基于多头注意力对抗机制的复杂场景行人轨迹预测

余 力[1)] 李慧媛[1)] 焦晨璐[1)] 冷友方[1)] 徐冠宇[2)]

[1)](中国人民大学信息学院 北京 100872)

[2)](北京理工大学信息与电子学院 北京 100081)

摘 要 行人轨迹预测对智慧城市建设、公共危机管理具有重要意义.复杂场景中的行人轨迹不仅包含行人个体运动时序性特征,还包含行人与周围其他运动实体之间的交互性特征.如何根据场景变化,对这种时序性和交互性特征进行深度刻画并进行轨迹预测,是复杂场景行人轨迹预测的关键问题.本文采用多头注意力机制和对抗生成方法,提出一种基于多头注意力机制的生成对抗网络模型(Multi-head Attention Generative Adversarial Model,MAGAM),对复杂场景下多行人轨迹进行建模.论文首先通过多头注意力机制融合行人的相对位移信息,从不同方面学习轨迹特征空间中各子空间特征的权重信息,实现对行人之间相互影响的交互性轨迹特征刻画;然后采用对抗生成机制和多轨迹生成策略,实现对复杂场景下不同个体移动轨迹的生成与预测.最后,本文在两个公开的数据集(ETH 和 UCY)进行了实验验证.实验结果表明,在 *ADE*、*FDE* 和 *AnlDE* 三个指标上,本文提出的 MAGAM 模型比基准模型误差平均降低了 26.90%、21.02%和 24.06%.本文对模型的预测结果进行可视化分析,直观展示了本论文模型的合理性.

关键词 复杂场景;轨迹预测;多头注意力;位置编码;对抗生成

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2022.01133

Trajectory Prediction in Complex Scenes Based on Multi-Head Attention Adversarial Mechanism

YU Li[1)] LI Hui-Yuan[1)] JIAO Chen-Lu[1)] LENG You-Fang[1)] XU Guan-Yu[2)]

[1)](*School of Information, Renmin University of China, Beijing* 100872)

[2)](*School of Information and Electronics, Beijing Institute of Technology, Beijing* 100081)

Abstract Pedestrian trajectory prediction plays a vital role in intelligent city construction and public crisis management. Distinct from the single trajectory prediction which rely on strong temporal correlation, in the complex scenes, the trajectory reflects not only the temporal characteristics of a single person, but the interactive features between human and other moving objects nearby. Therefore, how to deeply describe such temporality and interactivity, and then to generate accurate trajectory prediction results according to the change of the scene has become a major problem in the field of trajectory prediction today. In recent years, deep learning has attracted great attention and achieved success in the trajectory prediction tasks. However, most of these methods capture the influence between pedestrians from a single view, and they fail to consider the multiple factors which have an effect on the decision of pedestrians, such as going straight or turning. To this end, in this paper, we propose a multi-head attention generative adversarial model (MAGAM) which combines the multi-head attention mechanism and the generative adversarial network to model the pedestrian trajectory in the complex scenes. Specifically, the

收稿日期:2020-09-21;在线发布日期:2021-01-18.本课题得到国家自然科学基金(71271209)、中央高校基本科研业务基金、中国人民大学研究基金(2020030228)资助.**余 力**(通信作者),博士,副教授,博士生导师,主要研究领域为大数据分析、推荐系统.E-mail:buaayuli@ruc.edu.cn.**李慧媛**,博士研究生,主要研究方向为推荐系统、深度学习.**焦晨璐**,硕士,主要研究方向为深度学习、大数据.**冷友方**,博士研究生,主要研究方向为推荐系统、深度学习.**徐冠宇**,硕士研究生,主要研究方向为深度学习、大数据.

MAGAM model employs multi-head attention mechanism with relative displacement information to learn the attentive weight of subspace features in the whole trajectory feature space on different aspects, to realize the characterization of the interactive trajectory features that resulting from mutual influence between pedestrians. Moreover, the adversarial generation strategy and multi-trajectory generation strategy are used to achieve the reasonable generation of individual moving trajectory in the complex scenes. During the training process, the generator firstly extracts the personalized temporal features of pedestrians from historical observation sequences with long short-term memory (LSTM) based encoders. Secondly, the locations of pedestrians and temporal features are integrated into the multi-head attention model to learn the different weights and output the interactive state of the pedestrians. Thirdly, the interactive states and the Gaussian noise are fed into the LSTM-based decoders to generate multiple prediction trajectories. Then the discriminators are employed to judge whether the input trajectory belongs to the truth trajectory or generated trajectory as much as possible. By training with the adversarial mechanism, we could obtain the approximate truth results when modeling convergences. Finally, in order to estimate the performance of the proposed model, we conduct the experiments on two public datasets (ETH and UCY) which are widely used in the trajectory prediction tasks. We evaluate the prediction results based on three evaluation indicators: the average displacement error, the final displacement error, and the average no-linear displacement error. Compared with the existing trajectory prediction methods, the three metrics of the MAGAM model on all the datasets reduced by 26.90%, 21.02% and 24.06% on average. And the prediction results and the interactive scenes among pedestrians are visualized and analyzed which demonstrates the rationality of the results. Additionally, the performance of the MAGAM model including the average convergence accuracy, the average convergence time and the average prediction time is verified through related experiments, compared with the baselines, the MAGAM model gets the longest convergence time and prediction time.

Keywords complex scenes; trajectory prediction; multi-head attention; positional encoding; adversarial generation

1 引 言

随着大数据和物联网技术的快速发展,轨迹预测技术被广泛应用于自动驾驶[1]、空中交通管理[2]、疏浚工程监管[3]等领域,在智慧城市建设中发挥着越来越重要的作用.轨迹预测是指根据运动实体过去一段时间的路径,预测其未来的轨迹.运动轨迹在不同的场景下会有不同的特点,在特定场景下单一实体的运动轨迹呈现出较强的时间相关性,而在复杂场景中,由于同时存在的多个运动实体之间会相互影响,目标实体的轨迹不仅具有时间相关性,还有实体之间的社会交互性.

城市道路交通属于典型的复杂场景,该场景中存在行人与行人、行人与车辆(机动车和非机动车)的多种交互现象.如图 1 所示,在车辆往来的道路上,为防止车辆与行人发生碰撞,需要预测行人的运动轨迹.如果行人的轨迹是 a,则要控制车辆减速或刹车;如果行人轨迹是 b,则不需要对车辆进行操作,因为在该情况下车辆与行人不会发生碰撞.在高密度区域的城市交通中,行人轨迹预测问题更具复杂性.首先,对于行人个体而言,其行为特征具有一定的内在随机性,路径选择和行走速度等因人而异,使得行人的运动过程呈现多样性,难以得到精准的

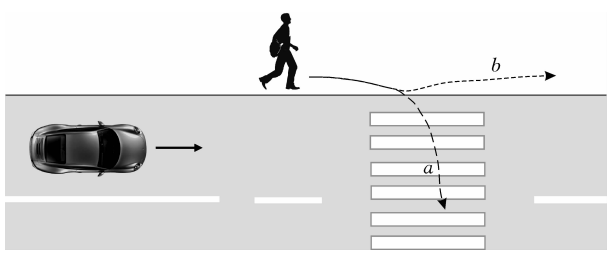

图 1 自动驾驶与行人轨迹预测场景

描述. 其次，行人的运动轨迹会受到周围运动实体的影响，例如在拥挤的道路上，行人会躲避对面走来的人，或者超越前方的行人等. 此外，行人的运动轨迹会受到社交关系的影响，如群体结伴出行时每个成员的行走路线高度相似. 这些因素给行人轨迹预测带来了很大的挑战.

为更好地进行行人轨迹预测，不仅需要考虑复杂空间中行人之间的交互行为建模，还需考虑如何生成多条合理可行的预测路径. 例如，当遇到障碍物时行人要绕行而不能穿过，到达聚会地点应停留而不是继续行走. 针对上述问题，有研究者提出 Social LSTM[4] 方法，利用社交池化融合行人之间的交互信息，采用 LSTM 预测下一时刻的行人轨迹，之后有学者提出 social GAN[5] 方法，结合生成对抗机制[6] 和改进的社交池化操作生成更加合理的轨迹. 近年来，随着注意力机制在翻译任务中的成功应用，有学者将其引入复杂场景轨迹预测任务以获得目标行人对于周围实体的关注程度[7-8]，但该方法只能得到周围实体对目标行人的“平均”影响力，不能细分目标行人在不同因素上对邻居行人的关注程度，如运动速度、加速度、方向、社会关系等.

本文采用多头注意力机制和对抗生成的方法，提出一种基于多头注意力机制的生成对抗网络模型(Multi-head Attention Generative Adversarial Model，MAGAM)预测复杂场景下的行人运动轨迹. 首先，为克服传统注意力机制只能从单一方面学习行人关注度的缺陷，MAGAM 引入融合了位置信息的多头注意力机制，来刻画同一空间中目标行人在多个影响因素下对周围行人的关注. 同时，MAGAM 采用生成对抗机制进行训练，生成器用来生成预测轨迹而判别器鉴别其真假，并用于参数优化. 在生成器中，基于长短记忆网络(Long Short-Term Memory，LSTM)从行人的历史观测轨迹序列中提取行人独特的时序运动特征，利用加入了位置特征的多头注意力机制从多个不同的角度捕捉行人之间的相互影响力，最后联合个人特征和多人交互特征以及高斯噪声的状态向量输入解码器，生成多条不同的轨迹，使得模型能够得到更符合实际的预测结果. 最后，本研究在公开数据集 ETH 和 UCY 上进行了实验.

本文首先在第 2 节对轨迹预测问题进行国内外研究现状综述；第 3 节对行人轨迹预测问题进行形式化定义；第 4 节详细介绍 MAGAM 模型及其组成模块；第 5 节进行实验验证与分析；最后对本文进行总结.

2　国内外研究现状

本部分主要从特定场景下单个体的轨迹预测和复杂场景下的多轨迹预测两个方面介绍国内外研究现状.

2.1　特定场景单个体轨迹预测

在轨迹研究领域中，对单个体的轨迹预测有很多应用场景，例如石庆研等人[2] 基于历史航行数据利用 LSTM 和整合移动平均自回归模型对飞机进行短期航迹预测. 徐婷等人[3] 利用大数据对挖泥船施工轨迹进行识别和预测，解决施工过程无法实时监控的问题. Wan 等人[9] 提出利用 K 近邻和贝叶斯算法，从人的地理位置信息中挖掘其偏好，对旅行者进行旅游路线的个性化推荐. 目前车辆行驶轨迹预测技术已经被广泛应用到自动驾驶的技术研发中，以提高驾驶安全性. Lu 等人[10] 基于出租车的行驶位置、行驶速度、载客量等信息，对出租车的行驶轨迹进行分析与预测，帮助构建更加智能的城市交通网络. 谢枫等人[11] 利用对城市道路交叉路口智能车周围的转弯车辆进行轨迹预测. Graves[12] 利用 LSTM 对手写字体轨迹进行学习与生成，实现用计算机替代人工手写的工作. 上述研究的共同点在于研究的场景较为简单，研究的对象均为单一个体，不受环境中其他个体的影响. 但在复杂场景中，研究对象与周围实体会有交互行为，因此单个体轨迹预测方法不适用于复杂场景的轨迹预测任务.

2.2　复杂场景多轨迹预测

目前关于复杂场景轨迹预测问题的研究方法总体可以分为基于统计模型的预测方法和基于深度学习的预测方法两类.

2.2.1　基于统计模型的预测方法

基于统计模型的预测方法主要通过概率统计模型来反映个体的运动特点. Helbing 等人[13] 提出社会力模型，利用引力和斥力来为行人和空间中移动实体的关系建模. 杨文彦等人[14] 考虑行人的个体差异性，从而改进社会力模型，对人车混行路口的行人行走轨迹进行预测. Trautman 等人[15] 提出交互式高斯过程，利用高斯过程预测每个行人的轨迹，然后根据社会力模型考虑行人之间的相互影响. 这些方法的优点是结构简单、计算效率高，缺点是模型对参数敏感、模型泛化能力差等.

2.2.2　基于深度学习的预测方法

近年来，随着深度学习的发展，以循环神经网络

(Recurrent Neural Network, RNN)[16]、LSTM[17]等为代表的深度学习方法可有效反映个体的时序特征,成为近年来复杂轨迹预测的重要方向. Alahi 等人[4]提出 Social LSTM 模型,利用场景空间网格分割和特征池化来建模行人之间的社会交互性,对行人轨迹进行预测. Zhang 等人[18]设计了一个数据驱动状态细化模块来计算邻居的当前意图,并从临近的行人中选择有用的信息.

为提高模型预测的准确率,有学者采用生成对抗网络(Generative Adversarial Networks, GAN)的方法为行人生成未来轨迹. Gupta 等人[5]将 GAN 引入行人轨迹预测任务,在构建行人之间的交互关系时以人与人之间的距离代替 social LSTM 模型中复杂的网格网络,降低计算开销,并生成多个符合真实场景的预测结果. 此后有学者围绕多轨迹生成和交互信息提取两个方向进行研究. 前者通过在轨迹预测中引入潜码分布,使得模型能够自动学习将运动模式与分布相联系,以生成更多实际可用的轨迹[19-21]. 对于交互信息提取,Kosaraju 等人[22]利用图注意网络编码行人与场景的物理互动;Zhang 等人[20]将行人之间的交互建模成有向边的图结构;Yang 等人[21]提出 TPPO 模型,根据行人移动方向与未来轨迹的相关性,利用社会注意力池化模块聚合邻居来为行人交互建模;张睿等人[8]在生成对抗网络中引入物理注意力机制和社会注意力机制,分别从场景环境和行人交互两个方面进行建模,以获得符合物理限制和社会规范的预测路径;Huang 等人[23]通过时空图注意机制来学习行人交互的时间相关性;毛琳等人[24]设计一种空时社交汇集机制解决 Social GAN 对行人长时社交关系考虑不足的问题,使模型既能保持行人短时社交敏感性,又能增强长时社交关系的记忆;Mohamed 等人[25]利用图网络结构提取行人之间的社交互动,有效地提高了模型的预测精度.

以上这些模型在对行人的社交关系建模时大多采用池化、注意力机制和图神经网络结构,对学习行人与其他实体之间的影响力有重要作用,但其对行人之间复杂关系的挖掘只聚焦于行人表现出的特征上,如距离或角度,没有更深度地从潜在空间中挖掘行人之间多方面的影响因素,如运动速度、群体的社会关系等,这些因素是同时存在的. 本文利用多头注意力机制从多个角度挖掘行人之间细粒度的影响因素,能够更加全面地学习行人之间的交互影响力.

3 轨迹预测问题定义

在视觉领域,载有行人运动轨迹的视频首先被切割成帧图片,通过图像识别技术对每一帧图片中的行人进行识别与定位,得到以帧图片构成的坐标系中每个行人的二维坐标数据(x,y). 因此,行人轨迹预测问题可以转化为时间序列预测问题,其中每个行人的轨迹值可以看作一组坐标序列. 本文的研究问题定义如下.

定义 1(观测轨迹序列). 已知某道路空间中有 N 个行人,其历史轨迹表示为 $\boldsymbol{T}=\{\boldsymbol{T}_1,\boldsymbol{T}_2,\cdots,\boldsymbol{T}_N\}$,其中,行人 i 的观测轨迹为

$$\boldsymbol{T}_i=\{(x_i^t,y_i^t)\in\mathbb{R}^2\mid t=1,2,\cdots,t_o\} \tag{1}$$

定义 2(预测轨迹序列). 已知行人的观测轨迹序列,预测当前场景下 N 个行人的未来运动轨迹 $\hat{\boldsymbol{T}}=\{\hat{\boldsymbol{T}}_1,\hat{\boldsymbol{T}}_2,\cdots,\hat{\boldsymbol{T}}_N\}$. 其中第 i 个行人的预测轨迹序列为

$$\hat{\boldsymbol{T}}_i=\{(\hat{x}_i^t,\hat{y}_i^t)\in\mathbb{R}^2\mid t=t_o+1,\cdots,t_o+t_p\} \tag{2}$$

其中,$t_o\in\mathbb{Z}$ 表示观测轨迹序列的长度;$t_p\in\mathbb{Z}$ 表示预测轨迹序列的长度,且 $t_p\geq 1$.

4 MAGAM 模型

复杂场景中的行人轨迹建模与预测的关键问题是如何对交互情况下的行人路径决策心理进行深度建模,生成多种符合社会规范的轨迹. 针对这些关键问题,本文提出一种基于多头注意力机制的生成对抗网络模型. 本节将从模型的整体框架出发,通过论述模型的具体结构来介绍对行人轨迹的建模和预测过程,最后介绍模型的训练.

4.1 模型整体框架

图 2 为 MAGAM 模型的整体框架. 模型采用生成对抗网络结构,包括一个生成器和一个判别器. 生成器以行人的观测轨迹序列 $\boldsymbol{T}$ 作为输入,学习各行人之间的复杂交互关系并生成预测轨迹 $\hat{\boldsymbol{T}}$. 判别器学习判断某一轨迹是真实轨迹还是由生成器生成的轨迹. 其中,生成器包括 3 个模块,分别为轨迹编码器模块、交互特征提取模块和轨迹解码器模块. 判别器中仅包含轨迹编码器模块. 图 2 中描述了行人数量为 3 时的轨迹训练和预测过程. 生成器以行人轨迹$\{\boldsymbol{T}_1,\boldsymbol{T}_2,\boldsymbol{T}_3\}$作为输入,利用三个由 LSTM 构成的编码器对每个行人的历史观测轨迹进行编码,得到

每个行人的隐层状态$\{h_{e1},h_{e2},h_{e3}\}$；然后聚合行人相对位移信息和隐层状态，通过多头注意力机制从不同方面捕捉各行人之间的交互影响力，得到行人的交互状态$\{MHA_1,MHA_2,MHA_3\}$；解码器以行人的隐层状态、交互状态和高斯噪声为输入，生成行人的未来轨迹$\{\hat{\boldsymbol{T}}_1,\hat{\boldsymbol{T}}_2,\hat{\boldsymbol{T}}_3\}$. 真实轨迹与生成轨迹输入判别器中，经过编码后由分类器对输入轨迹进行鉴别.

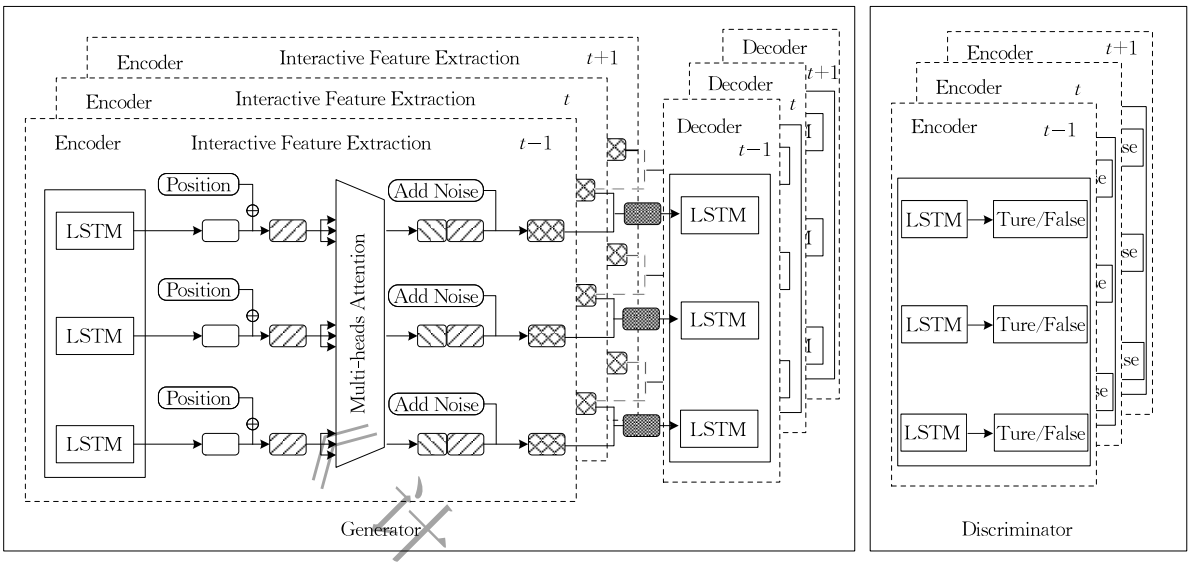

图 2　MAGAM 模型框架

4.2　轨迹编码器模块

本模块对每个行人的历史轨迹进行编码. 行人的轨迹本质上是一组由连续单位时刻下的位置组成的二维时间序列，具有明显的时间相关性. LSTM 是一种特殊的循环神经网络，解决传统 RNN 训练时后向传播中出现的梯度弥散问题[26]. MAGAM 利用 LSTM 对每个行人的历史轨迹序列进行建模，在每一个时刻，LSTM 根据上一个时刻的隐层状态和当前时刻的输入，通过输入门、遗忘门和输出门的非线性计算，得到当前时刻的状态. 因此，LSTM 能够有效地捕获行人在之前每一个时刻的重要相关特征.

对行人 i 的观测轨迹 $\boldsymbol{T}_i$，首先利用多层感知机(Multilayer Perceptron，MLP)将轨迹坐标(x_i^t,y_i^t)映射到特征空间，得到轨迹坐标的位置特征$\boldsymbol{e}_i^t$，

$$\boldsymbol{e}_i^t=\mathrm{MLP}(x_i^t,y_i^t;\boldsymbol{W}_e) \tag{3}$$

其中，$\boldsymbol{e}_i^t\in\mathbb{R}^d$为行人 i 在 t 时刻下的坐标嵌入表示，$i\in\{1,\cdots,N\}$，$t\in\{1,\cdots,t_o\}$，d 表示向量维度，为超参数；$\boldsymbol{W}_e$为 MLP 的参数矩阵. 然后，LSTM 从位置特征空间中学习行人的运动模式. 在 t 时刻，LSTM 编码过程表示为

$$\boldsymbol{h}_{ei}^t=\mathrm{LSTM}(\boldsymbol{e}_i^t,\boldsymbol{h}_{ei}^{t-1};\boldsymbol{W}_{enco}) \tag{4}$$

其中，$\boldsymbol{h}_{ei}^{t-1}$为 $t-1$ 时刻 LSTM 的隐层状态，$\boldsymbol{W}_{enco}$为 LSTM 编码器的权重矩阵. 因此 N 个行人在 t 时刻下的隐层状态$\boldsymbol{H}_e^t$可以表示为

$$\boldsymbol{H}_e^t=\{\boldsymbol{h}_{e1}^t,\boldsymbol{h}_{e2}^t,\cdots,\boldsymbol{h}_{eN}^t\} \tag{5}$$

4.3　交互特征提取模块

本模块旨在利用多头注意力机制深度捕捉行人之间的交互模式. 在上一节中，编码器学习了行人自身的运动特征，但是在复杂的环境中，行人的运动模式往往会受到其他运动实体(如行人、车辆等)的影响而表现出多样性，如超越前面行走的人、躲避从后方驶来的车辆. 行人与场景中的运动实体之间具有很强的交互性，因此在预测某个行人的轨迹时，不仅要考虑该行人自身的运动模式，还应考虑当前时刻下周边其他运动实体的轨迹和运动模式. 为了学习行人①之间的交互关系，需要分析当前时刻下目标行人的路径决策心理，即先根据周围行人的运动特征判断出最具有影响力的个体，然后根据这些个体调整自己下一时刻的路径. 例如，行人会选择躲避对向走来的距离最近的人，或者行走速度最快的人.

多头注意力机制[27](Multi-Head Attention mechanism，MHA)是近年来兴起的一种深度学习技术，最早由谷歌公司提出，并成功应用于机器翻译任务. 多头注意力机制不仅能够编码远距离的依赖关系，还能通过集成不同子空间的信息加强对目标的特征表示. 多头注意力机制可从多个角度挖掘对象间的相互依赖关系，本文将其引入行人轨迹预测任务中，

① 为了方便描述，在下文中用行人代表运动实体.

来深度捕获复杂场景下多个行人的交互关系. 如果将研究场景下的所有行人轨迹序列看作全局空间，将每一个行人的轨迹特征和当前状态视为一个子空间，那么在为每个行人路径决策建模时，多头注意力能够从多个不同的角度捕捉行人之间的相互影响力. 图 3 描述了利用多头注意力机制提取行人交互特征的过程. 由于行人决策过程受到周围其他行人位置的影响，MAGAM 在交互特征提取建模时加入了当前时刻下的行人相对位置信息. 不同于翻译任务中词向量的位置编码方式，本文对行人的相对位置编码如图 4 所示.

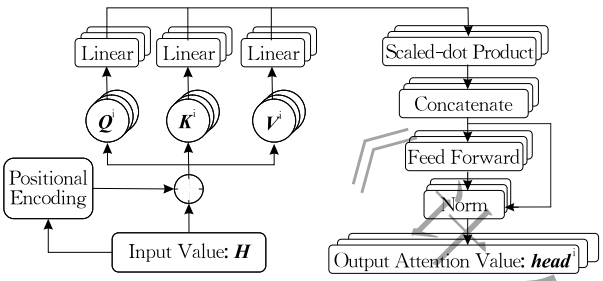

图 3 多头注意力机制的行人交互特征提取

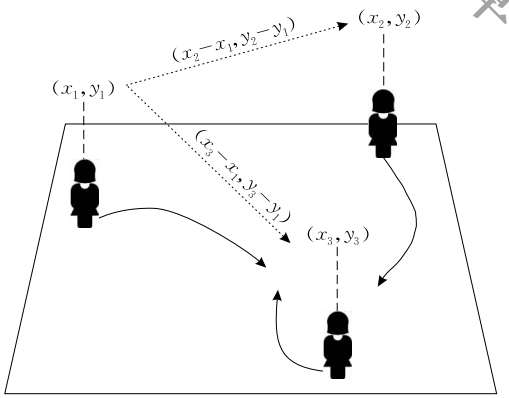

图 4 行人交互的位置编码

在对图 4 中行人 1 的空间位置进行编码时，要综合考虑行人 2 和行人 3 相对于行人 1 的位移. 因此，在t 时刻行人i 的相对位移向量 $\boldsymbol{p}_i^t$可以计算如下：

$$\boldsymbol{p}_i^t=\sum_{j\in \mathrm{N}}\alpha_{ji}^t FC(x_j^t-x_i^t,y_j^t-y_i^t;\boldsymbol{W}_p) \tag{6}$$

$$\alpha_{ji}^t=\frac{s_{ji}^t}{\sum_{j=1}^{N}s_{ji}^t} \tag{7}$$

$$s_{ji}^t=\frac{1}{\sqrt{(x_j^t-x_i^t)^2+(y_j^t-y_i^t)^2}} \tag{8}$$

其中，α_{ji}^t是一个标量，表示行人 j 对行人 i 的影响力，由于距离越远的行人影响越小，因此权重计算为该行人到目标行人欧氏距离的倒数. FC 为全连接层，$\boldsymbol{W}_p\in\mathbb{R}^{d\times 2}$将二维坐标映射到 d 维空间. 因此可以得到相对位移矩阵：

$$\boldsymbol{P}^t=\{\boldsymbol{p}_1^t,\boldsymbol{p}_2^t,\cdots,\boldsymbol{p}_N^t\} \tag{9}$$

多头注意力机制将行人的相对位移矩阵与状态矩阵聚合作为输入，通过线性变换将输入矩阵转换为三个维度相同的矩阵 $\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}\in\mathbb{R}^{N\times d}$.

$$\begin{aligned}\boldsymbol{M}^t&=Agg(\boldsymbol{P}^t,\boldsymbol{H}_e^t)\\&=\{Agg(\boldsymbol{p}_1^t,\boldsymbol{h}_{e1}^t),\cdots,Agg(\boldsymbol{p}_N^t,\boldsymbol{h}_{eN}^t)\}\end{aligned} \tag{10}$$

$$\boldsymbol{Q}^i=\boldsymbol{M}^t\boldsymbol{W}^{q,i} \tag{11}$$

$$\boldsymbol{K}^i=\boldsymbol{M}^t\boldsymbol{W}^{k,i} \tag{12}$$

$$\boldsymbol{V}^i=\boldsymbol{M}^t\boldsymbol{W}^{v,i} \tag{13}$$

其中，Agg 表示将 t 时刻的相对位移矩阵与状态矩阵的对应元素相加，$\boldsymbol{M}^t\in\mathbb{R}^{N\times d}$为聚合后得到的状态-位移矩阵. $\boldsymbol{W}^{q,i},\boldsymbol{W}^{k,i},\boldsymbol{W}^{v,i}\in\mathbb{R}^{d\times d}$分别为第 $i(i\in\{1,\cdots,l\})$个线性变换下 $\boldsymbol{Q}^i,\boldsymbol{K}^i,\boldsymbol{V}^i$对应的转移矩阵. l 个线性变换能够在 l 个不同的角度捕捉行人之间的相互影响力. $\boldsymbol{Q}^i,\boldsymbol{K}^i,\boldsymbol{V}^i$首先进行缩放点积注意力计算得到第 i 个方面的注意力值矩阵：

$$\begin{aligned}\boldsymbol{head}^i&=Attention(\boldsymbol{Q}^i,\boldsymbol{K}^i,\boldsymbol{V}^i)\\&=\mathrm{softmax}\left(\frac{\boldsymbol{Q}^i\cdot\boldsymbol{K}^i}{\sqrt{d}}\right)\cdot\boldsymbol{V}^i\end{aligned} \tag{14}$$

其中 softmax 函数根据 $\boldsymbol{Q}^i$计算 $\boldsymbol{K}^i$中每个行人的权重，d 是向量维度，分母中$\sqrt{d}$是对权重进行缩放，避免维度太高时点积的结果过大. 然后，拼接 l 个注意力值矩阵，通过线性变换聚合多个注意力下行人的相互影响力，

$$MHA^t(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=Concat(\boldsymbol{head}^1,\cdots,\boldsymbol{head}^l)\cdot\boldsymbol{W}^o \tag{15}$$

其中 MHA^t表示通过多头注意力机制得到的 t 时刻下行人交互状态矩阵. $Concat$ 表示连接操作，$\boldsymbol{W}^o\in\mathbb{R}^{ld\times d}$为变换矩阵.

4.4 轨迹解码器模块

本模块利用基于 LSTM 的解码器来预测行人 i 的未来轨迹. 为了体现空间中行人 i 与其他行人之间的时序交互关系，MAGAM 在解码之前，拼接由编码器和相对位移信息得到的状态-位移向量 $\boldsymbol{M}_i$与由多头注意力机制获取的行人交互状态 MHA_i. 此外，为了能够获得多条不相同且合理的预测路径，同时增强模型的泛化能力，降低轨迹数据稀缺和模型欠拟合带来的影响，MAGAM 在解码器的状态中加入随机高斯噪声 z. 考虑到行人之间的影响力会随着邻居行人状态变化而发生改变，MAGAM 将所有行人前一时刻 $t-1$ 和当前时刻 t 的交互向量进行拼接，利用池化操作获得行人的时序交互特征. 其具体过程见式(16)～(18).

$$\boldsymbol{g}_i^t = FC([\boldsymbol{M}_i^t, MHA_i^t]; \boldsymbol{W}_c) \tag{16}$$

$$\boldsymbol{h}_{di}^t = [\boldsymbol{g}_i^t, \boldsymbol{z}_i^t] \tag{17}$$

$$\boldsymbol{P}_i = \boldsymbol{FP}(\boldsymbol{h}_{d1}^{t-1}, \cdots, \boldsymbol{h}_{dN}^t) \tag{18}$$

其中，FC 为全连接层，通过权重 $\boldsymbol{W}_c$ 将增强状态向量映射到 d 维空间. 因此，解码器状态向量的构成包括三种信息，即行人的位移状态、交互状态和高斯噪声，解码器利用这些状态特征对目标行人的未来轨迹进行预测. 过程见式(19)和式(20).

$$\boldsymbol{h}_{di}^t = \mathrm{LSTM}(\boldsymbol{FC}(\boldsymbol{P}_i, \boldsymbol{h}_{di}^{t-1}), \boldsymbol{e}_i^{t-1}; \boldsymbol{W}_d, \boldsymbol{W}_{deco}) \tag{19}$$

$$\hat{T}_i^t = (\hat{x}_i^t, \hat{y}_i^t) = MLP(\boldsymbol{h}_{di}^t; \boldsymbol{W}_m) \tag{20}$$

其中，$\boldsymbol{h}_{di}^{t-1}$ 是 $t-1$ 时刻解码器的状态向量，$\boldsymbol{W}_d$ 和 $\boldsymbol{W}_{deco}$ 为参数矩阵. MLP 是以 ReLU 作为激活函数的多层感知机，预测行人 i 的未来轨迹 $(\hat{x}_i^t, \hat{y}_i^t)$，$\boldsymbol{W}_m$ 为 MLP 的参数矩阵.

4.5　生成器与判别器

如上文所述，MAGAM 模型的整体结构是一个生成对抗网络，生成器包括 3 个模块，基于 LSTM 模型的轨迹编码器和解码器模块以及基于多头注意力机制的行人交互特征提取模块. 生成器通过 3 个模块的计算能够为每个行人生成未来时刻的运动轨迹. 然后，判别器用来鉴别其输入的轨迹是真实轨迹还是由生成器生成的轨迹. 判别器中包含一个基于 LSTM 的编码器模块，其结构与生成器编码器相同. 行人的真实轨迹和生成轨迹输入判别器中，在经过编码器进行编码后，由 Softmax 分类器计算输入轨迹是真实轨迹的概率，进行轨迹分类.

4.6　模型训练

MAGAM 模型使用生成对抗机制进行模型训练. 在最大最小化博弈中，生成器 G 根据行人的历史轨迹 T 和先验噪声 z 尽可能产生与真实轨迹一致的生成轨迹 $G(T,z)$，以使判别器无法正确分类；而判别器 D 尽可能正确鉴别出两者. 生成器与判别器交替进行对抗训练，当模型达到最优时，生成器能够生成令判别器无法分辨真假的轨迹. MAGAM 在训练时的损失函数由两部分构成，对抗损失以及预测轨迹与真实轨迹的 $L2$ 损失

$$\mathcal{L} = \mathcal{L}_{\mathrm{GAN}} + \lambda \mathcal{L}_{L2} \tag{21}$$

$$\mathcal{L}_{\mathrm{GAN}} = \min_G \max_D \mathbb{E}_{T \sim p_{\mathrm{data}}([x,y])}[\log D(T)] + \mathbb{E}_{z \sim p(z)}[\log(1 - D(G(T,z)))] \tag{22}$$

$$\mathcal{L}_{L2}(G) = \min_k \| T - G(T,z)^k \|_2 \tag{23}$$

其中，λ 和 k 均为超参数，$\mathbb{E}$ 为期望. 在训练过程中，判别器要尽可能对真实轨迹的输出概率接近 1，对生成轨迹的输出概率接近 0，因此需要最大化判别器损失，而生成器则相反. 为了使生成器生成的轨迹多样化，与 Social GAN 类似，本文对生成轨迹进行 k 次采样，选择损失最小的输出进行反向传播.

5　实　验

本节首先介绍实验使用的数据集、实验设置、评价指标和基准模型，然后对比和讨论 MAGAM 与基准模型的误差结果，并对 MAGAM 预测结果进行可视化分析. 此外，本文探索 MAGAM 模型用于车辆轨迹预测的效果，并对预测结果进行可视化分析. 接着，本文对 MAGAM 与基准模型的计算耗时进行对比和讨论. 最后对本节实验进行总结.

5.1　数据集

本文采用两个公开数据集 ETH[5] 与 UCY[5] 来验证 MAGAM 模型的预测准确率. 数据均来自于真实的街道监控视频，视频中记录了鸟瞰视角下的道路行人轨迹以及多种行人交互场景，如行人避让对向的来人；多人同向而行，聚集或分散等. ETH 数据集包括两个场景，其中 Eth 视频记录了从斯特恩沃茨街上“ETH 中心”主楼顶层俯瞰人行道路的景象，视频时长共计约 8 min，标注了约 750 个不同的行人；Hotel 视频记录了从巴哈霍夫斯特街上的一家酒店 4 楼俯瞰人行道路的景象，视频共计约 13 min，标注了约 750 个不同的行人. UCY 数据集包含两个场景，ZARA 和 University. ZARA 数据集记录了从“ZARA”服装店门口通过的行人景象，视频时长共计 13 min，标记了约 350 个不同的行人；University 记录下塞浦路斯大学内一条学校道路上的景象，视频时长共计 7 min，标记到约 620 个不同的学生. 表 1 描述了数据集的详细情况.

表 1　ETH 与 UCY 数据集

数据集	行人数量	描述	平均路径长度	分辨率	密度	行数
Eth	750	Outdoor on a path	123.16	768×576	4.150	8908
Hotel	750	Outdoor at hotel entrance	16.04	768×576	5.935	6544
ZARA	350	At shopping streets	42.74	720×576	7.575	4249
University	620	From student pedestrians	3.22	720×576	10.780	14115

5.2　实验设置

本实验利用历史观测轨迹数据训练模型，然后对行人未来的轨迹进行预测. 首先将每个场景下的行人轨迹数据集以比例 5∶4∶1 划分为训练集、验证

集和测试集，并以 8 个时间步长进行分割，每个时间步长为 0.4 s. 实验中输入前 8 个时刻的观测轨迹，预测未来 8 个时刻的行人轨迹. 为了验证多样本生成机制的效果，实验设置了不同的样本生成个数，在测试时通过多次取样，分别选择样本个数为 1、10 和 20 进行模型评估. 模型参数设置方面，生成器的编码维度和解码维度均为 32，学习率为 1×10^{-3}. 多头注意力机制中的注意力数目为 4. 高斯噪声维度为 8. 训练中采用 Adam 优化器作为损失函数梯度下降的优化算法，其权重衰减系数为 1×10^{-3}. 训练的迭代次数设为 200，批处理大小为 32. 本次实验采用 Pytorch 框架，在 NVIDIA GPU GTX1080 Ti 设备上完成.

5.3 评价标准

本研究采用以下 3 个指标来评估 MAGAM 模型：

(1) 平均位移误差(Average Displacement Error, ADE). 表示所有行人的真实轨迹坐标与预测轨迹坐标之间的平均 $L2$ 误差，其计算公式如下：

$$ADE=\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{T}_i-\hat{\boldsymbol{T}}_i)^2 \tag{24}$$

$$(\boldsymbol{T}_i-\hat{\boldsymbol{T}}_i)^2=\frac{1}{t_p}\sum_{t=t_o+1}^{t_o+t_p}[(x_i^t-\hat{x}_i^t)^2+(y_i^t-\hat{y}_i^t)^2] \tag{25}$$

其中 t_o 表示观测轨迹序列长度，t_p 表示预测轨迹序列长度.

(2) 终点位置误差(Final Displacement Error, FDE). 表示所有行人的预测轨迹终点坐标与真实轨迹终点坐标之间的平均平方误差，公式如下：

$$FDE=\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{T}_i^{t_p}-\hat{\boldsymbol{T}}_i^{t_p})^2 \tag{26}$$

$$(\boldsymbol{T}_i^{t_p}-\hat{\boldsymbol{T}}_i^{t_p})^2=(x_i^{t_p}-\hat{x}_i^{t_p})^2+(y_i^{t_p}-\hat{y}_i^{t_p})^2 \tag{27}$$

(3) 平均非线性位移误差(Average no-linear Displacement Error, AnlDE). *AnlDE* 是 *ADE* 的变体，表示仅考虑轨迹中非线性区间的平均 $L2$ 误差. 在实验中大部分预测的误差会出现在非线性轨迹中，如绕行、掉头或是穿插行走等. 为了验证 MAGAM 模型对复杂场景中非线性轨迹的预测效果，本文基于设定的阈值筛选出非线性轨迹，计算其 $L2$ 误差.

5.4 基准模型

本文使用 3 个经典的轨迹预测模型作为对比，来验证 MAGAM 模型的有效性，分别为 LSTM、social LSTM 以及 social GAN.

(1) LSTM 模型从每个行人的观测轨迹中学习运动模式，预测其未来轨迹. 实验中将本文模型进行简化，移除判别器和交互特征提取模块，仅保留了 LSTM 的编码器，特征提取以及 LSTM 的解码器，为每个行人轨迹进行独立建模.

(2) social LSTM 模型(S_LSTM)由 Alahi 等人提出，利用“social”机制提取行人之间的交互特征. 实验中将每个行人的轨迹单独建模，并在每一个时间步长下分割场景，将处在附近网格内的行人的隐层状态进行池化.

(3) Social GAN 模型(S_GAN)由 Gupta 等人提出，利用对抗生成机制进行训练，并使用池化方法提取行人之间的交互关系. 实验中基于构建好的 S_LSTM 网络，将 LSTM 建模过程放入生成器，并在生成器中实现池化过程，最终利用判别器优化轨迹结果.

5.5 实验结果与分析

5.5.1 行人轨迹预测误差对比分析

本实验将 16 个时间单位下的轨迹序列作为一组，前 8 个时刻作为历史轨迹数据，模型输出后 8 个时刻的轨迹，根据输出的预测值和真实值计算 *ADE*、*FDE* 以及 *AnlDE* 指标，并与基准模型进行比对，详细实验结果如表 2 所示，具体有以下几方面的重要结论：

(1) 从整体性能上来看，MAGAM 模型相比 LSTM、S_LSTM 和 S_GAN 模型，在 *ADE* 指标上平均误差分别降低 30.61%、29.17% 和 20.93%，在 *FDE* 指标上降低了 24.74%、26.26% 和 12.05%，在 *AnlDE* 指标上降低了 34.38%、31.15% 和 6.67%. 在 *ADE*、*FDE* 和 *AnlDE* 三个指标上，MAGAM 模型比基准模型误差平均降低了 26.90%、21.02% 和 24.06%. 上述结果表明，相比于基准模型，MAGAM 在整体性能上有所提高.

(2) 从 *ADE* 指标对比可以看出，在大多数场景下 LSTM 模型的预测误差比 S_LSTM 模型高，表明在复杂场景中行人之间的“社交性”对行人轨迹有重要影响，在建模时考虑社交因素能够提高预测轨迹的准确性. S_GAN 和 MAGAM 模型的误差低于 LSTM 和 S_LSTM 模型，证明通过对抗生成网络的训练，模型能够生成更准确的行人轨迹. S_GAN 和 MAGAM 模型都考虑了交互性和对抗生成，但在对行人交互性建模时，S_GAN 模型利用了社交池化方法，在历史轨迹序列最后一个时刻将其他行人的状态特征经过池化进行融合，而 MAGAM 模型在每个时刻利用多头注意力机制在交互共享的基础上加入行人之间的相关性影响因子，从多个不同的方面计算行人之间的注意力程度. 在同样生成 20 个

表 2　基准模型之间的评价指标结果对比

指标	Dataset	LSTM	S_LSTM	S_GAN-VP20	MAGAM		
					Samples=1	*Samples*=10	*Samples*=20
ADE	Zara1	0.27	0.27	0.22	0.28	0.23o	**0.20**(13.04%)
	Zara2	0.33	0.30	0.27	0.23	0.20	**0.18**
	University	0.35	0.49	0.46	0.41	**0.34**	0.35
	Eth	0.93	0.87	0.69	0.71	0.67	**0.63**
	Hotel	0.55	0.47	0.50	0.54・	0.41(24.07%・)	**0.36**
	Average(↓)	0.49(30.61%)	0.48(29.17%)	0.43(20.93%)	0.43▽	0.37(13.95%▽)	**0.34**(8.11%)
FDE	Zara1	0.58	0.63	0.45	0.52	0.46	**0.41**
	Zara2	0.68	0.65	0.57	0.42	0.41	**0.36**
	University	**0.73**	1.04	0.96	1.06	1.01	0.75
	Eth	1.76	1.70	**1.20**	1.32	1.30	1.29
	Hotel	1.09	0.94	0.98	1.05	**0.80**	0.82
	Average(↓)	0.97(24.74%)	0.99(26.26%)	0.83(12.05%)	0.87	0.80	**0.73**
AnlDE	Zara1	0.35	0.31	0.20	0.27	0.21	**0.17**
	Zara2	0.33	0.28	0.18	0.19	0.18	**0.15**
	University	0.58	0.68	0.50	0.55	0.51	**0.46**
	Eth	1.81	1.69	1.29	1.28	**1.26**	1.27
	Hotel	0.14	0.08	0.07	0.07	0.06	**0.04**
	Average(↓)	0.64(34.38%)	0.61(31.15%)	0.45(6.67%)	0.47	0.44	**0.42**

预测轨迹的情况下，MAGAM 模型的预测误差比 S_GAN 模型低，说明多头注意力机制能够更好地捕捉行人之间的交互特征，从而提高模型的准确率. 同时，在 *FDE* 指标上以上模型也具有相同的结论.

(3) 在 University 场景中 LSTM 模型效果最好，其 *ADE* 和 *FDE* 指标值低于其他模型. 通过分析视频我们可以发现，University 场景中的学生行人大多数为成群聚集站立状态，其移动的幅度相对较小，长时间的行走轨迹相对偏少. 在该场景下社交关系的影响较大，S_LSTM 和 S_GAN 模型对于行人交互的建模可能会误将距离较近的朋友当成需要避让的行人，导致预测误差较大. 相比之下，MAGAM 模型的 *ADE* 和 *FDE* 误差与 LSTM 模型较为接近，说明 MAGAM 模型在利用多头注意力机制学习行人的交互特征时能够学习到行人的社会关系带来的影响，对于轨迹的刻画倾向于个体运动特征.

(4) 对比各模型在 *AnlDE* 指标上的结果可以看出，MAGAM 模型在非线性轨迹的预测中误差最小. 由于非线性轨迹普遍是由行人交互影响产生，多头注意力机制在建模时可以更深层次、多角度地挖掘行人之间的交互关系，从而提高预测效果. 纵观每个场景的非线性轨迹误差值，可以发现 Eth 场景的预测误差最大，Hotel 场景的误差最小. 对此，本文猜测这是因为 Eth 场景中行人之间的交互行为较多而 Hotel 场景中交互行为较少. 通过原视频可以发现，Eth 场景为大楼入口和人行道形成“丁”字形交叉口，出入大楼的行人与人行道上行走的行人有较多的交互. 而 Hotel 场景中道路宽阔，行人之间的距离较远，交互较少，行人的轨迹大多是线性的.

(5) 对比 MAGAM 模型在生成轨迹数量分别为 1、10 和 20 时的误差，可以发现，当样本生成轨迹数量从 1 增长为 10 时，MAGAM 模型的预测误差随之下降. 以 *ADE* 指标为例，MAGAM(*Samples*=10)比 MAGAM(*Samples*=1)的平均误差降低 13.95%；最大误差降低出现在 Hotel 数据集上，其误差降低 24.07%. 当生成轨迹数量从 10 增加至 20 时，MAGAM 模型的预测效果有进一步的提升，但误差下降程度相对缩小. 以 *ADE* 指标为例，MAGAM(*Samples*=20)比 MAGAM(*Samples*=10)的误差平均下降 8.11%. MAGAM 模型在不同生成轨迹数量的实验结果，一方面说明多样本生成机制能够产生更多合理的轨迹，有效降低预测误差，另一方面，在一定范围内提高生成轨迹数量能够提高预测效果，但数量过大反而会降低模型的准确率. 例如在 Hotel 场景中轨迹数量从 10 增加到 20，MAGAM 模型的 *FDE* 指标值增长. 因此生成轨迹数量的取值应当适中，且应当根据不同的场景情况进行调整.

5.5.2　行人轨迹预测结果可视化分析

图 5 展示了 MAGAM、LSTM、S_LSTM 和 S_GAN 模型的预测轨迹与真实的行人轨迹在 UCY 数据集中的 ZARA 场景下的对比图，其中左边为二维坐标系中的可视化对比结果，实线表示历史轨迹，虚线表示预测轨迹；右边为真实场景中的可视化对比结果. 在预测模型中，S_GAN 和 MAGAM 模型均产生 20 个轨迹，选择其中误差最小的轨迹进行可视化.

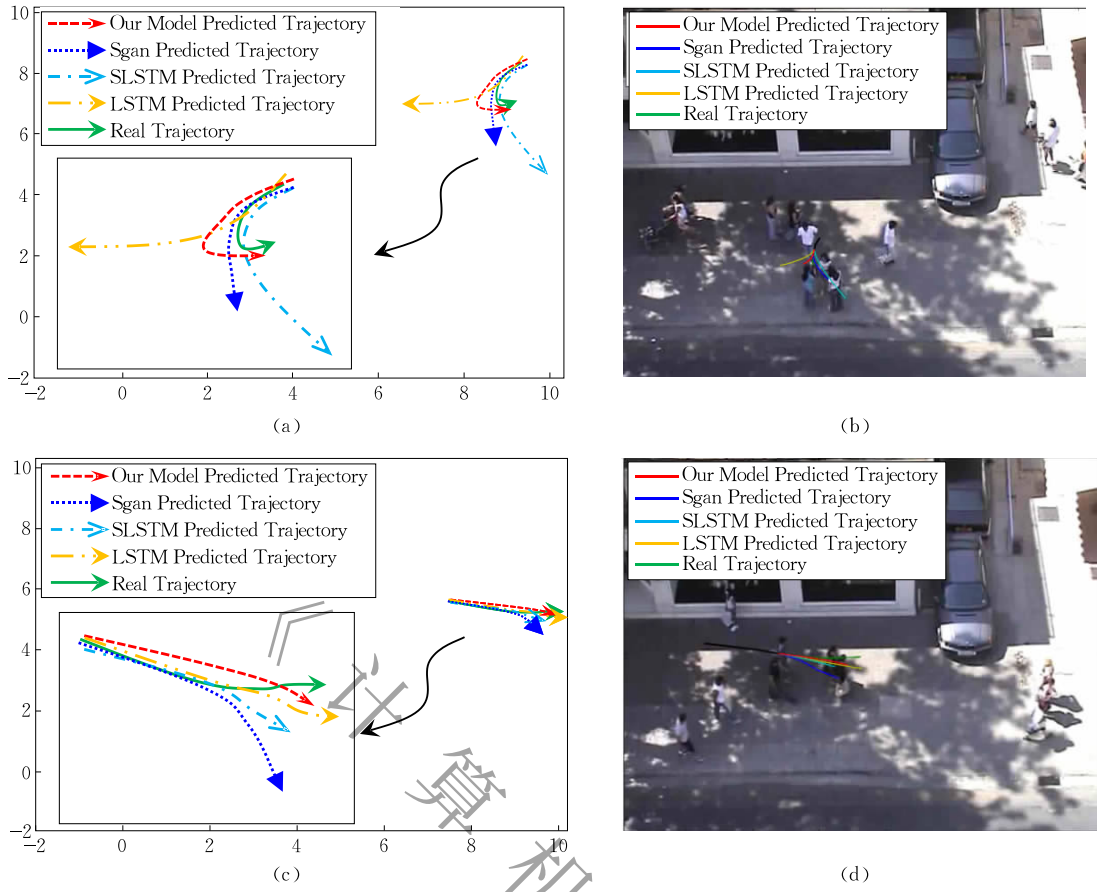

图 5 行人轨迹预测结果可视化

从图 5 中可以看出，不同模型预测轨迹的偏差首先体现在方向上，其次，从轨迹的长短可以侧面体现对行走速度的不同预测结果. 对比不同时刻的结果可以看出，不论是在行走方向还是行走速度上，MAGAM 模型预测轨迹与真实轨迹是最接近的. 图 5(a)与图 5(b)表明，MAGAM 模型可以成功地预测出目标行人遇到前面聚集的人群后减速停留的行为；S_GAN 模型预测该行人走进聚集的人群中，S_LSTM模型预测该行人为躲避人群而转向行走，而 LSTM 预测行人加速穿过聚集人群，后 3 种预测结果与实际情况不相符. 在图 5(c)与图 5(d)所示，目标行人加速超过前方行人时会朝着偏离前方行人的方向前进. 从实验结果可以发现，MAGAM 模型在预测方向偏移的情况下，能够产生与真实轨迹速度和方向最为相似的结果，而 S_GAN 虽然同样预测躲避与偏移方向，但其方向却是偏右，而真实是偏左.

为了验证 MAGAM 模型在复杂场景中对多行人交互行为的建模效果，本实验将相同时刻下 4 个行人的预测轨迹进行可视化展示，并利用热图对场景中的“交互性”进行可视化验证，如图 6 所示. 其中左侧为 4 个行人轨迹交互的热图，颜色越明亮的地方行人越聚集，右侧为 4 个行人预测轨迹分别在 $t=\{2,4,6,8\}$ 时刻的可视化结果. 如在图 6(a)中，在 $t=2$ 时，行人 3 与行人 4 相向而行，双方感知对面来人，模型预测到这两人的“避让”意图，即行人 3 与 4 均稍微偏向前其进方向的左侧；同时，行人 1 与行人 2 的预测路径稍微偏向前进方向的左侧，说明模型同样预测到两人的“避让”行为. 到 $t=4$ 时，行人 3 与行人 4 的距离靠近，且与行人 1 相向而行，此时行人 3 同时关注行人 4 与行人 1，但更加关注距离最近的行人 4；行人 1 与行人 2 完成了“避让”过程. 到 $t=6$ 时，行人 3 绕过了行人 4，遇到了迎面而来的行人 1，此时行人 3 对于行人 1 的关注最大，模型预测行人 3 与行人 1 之间的避让意图，两人的轨迹均向右侧偏移.

5.5.3 道路车辆轨迹预测可视化分析

现实中道路车辆拥挤情况更为常见，为保证安全行驶，驾驶员需要关注道路中的其它移动物体，车辆与其它移动实体之间同样存在交互行为. 为探索 MAGAM 模型的可扩展性，本节将 MAGAM 模型应用于复杂场景道路车辆轨迹预测. 道路车辆数据

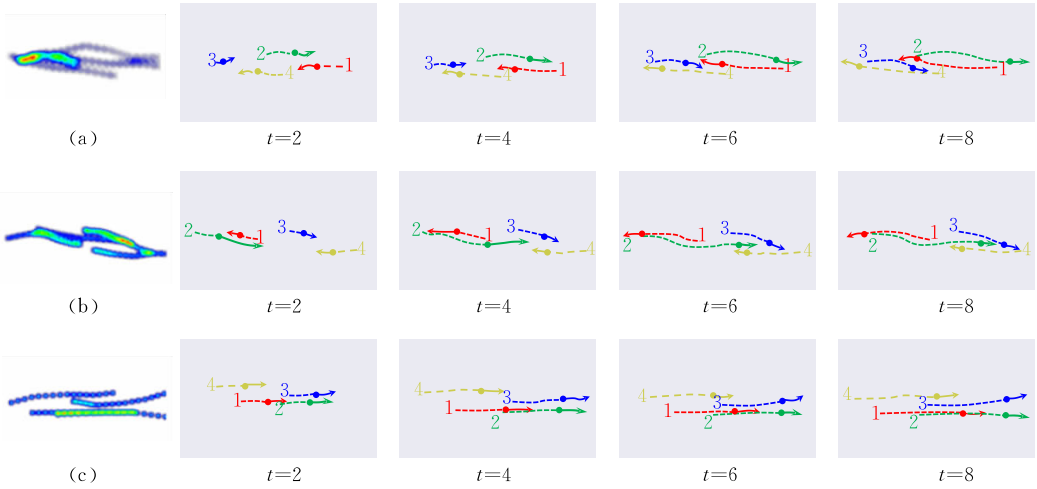

图 6　多行人交互场景可视化结果

来自上海市通阳路 250 弄通阳公寓街道的一段道路监控视频，该视频摄制于 2019 年 4 月 22 日的 14 时左右，视频总时长约 10 min. 将视频以 1.2s/帧进行切割，定位图像中识别到的车辆经纬度坐标并记录对应的车辆 ID.

实验结果如图 7 所示. 在图 7(a)中，目标车辆周围存在 3 个交互车辆，从历史行驶轨迹看出，该车先靠右行驶躲避其左前方的预变道车辆，随后向左偏移以躲避停靠在右侧的车辆. MAGAM 模型预测出该车将靠左变道行驶，躲避并超越前方车辆. 图 7(b)中自行车和电动车的预测轨迹显示其能够躲避停靠在路边的银色私家车，预测轨迹是合理的. 图 7(c)中骑着电动车的人关注前方的车辆和行人，模型预测电动车从车辆和行人之间穿过. 图 7(d)中自行车先靠左绕行避过停靠路边的私家车，随后靠右边骑行，躲避行驶而来的大货车.

图 7　道路车辆轨迹预测结果可视化

5.5.4　模型性能对比

本节将对 MAGAM 模型和基准模型的性能进行进一步实验比对和讨论. 为了在实验过程中保持基准模型的超参数的一致性，本文基于实验设置中模型参数值对基准模型的参数进行定义与赋值，包括迭代次数和学习率等. 在模型的训练和预测过程

中,本实验将以 3 种性能指标值作为模型性能对比依据.在模型训练时,实验将观测并记录模型的平均收敛精度和平均收敛耗时;在模型测试过程中,实验以平均测试耗时为测试性能指标.在这些指标中,平均收敛精度表示模型达到收敛状态时在验证集上测试的准确度;平均收敛耗时表示从模型训练开始到模型达到收敛状态所消耗的时间,单位为小时(h);平均预测耗时表示从模型开始测试到测试完成的所消耗的时间,单位为秒(s).

从表 3 可以看出,模型结构越复杂,模型训练时所消耗的时间越多,模型达到收敛状态的平均耗时越长.另外,随模型复杂度的增强,在同一个测试集下,模型测试的平均耗时也越长.这表示模型越复杂,需要花费更多的时间来进行模型训练以及模型测试.在平均收敛精度方面,从表中第一列可以看出,模型达到收敛状态后,模型测试的准确率自上而下逐渐提高,且模型 MAGAM 在验证集上得到最高的精度,对比其他基准模型,MAGAM 模型效果最好.

表 3 模型性能对比

模型	平均收敛精度	平均收敛耗时/h	平均预测耗时/s
LSTM	0.338	1.34	4.78
S_LSTM	0.310	1.76	4.99
S_GAN_VP20	1.022	2.21	39.85
MAGAM	1.073	4.06	120.03

5.5.5 实验讨论

为验证 MAGAM 模型的性能和可扩展性,本文不仅进行了行人轨迹预测实验,还将模型应用于道路车辆轨迹的预测任务中,均得到了合理的预测路径.此外,实验对 MAGAM 模型与基准模型的时间成本进行比较,MAGAM 模型总耗时最长,模型精度最高.因此若将该模型用于自动驾驶和行人轨迹预测,能够得到更高的准确度,但需要更高的成本.通过对 MAGAM 模型与基准模型的,我们得出以下结论:

首先,多头注意力机制能够更深度更全面地捕捉多个行人之间的交互影响力.多头注意力机制,利用多个注意力在不同的特征空间中计算行人之间影响力的权重,融合多重特征空间的权重作为行人的影响力,即当前行人在此刻场景中对其他行人的关注程度.为了使多头注意力机制更好地学习交互特征,在输入向量中加入了行人的相对位移信息,因此,输入向量中不仅包括个人运动特征,还包括了行人之间的信息.实验结果证明,多头注意力机制能够更好地刻画行人的路径决策心理,帮助提高 MAGAM 模型对于行人轨迹的预测准确率.

其次,利用对抗训练的多轨迹生成方法能够预测出符合实际的轨迹.MAGAM 利用生成对抗机制进行训练,模型在生成器中拟合预测轨迹,并在判别器中结合真实轨迹对生成轨迹进行校验,将误差回传到生成器进行参数优化以达到更好的收敛结果.为了产生多条合理可行的轨迹,模型对生成轨迹添加了高斯噪声,提高生成器和判别器对噪声的敏感度,并在一定程度上降低了生成的多条轨迹之间的相似度.通过实验结果可知,添加高斯噪声的多轨迹生成机制能够帮助降低预测轨迹和真实轨迹之间的误差,实现更准确更符合实际场景的轨迹预测.

6 总 结

针对复杂场景中行人轨迹的多样性与行人之间复杂的交互关系,本文提出了基于多头注意力对抗机制的行人轨迹预测模型.首先,利用 LSTM 提取每个行人的轨迹特征,然后通过多头注意力机制从融合了行人相对位移信息的位移-状态矩阵中学习多个行人之间的交互关系,再对加入了噪声的特征向量进行解码得到预测轨迹.MAGAM 模型基于对抗生成机制和高斯噪声来拟合生成多种可行的行人轨迹.为验证 MAGAM 模型的性能,本文不仅进行了行人轨迹预测实验,还将模型应用于道路车辆轨迹的预测任务中,均得到了合理的预测路径.但相比于 social LSTM 和 social GAN 模型,MAGAM 的计算成本较高,未来我们将探索优化方法降低计算消耗的时长.此外,在实际应用中,需要根据具体的场景以及目标物体移动轨迹的特征,对模型的局部算法结构进行改进和优化,从而实现对复杂场景中的预测目标更全面的特征挖掘.

参 考 文 献

[1] Large F, Vasquez D, Fraichard T, et al. Avoiding cars and pedestrians using velocity obstacles and motion prediction// Proceedings of the IEEE Intelligent Vehicle Symposium. Los Alamitos, USA, 2004: 375-379

[2] Shi Qing-Yan, Yue Ju-Cai, Han Ping, et al. Short-term flight trajectory prediction based on LSTM-ARIMA model. Journal of Signal Processing, 2019, 35(12): 2000-2009(in Chinese)
(石庆研,岳聚财,韩萍等. 基于 LSTM-ARIMA 模型的短期航班飞行轨迹预测. 信号处理, 2019, 35(12): 2000-2009)

[3] Xu Ting, Dai Wen-Bo, Lu Jia-Jun. Identification and prediction of ship construction path based on AIS big data. Port & Waterway Engineering, 2019, (12): 119-126(in Chinese)
(徐婷，戴文伯，鲁嘉俊. 基于自动识别系统大数据的船舶施工轨迹识别与预测. 水运工程，2019，(12)：119-126)

[4] Alahi A, Goel K, Ramanathan V, et al. Social LSTM: Human trajectory prediction in crowded spaces//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 961-971

[5] Gupta A, Johnson J, Li F-F, et al. Social GAN: Socially acceptable trajectories with generative adversarial networks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA, 2018: 2255-2264

[6] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets//Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, USA, 2014: 2672-2680

[7] Sun Ya-Sheng, Jiang Qi, Hu Jie, et al. Attention mechanism based pedestrian trajectory prediction generation model. Journal of Computer Applications, 2019, 39(3): 668-674(in Chinese)
(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. 计算机应用，2019，39(3)：668-674)

[8] Zhang Rui, Wu Bo-Xiong, Zhang Li-Yuan, Zhang Bo. Human trajectory prediction method for complex scenes. Computer Engineering and Applications, 2021, 57(6): 138-143(in Chinese)
(张睿，吴伯雄，张丽园，张博. 复杂场景下行人轨迹预测方法. 计算机工程与应用，2021，57(6)：138-143)

[9] Wan Lin, Hong Yu-Ming, Huang Zhou, et al. A hybrid ensemble learning method for tourist route recommendations based on geo-tagged social networks. International Journal of Geographical Information Science, 2018, 32(11): 2225-2246

[10] Lu Feng, Duan Ying-Ying, Zheng Nian-Bo. A practical route guidance approach based on historical and real-time traffic effects//Proceedings of the 2009 17th International Conference on Geoinformatics. Fairfax, USA, 2009: 1-6

[11] Xie Feng, Li Yong-Le, Su Zhi-Yuan, et al. A method for predicting turning vehicle trajectory in urban intersection. Journal of Military Transportation University, 2019, 21(11): 78-83(in Chinese)
(谢枫，李永乐，苏致远等. 一种城市交叉路口转弯车辆轨迹预测方法. 军事交通学院学报，2019，21(11)：78-83)

[12] Graves A. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850, 2013

[13] Helbing D, Molnar P. Social force model for pedestrian dynamics. Physical Review E, 1995, 51(5): 4282-4286

[14] Yang Wen-Yan, Zhang Xi, Chen Hao, Jin Wen-Qiang. A model of pedestrian trajectory prediction for autonomous vehicles based on social force. Journal of Highway and Transportation Research and Development, 2020, 37(8): 127-135(in Chinese)
(杨文彦，张希，陈浩，金文强. 基于社会力的自动驾驶汽车行人轨迹预测模型. 公路交通科技，2020，37(8)：127-135)

[15] Trautman P, Krause A. Unfreezing the robot: Navigation in dense, interacting crowds//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots & Systems. New York, USA, 2010: 797-803

[16] Giles C L, Kuhn G M, Williams R J. Dynamic recurrent neural networks: Theory and applications. IEEE Transactions on Neural Networks, 1994, 5(2): 153-156

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

[18] Zhang Pu, Ouyang Wan-Li, Zhang Peng-Fei, et al. SR-LSTM: State refinement for LSTM towards pedestrian trajectory prediction//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA, 2019: 12085-12094

[19] Tang C, Salakhutdinov R. Multiple futures prediction//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada, 2019: 15398-15408

[20] Zhang Li-Dan, She Qi, Guo Ping. Stochastic trajectory prediction with social graph network. arXiv preprint arXiv: 1907.10233, 2019

[21] Yang Biao, Yan Guo-Cheng, Wang Pin, et al. TPPO: A novel trajectory predictor with pseudo oracle. arXiv preprint arXiv:2002.01852, 2020

[22] Kosaraju V, Sadeghian A, Martín-Martín R, et al. Social-BiGAT: Multimodal trajectory forecasting using bicycle-GAN and graph attention networks//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver, Canada, 2019: 137-146

[23] Huang Ying-Fan, Bi Hui-Kun, Li Zhao-Xin, et al. STGAT: Modeling spatial-temporal interactions for human trajectory prediction//Proceedings of the IEEE International Conference on Computer Vision. Seoul, Korea (South), 2019: 6272-6281

[24] Mao Lin, Gong Xin-Fei, Yang Da-Wei, et al. Space-time social relationship pooling pedestrian trajectory prediction model. Journal of Computer-Aided Design & Computer Graphics, 2020, 32(12): 1918-1925(in Chinese)
(毛琳，巩欣飞，杨大伟等. 空时社交关系池化行人轨迹预测模型. 计算机辅助设计与图形学学报，2020，32(12)：1918-1925)

[25] 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 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA, 2020: 14424-14432

[26] Liu Jun, Wang Gang, Duan Ling-Yu, et al. Skeleton-based human action recognition with global context-aware attention LSTM networks. IEEE Transactions on Image Processing, 2017, 27(4): 1586-1599

[27] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, USA, 2017: 6000-6010

YU Li, Ph. D., associate professor, Ph. D. supervisor. His research interests include big data analysis, recommender systems.

LI Hui-Yuan, Ph. D. candidate. Her research interests include recommender systems, deep learning.

JIAO Chen-Lu, M. S. Her research interests include deep learning, big data.

LENG You-Fang, Ph. D. candidate. His research interests include recommender system, deep learning.

XU Guan-Yu, M. S. candidate. His research interests include deep learning, big data.

Background

In this paper, we research the pedestrian trajectory prediction in the complex scene, where the pedestrian trajectory is affected by a wide range of factors, such as people nearby and social relations. Since people have interactions with a nearby moving entities for a period of time, it can be formulated as a time series problem. We should predict the future trajectory for each pedestrian based on her/his historical trajectory and the interactions between the pedestrians. Therefore, the most significant problem of the pedestrian trajectory prediction in the complex scene is how to model the interaction relationship between the pedestrians. In recent years, many works have been made to extract the interactive influence between the pedestrians and generate the future trajectory. For example, Social LSTM employed social pooling to aggregate the interaction information of each person and predict the trajectory with LSTM. Social GAN improved the social pooling and predicted with GAN model. Some researchers employed attention mechanisms to learn the interaction relations. They have achieved success in this task, however, they model the interaction influence of pedestrians only from a single aspect, fail to consider the multiple aspect influential factors.

Therefore, this paper proposes a multi-head attention generative adversarial model (MAGAM) to model the multiple influential factors in the interaction process between pedestrians in complex scenes. The MAGAM model consists of a trajectory generator and a trajectory discriminator. In the trajectory generator, there are three modules: the trajectory encoder, the interactive feature extractor, and the trajectory decoder. The discriminator is made up of the trajectory encoder. In the MAGAM model, the multi-head attention mechanism with relative displacement information is employed to learn the attentive weight of subspace features in the whole trajectory feature space on different aspects, to realize the characterization of the interactive trajectory features that result from mutual influence between pedestrians. Moreover, the adversarial generation strategy and multi-trajectory generation strategy are used to achieve the reasonable generation of individual moving trajectory in the complex scenes. During the training process, the LSTM-based encoder and decoder are employed to extract the temporal features of the pedestrian and predict the future trajectories, respectively. The discriminators are employed to judge whether the input trajectory belongs to the truth trajectory or generated trajectory as much as possible. This paper conducted experiments on two public datasets (ETH and UCY). The experimental results show that the performance of the MAGAM model is superior compared with the baselines. And it could predict the reasonable trajectory for pedestrians.

This work is supported by the National Natural Science Foundation of China (No. 71271209), the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (No. 2020030228).