# 基于社会注意力机制的行人轨迹预测方法研究

李琳辉[1,2]，周彬[1]，连静[1,2]，周雅夫[1]

（1. 大连理工大学汽车工程学院，辽宁 大连 116024；2. 大连理工大学工业装备结构分析国家重点实验室，辽宁 大连 116024）

**摘 要**：为提高行人交互中轨迹预测速度、精度与模型可解释性，提出了一种基于社会注意力机制的 GAN 模型。首先，定义了一种新型社会关系，对行人间的影响进行社会关系建模，设计了基于注意力机制的网络模型，提高了网络预测速度和可解释性。然后，探索不同池化汇集机制对预测结果的影响，确定性能优异的池化模型。最后，搭建了轨迹预测网络，并在 UCY 和 ETH 数据集中进行训练。实验结果表明，所提模型预测精度优于现有方法，且实时性较现有方法提升 18.3% 。

**关键词**：行人轨迹预测；生成对抗网络；注意力机制；社会力模型；最优池化模型

**中图分类号**：TP391

**文献标识码**：A

**doi:** 10.11959/j.issn.1000−436x.2020100

# Research on pedestrian trajectory prediction method based on social attention mechanism

LI Linhui[1,2], ZHOU Bin[1], LIAN Jing[1,2], ZHOU Yafu[1]

1. School of Automotive Engineering, Dalian University of Technology, Dalian 116024, China

2. State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116024, China

**Abstract:** In order to improve the speed, accuracy and model interpretability of trajectory prediction in pedestrian interaction, a GAN model based on social attention mechanism was proposed. Firstly, a new type of social relationship on pedestrians was defined to model social relationships and a network model based on the attention mechanism was designed to improve the speed and interpretability of network prediction. Secondly, the influence of different pooling mechanisms on the prediction results was explored to determine the pooling model with excellent performance. Finally, a trajectory prediction network was built on this basis and trained on the UCY and ETH data sets. The experimental results show that the model not only has better prediction accuracy than the existing methods, but also improves the real-time performance by 18.3% compared with the existing methods.

**Key words:** pedestrian trajectory prediction, generative adversarial network, attention mechanism, social force model, optimal pooling model

## 1 引言

行人运动的数据分析对于道路安全、机器人导航[1]、安全监控等领域具有重要的意义。研究行人轨迹需要收集行人的数据并进行离线分析，了解行人行为和周围环境并以此做出合理决策。在具有实时决策功能的系统中对行人的行进路线进行预测，可以尽早发出警报或者采取相应的预防措施。

轨迹预测问题可视为序列决策问题，即根据行

**收稿日期**：2019−12−24；**修回日期**：2020−03−13

**通信作者**：连静，lianjing@dlut.edu.cn

**基金项目**：国家自然科学基金资助项目（No.61976039, No.51775082）；中央高校基本科研业务费专项基金资助项目（No.DUT19LAB36, No.DUT17LAB11）

**Foundation Items:** The National Natural Science Foundation of China (No.61976039, No.51775082), The Fundamental Research Funds for the Central Universities (No.DUT19LAB36, No.DUT17LAB11)

人过去时刻的位置预测未来时刻的轨迹。但该问题非常复杂。首先，行人的运动具有一定的随机性，在预测任务中生成一条确定性轨迹是不符合实际情况的。其次，每个行人并不是独立存在的，根据 Mehdi 等[2]的研究，70%的行人倾向于成群行走，他们在同一时空下进行交互，这使处理该问题变得更加困难。

现有的行人轨迹预测方法可分为两类：一类是基于模型的方法，另一类是基于数据驱动的深度学习的方法。Kitani 等[3]证明静态环境的语义信息（例如人行道的延伸区域和斑马线的位置等）有助于模型更加准确地预测未来时刻的行人轨迹。Lee 等[4]从顶视图像中学习场景的上下文来预测每个智能体将来的轨迹。Yamaguchi 等[5]采用基于优化的方法，通过手动设计涵盖运动相关方面的目标函数，实时优化函数参数。但此类基于模型的方法在处理轨迹预测问题时受到两方面限制：1) 需要手动设计模型函数来模拟人与人之间交互的场景，而不是通过数据拟合函数，这使模型函数只能适用于简单的交互场景（如引力/排斥力社会力模型）；2) 模型专注于建立当下相邻行人的交互，但无法对将来时刻发生的交互行为进行合理建模及预测。

近年来，随着计算机视觉和人工智能相关技术的飞速发展，基于深度学习的行人轨迹预测方法引起了研究人员的关注，基于循环神经网络（RNN, recurrent neural network）及长短时记忆（LSTM, long-short term memory）网络[6]的方法已经被证明在处理时序问题时的有效性，但基于 RNN 及 LSTM 的方法无法对行人之间的空间联系进行有效建模。在文献[7]中，行人集合被建模为时空图，其中边（时间和空间）与 RNN 相连，时间边捕捉单个行人的信息，空间边捕捉行人交互的信息，输出采用双变量高斯分布，该方法能较好地对时空信息进行有效建模，但计算复杂。Alahi 等[8]提出了社会长短时记忆（S-LSTM, social long-short term memory）网络模型，通过对周围行人进行网格化建模，将周围行人的不同特征进行隐藏池化，利用隐藏特征计算预测轨迹。Gupta 等[9]将基于生成对抗网络（GAN, generative adversarial network）[10]的方法引入行人轨迹预测领域，并使用最大池化方法生成多条社会可接受的轨迹，但该方法提取的特征是经过池化模型后的最大特征，模型忽略对行人交互有用的其他特征信息。Sadeghian 等[11]通过融合环境中场景的上下文信息以及行人的历史轨迹，使用 GAN 网络生成多条物理条件下的可接受轨迹。Vineet 等[12]将图注意力（GAT, graph attention）网络[13]引入轨迹预测领域，通过使用图注意力网络增强轨迹预测的推理能力。但当前基于数据驱动方法的网络模型存在网络冗余度增大，可解释性降低，实时性不高及精度不足等问题。

针对上述方法存在的问题，本文从社会力模型、注意力机制、最优池化机制及 GAN 网络这 4 个角度入手，提出了一种基于社会注意力机制的 GAN（SA-GAN, social-attention GAN）模型，其结构如图 1 所示。首先，定义一种新型的社会关系，用来描述行人之间的相互影响，实现交互场景中行人相互关系的社会关系建模，并设计注意力模型对行人的社会关系进行注意力建模，减少模型冗余度，提高预测速度和可解释性。然后，探索不同池化汇集机制对于轨迹预测结果的影响，选择相应场景下性能优异的池化汇集模型，提高预测精度。最后，构建轨迹预测模型，设计相应的损失函数和模型参数，最大程度上避免 GAN 模型在训练时发生模式崩溃。实验表明，基于本文方法进行的轨迹预测，在提高预测速度的同时有效提高了预测精度，并达到减少模型冗余度和提高模型可解释性的效果。

## 2 轨迹预测模型构建

### 2.1 轨迹预测问题定义

轨迹预测问题可以表示为根据行人过去的状态和信息来估计将来时刻的状态。第 $i$ 个行人在 $\tau$ 时刻的状态为该行人的位置，用二维坐标表示为 $\boldsymbol{X}_i^{\tau}=(x_i^{\tau},y_i^{\tau})\in\mathbb{R}^2$，所有行人轨迹的集合为

$$\boldsymbol{X}_i^{1:t}=\{(x_i^k,y_i^k)\in\mathbb{R}^2\mid k=1,2,\cdots,t\}\ \forall i\in[N] \tag{1}$$

$$[N]\in\{1,2,\cdots,N\} \tag{2}$$

其中，$N$ 是行人轨迹的数目，$t$ 是行人轨迹中所输入观测轨迹长度。用 $Y_i^{1:T}$ 表示第 $i$ 个行人在($t$+1)～($t$+$T$)时刻数据集中真实的轨迹序列，$\widetilde{Y}_i^{1:T}$ 表示模型预测输出的轨迹序列，$T$ 表示行人轨迹中输出的轨迹的长度，则

$$\begin{aligned}Y_i^{1:T}=\{(x_i^k,y_i^k)\in\mathbb{R}^2\mid k=\\ t+1,t+2,\cdots,t+T\},\ \forall i\in[N]\end{aligned} \tag{3}$$

$$\tilde{Y}_i^{1:T}=\{(\tilde{x}_i^k,\tilde{y}_i^k)\in\mathbb{R}^2\mid k=t+1,\ t+2,\cdots,t+T\},\forall i\in[N] \tag{4}$$

轨迹预测的目标是将每个行人的 $1\sim t$ 时刻的轨迹输入模型，通过学习模型的参数 $W^{'}$，预测每一位行人在 $(t+1)\sim(t+T)$ 时刻的轨迹 $\tilde{Y}_i^{1:T}$。

$$\tilde{Y}_i^{1:T}=f(X_i^{1:t};W^{'}) \tag{5}$$

其中，模型参数 $W^{'}$ 是模型中所有使用的深层神经网络的权重。使用反向传播算法对模型所有权重参数进行训练，对行人预测轨迹和真实轨迹通过最小化损失函数 $W^{*}$ 实现随机梯度下降。

### 2.2　轨迹预测生成模型

如图 1 所示，SA-GAN 的整体网络架构由几个关键模块组成，分别为轨迹生成器 G、社会注意力模型 SA（包括社会关系 S 和注意力模块 A）、池化模块 P 和轨迹判别器 D。轨迹生成器 G 接收行人 $i$ 过去时刻的轨迹 $X_i^{1:t}$，并将行人的轨迹特征经过编码器编码，得到隐藏向量 $\boldsymbol{H}_i^{1:t}$，然后与行人之间的社会关系一起输入注意力模块 A 中，突显出最重要的信息后输出。池化模块 P 将注意力模块的输出 $A_i^t$ 作为输入用来生成池化向量 $\boldsymbol{P}_i^{1:t}$，学习行人之间的互动以及行人对每位行人未来路径的影响，解码器接收隐藏向量 $\boldsymbol{H}_i^{1:t}$、池化向量 $\boldsymbol{P}_i^{1:t}$ 和噪声 $z$ 作为生成器生成轨迹的条件生成预测轨迹。轨迹判别器 D 则接收生成器 G 生成的预测轨迹 $\tilde{Y}_i^{1:T}$ 和数据集样本中的真实轨迹 $Y_i^{1:T}$，通过强制生成器模型生成更逼真的样本（轨迹）来确定模型的有效性。

## 3　社会注意力模型搭建

将轨迹预测问题视为序列决策问题，单个行人 $i$ 在时刻 $t$ 的状态不仅与当前位置信息和过去状态有关，也与周围其他行人的影响有关，为准确描述周围行人对行人 $i$ 的影响，本文基于行人之间的社会关系对行人的影响进行社会建模，行人之间的社交关系 $S_{ij}$ 定义如下。

1) 行人之间的相对距离 $R_{ij}$。

$$R_{ij}=|\boldsymbol{X}_i^t-\boldsymbol{X}_j^t| \tag{6}$$

2) 行人 $j$ 与行人 $i$ 的方位角 $\theta_{ij}$，即行人 $i$ 的速度向量和行人 $j$ 与行人 $i$ 位移向量的夹角。

$$\theta_{ij}=\cos\left((\boldsymbol{X}_i^t-\boldsymbol{X}_i^{t-1}),(\boldsymbol{X}_i^t-\boldsymbol{X}_j^t)\right) \tag{7}$$

3) 行人之间的最近距离 $D_{ij}$，即 2 个行人将要达到的最小距离。

$$D_{ij}=\min(|\boldsymbol{X}_i^t-\boldsymbol{X}_j^t|) \tag{8}$$

2 个行人之间的方位角用来提取行人轨迹状态的方向信息。相对距离的大小直接影响行人间的交互，相对距离越小，对行人交互的影响越大。行人间

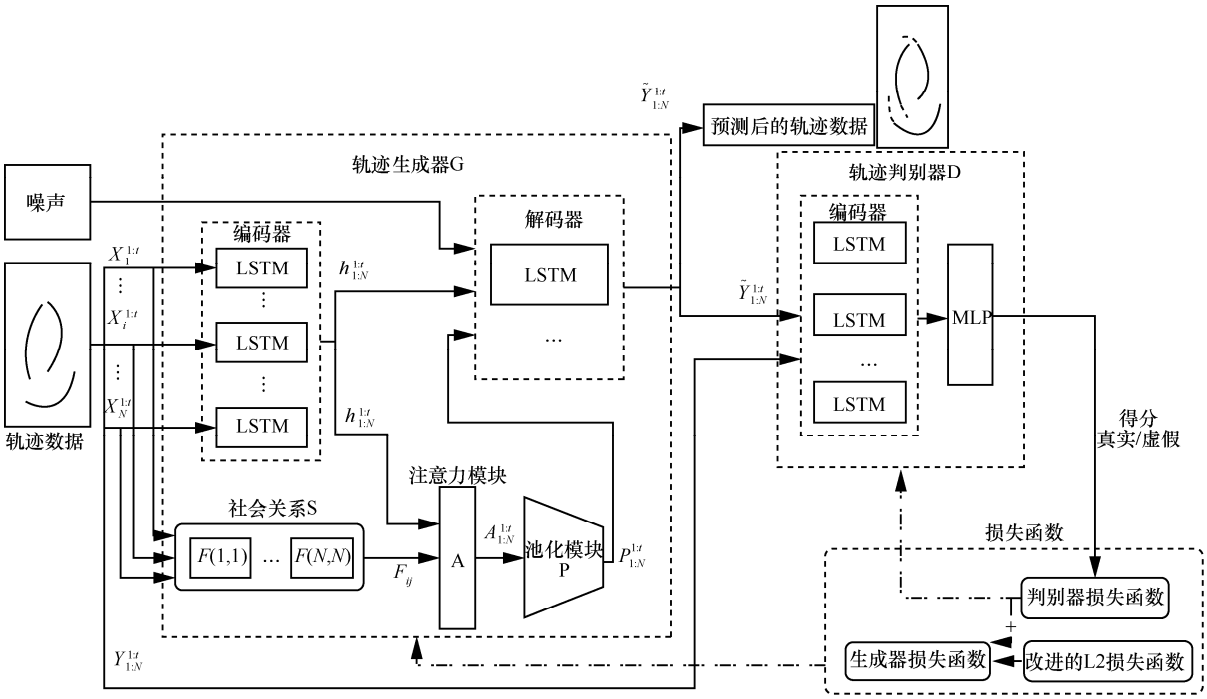



图 1　SA-GAN 网络架构

的最短距离可以使模型有效地提取行人交互时避免碰撞的特征信息。社交关系 $S_{ij}$ 经过嵌入层 $\varphi$ 嵌入，将输出送到多层感知器（MLP, multilayer perceptron），其中 $W_{\varphi}$ 和 $W_{\text{MLP}}$ 分别为嵌入层和 MLP 的权重参数。行人 $i$ 与行人 $j$ 之间进行交互后的特征向量 $\boldsymbol{F}_{ij}$ 为

$$\boldsymbol{F}_{ij}=\text{MLP}(\varphi(S_{ij};W_{\varphi});W_{\text{MLP}}) \tag{9}$$

注意力机制[14]源于人类的视觉注意力机制，人类利用有限的注意力资源从大量的信息中提取出有效信息，是人类长期进化出的生存机制。基于深度学习的注意力机制最先应用于图像领域，其机制与人类的视觉注意力机制相似，都是从众多的信息中获取对任务目标最有效、最关键的信息。之后注意力机制被广泛应用于语音识别、机器翻译等自然语言处理领域，借助编码器–解码器模型[15]在相关领域取得了很好的成果[16]。本文注意力模块与人类视觉注意力机制类似，都更加注意所在意的信息。在观察一幅图像时并非一次性就了解图像的全部，而是先将注意力集中到图像的局部特征。与观察图像类似，当处理人与人交互的复杂场景时，通常行人会结合当前的状态和周围行人的影响，着重注意对自己有影响的部分信息，以迅速做出相应决策和改变轨迹[17]。为评估其他行人对行人 $i$ 的影响，通过编码器模块输出的隐藏向量 $\boldsymbol{H}_i^t$ 与行人交互后的特征向量 $\boldsymbol{F}_{ij}$ 结合，输入注意力模块得到行人 $i$ 的注意力向量 $\boldsymbol{A}_i^t$。

$$\mathbf{Con}=[\boldsymbol{F}_{ij},\boldsymbol{H}_i^t] \tag{10}$$

$$\boldsymbol{A}_i^t=\text{ATT}(\mathbf{Con};W_{\text{att}}) \tag{11}$$

其中，**Con** 模块进行向量的拼接；注意力模块 ATT 使用对文献[18]所提结构进行改进的软注意力机制设计使用多层感知器，使整个体系可以通过反向传播算法进行端到端的训练；$W_{\text{att}}$ 为注意力模块的权重。添加社会注意力机制后的模型不仅可以处理复杂场景中的行人交互，为轨迹预测增加可解释性，同时抑制输入数据的冗余，使预测模型将重点放在重要特征上，加快模型收敛时间，节省计算资源。

## 4 网络结构设计

通过分析当前先进的轨迹预测模型，本文选择基于生成对抗网络的网络架构[19]来生成预测轨迹，并进一步探究提高轨迹预测性能的方法，在所提改进的社会注意力模块的基础上，探索最优池化汇集方法。

### 4.1 建立池化模型

在搭建好社会注意力模型后，将含有注意力信息的注意力特征向量 $\boldsymbol{A}_i^t$ 汇总到池化模块中，池化模块 P 总结行人做出决定所需要的所有信息，得到输出的池化向量 $\boldsymbol{P}_i^t$，以实现多人共同推理，跨 LSTM 实现信息共享的功能。

### 4.2 轨迹生成器结构设计

轨迹生成器 G 基于编码器–解码器框架，采用类似于条件 GAN 的架构（如图 1 所示）来缓解 GAN 训练模式崩溃的问题，由 encoder、social module、attention module、pooling module 和 decoder 组成。首先使用单层 MLP 作为嵌入层，将行人 $i$ 在 $t$ 时刻的位置 $\boldsymbol{X}_i^t$，从坐标空间映射到特征空间得到特征向量 $\boldsymbol{E}_i^t$，将嵌入的特征向量 $\boldsymbol{E}_i^t$ 经过 $\text{LSTM}_{\text{encoder}}$ 编码器编码处理，学习轨迹的时间特征，得到 LSTM 层的隐藏向量 $\boldsymbol{H}_i^t$。

$$\boldsymbol{E}_i^t=\text{MLP}(\boldsymbol{X}_i^t;W_{\text{E}}) \tag{12}$$

$$\boldsymbol{H}_i^t=\text{LSTM}_{\text{encoder}}(\boldsymbol{E}_i^t,\boldsymbol{H}_i^{t-1};W_{\text{encoder}}) \tag{13}$$

其中，MLP(·) 是采用 ReLu 激活函数的嵌入层，$W_{\text{E}}$ 是嵌入层的权重参数，$W_{\text{encoder}}$ 是编码器模块的权重参数，参数由场景中的所有行人共享。

在解码及生成样本分布的过程中，设计一个类似于编码器的 LSTM 结构作为解码器，不同于其他 LSTM 结构隐藏状态的随机初始化，decoder 的初始化是采用经过社会注意力处理的池化向量 $\boldsymbol{P}_i^t$、编码器隐藏向量 $\boldsymbol{H}_i^t$ 与随机噪声 $z$ 合成的特征向量 $\mathbf{Con}_i^t$。decoder 作为具有输入条件的轨迹生成器，输入条件是上一时刻的轨迹隐藏状态和解码器初始化的隐藏状态。

$$\mathbf{Con}_i^t=[\text{MLP}(\boldsymbol{P}_i^t,\boldsymbol{H}_i^t),z] \tag{14}$$

$$\boldsymbol{H}_{i-d}^t=\text{LSTM}_{\text{decoder}}(\mathbf{Con}_i^t,E(x_i^{t-1},y_i^{t-1});W_{\text{decoder}}) \tag{15}$$

$$\tilde{\boldsymbol{Y}}_i^t=(\tilde{x}_i^t,\tilde{y}_i^t)=\varphi(\boldsymbol{H}_{i-d}^t) \tag{16}$$

在进行 $T=(t_{\text{obs}}+1,t_{\text{pred}})$ 时刻的轨迹预测时，循环式(17)～式(20)所示过程。

$$\boldsymbol{P}_i=P(\boldsymbol{H}_{i-d}^t,\cdots,\boldsymbol{H}_{n-d}^t) \tag{17}$$

$$\boldsymbol{H}_{i-d}^{t+1}=\text{MLP}(\boldsymbol{P}_i,\boldsymbol{H}_{i-d}^t;W_{\text{MLP}}) \tag{18}$$

$$\boldsymbol{H}_{i-d}^{t+1}=\mathrm{LSTM}_{\mathrm{decoder}}(\boldsymbol{H}_{i-d}^{t+1},E(\tilde{x}_i^t,\tilde{y}_i^t;W_E);W_{\mathrm{decoder}}) \quad (19)$$

$$\tilde{\boldsymbol{Y}}_i^{t+1}=(\tilde{x}_i^{t+1},\tilde{y}_i^{t+1})=\varphi(\boldsymbol{H}_{i-d}^{t+1};W_\varphi) \quad (20)$$

其中，$W_{\mathrm{decoder}}$、$W_{\mathrm{MLP}}$ 和 $W_\varphi$ 为 LSTM 解码器、MLP 及输出层的网络参数。

### 4.3　轨迹判别器结构设计

如图 1 所示，轨迹判别器由 LSTM 层以及 MLP 组成，LSTM 层对输入的轨迹进行编码，输入真实轨迹和生成器生成的虚假轨迹，输出编码器的隐藏状态 $H_{\mathrm{D}}$，通过 MLP 对轨迹的真假性进行打分。

$$H_{\mathrm{D}}=\mathrm{LSTM}_{\mathrm{encoder}}(\tilde{\boldsymbol{Y}}_i^{1:T};W_{\mathrm{D}}) \quad (21)$$

$$\mathrm{Score}=\mathrm{MLP}(H_{\mathrm{D}};W_m) \quad (22)$$

### 4.4　损失函数及模型参数的确定

由于 GAN 训练很困难,当轨迹生成器 G 和轨迹判别器 D 之间不平衡时，很容易造成梯度消失或者模式崩溃（采样的合成数据没有多样性），特别是在预测行人轨迹时，避免模式崩溃至关重要[20]。现有的大多数轨迹预测的方法都使用了真实轨迹与预测轨迹的 L2 范数来估计未来的状态[19]，使用 L2 损失虽然可以加快损失函数的收敛速度，但是其趋向预测一条未来平均轨迹（即每个行人将来可行轨迹的均值），本文采用 GAN 对抗损失函数以及改进的 L2 损失函数来避免这一问题。SA-GAN 损失函数 $W^*$ 的基础是对抗损失，在其基础上结合多样性损失，增加轨迹生成的多样性。

$$W^*=L_{\mathrm{GAN}}+\lambda L_* \quad (23)$$

其中，$L_{\mathrm{GAN}}$ 作为对抗损失用来优化生成器和判别器，$L_*$ 为多样性损失，$\lambda$ 为损失系数。

$$L_{\mathrm{GAN}}=\min_G\max_D \mathrm{E}_{P_{\mathrm{data}}(x_i^{1:t})}[x_i^{1:t}\log(\tilde{\boldsymbol{Y}}_i^{1:T})+\mathrm{E}_{P_{\mathrm{gen}}(\tilde{Y}_i^{1:T})}[(1-x_i^{1:t})\log(1-\tilde{\boldsymbol{Y}}_i^{1:T})] \quad (24)$$

$$L_*=\|\tilde{\boldsymbol{Y}}_i^{1:T}-\boldsymbol{X}_i^{1:t}\|_2^2 \quad (25)$$

本文模型架构基于 PyTorch 搭建，网络结构的 MLP 中的 FC 层之后都与 L1 正则化层和激活函数 ReLU 层相连。模型中统一使用 2×64 的嵌入层做轨迹编码处理，LSTM 网络隐藏层单元个数设置为 128，采用 8 维的高斯噪声[19]作为采样分布，使用 Adam 优化器对轨迹生成器和轨迹判别器进行迭代训练，最小批量大小为 64，生成器和判别器的学习率为 0.000 5，训练次数 10 000 次。

## 5　实验与结果分析

### 5.1　实验过程

为评估本文提出的 SA-GAN 模型性能，本文实验使用 2 个公开数据集，分别为 ETH[21]和 UCY[22]。数据集包含各种类型的社会交互场景下行人的轨迹坐标，其中包含行人交互、避免碰撞和群体行人的轨迹坐标，以 2.5 frame/s 的速度进行手动采样标记。其中 ETH 包含 2 个数据集（ETH 和 Hotel），UCY 包含 3 个数据集（Zara1、Zara2 和 Univ）。为评估算法的性能，在 5 个数据集中进行验证，采用交叉验证的方法，分别在其中 4 个数据集中进行训练，在另外一个数据集中测试验证。将算法在数据集上训练的结果采用各种基准进行比较分析，并且对注意力机制以及不同池化模型进行定性分析。实验运行在 Ubuntu 18.04 LTS 的操作系统中，GPU 为 NVIDIA Titan X，采用 PyTorch 1.1.0，CUDA 10.1 以及 Cudnn v7.5.0 的深度学习框架。

本文采用最具代表性的线性回归器、LSTM、S-LSTM、S-GAN、Sophie 和 Social-BiGAT 网络作为比较基准，说明如下。

线性回归器，可以采用最小化 MSE 来优化回归器的参数。

LSTM，基于数据驱动的轨迹预测方法。

S-LSTM，将每个行人与一个 LSTM 单元相关联，采用社交池化机制收集相邻行人的隐藏状态并进行预测。

S-GAN，基于 GAN 的方法进行轨迹的生成预测。

Sophie，基于物理和社会注意力的轨迹预测生成模型。

Social-BiGAT，基于图注意力网络的轨迹预测生成模型。

### 5.2　实验分析

与之前研究方法类似[9]，使用以下指标对每个行人的轨迹进行预测评估。

1) 平均位移误差（ADE, average displacement error）。实际轨迹与预测轨迹序列在每一个时间步长内的 L2 欧氏距离。

$$\mathrm{ADE}=\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{X}_i^t-\tilde{\boldsymbol{Y}}_i^t\| \quad (26)$$

2) 最终位移误差（FDE, final displacement error）。实际轨迹与预测轨迹序列在最终位置的 L2 欧

氏距离。

$$\mathrm{FDE} = \| \boldsymbol{X}_i^n - \tilde{\boldsymbol{Y}}_i^n \| \tag{27}$$

评估测试数据集在模型预测 $t_{\mathrm{pred}} = 8$ 帧（3.2 s）和 $t_{\mathrm{pred}} = 12$ 帧（4.8 s）时平均 ADE 和平均 FDE。

**5.2.1** 不同池化类型分析

池化汇集机制可以对不同行人的交互信息进行综合分析，在行人轨迹预测任务中可以实现跨 LSTM 网络的信息共享，从而解决无法有效对行人间的相互影响进行建模的问题。但不同池化类型对轨迹预测问题的影响，在轨迹预测任务的研究中，并没有深入研究探讨。

分别采用 Gumbel Pooling、Max Pooling、Average Pooling 及 Random Pooling 的池化汇集方法，在不同数据集的轨迹预测任务中对平均 ADE 和平均 FDE 进行对比分析，分析结果如表 1 所示。

训练参数的使用和 5.1 节保持一致，对比分析了当预测步长为 8 帧时，SA-GAN 采用不同池化汇集类型的平均 ADE 和平均 FDE。Gupta 等[9]使用的池化方法是最大池化，但由表 1 可以看出，当使用最大池化的方法时，该方法仅在 ETH 数据集中有较好的表现，但是在其他数据集中的表现一般。这可能是由于采用了最大池化的方法，对有效信息过滤太多，无法有效保留有用的特征信息，特别是采用了社会注意力机制，输入池化层的信息含有更多的有效特征，进行最大池化反而结果不好。由表 1 可知，Gumbel Pooling、Average Pooling 和 Random Pooling 具有相似的效果，其中，Average Pooling 的方法在 Hotel 数据集中有着很好的表现，由于 Hotel 数据集行人较少，大多数行人多为线性轨迹，当采用 Average Pooling 的池化方法时，能够对该场景进行更为有效的建模。Gumbel Pooling 的方法处理 ETH 数据集的任务表现性能最优。综上所述 Average Pooling 的方法适合处理人流量较为稀疏的场景，Gumbel Pooling 处理人流密集场景的行人轨迹效果较好。所以，本节模型进行精度分析和速度分析时，采用的方法是基于 Gumbel Pooling 的池化汇集方法。

**5.2.2** 模型预测精度分析

为在评估模型时保持基准的一致性，对于生成系列轨迹的 GAN 模型，使用文献[9]中所提方法，使模型生成 $N$ 个样本，并采用生成的最接近真实样本的轨迹进行评估，本文采用的数据样本采集数量为 $N$=20。表 2 显示了各种行人轨迹预测算法在预测轨迹长度分别为 $t_{\mathrm{pred}} = 8$ 帧（3.2 s）和 $t_{\mathrm{pred}} = 12$ 帧（4.8 s）时，平均 ADE（平均位移误差）和平均 FDE（最终位移误差）的指标大小。

对比分析各方法，模型在预测 8 帧及预测 12 帧轨迹时的平均 ADE 和平均 FDE 都随着预测距离的增加而增加，所有网络的精度指标都明显下降，这与预测时间越久，轨迹预测难度越大的事实相对应。整个测试实验中，各个模型在预测 8 帧和预测 12 帧轨迹误差都有着相似的趋势。

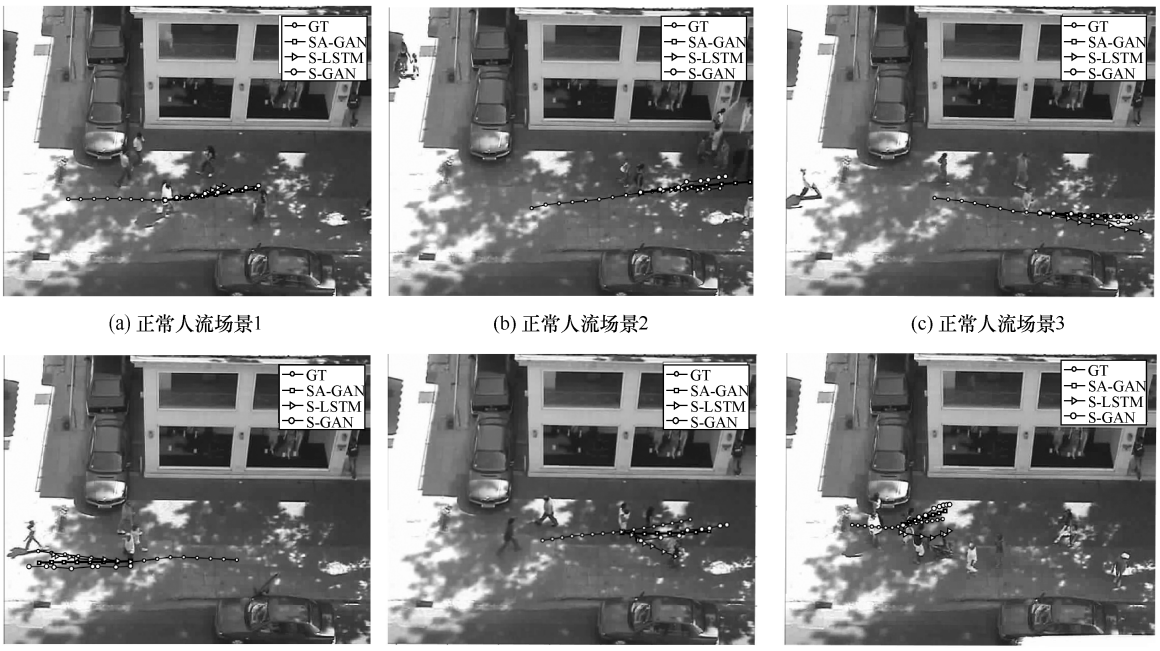
线性模型由于无法对不同行人之间复杂的社会交互进行有效建模，在整个轨迹预测任务中表现最差，但在 Hotel 数据集中 ADE 和 FDE 指标都很好，这也证实了 Hotel 场景作为一个稀疏的场景，行人之间的交互较少，轨迹大部分都是线性的。与 LSTM 以及基于网格划分进行池化汇集的 S-LSTM 比较，基于 GAN 的模型趋向于生成多条随机的轨迹，比之前生成一条确定性的轨迹，有着更高的精度和社会可接受性。基于物理和社会注意力机制的 Sophie 模型，虽在部分数据集中有着不错的表现，但平均效果仍略低于 SA-GAN 模型。值得注意的是，基于图注意力网络的 Social-BiGAT 模型，虽然

**表 1** **不同池化汇集类型的 SA-GAN 模型 ADE 和 FDE 对比分析**

| 数据集 | ADE/m | | | | FDE/m | | | |
|---|---|---|---|---|---|---|---|---|
| | Gumbel Pooling | Max Pooling | Average Pooling | Random Pooling | Gumbel Pooling | Max Pooling | Average Pooling | Random Pooling |
| ETH | 0.58 | **0.55** | 0.64 | 0.60 | 1.17 | **1.05** | 1.24 | 1.14 |
| Hotel | **0.31** | 0.45 | 0.29 | 0.34 | 0.62 | 0.86 | **0.55** | 0.65 |
| Univ | **0.34** | 0.52 | 0.37 | 0.37 | **0.69** | 0.93 | 0.75 | 0.74 |
| Zara1 | **0.21** | 0.54 | 0.22 | **0.21** | 0.45 | 0.98 | 0.42 | **0.41** |
| Zara2 | **0.21** | 0.42 | 0.22 | **0.21** | 0.43 | 0.72 | 0.44 | **0.42** |
| 平均 | **0.33** | 0.496 | 0.348 | 0.346 | **0.672** | 0.908 | 0.68 | **0.672** |

**表 2　　各种网络在各数据集下 ADE 和 FDE 对比分析**

| 预测时间 | 数据集 | 评价基准 ADE/FDE | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Linear | LSTM | S-LSTM | S-GAN | Sophie | Social-BiGAT | SA-GAN |
| $t_{pred}=8$ | ETH | 0.84/1.60 | 0.70/1.45 | 0.73/1.48 | 0.67/1.47 | — | — | **0.58/1.17** |
| | Hotel | 0.35/**0.60** | 0.55/1.17 | 0.49/1.01 | 0.52/1.05 | — | — | **0.31**/0.62 |
| | Univ | 0.56/1.01 | 0.36/0.77 | 0.41/0.84 | **0.34**/0.75 | — | — | **0.34/0.69** |
| | Zara1 | 0.41/0.74 | 0.25/0.74 | 0.27/0.56 | 0.26/**0.45** | — | — | **0.21/0.45** |
| | Zara2 | 0.53/0.95 | 0.31/0.65 | 0.33/0.79 | 0.29/0.58 | — | — | **0.21/0.43** |
| | 平均 | 0.53/0.98 | 0.43/0.91 | 0.45/0.91 | 0.42/0.86 | — | — | **0.33/0.67** |
| $t_{pred}=12$ | ETH | 1.33/2.94 | 1.09/2.41 | 1.09/2.35 | 0.92/1.73 | 0.70/1.43 | **0.69**/1.29 | 0.72/**1.28** |
| | Hotel | **0.39/0.72** | 0.86/1.91 | 0.79/1.76 | 0.67/1.37 | 0.76/1.67 | **0.49/1.01** | 0.50/**1.01** |
| | Univ | 0.82/1.59 | 0.61/1.31 | 0.67/1.40 | 0.76/1.52 | **0.54**/1.24 | 0.55/1.32 | 0.58/**1.19** |
| | Zara1 | 0.62/1.21 | 0.41/0.88 | 0.47/1.00 | 0.35/0.68 | **0.30/0.63** | 0.30/0.62 | 0.42/0.83 |
| | Zara2 | 0.77/1.48 | 0.52/1.11 | 0.56/1.17 | 0.42/**0.84** | 0.38/0.78 | **0.36/0.75** | 0.39/0.85 |
| | 平均 | 0.79/1.59 | 0.70/1.52 | 0.72/1.54 | 0.62/1.23 | 0.54/1.15 | **0.48/1.00** | 0.52/1.03 |




(a) 正常人流场景1　(b) 正常人流场景2　(c) 正常人流场景3

(d) 正常人流场景4　(e) 密集人流场景1　(f) 密集人流场景2

图 2　各模型预测轨迹可视化对比

在 Zara1 和 Zara2 数据集中的性能优于本文模型，但是由于该模型采用了 VGG 编码器对原始图像进行处理，包含多层卷积池化操作，会显著增加模型的参数量，将影响模型实时性。此外，相比于隐式的使用 GAT 网络进行特征汇集，显式地通过函数定义行人之间的交互关系，具有更高的模型可解释性。由表 2 可以看出，SA-GAN 在 ETH、Hotel 的测试实验中预测误差显著降低，而在 UCY 的数据集中，本文提出的 SA-GAN 模型有较好的精度，这是因为该模型不仅在池化模块中加入了注意力机制，而且采用社会关系模型，对社会关系进行社会层面的建模，加上 GAN 的优点，整个模型在本质上趋向于产生多条符合社会规范的轨迹，这些轨迹在一些复杂的场景和非线性轨迹化的情况中具有良好的性能。

图 2 中给出了 ETH 数据集的 Zara1 场景中 GT（实际轨迹）、S-LSTM、S-GAN 和 SA-GAN 的行人轨迹可视化对比，其中实际轨迹包括模型观测到的

8 帧数据以及将来时刻的 8 帧数据。

图 2(a)～图 2(d)为正常人流量场景中各方法轨迹对比，图 2(e)～图 2(f)为人流密集场景中各方法轨迹对比。可以看出，基于本文提出的方法在各场景中都可以达到与真实轨迹相近的预测轨迹，尤其是在场景简单的环境中，预测轨迹与真实轨迹基本贴合，图 2(a)中 SA-GAN 成功预测未来轨迹，并避免与迎面而来的行人相撞，基于 GAN 的方法在预测行人速度和轨迹上与原始轨迹基本一致，但基于 S-LSTM 的方法，预测到行人的速度会随着预测距离的增加而降低。在图 2(b)中，可以看出最优池化和社会注意力机制的作用，S-GAN 由于采用最大池化进行汇集，只关注对于行人影响最大的特征，而图 2(b)中迎面走来多个行人，预测轨迹显示，单一地提取对行人影响最大的特征信息，因此产生了与真实轨迹相差较大的预测轨迹，而采用社会注意力机制和最优池化模型对周围行人特征信息进行社会注意力建模和最优池化采样，生成的轨迹与真实轨迹基本吻合。图 2(c)场景为并排行走的行人间轨迹预测，各方法都表现出较强的交互能力，与 S-LSTM 方法对比发现，本文的方法通过在社会关系建模中考虑周围行人的方向和速度，可以实现对轨迹的精准预测，S-LSTM 尽可能避免与旁边行人进行碰撞但没有考虑行人之间的行进方向，而产生远离行人的预测轨迹。在图 2(e)和图 2(f)中，密集环境中基于 S-LSTM 的预测基本失效，而基于 GAN 的方法在避免碰撞方面表现出优势。在图 2(e)和图 2(f)中，SA-GAN 都成功预测到行人快速超过前方行人，并避免与相向行人碰撞。

#### 5.2.3 模型预测速度分析

模型的预测速度对于模型预测的时效性和效率都有至关重要的影响，尤其是在自动驾驶任务中，更少的轨迹预测时间可以给自动驾驶系统更多的反应时间从而做出合理的决策，可以更好地保护行车安全和道路安全。本节对基于数据驱动的方法进行速度时效性方面的对比分析。

模型训练参数的选择和 5.1 节一致，所有模型都基于 Ubuntu 18.04 系统，在单个 GPU 处理器 Tatian X 中进行预测。为验证预测的时效性，对模型分别预测 8 帧和 12 帧轨迹所用的时间进行分析，结果如表 3 所示。由表 3 可得，预测时间随着预测帧数的提高而增加。基于 LSTM 的轨迹预测模型拥有最快的预测速度，这是由于该方法采用最为简单的 LSTM 模型，但该方法预测的精度低，无法进行交互建模，避免碰撞。S-LSTM 需要对周围栅格地图建模，模拟人与人之间进行交互，所以耗时最长。其他方法较于 S-LSTM 的方法都有了速度上的明显提升，是因为不再多次计算栅格图的隐藏状态，从而节约了预测时间。基于 GAN 的方法中，SA-GAN 拥有更快的预测速度，这是由于其采用社会注意力机制、最优池化汇集机制和改进设计的条件 GAN 架构，使模型能够实时地对特征信息进行融合分析，分析出轨迹中重要的特征信息，并以此做出相应决策。该方法较之前的研究，预测速度平均提升了 18.3%。

**表 3 轨迹预测模型预测速度分析**

| 预测模型 | 预测 8 帧时间/min | 预测 12 帧时间/min |
|---|---|---|
| LSTM | 0.029 | 0.042 |
| S-LSTM | 0.504 | 0.825 |
| S-GAN | 0.202 | 0.341 |
| SA-GAN | 0.163 | 0.282 |

## 6 结束语

本文提出一种基于 GAN 的轨迹生成框架，用于实现行人交互场景下的轨迹预测。通过定义社会注意力机制，对行人之间的交互模式进行社会建模，注意力机制提取对行人轨迹预测最为重要的信息，通过探索最优池化汇集机制实现跨 LSTM 信息共享。为捕获行人未来路径的不确定性，引入生成对抗网络预测轨迹的分布。通过在常用公开数据集中对各个基准的实验评估表明，与之前的相关研究相比，SA-GAN 在测试数据集中，不仅提高了预测的准确性和实时性，而且在减少模型冗余度、提高模型可解释性以及避免碰撞方面具有优越的性能。

现有的轨迹预测方法，无论是基于循环递归网络还是 GAN 结构，由于采用递归架构，导致训练过程中参数使用效率较低，训练代价较为昂贵。未来研究中，在 SA-GAN 的基础上基于现有方法考虑采用时间卷积[24-25]（TCN, temporal convolutional network）和图注意力结构来优化模型，这将有助于突破架构限制，在此基础上，加入轻量化的人物行为姿态和更为丰富的语义信息处理网络，对复杂交互场景中的行人轨迹、场景和行人姿态进行联合建模，将有望实现行人轨迹和行人姿态的联合预测，从而提高预测的速度、精度和可解释性。

## 参考文献：


[1] DAI M, WANG J, YIN G, et al. Dynamic output-feedback robust control for vehicle path tracking considering different human drivers' characteristics[C]//Technical Committee on Control Theory. Piscataway: IEEE Press, 2017: 1157-1162.

[2] MEHDI M, NIRIASKA P, SIMON G, et al. The walking behaviour of pedestrian social groups and its impact on crowd dynamics[J]. Plos One, 2010, 5(4):e10047.

[3] KITANI K, ZIEBART B, BAGNELL J A, et al. Activity forecasting [C]// Proceedings of the 12th European conference on Computer Vision - Volume Part IV. Berlin: Springer, 2012: 201-214.

[4] 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. Piscataway: IEEE Press, 2017: 336-345.

[5] YAMAGUCHI K, BERG A C, ORTIZ L E, et al. Who are you with and where are you going?[C]//Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2011: 1345-1352.

[6] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.

[7] VEMULA A, MUELLING K, OH J. Social attention: modeling attention in human crowds[C]//2018 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2018: 1-7.

[8] 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. Piscataway: IEEE Press, 2016: 961-971.

[9] GUPTA A, JOHNSON J, FEI-FEI L, et al. Social GAN: socially acceptable trajectories with generative adversarial networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018:2255-2264.

[10] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2672-2680.

[11] SADEGHIAN A, KOSARAJU V, SADEGHIAN A, et al. SoPhie: an attentive GAN for predicting paths compliant to social and physical constraints[J]. ArXiv Preprint, ArXiv: 1806.01482, 2018.

[12] VINEET K, AMIR S, ROBERTO M, et al. Social-BiGAT: multimodal trajectory forecasting using bicycle-GAN and graph attention networks[J]. ArXiv Preprint ArXiv:1907.03395, 2019.

[13] PETAR V, GUILLEM C, ARANTXA C, et al. Graph attention networks[J]. CoRR, abs/1710.10903, 2018.

[14] MNIH V, HEESS N, GRAVES A, et al. Recurrent models of visual attention[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. 2014: 2204-2212.

[15] KINGMA D P, WELLING M. Auto-encoding variational Bayes[J]. ArXiv preprint, ArXiv:1312.6114, 2013.

[16] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2017: 5998-6008.

[17] 孙亚圣, 姜奇, 胡洁, 等. 基于注意力机制的行人轨迹预测生成模型[J]. 计算机应用, 2019, 39(3): 668-674.
SUN Y S, JIANG Q, HU J, et al. Attention mechanism based pedestrian trajectory prediction generation model[J]. Journal of Computer Applications, 2019, 39(3): 668-674.

[18] XU K, BA J, KIROS R, et al. Show, attend and tell: neural image caption generation with visual attention[J]. ArXiv Preprint, ArXiv:1502.03044, 2015.

[19] MIRZA M, OSINDERO S. Conditional generative adversarial nets[J]. ArXiv Preprint, ArXiv: 1411.1784, 2014.

[20] XU Q, HUANG G, YUAN Y, et al. An empirical study on evaluation metrics of generative adversarial networks [J]. ArXiv Preprint, ArXiv:1806.07755, 2018.

[21] PELLEGRINI S, ESS A, VAN GOOL L. Improving data association by joint modeling of pedestrian trajectories and groupings[C]//European Conference on Computer Vision. Berlin: Springer, 2010: 452-465.

[22] LERNER A, CHRYSANTHOU Y, LISCHINSKI D. Crowds by example[J].Computer Graphics Forum. 2007, 26(3): 655-664.

[23] ABDUALLAH M, KUN Q, MOHAMED E, 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. Piscataway: IEEE Press, 2020.

[24] HE Z, CHOW C, ZHANG J. STCNN: a spatio-temporal convolutional neural network for long-term traffic prediction[C]// 2019 20th IEEE International Conference on Mobile Data Management. Piscataway: IEEE Press, 2019: 226-233.


[作者简介]


**李琳辉**（1981– ），男，河南辉县人，博士，大连理工大学副教授，主要研究方向为智能车辆环境感知、规划决策与导航控制等。

**周彬**（1997– ），男，山东临沂人，大连理工大学硕士生，主要研究方向为智能车辆规划决策、轨迹预测等。

**连静**（1980– ），女，吉林公主岭人，博士，大连理工大学副教授，主要研究方向为新能源汽车智能化、轨迹预测等。

**周雅夫**（1962– ），男，辽宁大连人，大连理工大学教授，主要研究方向为新能源汽车动力控制、新能源汽车网联化等。