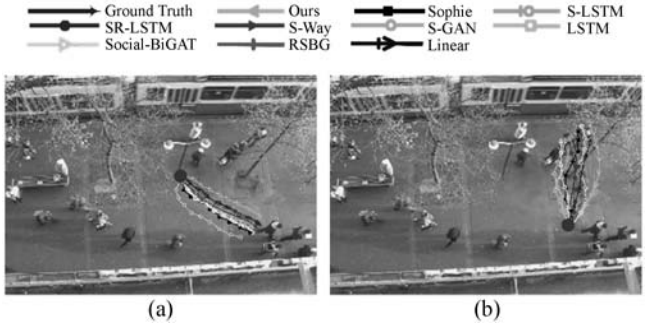
方法		评价基准 ADE/FDE					
		DataSet					AVG
		ETH	Hotel	Univ	ZARA1	ZARA2	
确定性模型	Linear	1.33/2.94	0.39/0.72	0.82/1.59	0.62/1.21	0.77/1.48	0.79/1.59
	LSTM	1.09/2.41	0.86/1.91	0.61/1.31	0.41/0.88	0.52/1.11	0.70/1.52
	S-LSTM	1.09/2.35	0.79/1.76	0.67/1.40	0.47/1.00	0.56/1.17	0.72/1.54
	SR-LSTM	0.63/1.25	0.37/0.74	0.51/1.10	0.41/0.90	0.32/0.70	0.45/0.94
随机性模型	S-GAN	0.92/1.73	0.67/1.37	0.76/1.52	0.35/0.68	0.42/0.84	0.62/1.23
	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-Ways	0.39/0.64	0.39/0.66	0.55/1.31	0.44/0.64	0.51/0.92	0.46/0.83
	Social-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
	RSBG	0.80/1.53	0.33/0.64	0.59/1.25	0.40/0.86	**0.30/0.65**	0.48/0.99
本文方法		**0.36/0.63**	**0.36/0.64**	0.56/1.21	0.42/0.86	0.35/0.75	**0.41/0.82**

3.3　轨迹可视化

图 4 给出了模型在 UCY-ZARA1 数据集上预测轨迹的可视化结果。历史观察轨迹用灰色直线表示,预测分布用黑色直线表示,真实轨迹用灰色带箭头的虚线表示。结果显示,大多数情况下,预测分布对地面真实轨迹有很好的覆盖性,并产生了合理的预测轨迹。

图 4　模型在 ZARA1 数据集下的预测分布

取预测分布中最接近真实轨迹的样本进行评估,图 5 显示了 Hotel 数据集中本文模型与表 1 各模型预测轨迹对比的可视化结果。第一列表示当人们穿过人群或接近有轨电车时,模型能够正确预测实际的情况。可以看出,模型预测的路径能够更好地捕捉复杂的动态,显示出更多的自然运动,而不需要不断调整运动方向,并且我们的模型更接近于地面真实轨迹。第二列展示了模型无法捕捉到达不同目的地的正确的情况。



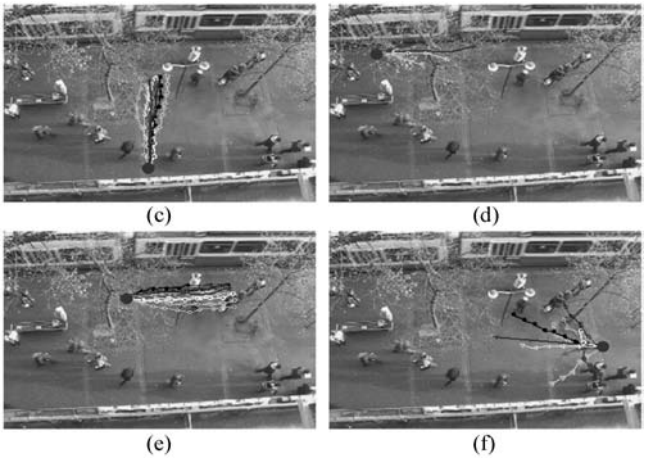


图 5　Hotle 数据集下各模型比较结果

3.4　社会意识特征选择

图 6 显示了社会注意力模型如何选择特征,其中每一行与隐藏特征的某个维度相关。第一列显示了由隐藏特征捕获的轨迹模式,这些特征从原点开始,由 LSTM 提取单个行人的隐藏状态,通过社会注意力模块进行成对的信息选择。其他列显示了对特征具有高响应的一些示例。在这些成对轨迹样本中,左边和右边分别是行人 i 和 j 的轨迹,高响应意味着选择了相应的特征。

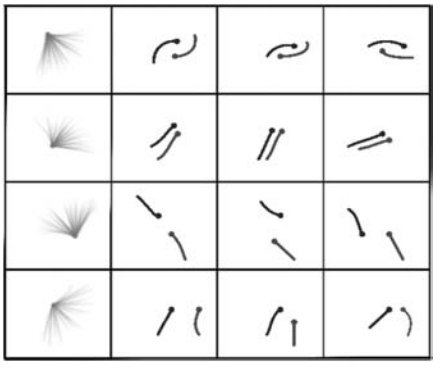


图 6　社会意识特征选择示意图

图 6 中:1) 第一行:轨迹接近但方向相反,行人 i 最关心对方是否会走向他的轨迹特征。2) 第二行:轨迹非常接近且并行行走,选择的特征遵循其行走方向。3) 第三行:与第一行相似。4) 第四行:与第二行相似。

4 结 语

本文提出了一种基于 GAN 的社会和场景感知行人轨迹预测模型。该模型定义一个社会注意力模块处理行人交互问题,将 LSTM 视为特征提取器,根据消息传递机制自适应地重新定义所有行人的当前特征,并利用注意力机制形成具有社会感知的信息选择机制。此外,对于物理场景约束使用语义池机制,使行人意识到周围空间的物理元素并作出相应的改变。由于 GANs 容易发生模式崩溃,模型还使用 Info-GAN 进行对抗性训练。

实验结果表明,本文方法可生成了多条合理的轨迹,并在一定程度上提高了之前方法预测轨迹的精度。因为汽车、自行车或滑板手等多个元素通常共享同一个环境,未来的工作将研究不同的数据集(例如,斯坦福无人机数据集[21])下的行人轨迹预测方法,这些数据集可以捕捉到更复杂的动态以及多个代理设置。

参 考 文 献

[1] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//27th International Conference on Neural Information Processing Systems,2014:2672 - 2680.

[2] Chen X, Duan Y, Houthooft R, et al. InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets[J]. Advances in Neural Information Processing Systems,2016:2172 - 2180.

[3] Berg V D, Guy S J, Lin M, et al. Reciprocal n-body collision avoidance[M]//Springer Tracts in Advanced Robotics. Switzerland: Spring,2011:3 - 19.

[4] Yamaguchi K, Berg A C, Ortiz L E, et al. Who are you with and where are you going? [C]//IEEE Conference on Computer Vision and Pattern Recognition,2011:1345 - 1352.

[5] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation,1997,9(8):1735 - 1780.

[6] Alahi A, Goel K, Ramanathan V, et al. Social LSTM: Human trajectory prediction in crowdeds spaces[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016:961 - 971.

[7] Zhang P, Ouyang W L, Zhang P F, et al. SR-LSTM: State refinement for LSTM towards pedestrian trajectory prediction [EB]. arXiv:1903.02793v1,2019.

[8] Gupta A, Johnson J, Li F, et al. Social GAN: Socially acceptable trajectories with generative adversarial networks [C]//IEEE Conference on Computer Vision and Pattern Recognition,2018:2255 - 2264.

[9] Mangalam K, Girase H, Agarwal S, et al. It is not the Journey but the destination: Endpoint conditioned trajectory prediction[EB]. arXiv:2004.02025v3,2020.

[10] Sadeghian A, Kosaraju V, Sadeghian A, et al. SoPhie: An attentive GAN for predicting paths compliant to social and physical constraints[EB]. arXiv:1806.01482,2018.

[11] Haddad S, Wu M Q, Wei H, et al. Situation-aware pedestrian trajectory prediction with spatio-temporal attention model[EB]. arXiv:1902.05437,2019.

[12] Velickovic P, Cucurull G, Casanova A, et al. Graph attention networks[M]//Introduction to Graph Neural Networks. Synthesis Lectures on Artificial Intelligence and Machine Learning. Springer,2018:39 - 41.

[13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//31st International Conference on Neural Information Processing System,2017:6000 - 6010.

[14] Lisotto M, Coscia P, Ballan L. Social and scene-aware trajectory prediction in crowded spaces[EB]. arXiv: 1909.08840,2019.

[15] Mohamed A, Qian K, Elhoseiny M, et al. Social-STGCNN: A social spatio-temporal graph convolutional neural network for human trajectory prediction[C]//IEEE Conference on Computer Vision and Pattern Recognition,2020.

[16] Liang J W, Jiang L, Murphy K, et al. The garden of forking paths: Towards multi-future trajectory prediction[EB]. arXiv:1912.06445v3,2020.

[17] Pellegrini S, Ess A, Gool L. Improving data association by joint modeling of pedestrian trajectories and groupings[C]// European Conference on Computer Vision,2010:452 - 465.

[18] Lerner A, Chrysanthou Y, Lischinski D. Crowds by example [J]. Computer Graphics Forum,2007,26(3):655 - 664.

[19] Amirian J, Hayet J B, Pettre J. Social ways: Learning multi-modal distributions of pedestrian trajectories with GANs [EB]. arXiv:1904.09507v2,2019.

[20] Vineet K, Sadeghian A, Martin R, et al. Social-BiGAT: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks[EB]. arXiv:1907.03395,2019.

[21] Sun J H, Jiang Q H, Lu C W. Recursive social behavior graph for trajectory prediction[EB]. arXiv:2004.10402v1, 2020.

[22] Robicquet A, Sadeghian A, Alahi A, et al. Learning social etiquette: Human trajectory understanding in crowded scenes [C]//European Conference on Computer Vision,2016:549 - 565.