

中图法分类号: TN9; TP18; TP389.1 文献标识码: A 文章编号: 1006-8961(2025)12-3900-14

论文引用格式: Wu E H and Ji Q G. 2025. Path stepwise estimation network combining social constraint and trajectory endpoints. Journal of Image and Graphics, 30(12):3900-3913(吴恩泓, 纪庆革. 2025. 结合社会约束与轨迹终点的逐步估计网络. 中国图象图形学报, 30(12):3900-3913)[DOI:10.11834/jig.240754]

结合社会约束与轨迹终点的逐步估计网络

吴恩泓, 纪庆革*

中山大学计算机学院, 广州 510006

摘要: 目的 多数的行人轨迹预测方法专注于序列化数据的特征, 忽略了对行人轨迹的社会语义进行学习。因此, 本文着重研究行人轨迹中的社会特征与人类行走特征, 提出结合社会约束与轨迹终点的路径逐步估计网络 (path stepwise estimation network combining social constraints and trajectory endpoints, PSEN)。方法 根据人在行走中对路径规划的3个特征: 1) 社会约束, 将人群按照社交约束, 依据运动学信息进行分类, 并根据社交权重得到被预测行人与类内其他行人的社交注意力, 从而影响后续的路径估计网络; 2) 通过模拟行人会先确定终点, 有目的地规划自己行走的路径这一特征, 设计一个终点估计网络, 通过时空序列对终点进行预测, 对完整的路径规划提供参考价值; 3) 行人不断根据周边环境与终点进行局部路径微调并重新分配注意力的特征, 搭建终点与路径微调网络, 实现自动根据环境进行微调路径规划的效果。结果 实验在ETH/UCY (Eidgenössische Technische Hochschule Zürich pedestrian and University of Cyprus pedestrian) 数据集上与6种基线方法进行比较, 在SDD (Stanford drone dataset) 数据集上与5种基线方法进行对比。在ETH/UCY 整个数据集中, 平均位移误差 (average displacement error, ADE) 和最终位移误差 (final displacement error, FDE) 平均降低5.1%和7.5%, 在SDD数据集中, ADE和FDE平均降低1%和2%。针对行人较为密集的场景, 如ZARA1、ZARA2和UNIV数据集的预测效果均提升10%以上。在ETH/UCY数据集上进行消融实验, 证明PSEN各模块均能够提高行人轨迹预测任务的效果, ADE和FDE分别平均降低19%和31%。结论 本文提出的结合社会约束与轨迹终点的路径逐步估计网络 (PSEN), 综合了真实世界中行人场景的3个特点, 在ETH/UCY和SDD数据集上取得了更优异效果。

关键词: 行人轨迹预测; 序列化预测; 循环神经网络 (RNN); 条件变分自动编码器 (CVAE); 社交约束; ETH/UCY数据集

Path stepwise estimation network combining social constraint and trajectory endpoints

Wu Enhong, Ji Qingge*

School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China

Abstract: **Objective** Pedestrian trajectory prediction constitutes a critical research challenge in autonomous driving systems, intelligent security surveillance, and human-robot interaction frameworks. The capability to accurately anticipate pedestrian movement patterns directly influences the operational safety of autonomous vehicles, the responsiveness of surveillance systems, and the adaptability of social robots in dynamic environments. While existing approaches predominantly focus on leveraging sequential data patterns and optimizing model architectures through recurrent neural networks, they

收稿日期: 2024-12-26; 修回日期: 2025-03-29; 预印本日期: 2025-04-05

* 通信作者: 纪庆革 issjqg@mail.sysu.edu.cn

基金项目: 国家自然科学基金项目 (62276280)

Supported by: National Natural Science Foundation of China (62276280)

often overlook the intrinsic social-semantic characteristics embedded in real-world pedestrian interactions. Current methodologies tend to treat trajectory prediction as a purely sequential modeling task, overlooking three fundamental aspects: 1) the social constraints governing crowd movement patterns, 2) the intentional, destination-oriented nature of human locomotion, and 3) the dynamic adaptation mechanisms pedestrians employ during path navigation. This oversight leads to sub-optimal performance, particularly in dense pedestrian scenarios where social interactions and environmental adaptability notably influence movement decisions. Aiming to address these limitations, this paper proposes path stepwise estimation network (PSEN), a novel framework that systematically integrates social relationship modeling, endpoint-aware trajectory planning, and environment-adaptive path refinement. The proposed model bridges the gap between conventional sequence prediction paradigms and the complex socio-spatial dynamics inherent in real-world pedestrian navigation scenarios.

Method This paper incorporates the characteristics of path planning observed in daily human walking, which can be broadly divided into three key aspects. First, social restrictions are considered. The crowd is categorized based on movement direction, speed, and distance to demonstrate these reflections. Intra-class feature learning is then performed on the classified groups. The social relationships between predicted pedestrians and other pedestrians are calculated using social weights to obtain social attention, which affects the subsequent path estimation network. Second, an endpoint estimation network is introduced by stimulating the feature that pedestrians typically identify a destination and then purposefully plan their walking path. This network leverages the strengths of serialized prediction tasks by using spatiotemporal sequences to predict an endpoint. The estimated endpoint serves as a reference condition within the overall network model, guiding the complete path planning process. Third, this paper address the fact that pedestrians constantly fine-tune their local paths and adjust their focus based on environmental context and destination. Aiming to model this behavior, an endpoint and path fine-tuning network is constructed using conditional variational autoencoder (CVAE) and multilayer perceptron (MLP). This module takes the output of the endpoint estimation network as a condition and uses the output from the social restriction module, along with the historical trajectory, as inputs for feature learning. After every three frames of prediction, the social restriction and endpoint module outputs are updated according to the current environment of the pedestrians. This update allows the model to automatically fine-tune the planned path in response to dynamic surroundings.

Result The experiments are conducted by comparing the proposed method with six baseline methods on the ETH/UCY dataset, five baseline methods on the SDD dataset, and four baseline methods on the NBA SportVU dataset. The evaluation metrics used are average displacement error (ADE) and final displacement error (FDE). On the entire ETH/UCY dataset, ADE and FDE are reduced by an average of 5.1% and 7.5%, respectively. On the SDD dataset, reductions of 1% in ADE and 2% in FDE are observed on average. When analyzing individual datasets, the performance improvements are highly pronounced in scenarios with denser pedestrian traffic. Notably, in the ZARA1, ZARA2, and UNIV datasets, the proposed method achieves improvements of over 10% in prediction accuracy. Ablation experiments are also conducted on the ETH/UCY dataset to evaluate the contributions of individual components of the PSEN framework. The experimental results demonstrate that each module of PSEN notably improves the effectiveness of pedestrian trajectory prediction, achieving average reductions of 19% and 31% in ADE and final displacement error FDE, respectively. Ablation experiments are performed in parameters such as social distance, social attention weights, and the number of frames used in stepwise trajectory generation. These experiments confirm that all aspects of the network design positively impact pedestrian trajectory prediction. However, the model does not perform as well on the NBA sportVU dataset. This dataset is characterized by 10 players moving at high speeds, with trajectory endpoints changing dynamically based on in-game situations and players' intentions. Different from ETH/UCY and SDD datasets, where movement is predictable and socially constrained, the varying roles and tactical decisions of agent in NBA dataset play a crucial role in path planning, making prediction highly challenging. Therefore, achieving accurate predictions by relying solely on time-position information is difficult because the characteristics of pedestrians in this setting notably differ from those in typical pedestrian scenes. In sports scenes, athletes actively seek collisions and obstructions as part of their strategic movement. PSEN does not consider the role-specific behaviors of agents, limiting its effectiveness in such environments. **Conclusion** The PSEN model proposed in this paper integrates the serialization task with three key features of real-world pedestrian scenes. By combining recurrent neural networks with a CVAE, PSEN effectively reflects the complex features of pedestrian trajectory prediction in realistic sce-

narios. The model achieves superior performance on the ETH/UCY and SDD datasets, providing a new direction for subsequent tasks in pedestrian trajectory prediction. However, this study focuses only on interactions among pedestrians and does not consider the relationship between pedestrians and other objects, such as vehicles and obstacles. In novel environments, or in scenes where pedestrians are sparse but other dynamic or static objects are abundant, the performance of the model may degrade. Further research is needed in terms of the relationships between pedestrians and objects, along with their associated feature information.

Key words: pedestrian trajectory prediction; sequential prediction; recurrent neural network (RNN); conditional variational autoencoder(CVAE); social constraints; ETH/UCY dataset

0 引言

行人轨迹预测任务在许多领域(如自动驾驶(Luo等,2018)、安防系统以及刑侦技术等)都具有重要作用(Raksincharoensak等,2016),一个性能优异的行人轨迹预测模型能够显著推动上述领域的发展。国内外研究人员也致力于这些方面的工作。行人轨迹预测任务的发展大致分为3个阶段。第1阶段使用经典的机器学习理论结合运动学信息进行预测;第2阶段使用神经网络对行人轨迹预测中的时间特征进行预测;第3阶段分析行人轨迹预测数据中的语义信息,针对语义信息进行提取,预处理之后再构建合适的模型进行学习。

对行人轨迹预测任务的研究主要分为3个阶段。

第1阶段采用运动学信息(速度、加速度等)进行预测(Yamaguchi等,2022),关注数据中表现出的运动学信息(Cheng等,2018),使用经典的机器学习理论(Helbing和Molnár,1995),通过对现实世界进行建模,使用投影、转换和聚类等方式对目标数据进行处理,最后通过一系列数学原理或人为制定的规则得到机器模型的处理结果(Tay和Laugier,2008;Antonini等,2006;Hochreiter和Schmidhuber,1997)。预测方式为结合前3~5s相关运动学信息,针对后3s的轨迹进行预测。其中,后者的预测方程简化为

$$Y(x) = \frac{1}{\beta} \sum_{i=0}^n (G(\tau a^i \cos \theta, v(x)) + x) \quad (1)$$

$$Y(y) = \frac{1}{\beta} \sum_{i=0}^n (G(\tau a^i \sin \theta, v(y)) + y) \quad (2)$$

式中, Y 为预测的结果, x 与 y 为坐标, β 与 τ 为超参, a^i 为 i 时刻的加速度, θ 为方向, v 为速度, G 为Gaussian Processes方法(Tay和Laugier,2008)。

通过建模信息与人为选定的超参数,从而对行

人未来的轨迹进行预测。基于经典机器学习的行人轨迹预测算法,其优点是可解释性强,针对稀疏场景以及无太多杂乱物体的简单场景效果强,缺点是泛化能力弱,只根据制定的相关规则进行预测,对新场景、复杂场景的预测效果不佳。

第2阶段结合时间特征的循环神经网络生成模型。在神经网络以及深度学习技术兴起之后,国内外学者根据行人轨迹预测任务的特性,将行人轨迹预测任务作为序列化生成任务的下游任务。Alahi等人(2016)最早将神经网络运用于行人轨迹任务研究,使用标准的长短期记忆网络(long short-term memory, LSTM)生成未来8帧的移动轨迹。后续的研究人员在此基础上利用数据中的时空信息特征,使用更多更先进的模型进行研究。如:使用生成对抗网络(generative adversarial network, GAN)(Sadeghian等,2019; Amirian等,2019)、Transformer(Yu等,2020)、条件变分自动编码器(conditional variational autoencoder, CVAE)(Xu等,2022c)和图卷积网络(graph convolutional network, GCN)(Ivanovic和Pavone,2019;陈浩东和纪庆革,2023;Huang等,2019;Shi等,2021;朱鹏飞和张德平,2023)等。随着模型的发展,行人轨迹预测模型的性能也稳步提升,但在模型发展越来越庞大的过程中,也要求具有足够多的数据,才能够使得模型收敛(Zhao和Wildes,2021),但行人轨迹预测任务并没有那么庞大的数据量。

由于上述原因,行人轨迹预测研究发展到第3阶段,学者们不再直接将数据输入模型,而是开始研究数据的语义信息,进行提取后再输入模型,经典的工作如Social-GAN(Gupta等,2018)首次提出行人轨迹预测数据中的社交信息,并以此提高预测精度,后续工作人员结合社交信息的概念,使用不同的语义信息提取技术和与其相匹配的模型,提出STGCN

(spatial-temporal graph convolutional network) (Mohamed 等, 2020)、Sophie (Sadeghian 等, 2019)、Trajectrion++ (Salzmann 等, 2020)、SimAug (Liang 等, 2020)等方法。另外,部分学者在研究数据的同时充分借鉴在其他任务中表现优异的特性。如:结合傅里叶变换对数据进行预处理(Wong 等, 2022)、使用刚性结构理论对数据和模型间的关系进行融合、基于领域信息和注意力的无参考点云评估(陈晓雷, 2024)、基于视频第一视角的分析方法(Liang 等, 2019)、使用短时记忆进行分析(张强 等, 2023)。

目前,国内外对于行人轨迹预测任务的研究中,将更多的注意力放在研究行人轨迹数据上的特点(Bi 等, 2020)、模型结构优化以及人与场景之间的数据特点,因此对于行人本身对路径规划与决策方面的研究比较缺失,本文拟针对该方向进行研究,针对行人移动过程中的3个特点进行研究,即注意周边特定距离内的其他人的运动学信息、行人移动具有目的性、行人所做出的路径规划和决策会实时更新,提出结合社会约束与轨迹终点的路径逐步估计网络(path stepwise estimation network combining social constraints and trajectory endpoints, PSEN)。

本文主要贡献如下:1)提出基于社会约束与轨迹终点的路径逐步估计网络PSEN。与现有工作相比,PSEN对于路径规划是根据长期指标即终点与短期指标即周边环境实时更新的,实现更准确的预测;2)设计了终点估计模块与社会约束模块,通过在整体模型中嵌入并实时更新两个模块的输入,从而在不同的场景下体现终点与社会约束对行人的指导比例,直观解释了终点与社交约束对行人轨迹的指导关系;3)PSEN在公开数据集ETH/UCY (Eidgenössische Technische Hochschule Zürich pedestrian and Universitf of Cyprus pedestrian)和SSD (Stanford drone dataset)上的实验结果表明,PSEN取得了最优结果,消融实验表明PSEN逐步估计网络以及终点模块和社会约束模块的融合输入能有效提升预测准确度,定性分析表明PSEN能够体现终点与社会约束对行人的指导作用。

1 相关工作

本节介绍当前研究工作中对行人场景语义信息的理解及针对多语义信息的学习模型。

1.1 行人场景语义信息

行人场景语义信息指的是在行人轨迹预测任务中,相关的数据集所蕴含的表示当前场景特点、行人本身特点的相关特征信息。现有工作中,对行人场景语义信息进行了多方位的分析。Alahi 等人(2016)研究了行人间欧氏距离在行人场景中所表示的语义信息,并针对该语义信息使用多层感知机(multilayer perceptron, MLP)进行特征提取,作为轨迹预测的输入之一。Ivanovi 和 Pavone(2019)在针对行人间欧氏距离的同时也考虑行人的运动学信息,如速度、加速度等,并使用Transformer框架对该语义信息的特征进行提取,获得全局社交特征信息并作为轨迹生成的重要参考。Mohamed 等人(2020)则针对行人间的社交特征,使用图重新对行人进行建模,从而获得一个具有社交特征信息的数据结构,使用图卷积网络结合行人场景语义信息进行预测生成。Wong 等人(2022)通过对行人间进行数学建模,运用相关度高的模型,如刚体来描述短时间内行人场景中的语义信息。Yamaguchi 等人(2022)采用点云重建的方式提取行人语义信息。Long 等人(2023)通过采用不同的摄像头重新制作数据集,进而提取出更多的行人语义信息。Zhao 和 Wildes (2021)首次提出行人目标这一特征语义,并将其运用于网络中。Chiara 等人(2022)提出目标驱动型的自注意力网络,针对行人的本身特点即行走具有目的,进行了研究并建模。Meng 等人(2022)提出路径符合有限数量的特征,行人可能更多会照着前人的轨迹行走,因此通过给历史路径投入注意力,有效提升了预测的准确性,Robicquet 等人(2016)对拥挤场景下行人的注意力进行了研究,余力等人(2022)在复杂场景下提出多头注意力机制对行人进行预测。

在前人的相关工作中,针对行人场景语义信息的研究大多都是针对场景的特点,如是否存在人行道等进行场景限制特征的提取,或是针对数据本身的语义特征进行研究,如行人之间的距离、运动信息等,而针对真实行人的行走特点的研究较少。因此,本文提出的PSEN在前人的工作基础之上,着重研究行人行走的特点,提取新的语义信息特点,根据行人具有目的性,即行人针对路径的终点是已有规划,并且该终点会影响行人的路径规划,提出路径终点估计模块,对行人具有目的性进行建模,根据行人与周围行人间的社交约束与距离强相关,即行人会将

更多的注意力放在离自身更近的其他行人,提出社交约束模块,根据行人对周边环境的注意力差异进行建模。

1.2 多语义融合生成

行人轨迹预测作为序列任务的下游领域,具有其独特的性质。行人所处的场景语义信息极为丰富,不能仅仅因为输入与输出均为时间序列,就简单地将行人轨迹预测任务完全等同于一般的序列化任务进行处理。在真实世界中,行人在迈向目的地的过程中,会综合考虑人与人、人与物之间的关系以及基本的社交规范等多种信息,进而做出决策(Mangalam等,2020)。因此,行人轨迹预测任务中蕴含着多语义信息。有效地结合多语义信息,能够显著提升行人轨迹预测的精度。

目前,结合多语义信息的工作成果主要分为5类。第1类是将人与人之间的社交距离融入模型,作为社交信息与时空信息相结合的方法。典型工作如 Social-GAN (Gupta等,2018)、Sophie (Sadeghian等,2019)、Multi-agent (Wang和Chen,2023)等。此类方法通过对社交距离的考量,将其转化为模型可处理的社交信息,与时空信息共同作用,以更好地预测行人轨迹。第2类是将行人按照速度、方向等运动信息进行划分,给予不同的社交权重,再与时空信息相结合的方法。例如 AgentTransformer (Yuan等,2021)等典型工作方法,通过对行人进行运动信息的分类,为不同的行人赋予不同的社交权重,从而更准确地捕捉行人之间的社交关系和运动趋势。第3类方法根据场景限制与行人之间的社交规范,通过场景结合时空特征。典型工作如 STGCN (Mohamed等,2020),充分考虑场景的限制和社交规范,将场景信息与时空特征相结合,提高轨迹预测的准确性。如 Yu等人(2020)针对时空与社交特征,通过 Transformer 框架进行学习两者之间的特征关系,继而针对行人轨迹进行预测。第4类方法针对多语义信息进行预处理或采用适当的信息提取方式,再结合符合特性的神经网络模型进行训练。典型工作如 View Vertically (Wong等,2022)等,通过对多语义信息的预处理和有效的信息提取,结合合适的神经网络模型,提升行人轨迹预测的性能,如 Xu等人(2022c)使用时间潜在信息,提出 SocialVAE (social variational autoencoder),针对时间特征进行预处理后才投入网络进行学习。Shi等人(2023)通过改造

行人轨迹多语义的信息表现形式,使其更易于被 Transformer 框架进行学习,也提高了预测的精确度。Yamaguchi等人(2011)通过研究行人行走意图提高预测的准确度。第5类方法发掘行人本身的特征语义并进行有限的学习,融合其他环境信息,如时空、社交规范等进行学习,如 Zhao和Wildes(2021)通过调整行人终点在学习过程中的权重,从而提高预测的准确性,说明行人本身的特点也能够与其他场景特征信息进行融合生成。Meng等人(2022)提出将当前场景内所有的历史轨迹与其他场景信息进行融合,取得了不错的成果。

根据上述工作的实验结果(Sadeghian等,2019; Gupta等,2018; Yuan等,2021; Mohamed等,2020; Chiara等,2022),结合多语义信息的模型确实在一定程度上提高行人轨迹预测的精度,但其中也存在一些问题。首先,增加多语义往往会导致模型参数量的增加。例如 Social-GAN、AgentTransformer 等进行全局社交的模型中,存在与真实世界不符合的情况。在这些模型中,行人可能会受到远方处于另外一条人行道上行人的影响,这在稀疏人群的情况下会导致预测精度下降。其次,对于像仅仅使用终点进行路径估计的模型,虽然在短时间的预测上有不错的效果,但在预测12帧以上的轨迹时,精度会大幅下降(Liu等,2021)。

本文提出的 PSEN 针对行人场景中行人本身的特点,结合行人场景中的语义信息进行研究,提出新的语义信息特征提取模块与多语义融合模块,上述前人工作针对场景语义信息在预测的过程中不会改变,且一次性预测目标行人的完整轨迹,而 PSEN 不同之处是,根据行人本身的特点,即行人对路径规划进行实时更新,每次周围环境的变化和与终点的距离都将影响行人下一刻的路径规划,为不同的交互信息分配注意力权重,运用逐步估计理念,通过多次短期路径预测,对于场景语义信息会进行实时更新,最终生成完整预测路径,充分结合行人本身的行走特点与场景语义信息。

2 研究方法

本文深入研究真实世界中行人场景中的行人特点,在前人研究工作的基础上,创新性地提出基于社会约束与轨迹终点估计的路径逐步估计网络

(PSEN)。

2.1 PSEN 模型

PSEN 模型主要由 3 个模块构成。分别为社会约束模块、终点估计模块与路径逐步估计模块。对于行人注意周边特定距离内的其他人的运动学信息的特点,由社会约束模块进行建模并提取相关语义信息,得到社交约束 Mask 向量,将 Mask 向量与目标信息的社会信息进行特征提取后输入标准 Transformer 模型,得到社会约束模块的输出向量。对行人移动具有目的性的特点,由终点估计模块根据时空信息进行建模,通过多层感知(MLP)层,提取历史轨迹与行人约束模块的输出特征,将提取出的特征输入长短期记忆(LSTM)网络得到终点估计模块的输出。对行人所做出的路径规划和决策会实时更新的特点,由路径逐步估计模块对其特点进行建模,将社会约束模块与终点估计模块的输出,作为 CVAE 的输入,并将每次预测后的结果作为两个模块的一部分输入,进行更新,逐步生成全部轨迹。网络模型结构如图 1 所示,其中网络的输入是目标历史真实

轨迹向量 $X_{(n)}$, 输出是生成的轨迹向量 $Y_{(n)}$, 由于该网络为逐步估计, 因此其会自动更新网络的输入, 网络的多次输出合并是完成的生成轨迹。

这 3 个模块通过以下各式进行多语义的融合预测, 具体为

$$Y_{(n)} = \delta \sum_{i=1}^n X_{PE(i)} \quad (3)$$

$$X_{P(n)} = F(X_{SC(j)}, X_{EE(n)}, \sum_{i=1}^n X_{(i)}) \quad (4)$$

$$X_{EE(n)} = E(\sum_{i=1}^k X_{(i)}, X_{SC(j)}) \quad (5)$$

$$X_{SC(j)} = \frac{1}{\theta} \sum_{i=1}^{n-1} S(X_{(j)}, X_{(j-i)}) \quad (6)$$

式中, $Y_{(n)}$ 表示预测的 n 帧数据, $X_{SC(j)}$ 为社会约束模块的输出, j 为对应的第 j 个行人, $X_{EE(n)}$ 为终点估计模块的输出, n 为该行人的历史真实轨迹, δ, θ 为超参数, F 为路径估计模块中的网络模型, E 为终点估计模块的网络模型, S 为社会约束模块的网络模型, k 为历史真实轨迹帧数。

PSEN 综合利用不同模块的信息, 提高行人轨迹

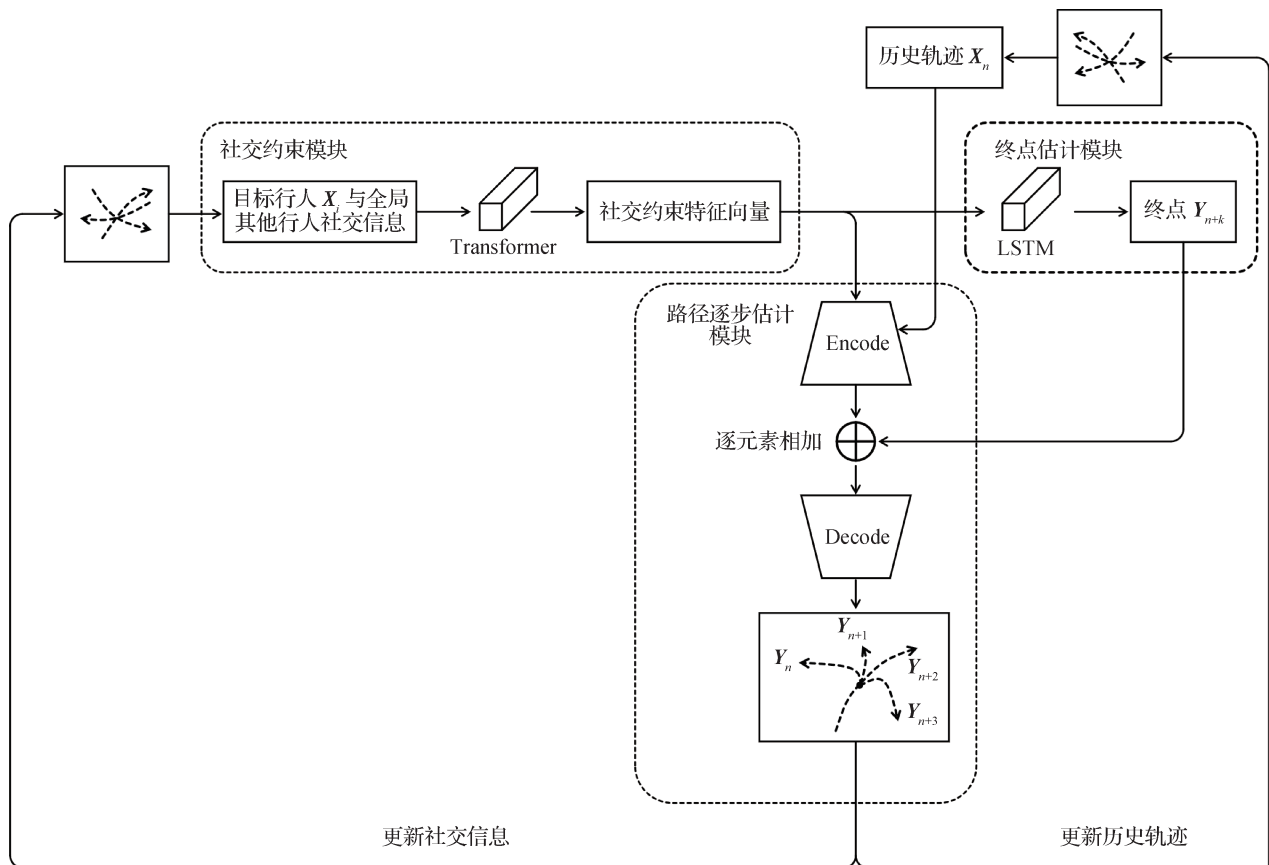


图 1 结合社会约束与终点估计的路径逐步估计网络

Fig. 1 Path stepwise estimation network combining social constraints and endpoints (PSEN)

预测的准确性和可靠性。这种多语义融合的方法能够充分发挥各个模块的优势,同时弥补单一模块的不足,为行人轨迹预测提供了一种更为有效的解决方案。

2.2 社会约束模块

社会约束模块的构建基于行人之间的社交距离。在真实世界中,行人会根据与他人的社交距离调整自身行为。该模块充分考虑这一特点,对不同社交距离下行人的行为进行建模,以准确捕捉行人之间的社交互动和行为约束。行人的社会约束特征呈现出以下3个显著且具有重要研究价值的特点: 1)行人往往会将主要注意力聚焦于距离较近的其他行人。在日常行走过程中,行人会更为关注身边其他行人的相关信息,这些信息犹如一个重要的参考系,为自身的行走决策提供依据。例如,行人会留意身旁行人的行走速度、方向等,以便调整自己的行进路线和步伐节奏。2)对于远距离且不在同一社交约束范围内的行人,行人通常不会投入过多注意力,甚至可能投入极少的注意力。以人行道场景为例,行人在正常行走或站立时,对于马路对面人行道中的行人,一般仅给予极少的关注。当行人的目标与远距离、不在同一社交规范的行人产生关系时,他们才会开始投入注意力,并且随着社交距离的逐渐拉近,投入的注意力会不断增加。如在过马路的场景中,行人会更加关注即将相遇的对面行人的行动,以便做出安全且合理的决策。3)行人会根据自身的运动特征与其他行人的运动特征的异同进行相应调整。若自身处于站立状态且在短时间内有极高概率持续站立,其他行人通常会对其进行避让。反之,若迎面而来的行人速度更快,当前行人也会做出避让等反应。由此可见,行人在行走过程中确实会受到社交

约束内的其他行人运动信息的影响,这种影响在行人的行动决策中起着至关重要的作用。

如图2所示,行人对于社交距离越近的人分配的注意力越多,其行为轨迹受其影响越大,由此行人根据注意力的权重,会进行分类,并根据上述3个特点设计社会约束模块,如图3所示。

社会约束模块的输入为行人的全局社交信息,使用注意力机制,通过历史数据信息获取目标行人与其他所有行人的最新位置信息与运动信息如速度、加速度等,然后将其他行人的社交信息根据式(7)和式(8)得到社交距离 Mask 向量,将 Mask 向量与目标信息的社会信息进行特征提取后输入标准 Transformer 模型,得到社交约束模块的特征向量。

$$h_j = \frac{\lambda}{d(\vec{x}_j, \vec{x}_i)} x_j, \quad d(\vec{x}_j, \vec{x}_i) < \eta \quad (7)$$

$$h_j = \frac{1}{d(\vec{x}_j, \vec{x}_i)} x_j, \quad d(\vec{x}_j, \vec{x}_i) > \eta \quad (8)$$

式中, x_j 表示除目标外其他行人的信息, x_i 表示目标行人的信息, $d(\vec{x}_j, \vec{x}_i)$ 表示两个对象的欧氏距离, η 为设定的社交约束超参数, λ 为影响力权重超参数。



图2 根据社会约束模块对行人分类示意图

Fig. 2 Schematic diagram of pedestrian classification according to social restriction module



图3 社会约束模块

Fig. 3 Social restriction module

2.3 终点估计模块

终点估计模块的提出源于行人在行走过程中表

现出的有意识且有目的的特点。行人通常具有明确的目的地,会朝着目标终点前进。然而,众多现有的

相关研究工作却未能充分意识到这一关键特性, 仅将行人轨迹预测单纯地视做时间序列化的生成任务进行处理。终点估计模块通过分析行人的行为和环境信息, 对行人的终点进行合理估计, 为轨迹预测提供重要的目标导向, 引入一个明确的终点对预测过程加以指导, 使预测结果的精准度得到显著提升。

终点估计模块的计算可表示为

$$Y_{n+k} = F(M(X_n, X_{SC(n)})) \quad (9)$$

式中, X_n 表示历史轨迹, $X_{SC(n)}$ 表示社会约束模块的输出, M 表示 MLP 网络, F 表示 LSTM 网络, 其中, n 表示前 n 帧真实轨迹, 它反映了行人在过去一段时间内的位置信息。而 k 则为预测的轨迹帧数, 代表着对行人未来运动轨迹的预测长度。

具体而言, 终点估计模块如图 4 所示, 在终点估计模块中, 首先将历史轨迹与行人约束模块的输出进行整合, 随后一同送入多层感知机 (MLP), 以实现特征的提取。在此基础上, 将提取出的特征输入长短期记忆 (LSTM) 网络, 最终得到终点 Y 。这一设计充分考虑了行人运动的时间序列特性以及各种约束因素, 旨在更加准确地估计行人运动的终点, 为行人轨迹预测提供更加可靠的依据。

终点估计模块的输出将会作为路径估计模块的输入, 用于提高预测的精准度。将在 4.1.2 节中进

行消融实验, 用以确认终点估计模块能够提高预测的精准度。

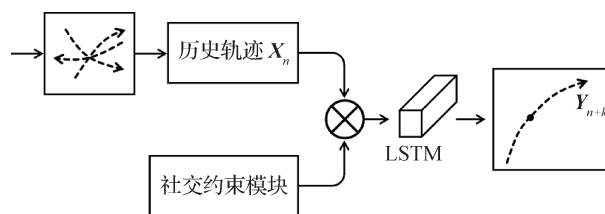


图4 终点估计模块

Fig. 4 Endpoint estimation module

2.4 路径逐步估计模块

真实情况下, 行人会根据周围的障碍物、其他行人的位置和运动状态等因素不断调整自己的行走路径。根据行人在行走过程中会实时关注周边环境并据此进行调整的特点, 提出路径逐步估计模块, 该模块能够通过对外部环境的感知和分析, 对行人的路径进行实时估计, 以更好地模拟行人的动态行为。在实际情境中, 行人在行走时其终点通常是相对固定的, 而路线以及注意力则会随着周围环境情况的变化而发生变换。因此, 终点模块的输出可以作为路径规划的条件变量, 对路径的生成起指导作用。路径估计模块如图 5 所示, 充分参照了条件变分自编码器 (CVAE) 模型的显著特点。

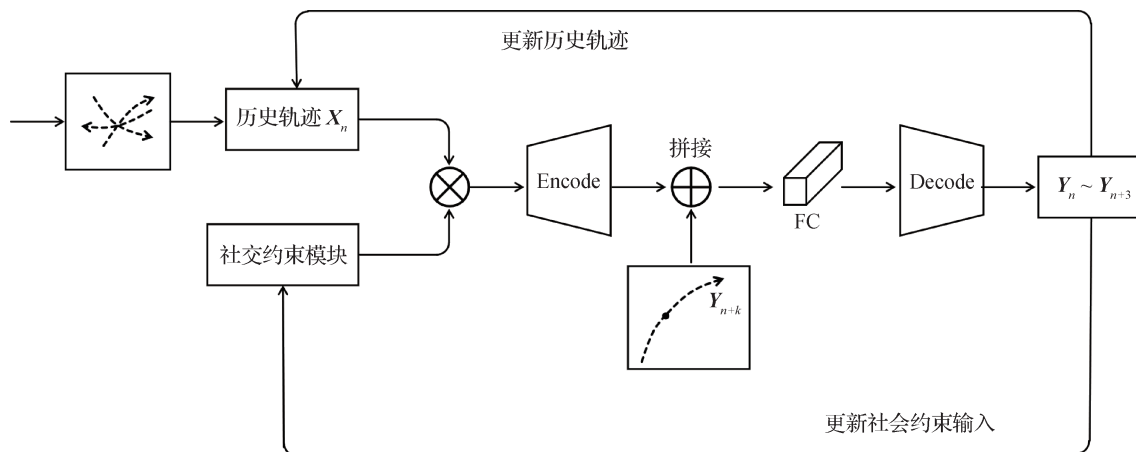


图5 路径估计模块

Fig. 5 Path estimation module

具体而言, 首先将历史轨迹与社会约束模块的输入共同通过多层感知机 (MLP) 进行全面的特征提取操作, 随后将提取后的特征输入编码器 (encoder), 进而得到潜在空间的特征向量。同时,

将终点模块的输出作为特定条件与潜在空间特征向量进行拼接后输入全连接层 (fully connected, FC)。最后, 经过解码器 (decoder) 生成指定帧数的路径, 并将历史轨迹中最新的轨迹更新为生成轨迹, 随后

重新计算社会约束模块的输出,以此进行迭代,逐步生成完整路径。在本文的设定中,参数 k 的值为3,这意味着每次能够生成3帧预测轨迹。接着,将生成的轨迹更新为历史轨迹中最新的轨迹信息,并同步更新社会约束模块的输入,以便进行新一轮轨迹生成。如此,通过逐步迭代的方式生成完整的预测路径。

3 实验设计

3.1 数据集

为了全面和严谨地验证基于社会约束与终点估计的路径估计网络,本文选取了行人轨迹预测领域中3个具有广泛认可度的公开数据集(即ETH、UCY和SDD),同时也选取了特定场景的数据,即NBA SportVU。其中,ETH数据集涵盖了ETH和HOTEL两个子集。UCY数据集则由UNIV、ZARA1和ZARA2这3个子集组成。ETH/UCY数据集所涉及的场景包括街道以及校园,在这样的环境下,有限的空间使得行人之间的社会交互表现得更为显著。而SDD数据集不仅包含校园中的常见道路,还囊括了十字路口等区域,这为行人提供了更为丰富的路径选择,极大地增强了数据的多样性。NBA SportVU数据集包含2015—2016 NBA赛季的球员轨迹,其中包含球员的身份、球队身份和时间段等信息。在ETH/UCY和SDD数据集中,行人位置坐标的采样时间间隔均设定为0.4 s。对于ETH/UCY数据集上的实验,本文采用了科学合理的留一法策略。具体而言,将8个时间步长(共计3.2 s)的历史轨迹作为输入数据,通过模型预测未来12个时间步长(4.8 s)的行人路径。这种实验设计和参数设置,旨在全面、准确地评估固定比例代表制在行人轨迹预测任务中的性能表现,为该领域的研究提供有力的参考依据。

3.2 评价标准

在本文的研究中,使用行人轨迹预测任务中的公共评价标准。其一为平均位移误差(average displacement error, ADE),通过计算预测的每一帧数据与实际真实数据之间的误差,能够有效地考量模型在路径规划方面的能力。具体而言,ADE指标可以细致地反映出模型在不同时间节点上对路径的预测精度,为评估模型的整体性能提供了关键的量化依据。其二为最终位移误差(final displacement error,

FDE),该指标聚焦于计算预测的最后一帧数据与真实数据的误差,从而对模型的终点预测能力进行深入考量。FDE能够直观地展现模型在预测最终位置时的准确性,对于衡量模型在实际应用中的可靠性具有重要价值。这两个评价标准的具体表达式为

$$\min ADE_K = \frac{\sum_{i=1}^N \min_{k=1}^K \sum_{t=1}^{T_2} \|\hat{X}_{2,t}^{ik} - X_{2,t}^i\|_2}{N \times T_2} \quad (10)$$

$$\min FDE_K = \frac{\sum_{i=1}^N \min_{k=1}^K \|\hat{X}_{2,t}^{ik} - X_{2,t}^i\|_2}{N} \quad (11)$$

式中, K 表示总帧数, \hat{X} 表示预测, X 表示真实轨迹。

3.3 实验细节与训练配置

本文实验所采用的设备参数如表1所示。按照基于社会约束与终点估计的路径估计网络的设计,实验对于模型中的非经典网络模型的超参数设定如表2所示。在训练阶段,PSEN的代码实现是基于PyTorch框架,设置的批大小为64,训练的轮数为300轮,优化器为Adam,学习率设置为0.001。

表1 设备参数

Table 1 Equipment parameters

设备类型	名称	数量
CPU	12th Gen Intel (R) Core i5-12450H	1
内存	32 GB	1
SSD	Micron_2450_MTFDKBA1T0TFK	1
GPU	NVIDIA GeForce RTX 4050 (6 G)	1

表2 网络模型中的超参数设定

Table 2 Hyperparameter settings in network models

参数	作用	数值
N	历史轨迹帧数	8
K	预测生成的帧数	12
η	社会约束模块中的社交距离	5
LSTM- N	LSTM网络层	12
Transformer- N	Transformer网络层	6
λ	社会约束模块中的注意力权重	8
k	路径估计模块中的逐步轨迹生成帧数	3
Encode- N	编码器层数	4
Decode- N	解码器层数	4
Z- N	潜在空间表达层	2

4 实验结果

4.1 定量分析

4.1.1 对比实验

为了验证 PSEN 的模型性能,在主流的公开行人轨迹预测数据集 ETH/UCY 和 SDD 上与当前主流方法进行对比实验,结果如表 3 和表 4 所示,并在 NBA SportVU 数据集进行对比,结果如表 5 所示。

总体而言,PSEN 在 ETH/UCY 和 SDD 上都取得了比当前主流方法更优异的结果。其中在 UNIV、ZARA1 和 ZARA2 数据集中,PSEN 均取得最优结果,与排名第 2 的结果相比,ADE 和 FDE 指标分别提升 4.3% 和 9.5%、11.1% 和 6.3%、7.7% 和 13.0%。这是因为这 3 个数据集的行人比较密集,能够充分发挥 PSEN 的社会约束模块与逐步估计模块的能力。在 ETH 数据集中,性能与最优方法存在差距,原因在于 ETH 数据集中行人较为稀疏,无法充分发挥 PSEN 中社会约束模块的能力,导致 PSEN 出现模型退化情况,PSEN 的社会约束模块没有得到充分利用,由终点估计模块和路径逐步估计模块进行预测,导致 PSEN 的性能优势不够明显,但与最优方法仅存在 0.04 的精度差距,说明 PSEN 本身网络模型的先进性。PSEN 在 SDD 数据集上 ADE 为最优结果,FDE 指标为次优结果,与最优结果相差 0.77 的精度,这依赖于 PSEN 的路径逐步估计模块,在每次迭

代中都会对路径根据更新后的社交约束和历史轨迹进行重新估计,从而确保每次迭代都更加贴合真值。

PSEN 在 NBA SportVU 数据集上表现不佳,通过分析数据集特点与 PSEN 的特点,本文认为 NBA 数据集中的特点为 10 个人在高速运动且路径终点会根据比赛中的情况、运动员的意识随时变化。与 ETH/UCY 和 SDD 数据集不同,NBA 数据集中每一个 agent 由于存在不同角色的差异,如前锋、中锋等。角色的差异对于 agent 规划路径而言是一个重要的参考指标,因此仅依靠时间一位置信息难以做到精准的预测,且该场景下与正常行人场景下的行人特点不一样,运动场景中运动员会主动寻求碰撞、阻拦等。本文方法 PSEN 仅模拟了正常行人情况下的真实路径规划特点,对运动场景中的运动员特点没有进行建模,因此针对 ETH/UCY、SDD 数据集中每一个 agent 不存在角色差异或仅依靠运动学信息划分 agent 的运动状态的情况下,能够有效预测,但对于 NBA 数据集没有考虑 agent 的角色差异,因此性能不够出色。

4.1.2 消融实验

本文针对 PSEN 在 ETH/UCY 数据集进行超参数与模块的消融实验,以验证基于社交约束与终点估计的路径逐步估计网络中各个模块的重要性。用于消融实验的网络模型如下:

PSEN:标准的基于社交约束与终点估计的路径估计网络。

表 3 ETH/UCY 数据集上的实验结果

Table 3 Experiment results on ETH/UCY datasets

方法	ETH		HOTEL		UNIV		ZARA1		ZARA2		平均	
	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE
Sophie(Sadeghian 等,2019)	0.70	1.43	0.76	1.67	0.54	1.24	0.30	0.63	0.38	0.78	0.54	1.15
Goal GAN(Dendorfer 等,2021)	0.59	1.18	0.19	0.35	0.60	1.19	0.43	0.87	0.32	0.65	0.43	0.85
Causal-STGCNN(Chen 等,2021)	0.64	1.00	0.38	0.45	0.49	0.81	0.34	0.53	0.32	0.49	0.43	0.66
PECNet(Mangalam 等,2021)	0.54	0.87	0.18	0.24	0.35	0.60	0.22	0.39	0.17	0.30	0.29	0.48
SocialVAE(Xu 等,2022c)	0.47	0.76	0.14	0.22	0.25	0.47	0.20	0.37	0.14	0.28	0.24	0.42
EqMotion(Xu 等,2023)	0.40	0.61	0.12	0.18	0.23	0.43	0.18	0.32	0.13	0.23	0.22	0.35
TUTR(Shi 等,2023)	0.40	0.61	0.11	0.18	0.23	0.42	0.18	0.34	0.13	0.25	0.21	0.36
RAN(Dong 等,2024)	0.41	0.59	0.13	0.21	0.25	0.46	0.22	0.41	0.16	0.31	0.23	0.40
PSEN(本文)	0.44	0.65	0.12	0.18	0.22	0.38	0.16	0.30	0.12	0.20	0.21	0.34

注:加粗字体表示各列最优结果。

表4 SDD数据集上的实验结果

Table 4 Experiment results on SDD dataset

方法	ADE	FDE
Sophie(Sadeghian等,2019)	16.27	29.38
Multiclass-SGCN(Li等,2022)	14.36	25.99
GroupNet(Xu等,2022a)	14.36	25.99
TUTR(Shi等,2023)	7.79	12.73
RAN(Dong等,2024)	10.97	19.95
PSEN(本文)	7.50	13.50

注:加粗字体表示各列最优结果。

表5 NBA SportVU数据集上的实验结果

Table 5 Experiment results on NBA SportVU dataset

方法	ADE	FDE
SocialVAE(Xu等,2022c)	0.58	0.95
PECNet(Mangalam等,2021)	14.36	25.99
Trajectron++(Salzmann等,2020)	14.36	25.99
Retrospective-Memory-based(Xu等,2022b)	1.25	1.47
PSEN(本文)	8.18	19.10

注:加粗字体表示各列最优结果。

PSEN-End:基于终点估计的路径估计网络,将社交约束模块去除。

PSEN-Social:基于社交约束的路径估计网络,将终点估计网络去除。

PSEN- $n(i)$:在基于社交约束与终点估计的路径估计网络中,其社交约束模块中的社交距离为 i 。

PSEN- $a(i)$:在基于社交约束与终点估计的路径估计网络中,其社交约束模块中的注意力权重为 i 。

PSEN- $k(i)$:在基于社交约束与终点估计的路径估计网络中,其路径估计网络模块中逐步轨迹生成帧数为 i 。

实验结果如表6所示。通过表6的结果可分析得出,社交约束模块与路径估计模块在路径估计网络中都能够提升模型的效果。其中,社交距离如果过大会导致模型的运算量上升且性能下降,过小则会导致模型的预测准确度下降;注意力权重过大会导致过分受环境影响从而降低准确度,过小则有可能导致碰撞,且导致ADE增大;轨迹生成帧数过多则无法充分利用逐步估计的优势导致模型预测准确度下降,过小则导致模型运算量过大,但并不能够有很大的效果提升。

4.2 定性分析

在定量分析的对比实验中,在ETH/UCY数据集与SDD数据集上,其指标ADE和FDE总体上表现优异,在UNIV和ZARA1与ZARA2中均取得了最优的结果,SDD上ADE指标也取得了最优的结果。同时消融实验表明,社交约束模块与终点估计模块能够很好地分析行人的社交与规划意图,路径逐步估计模块能够生成更加贴合实际的路径,从而降低ADE。本节对PSEN在各个行人场景中的表现,展开定性分析。如图6所示,其中绿色虚线为真实路径,红色区域为预测的路径范围,黄色区域为模型规定的社交影响范围。图6(a)展示了PSEN在行人较为稠密的场景中的表现效果。PSEN能够有效地将人群按照社会约束规则进行分类,并且在社交圈内的行人其行走方向与速度均会与社交圈内的其他行人有一定关系,说明PSEN学习到了行人将注意力分

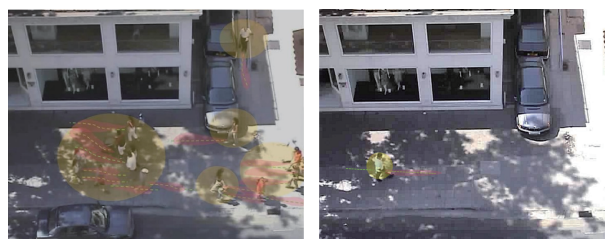
表6 消融实验结果

Table 6 Results of ablation experiment

方法	ETH		HOTEL		UNIV		ZARA1		ZARA2	
	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE
PSEN	0.44	0.86	0.12	0.18	0.22	0.38	0.16	0.30	0.12	0.20
PSEN-End	0.72	1.32	0.56	0.83	0.55	1.12	0.47	1.05	0.50	0.98
PSEN-Social	0.48	0.95	0.20	0.25	0.30	0.55	0.27	0.72	0.27	0.68
PSEN- $n(2)$	0.45	0.91	0.12	0.20	0.23	0.47	0.20	0.42	0.22	0.29
PSEN- $n(8)$	0.43	0.87	0.13	0.22	0.22	0.41	0.19	0.33	0.18	0.27
PSEN- $a(2)$	0.47	0.90	0.15	0.21	0.23	0.48	0.20	0.42	0.21	0.27
PSEN- $a(15)$	0.44	0.86	0.12	0.21	0.24	0.44	0.19	0.35	0.19	0.26
PSEN- $k(1)$	0.43	0.85	0.13	0.19	0.21	0.44	0.18	0.34	0.18	0.27
PSEN- $k(6)$	0.45	0.88	0.13	0.21	0.24	0.45	0.19	0.33	0.19	0.27

配到周边的社交信息这一特点,从而影响轨迹,当社交圈内无其他行人时,影响的行人则充分发挥终点估计模块的作用,由终点指导路径的生成。图6(b)展示了PSEN在行人稀疏的场景中的表现效果。在行人稀疏的场景中,PSEN的注意力权重将会侧重于终点,并由终点对路径预测起到指导作用,这与现实世界的行人行走的特点相符,在独自行走的时候,行人会更多地关注于自己的目的地。

根据以上定性分析,PSEN能够学习到现实世界中行人行走的3个特点:行人具有目的性、注意力根据周边社交环境变化而变化以及在行走过程中实时更新相关参数。



(a) 人群稠密场景

(b) 人群稀疏场景

图6 预测结果示例图

Fig. 6 Example diagram of experimental prediction

(a) crowded scenes; (b) sparsely populated scenes)

5 结论

本文提出基于社会约束与轨迹终点的路径逐步估计网络(PSEN),旨在将数据与行人特点通过3个模块进行有机结合,即社会约束模块、终点估计模块和逐步路径估计模块。其中,社会约束模块在训练过程中自动调整行人与全体行人社交关系的注意力权重;终点估计模块将历史轨迹与社会约束模块的输出进行融合,预测行人的终点;逐步路径估计网络将迭代更新各个模块的输入,生成预测的轨迹。在baseline对比实验中,PSEN在行人轨迹预测任务中的公开数据集ETH/UCY与SDD上取得了不错的效果,其中在UNIV、ZARA1、ZARA2数据集上,模型的性能指标相比所有基准模型(baseline)基本都得到优化,平均位移误差(ADE)和最终位移误差(FDE)分别平均降低7.73%和9.73%。通过消融实验,表明终点估计模块能够有效提升模型性能,在ETH/UCY数据集上平均位移误差(ADE)和最终位移误差(FDE)分别平均降低19%和31%,且在基于社会约

束与终点轨迹的路径估计网络中,社会约束模块对于预测具有最重要的作用,对于模型性能的提升效果最大。

最后,本文所提出的基于社会约束与终点估计的路径网络,仅仅是研究和分析了行人间的特点,并没有对行人与物体,如车辆、障碍物、建筑物等之间的关系进行分析,因此泛化性不够,在新的场景中或在行人稀疏,但其余物体丰富,如车辆数量众多,但只有一两个行人的十字路口的场景下,表现效果可能不甚理想,后续需要进一步针对人与物之间的关系和特征信息进行研究工作。

参考文献(References)

- Alahi A, Goel K, Ramanathan V, Robicquet A, Li F F and Savarese S. 2016. Social LSTM: human trajectory prediction in crowded spaces//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 961-971 [DOI: 10.1109/CVPR.2016.110]
- Amirani J, Hayet J B and Pettré J. 2019. Social ways: learning multi-modal distributions of pedestrian trajectories with GANs//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA: IEEE: 2964-2972 [DOI: 10.1109/CVPRW.2019.00359]
- Antonini G, Bierlaire M and Weber M. 2006. Discrete choice models of pedestrian walking behavior. Transportation Research Part B: Methodological, 40 (8) : 667-687 [DOI: 10.1016/j.trb.2005.09.006]
- Bi H K, Zhang R S, Mao T L, Deng Z G and Wang Z Q. 2020. How can I see my future? FvTraj: using first-person view for pedestrian trajectory prediction//Proceedings of the 16th European Conference on Computer Vision—ECCV 2020. Glasgow, UK: Springer: 576-593 [DOI: 10.1007/978-3-030-58571-6_34]
- Chen G Y, Li J L, Lu J W and Zhou J. 2021. Human trajectory prediction via counterfactual analysis//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 9804-9813 [DOI: 10.1109/ICCV48922.2021.00968]
- Chen H D and Ji Q G. 2023. Scene-constrained spatial-temporal graph convolutional network for pedestrian trajectory prediction. Journal of Image and Graphics, 28(10): 3163-3175 (陈浩东, 纪庆革. 2023. 用于行人轨迹预测的场景限制时空图卷积网络. 中国图象图形学报, 28(10): 3163-3175) [DOI: 10.11834/jig.221027]
- Chen X L, Zhang Y R, Hu S Y and Du Z L. 2024. No-reference point cloud quality assessment based on neighbor information and attention. Journal of Image and Graphics, 29(10): 2979-2991 (陈晓雷, 张育儒, 胡森涌, 杜泽龙. 2024. 基于邻域信息和注意力的无参考点云质量评估. 中国图象图形学报, 29(10): 2979-2991)

- [DOI: 10.11834/jig.230669]
- Cheng J Y, Cheng H, Meng M Q H and Zhang H. 2018. Autonomous navigation by mobile robots in human environments: a survey//Proceedings of 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO). Kuala Lumpur, Malaysia: IEEE: 1981-1986 [DOI: 10.1109/ROBIO.2018.8665075]
- Chiara L F, Coscia P, Das S, Calderara S, Cucchiara R and Ballan L. 2022. Goal-driven self-attentive recurrent networks for trajectory prediction//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans, USA: IEEE: 2517-2526 [DOI: 10.1109/CVPRW56347.2022.00282]
- Dendorfer P, Ošep A and Leal-Taixé L. 2021. Goal-GAN: multimodal trajectory prediction based on goal position estimation//Proceedings of the 15th Asian Conference on Computer Vision. Kyoto, Japan: Springer: 405-420 [DOI: 10.1007/978-3-030-69532-3_25]
- Dong Y H, Wang L, Zhou S P, Hua G and Sun C Y. 2024. Recurrent aligned network for generalized pedestrian trajectory prediction [EB/OL]. [2024-12-01]. <https://arxiv.org/pdf/2403.05810.pdf>
- Gupta A, Johnson J, Li F F, Savarese S and Alahi A. 2018. Social GAN: socially acceptable trajectories with generative adversarial networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2255-2264 [DOI: 10.1109/CVPR.2018.00240]
- Helbing D and Molnár P. 1995. Social force model for pedestrian dynamics. *Physical Review E*, 51(5): 4282-4286 [DOI: 10.1103/PhysRevE.51.4282]
- Hochreiter S and Schmidhuber J. 1997. Long short-term memory. *Neural Computation*, 9(8): 1735-1780 [DOI: 10.1162/neco.1997.9.8.1735]
- Huang Y F, Bi H K, Li Z X, Mao T L and Wang Z Q. 2019. STGAT: modeling spatial-temporal interactions for human trajectory prediction//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 6271-6280 [DOI: 10.1109/ICCV.2019.00637]
- Ivanovic B and Pavone M. 2019. The Trajectron: probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 2375-2384 [DOI: 10.1109/ICCV.2019.00246]
- Li R C, Katsigiannis S and Shum H P H. 2022. Multiclass-SGCN: sparse graph-based trajectory prediction with agent class embedding//Proceedings of 2022 IEEE International Conference on Image Processing (ICIP). Bordeaux, France: IEEE: 2346-2350 [DOI: 10.1109/ICIP46576.2022.9897644]
- Liang J W, Jiang L and Hauptmann A. 2020. SimAug: learning robust representations from simulation for trajectory prediction//Proceedings of the 16th European Conference on Computer Vision—ECCV 2020. Glasgow, UK: Springer: 275-292 [DOI: 10.1007/978-3-030-58601-0_17]
- Liang J W, Jiang L, Niebles J C, Hauptmann A G and Li F F. 2019. Peeking into the future: predicting future person activities and locations in videos//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5718-5727 [DOI: 10.1109/CVPR.2019.00587]
- Liu Y J, Yan Q and Alahi A. 2021. Social NCE: contrastive learning of socially-aware motion representations//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 15098-15109 [DOI: 10.1109/ICCV48922.2021.01484]
- Long J C, Mei J and Ma G F. 2023. Egocentric two-frame pedestrian trajectory prediction algorithm based on a panoramic camera. *IEEE Transactions on Instrumentation and Measurement*, 72: #5001013 [DOI: 10.1109/TIM.2022.3225018]
- Luo Y F, Cai P P, Bera A, Hsu D, Lee W S and Manocha D. 2018. PORCA: modeling and planning for autonomous driving among many pedestrians. *IEEE Robotics and Automation Letters*, 3(4): 3418-3425 [DOI: 10.1109/LRA.2018.2852793]
- Mangalam K, An Y, Girase H and Malik J. 2021. From goals, waypoints and paths to long term human trajectory forecasting//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 15213-15222 [DOI: 10.1109/ICCV48922.2021.01495]
- Mangalam K, Girase H, Agarwal S, Lee K H, Adeli E, Malik J and Gaidon A. 2020. It is not the journey but the destination: endpoint conditioned trajectory prediction//Proceedings of the 16th European Conference on Computer Vision—ECCV 2020. Glasgow, UK: Springer: 759-776 [DOI: 10.1007/978-3-030-58536-5_45]
- Meng M C, Wu Z Y, Chen T, Cai X R, Zhou X S, Yang F and Shen D G. 2022. Forecasting human trajectory from scene history//Proceedings of the 36th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: 24920-24933
- Mohamed A, Qian K, Elhoseiny M and Claudel C. 2020. Social-STGCNN: a social spatio-temporal graph convolutional neural network for human trajectory prediction//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 14412-14420 [DOI: 10.1109/CVPR42600.2020.01443]
- Raksincharoensak P, Hasegawa T and Nagai M. 2016. Motion planning and control of autonomous driving intelligence system based on risk potential optimization framework. *International Journal of Automotive Engineering*, 7(1): 53-60 [DOI: 10.20485/jsaeijae.7.AVEC14_53]
- Robicquet A, Sadeghian A, Alahi A and Savarese S. 2016. Learning social etiquette: human trajectory understanding in crowded scenes//Proceedings of the 14th European Conference on Computer Vision—ECCV 2016. Amsterdam, the Netherlands: Springer: 549-565 [DOI: 10.1007/978-3-319-46484-8_33]
- Sadeghian A, Kosaraju V, Sadeghian A, Hirose N, Rezatofighi H and

- Savarese S. 2019. SoPhie: an attentive GAN for predicting paths compliant to social and physical constraints//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1349-1358 [DOI: 10.1109/CVPR.2019.00144]
- Salzmann T, Ivanovic B, Chakravarty P and Pavone M. 2020. Trajec-tron++: dynamically-feasible trajectory forecasting with heteroge-neous data//Proceedings of the 16th European Conference on Com-puter Vision—ECCV 2020. Glasgow, UK: Springer: 683-700 [DOI: 10.1007/978-3-030-58523-5_40]
- Shi L S, Wang L, Long C J, Zhou S P, Zhou M, Niu Z X and Hua G. 2021. SGCN: sparse graph convolution network for pedestrian tra-jectory prediction//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 8990-8999 [DOI: 10.1109/CVPR46437.2021.00888]
- Shi L S, Wang L, Zhou S P and Hua G. 2023. Trajectory unified trans-former for pedestrian trajectory prediction//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 9641-9650 [DOI: 10.1109/ICCV51070.2023.00887]
- Tay M K C and Laugier C. 2008. Modelling smooth paths using Gaussian processes//Proceedings of the 6th International Conference on Field and Service Robotics. Berlin, Heidelberg: Springer: 381-390 [DOI: 10.1007/978-3-540-75404-6_36]
- Wang Y and Chen S W. 2023. Multi-agent trajectory prediction with spatio-temporal sequence fusion. *IEEE Transactions on Multime-dia*, 25: 13-23 [DOI: 10.1109/TMM.2021.3120535]
- Wong C H, Xia B H, Hong Z M, Peng Q M, Yuan W, Cao Q, Yang Y B and You X. 2022. View vertically: a hierarchical network for tra-jectory prediction via Fourier spectrums//Proceedings of the 17th European Conference on Computer Vision. Tel Aviv, Israel: Springer: 682-700 [DOI: 10.1007/978-3-031-20047-2_39]
- Xu C X, Li M S, Ni Z Y, Zhang Y and Chen S H. 2022a. GroupNet: multiscale hypergraph neural networks for trajectory prediction with relational reasoning//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 6488-6497 [DOI: 10.1109/CVPR52688.2022.00639]
- Xu C X, Mao W B, Zhang W J and Chen S H. 2022b. Remember inten-tions: retrospective-memory-based trajectory prediction//Proceed-ings of 2022 IEEE/CVF Conference on Computer Vision and Pat-tern Recognition. New Orleans, USA: IEEE: 6478-6487 [DOI: 10.1109/CVPR52688.2022.00638]
- Xu C X, Tan R T, Tan Y H, Chen S H, Wang Y G, Wang X C and Wang Y F. 2023. EqMotion: equivariant multi-agent motion predic-tion with invariant interaction reasoning//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Van-couver, Canada: IEEE: 1410-1420 [DOI: 10.1109/cvpr52729.2023.00142]
- Xu P, Hayet J B and Karamouzas I. 2022c. SocialVAE: human trajec-tory prediction using timewise latents//Proceedings of the 17th Euro-pean Conference on Computer Vision. Tel Aviv, Israel: Springer: 511-528 [DOI: 10.1007/978-3-031-19772-7_30]
- Yamaguchi K, Berg A C, Ortiz L E and Berg T L. 2011. Who are you with and where are you going//CVPR 2011. Colorado Springs, USA: IEEE: 1345-1352 [DOI: 10.1109/CVPR.2011.5995468]
- Yamaguchi N, Tokumaru H, Fukuda O, Okumura H and Yeoh W L. 2022. Bicycle-based collision prevention system using pedestrian trajectory prediction//2022 10th International Symposium on Com-puting and Networking Workshops (CANDARW). Himeji, Japan: IEEE: 151-153 [DOI: 10.1109/CANDARW57323.2022.00013]
- Yu C J, Ma X, Ren J W, Zhao H Y and Yi S. 2020. Spatio-temporal graph transformer networks for pedestrian trajectory prediction//Pro-ceedings of the 16th European Conference on Computer Vision—ECCV 2020. Glasgow, UK: Springer: 507-523 [DOI: 10.1007/978-3-030-58610-2_30]
- Yu L, Li H Y, Jiao C L, Leng Y F and Xu G Y. 2022. Trajectory predic-tion in complex scenes based on multi-head attention adversarial mechanism. *Chinese Journal of Computers*, 45(6): 1133-1146 (余力, 李慧媛, 焦晨璐, 冷友方, 徐冠宇. 2022. 基于多头注意力对抗机制的复杂场景行人轨迹预测. *计算机学报*, 45(6): 1133-1146) [DOI: 10.11897/SP.J.1016.2022.01133]
- Yuan Y, Weng X, Ou Y L and Kitani K. 2021. AgentFormer: agent-aware transformers for socio-temporal multi-agent forecasting//Pro-ceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 9793-9803 [DOI: 10.1109/ICCV48922.2021.00967]
- Zhang Y, Lu H Z, Wang J, Zhang L P and Hu M F. 2023. Short-term memory and CenterTrack based vehicle-related multi-target track-ing method. *Journal of Image and Graphics*, 28(10): 3107-3122 (张瑶, 卢焕章, 王珏, 张路平, 胡谋法. 2023. 短时记忆与CenterTrack的车辆多目标跟踪. *中国图象图形学报*, 28(10): 3107-3122) [DOI: 10.11834/jig.220026]
- Zhao H and Wildes R P. 2021. Where are you heading? Dynamic tra-jectory prediction with expert goal examples//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Mon-treal, Canada: IEEE: 7609-7618 [DOI: 10.1109/ICCV48922.2021.00753]
- Zhu P F and Zhang D P. 2023. Pedestrian trajectory prediction based on spatio-temporal graph. *Computer Systems and Applications*, 32(12): 284-291 (朱鹏飞, 张德平. 2023. 基于时空图的行人轨迹预测. *计算机系统应用*, 32(12): 284-291) [DOI: 10.15888/j.cnki.csa.009335]

作者简介

吴恩泓,男,硕士研究生,主要研究方向为计算机视觉和计算机图形学。E-mail: wuenh@mail2.sysu.edu.cn

纪庆革,通信作者,男,副教授,硕士生导师,主要研究方向为计算机视觉、计算机图形学、虚拟现实、计算机仿真和区块链。E-mail: issjqg@mail.sysu.edu.cn