

行人轨迹预测方法综述

李琳辉^{1,2}, 周彬¹, 任威威¹, 连静^{1,2}

(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: 随着深度学习技术的突破和大型数据集的提出, 行人轨迹预测的准确度已经成为人工智能领域的研究热点之一。主要对行人轨迹预测的技术分类和研究现状进行详细的综述。根据模型建模方式的不同, 将现有方法分为基于浅层学习的轨迹预测方法和基于深度学习的轨迹预测方法, 分析了每类方法中具有代表性的算法的效果及优缺点, 归纳了当前主流的轨迹预测公开数据集, 并在数据集中对比了主流轨迹预测方法的性能, 最后对轨迹预测技术面临的挑战与发展趋势进行了展望。

关键词: 轨迹预测; 深度学习; 序列决策

中图分类号: TP520

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202140

Review of pedestrian trajectory prediction methods

LI Linhui^{1,2}, ZHOU Bin¹, REN Weiwei¹, LIAN Jing^{1,2}

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: With the breakthrough of deep learning technology and the proposal of large data sets, the accuracy of pedestrian trajectory prediction has become one of the research hotspots in the field of artificial intelligence. The technical classification and research status of pedestrian trajectory prediction were mainly reviewed. According to the different modeling methods, the existing methods were divided into shallow learning and deep learning based trajectory prediction algorithms, the advantages and disadvantages of representative algorithms in each type of method were analyzed and introduced. Then, the current mainstream public data sets were summarized, and the performance of mainstream trajectory prediction methods based on the data sets was compared. Finally, the challenges faced by the trajectory prediction technology and the development direction of future work were prospected.

Key words: trajectory prediction, deep learning, sequence decision

1 引言

人们生活中出现了越来越多的智能自主系统, 这些系统感知、理解和预测人类行为的能力变得越来越重要。具体来说, 预测智能体 (agent) 的未来

位置并根据这些预测进行规划是自动驾驶车辆、服务机器人和高级监控系统 (包括智能交通或者智能城市) 的关键任务^[1]。理解人体运动是智能系统与人类共存和互动的一项关键技能, 其涉及表征、感知和运动分析等方面。而预测在人体运动分析中起

收稿日期: 2020-07-27; 修回日期: 2020-08-28

通信作者: 连静, lianjing80@126.com

基金项目: 国家自然科学基金资助项目 (No.51775082, No.61976039); 中央高校基本科研业务费专项资金资助项目 (No.DUT19LAB36, No.DUT20GJ207); 大连市科技创新基金资助项目 (No.2018J12GX061)

Foundation Items: The National Natural Science Foundation of China (No.51775082, No.61976039), The China Fundamental Research Funds for the Central Universities (No.DUT19LAB36, No.DUT20GJ207), Science and Technology Innovation Fund of Dalian (No.2018J12GX061)

着重要的作用,随着时间的推移,模型可对涉及多个智能体的场景进行预测,并以主动的方式对这些场景信息进行整合,即增强主动感知、预测性规划、模型预测性控制或人机交互的效果。因此,近年来行人轨迹预测在多个领域中受到越来越多的关注,例如自动驾驶汽车、服务机器人、智能交通、智慧城市等领域。

保证交通场景中道路使用者的自身安全是自动驾驶车辆被普及应用的前提条件^[2]。行人作为交通场景中的主要参与者,对其未来运动轨迹的合理推理和预测对于自动驾驶和道路安全至关重要。行人轨迹预测的难点在于以下3点。首先,行人运动灵活,精准预测其未来的运动轨迹几乎是一个不可能完成的任务,但是通过观察其历史时刻的运动轨迹,利用一些算法可以大致推测出行人未来的运动轨迹(例如最简单的匀速直线运动)。但是在实际道路场景中,相比于自行车、汽车等运动学模型,行人运动更加灵活多变,很难对行人建立合理的动力学模型(因为行人可以随时转弯、停止和运动等),有时甚至连人类驾驶员都很难预测出行人未来的运动轨迹。其次,在实际交通场景中,某一行人未来的运动不仅受自己意图的支配,也受周围行人的影响(例如避障、交互等),根据 Moussaid M 等人^[3]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,这种行人间的交互非常抽象,往往很难在算法中进行精确的建模。最后,常规算法模型的实现是找到一个输入到输出的函数映射,对于轨迹预测的模型来说,其对应着一个序列到另一个序列的映射,常规的模型或训练方式容易使模型预测结果陷入平均状态(预测结果倾向于预测一种折中的轨迹),显然使用常规的训练模型无法很好地对行人轨迹进行有效的预测。

行人轨迹的高动态性、随机性以及与交通环境智能体之间复杂的交互使轨迹预测问题充满挑战,但是对行人轨迹进行长时间的预测仍有必要,这对实现自动驾驶汽车的主动规划和决策^[4]影响巨大。目前国内外有关行人轨迹预测的研究与日俱增,对该领域的相关技术和文献进行综述十分必要。本文对行人轨迹预测的问题描述、研究挑战以及问题处理流程进行了梳理,对行人轨迹预测的研究进展,尤其是近几年的研究成果进行了综述。本文第2节给出了轨迹预测的问题描述。第3节对现有的轨迹预测方法进行了分类,并在此基础上简要地总结了

基于浅层学习及机器学习的方法,同时以发现问题为导向,以解决问题为目标,对基于深度学习的轨迹预测方法进行了详细的分析介绍。第4节介绍了目前业内主流的轨迹预测公开数据集,将一些轨迹预测方法在数据集中进行测试后,对取得优异测试结果的方法进行分析和比较。第5节对行人轨迹预测技术在今后发展中面临的挑战和发展趋势进行了总结和展望。

2 轨迹预测的问题描述

预测问题的本质是根据得到的智能体特征信息以及环境信息,推断出智能体未来时刻可能的状态。而轨迹预测问题作为预测问题的子问题,可以被视作序列决策问题,即通过引入行人位置信息、自我运动历史信息等,根据行人在过去时间段内的轨迹来预测行人在未来固定时刻的位置坐标。

行人的运动行为过程通常由行人的运动意图或外部环境直接(间接)决定。抽象来说,通常将行人轨迹预测问题分为3个方面,如图1所示。首先是外部激励,大多数预测方法很大程度上依赖于可观察到的外部激励信息,例如行人历史运动轨迹和运动学特征(位置、速度、角速度、人体关节角度等其他属性),同时还存在一些其他形式的环境信息(例如道路的几何场景、树木等静态物理信息以及人-车辆、人-人等移动交互信息),而对于端到端的方法而言,实际上其主要依赖于原始传感器数据。此外,行人轨迹预测的复杂性还来自不同的社会行为(如人群平行行走或结伴行走、人群中避免碰撞以及行人从不同方向汇集等),因此通俗来讲,在预测过程中一般还需要考虑行人与环境(如障碍物)以及行人与行人之间的交互。其次,对于行人轨迹预测的建模方法来说,有基于运动学和动力学方法的参数化建模方法,有基于浅层学习和深度学习的建模方法,本文重点对后者进行归纳,对模型的优缺点和相似性进行对比分析,并对当前轨迹预测领域存在的问题与未来发展趋势进行剖析和展望。最后,不同的建模方法使用的损失函数不同,进而会产生不同的参数化或结构化的预测形式,例如预测代理状态上的双变量高斯分布形式、行人预测的单轨迹或多轨迹形式,除此之外,还有基于图模型的运动模式以及预测网格的概率分布形式等。

具体而言,典型的轨迹预测问题的处理流程包括:前期数据集的获取,数据集的预处理和输入,



图 1 行人轨迹预测问题的组成元素

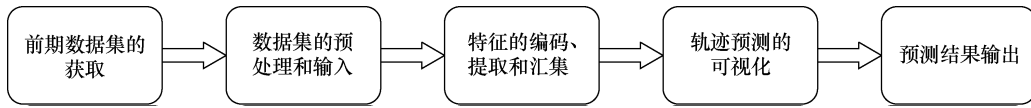


图 2 轨迹预测问题的处理流程

特征的编码、提取和汇集，轨迹预测的可视化和预测结果输出，如图 2 所示。现有的轨迹预测方法一般默认相应场景的数据集已经完成，由数据集直接提供轨迹信息或者场景信息。具体的常见公开数据集见第 4 节。

3 轨迹预测的技术分类

自 20 世纪 90 年代以来，大量基于浅层学习的轨迹预测模型被提出，这些模型对计算设备的算力要求比较高，且缺乏统一的评价标准，被用于测试该模型的数据集的质量也参差不齐，因此基于浅层学习的轨迹预测模型的综述很少被研究者提出。近十几年来，机器学习技术尤其是深度学习技术逐步兴起，由于递归神经网络在处理时序数据上的优异性能，基于递归神经网络的时序预测模型层出不穷。下面将对国内外的行人轨迹预测方法进行分类及综述。

根据预测模型的建模方式，本文大致将行人轨迹预测方法分为基于浅层学习的轨迹预测方法和基于深度学习的轨迹预测方法。浅层学习中的基于运动学的方法是比较早被应用在轨迹预测领域的，这类方法一般需要对行人的运动学特征（速度、位置和角速度等）进行建模，并将其与贝叶斯滤波器、马尔可夫网络或者贝叶斯网络结合起来，将当前状态传播到未来状态进行预测。根据是否考虑行人之间的交互影响，可将基于深度学习的轨迹预测的模型划分为单轨迹预测模型和交互轨迹预测模型；根据是否生成确定性的行人轨迹，又可将其划分为确定性轨迹预测模型和可接受轨迹预测模型。近十几年，随着模型的提出和优化，目前基于深度学习的轨迹预测方法趋于模块化，其需要考虑和利用的信息逐渐完善，预测的精度和实时性也逐渐提高。综上所述，行人轨迹预测方法分类如图 3 所示。

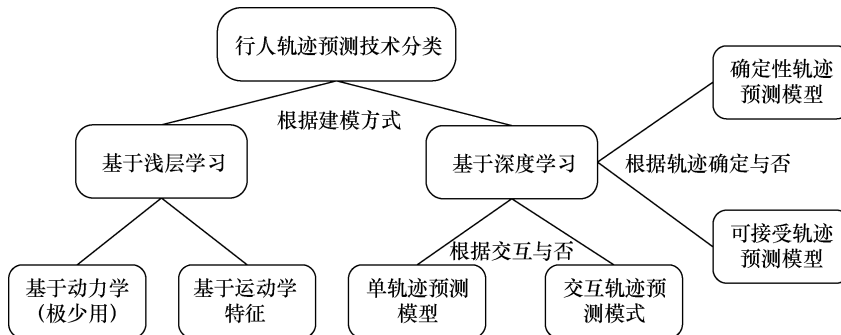


图 3 行人轨迹预测方法分类

实际上,一个好的轨迹预测模型往往需要综合利用不同的信息(轨迹信息、语义信息、社交信息),因此很难从诸多信息中的单一角度评估预测模型的好坏。为了对现有的轨迹预测方法进行更加细致的梳理和归纳,本文将以预测模型的建模方式为主干,对轨迹预测方法进行总结,同时以发现问题为导向,以解决问题为目标,对当前业界主流的轨迹预测前沿方法进行详细的介绍。

3.1 基于浅层学习的轨迹预测方法

最基本、最早出现的轨迹预测方法将基本的运动学模型(恒定速度/加速度/转弯)和贝叶斯滤波器或其扩展组合起来,将当前状态传播到未来状态^[5]。Schneider N 等人^[6]将基于单一的运动学模型的方法与基于多模型交互的方法进行了比较,结果表明,基于多模型交互的方法可以在机动过程中降低 30 cm 的横向位置估计误差,但是该模型的方法以恒定速度为假设,这显然不符合实际情况,还使得基于贝叶斯滤波器的模型难以捕获行人的切换动态,且其用来测试的数据集样本数量和运动类型有限,不足以支撑更复杂的运动模型的建立与预测。Pavlovic V 等人^[7]采用切换线性动力学系统(switched linear dynamical system, SLDS)模型^[8]来描述非线性和时变动力学,该模型基于马尔可夫链进行概率转移,在多种线性运动学模型间进行切换,从而预测实际情况中的非线性运动。SLDS 模型每次预测都具有不同的动态特性,并以指定的先验概率和转移概率对切换状态的隐马尔可夫链进行调节。然而,基于该模型的运动特征信息有时不足以支持模型进行状态的切换,且其对于一些更加复杂的运动模型来说效果有限,需要通过构建更大的运动捕捉数据集来满足更加复杂的运动模型测试的精确性要求。Kooij J F P 等人^[9]建立了基于上下文的动态贝叶斯网络(dynamic bayesian network, DBN)模型,用于行人路径预测。该模型将上下文信息(即行人头部方向、情况紧急程度和环境空间布局)作为潜在状态合并到 SLDS 模型的顶部,从而控制 SLDS 模型的切换状态,与只使用 SLDS 模型相比,该模型可以进行更准确的预测。但是,无论是基于 SLDS 模型的预测方法,还是基于 DBN 模型的预测方法,在进行模型预测推理和数学模型搭建的过程中始终需要大量的计算,这对于计算设备来说是一种巨大的算力消耗,而且不能很好地体现额外的场景(例如红绿灯、人行横道)以及基于 SLDS 模型的

基本运动类型(例如转弯情况)等信息。

Helbing D 等人^[10]提出了一种具有吸引力和排斥力的行人运动模型,被称为社会力模型。该模型被广泛应用于机器人和活动理解等^[11]领域。Alahi A 等人^[12]提出通过从人群里的行人轨迹中学习他们的相对位置来呈现社会亲和力特征,而 Yi S 等人^[13]提出使用人类属性来改善人群中的预测。

随着机器学习的快速发展,基于运动学的方法在做预测时有时会使用一些基于机器学习的跟踪算法来改进跟踪和预测,例如卡尔曼滤波(Kalman filter, KF)、马尔可夫模型(Markov model, MM)和高斯过程^[14-16](Gaussian Process, GP)等。KF 模型的优势在于处理行人轨迹预测问题时能够有效地处理无噪声点的轨迹数据,对短时间(1步/2步)内的预测精确度较高。相反,对长时间(10s/5步以上)的预测误差较大,模型复杂度增高,严重影响预测精度,且 KF 模型随着噪声的增大变得愈发敏感,预测精度也近似呈线性降低。MM 对行人运动过程的状态预测效果良好,但其对行人轨迹的波动较为敏感,且不适用于中长期的行人轨迹预测,一阶 MM 仅考虑了当前行人运动轨迹点对未来轨迹点的影响,历史轨迹点的数据信息无法被尽可能地利用,而高阶 MM 大大增加了模型计算的复杂度。GP 模型通过假设隐变量服从高斯分布,为概率预测提供非参数模型,基于 GP 模型预测的轨迹是从历史轨迹数据中学习得到的^[14],并通过在 GP 模型中指定适当的核函数(协方差函数)来明确行预测轨迹建模涉及的不确定性。对于一维高斯函数来说,随机变量 X 服从一个位置参数为 μ 、尺度参数为 σ 的概率分布(正态分布),记为 $X \sim N(\mu, \sigma^2)$,其中均值 μ 、方差 σ^2 为高斯分布的超参数, μ 决定了分布曲线的位置, σ^2 决定了分布曲线的幅度,参数需要通过数据训练得到。GP 模型能够较为有效地预测具有噪声点的轨迹数据,同时可以很好地避免轨迹数据离散性质的不足,并在此基础上有效地表达行人运动轨迹分布的统计特征。但是构造高斯过程十分复杂,需要付出较高的时间代价,且高斯模型是非稀疏模型,需要完整的样本或特征信息进行行人轨迹的预测,并且随着数据的增多,构造高斯过程的计算量大大增加。而且当单独使用 GP 模型时,在不同时间点,GP 模型预测误差的差别也比较大。

基于运动学并结合机器学习跟踪算法的这类

方法有一个很明显的特点，即需要根据历史时序的数据建立时序递推数学式： $X^t = f(X^{t-1})$ 或者 $P(X^t | X^{t-1})$ 。这类方法通常具有严格的数学证明和假设，能处理一些常规的问题，但是对于一些复杂的问题就变得“束手无策”了。这是因为基于运动学的方法中会引入一些先验假设，例如隐变量服从高斯分布、线性的状态转换方程以及观测方程等，而最终这些假设也限制和约束了方法的整体性能。

总而言之，基于浅层学习的轨迹预测方法在早期取得了一些成果，为行人轨迹预测的发展做出了诸多贡献。但是由于基于运动学（恒定速度/加速度/转弯）方法的局限性、运动特征信息提取的不足、特定场景信息的缺失、模型构建的复杂性以及当时的大型数据集样本及行人运动类型有限等问题，最终的预测效果与实际情况存在一定差距，传统的方法难以对较为复杂的行人运动模型及场景进行较为精确的预测。

3.2 基于深度学习的轨迹预测方法

基于浅层学习的轨迹预测方法需要对模型进行复杂、严谨的建模，而基于深度学习的轨迹预测方法一般不需要假设固定的数学模型，其网络可以凭借大规模的数据集来学习更加合理的映射关系。近几年，随着深度学习的兴起，各种用于处理时序数据的模型如雨后春笋般被提出，使得基于神经网络的轨迹预测算法流行起来，且这些算法的预测效果较传统算法有了较大的提升。

基于深度学习的轨迹预测方法有两个关键要素。一是网络结构的选择。不同的研究者对轨迹预测问题和场景辅助信息的使用有不同的理解，其选择的网络结构也可能不相同，对轨迹预测特征进行提取、汇集及预测时，结果也会有所差异。二是损失函数的设计。在网络优化的过程中，最终模型预测效果的好坏取决于能否设计出合理有效的损失函数。损失函数不仅会影响神经网络的优化过程，而且决定着大型公开数据集能否得到充分有效的利用。总体来说，如何选择、设计出合理有效的网络结构和损失函数是目前基于深度学习的轨迹预测方法的研究重点。

近年来，用于序列预测的循环神经网络（recurrent neural network, RNN）及其变体（包括长短期记忆（long short-term memory, LSTM）网络^[17]和门控循环单元（gated recurrent unit, GRU）^[18]），在序列预测任务中（如语音识别^[19-20]、标题生成^[21]、

机器翻译^[11]）取得了巨大的成功。由于循环神经网络在处理序列数据时存在长时依赖的问题，即当预测点与依赖的相关信息距离比较远时，网络就难以学到相关信息。为解决 RNN 的这一缺陷，自然而然地将 LSTM 引入轨迹预测领域。而序列模型 Seq2Seq 为序列生成任务提供了一个编码器-解码器（encoder-decoder）框架，该框架以输入序列为条件，学习输出序列（即 $P(X^t | X^{t-1})$ ）的条件分布，该框架可以将 LSTM 同时用于编码器和解码器中。编码器 LSTM 首先遍历输入序列，将其编码为固定长度的语义向量，然后将语义向量传递给解码器 LSTM，随后生成输出序列，近几年的模型大多采用了该框架^[22]。编码器-解码器框架虽然较为经典，但是也存在一定的不足之处，即在编码和解码两者之间建立联系的唯一桥梁就是上述固定长度的语义向量，整个序列携带的信息都将被压缩到此向量中，这会出现两个问题：其一，输入序列的信息不能充分地被语义向量表示；其二，后输入的序列信息会覆盖之前输入的信息。这两个问题会随着输入序列长度的增加愈发严重，最终造成解码准确度下降。为解决此问题，注意力模型诞生了，此模型不再要求编码器把整个序列携带的信息都压缩到语义向量中，而是将输入序列的信息编码成一个向量序列，解码时选择性地对其子集进行处理，以保证充分利用序列所携带的信息。在此基础上，Karatzolou A 等人^[23]通过对人类轨迹采用 Seq2Seq 序列模型扩展了 LSTM，并探讨了基于注意力的 Seq2Seq 序列模型对模型性能的影响，即基于注意力的 Seq2Seq 序列模型通过对序列信息的充分利用，解决了 Seq2Seq 序列模型的性能随着输入序列增长而下降的问题，验证了 Seq2Seq 序列模型在轨迹建模和运动模式预测中的有效性。

基于先前的研究，Alahi A 等人^[24]提出了用于轨迹预测的社会长短期记忆（social long short-term memory, S-LSTM）网络模型，该模型对行人之间可能发生相互冲突的社交互动进行了建模。每条轨迹被建模为一个 LSTM 层，并且不同的 LSTM 层可以通过社交池化（social-pooling）层共享信息，从而生成无冲突的轨迹，具体的计算思路是将该行人周围的区域划分成 $N \times N$ 个网格，每个网格大小相同，落入这些网格中的行人将会参与交互计算。该模型成功预测了不同社交互动引起的非线性行为（例如人群同时移动），但是此模型仅建立了单一

的模型设置（行人共享空间），在此基础上还可建立诸多对象（例如行人、自行车、滑板车、手推车等）的社交池化层来共享空间信息，另外也可通过加入场景中的图像信息建立人与空间的交互信息。除此之外，基于社交池化的模型在每次训练预测时都要对社交向量进行计算，这使得模型预测的实时性不高，该参考文献同样提出了将占用地图式长短期记忆网络（occupancy map long short-term memory, O-LSTM）结构作为社交池化层的简化版，以此来提升预测速度。Kitani K M 等人^[25]已经证明，将静态环境的语义特征信息（人行道的位、草地区域的延伸等）输入模型有助于更准确地预测未来时刻的行人轨迹。参考文献^[25]也通过使用语义场景信息对人-空交互进行建模，并以此来推断场景中的可穿越区域，从而预测行人的未来轨迹。在 O-LSTM 的基础上，Xue H 等人^[26]采用分层 LSTM 结构，提出了社交场景长短期记忆网络（social scene long short-term memory, SS-LSTM）模型。在 O-LSTM 考虑行人交互的基础上，该模型额外考虑行人所处的场景信息，做出轨迹预测，且相较于其他基于 LSTM 的模型，其在 ETH、UCY 数据集中有更好的表现。未来，增加行人交互信息所占的比重（比如引入行人之间的距离）或增加新的注意力机制（空间-时间）也会对其预测效果产生新的影响。

3.2.1 基于生成式网络架构的预测方法

虽然基于简单的 LSTM 以及 Seq2Seq 序列模型的轨迹预测网络在精度方面能够取得较为理想的结果，但是它们的预测结果为单一的预测轨迹输出，是模型预测轨迹的平均状态，并且与数据集结果高度拟合，这与行人的高动态性和随机性冲突。为解决模型的架构带来的问题，基于生成模型的一系列预测方法（如生成对抗网络（generative adversarial network, GAN）^[27]、变分自编码器（variational auto-encoder, VAE）^[28]）被相继提出。

为解决先前模型中的平均轨迹的问题，并对池化汇集模型进行改进，Gupta A 等人^[29]将基于生成对抗网络的方法引入行人轨迹预测领域，基于先前提到的 Seq2Seq 序列模型和社交池化层，通过引入噪声 z 、改进多样性损失函数，模型趋向于生成多样性的轨迹，并且使用最大池化方法对全局的行人进行交互分析，最终提出一种社会生成对抗网络（social generative adversarial network, Social-GAN）

模型。该模型的优势在于强调预测轨迹在社会规则上的规范性、合理性，即相对于其他预测模型，该模型生成的路径更加合理。同时该模型解决了预测结果与现实不符的单一预测轨迹输出问题，且相比于双向长短期记忆网络（bidirectional long short-term memory, B-LSTM）、S-LSTM 等模型，该模型的预测速度有了较大提升。但是该模型在进行池化汇集时提取的特征是经过最大池化后的最大特征，忽略了对行人交互有用的其他特征信息，并且该模型采用了传统的 GAN 架构，网络训练不稳定，容易崩溃。Amirian J 等人^[30]通过引入 Info-GAN 架构来改进传统 GAN 的架构，以此来解决 Social-GAN 模型训练易崩溃和掉落的问题。作者通过舍弃 L2 代价函数，引入基于交互信息的损失函数（information loss），增强了模型对多条合理轨迹的预测能力，并引入注意力机制^[31]，使模型自主分配对交互信息的关注。该文献表明，新提出的 Info-GAN 架构可以极大地改善多模式行人轨迹预测，避免类似训练崩溃的问题。在此基础上，Sadeghian A 等人^[32]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 架构生成多条物理条件下的可接受轨迹。Kosaraju V 等人^[33]采用基于 Cycle-GAN 的网络架构和训练方法，保证 GAN 在生成轨迹时对噪声的敏感性，从而有助于生成多样性的轨迹。该文献使用视觉几何组（visual geometry group, VGG）网络提取场景图像特征，使用 LSTM 提取行人轨迹特征，根据提取到的特征差异，分别使用基于点积（scaled dot-product）和图注意力网络（graph attention network, GAT）的多种注意力机制，以得到对于各种输入最合理的注意力向量。该模型可以理解较复杂的行人运动社会本质，不仅能够为特定的行人生成多个轨迹，而且可以通过多模式同时为多个行人预测更真实的运动轨迹。

Cheng H 等人^[34]提出了一种不同于上述 GAN 架构的轨迹预测生成模型，该模型使用带条件的变分自编码器（conditional variational auto-encoder, CVAE）来实现对轨迹多样性的预测，为了刻画噪声与已知轨迹和预测轨迹之间的分布情况 $P(z|X,Y)$ ，该模型通过引入变分估计，将噪声的分布情况简化成含有参数（均值、方差）的高斯分布，并将假定分布情况后的后验分布作为预测模型，再对此预测模型进行训练，隐式地从轨迹 X 和 Y 中建立与假定

分布之间的联系。由于基于 GAN 的方法不可避免地涉及采样操作，该方法在噪声分布 $z \sim \mathcal{N}(\mu, \sigma)$ 中直接采样，是一个离散的过程，且该方法在进行反向传播时不可微，导致网络无法更新。参考文献[34]找到了新的解决方案，相比于噪声分布 $z \sim \mathcal{N}(\mu, \sigma)$ 直接采样，该文献中的噪声分布为 $z = \mu + \sigma \odot \varepsilon$ ，其中 ε 满足高斯分布，在传播中可以理解为常数，从而避免了采样问题。该解决方案的预测结果体现了一个目标在某一时刻的轨迹多样性，并按随机变量采样的原理来拟合高斯分布，按照似然估计的原理进行排序。Yang B 等人^[35]也采用这种方法对噪声进行处理。

Liang J W 等人^[36]同样使用 LSTM 来接收历史信息并预测行人的未来轨迹。与其他模型不同的是，这个模型不仅接收行人的历史位置轨迹信息，同时也提取行人外观、人体骨架、周围场景布局以及周围行人的位置关系，通过增加输入信息提升预测性能。除了预测具体的轨迹，该模型还会做粗粒度预测（决策预测），输出行人未来时刻可能在的区域。自此，基于深度学习的轨迹预测方法开始往多任务以及模块化方向发展。而 Sun J H 等人^[37]将轨迹预测问题视为分类和回归问题的结合体，对于相同的一条历史轨迹来说，不同的目标可能会存在不同的未来轨迹（例如十字路口的两辆车，一辆可能直行，另一辆可能左转，即意图不同），而模型会预测不同意图的多个终点，并基于这些终点生成不同的候选轨迹。为减少模型预测轨迹的搜索空间，作者将轨迹建模为三次曲线，通过生成曲线簇来生成候选轨迹集合，并对候选轨迹进行分类和回归运算，分类模块对每个候选轨迹进行二分类，回归模块对候选轨迹进行修正，以得到更加精准的预测结果。

总而言之，基于生成式网络架构的预测方法在进行行人轨迹预测时能保证较高的精度，提升了模型的预测速度，并大大提高了预测轨迹在社会规则上的规范性，使复杂模型的建立更为合理有效。

3.2.2 基于图网络的预测方法

参考文献[38-39]介绍了图卷积网络（graph convolutional network, GCN）的研究现状，将卷积神经网络（convolutional neural network, CNN）的概念扩展到图中，图中定义的卷积运算将目标节点属性与相邻节点属性进行加权聚合^[40]。GCN 总体上与

CNN 相似，但是在图上进行卷积操作时需要邻接矩阵进行相关的定义和计算。参考文献[28,41-42]将 GCN 扩展到其他应用，例如矩阵补全和变分自动编码器。

Gupta A 等人^[29]和 Sadeghian A 等人^[32]利用具有交互机制的 GAN 或变分自动编码器的方法来考虑场景中的所有个体。但是这两种方法都无法学习人类行为的真正多模态分布，只能学习具有高方差的单一行为模式。此外，两种方法都受到它们学习社交行为的建模方式的限制，前者通过对场景中的所有行人使用相同的社交矢量来提取信息，而后者则需要手动定义排序操作，这种操作复杂且实时性较差，可能在所有情况下都无法达到最佳效果。在轨迹预测问题中，可以将行人之间的交互表达为图形^[43]，图形中的节点代表行人，而边缘代表行人之间的互动；较高的边缘权重对应更重要的交互。参考文献[43]通过使图完全连接，以高效的方式对人类之间的局部和全局交互进行建模，而无须使用可能丢失重要特征的模型，如合并或排序等。在参考文献[44]中，行人集合被建模为时空图（spatio-temporal graph, STG），时空图中的边（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，时空图输出双变量高斯分布的参数，该方法能较好地对时空信息进行有效建模，但该方式计算较为复杂。Haddad S 等人^[45]基于 LSTM 提出新型的时空图，旨在实现在拥挤的环境下，通过将行人-行人、行人-静态物品两类交互纳入考虑，对行人的轨迹做出预测。

Kosaraju V 等人^[33]将基于图注意力的社会化生成对抗网络（social bicycle-GAN and graph attention network, Social-BiGAT）^[13]引入轨迹预测领域，该网络可以更好地对场景中行人的社交互动进行建模，并生成逼真的多模态轨迹预测模型。图注意力网络引入了注意力机制，可更好地实现邻居聚合，该网络通过处理模型编码的轨迹信息，增强了轨迹预测的推理能力，也赋予了模型一定的可解释性。图注意力网络允许在可以被表示为图的任何类型的结构化数据上应用基于注意力机制的架构。图注意力网络基于图卷积网络的先验信息构建，允许模型隐式地为图中的节点分配不同的权重。Mohamed A 等人^[46]将行人的轨迹建模为时空图以替换聚集层，他们设计了一个特定的加权邻接矩阵，并通过一个核函数将行人之间的

社交互动嵌入邻接矩阵中,参考文献[46]使用图卷积神经网络和时间卷积网络对时空图进行处理,解决递归神经网络训练时参数过多和效率低下的问题。

总而言之,随着深度学习的快速发展,基于深度学习的轨迹预测方法成为研究热点。理论上而言,基于深度学习的方法基本可以解决利用传统的浅层学习方法的行人轨迹预测问题,但当神经网络足够深、功能足够强大时,如果数据集数据量过小,就非常容易产生过拟合的问题,从而影响预测精度。随着数据集规模越来越大、社交交互场景信息越来越丰富、数据集所涵盖的边界情况(*corner case*)越来越完善,实验场景就会越来越接近于现实场景,增加数据量、合理有效地设计神经网络结构及损失函数,对于行人轨迹预测的效率及精度都会有很大的改善。近几年图卷积网络被用于轨迹预测,取得了亮眼的成绩,但是基于图网络算法的模型存在模型建图效率低下、邻接矩阵难以确定等问题。相信随着图网络技术的发展和成熟,基于图网络的轨迹预测方法发展前景会更好。

4 数据集及性能比较

4.1 数据集介绍

ETH^[47]和 UCY^[48]公开数据集包含在各种类型的社交交互场景下行人的全局轨迹坐标,这些数据集包括行人交互、非线性轨迹、避免碰撞、站立及群体行人的轨迹坐标等,同时包含从固定的俯视图记录的5个独特的室外环境信息。每个室外环境中单个场景的人群密度不同,所有视频的每秒帧数为25,以2.5 fps的速度对行人轨迹进行采样标记。其中ETH由ETH和Hotel两个数据集组成,UCY由Zara1、Zara2和Univ这3个数据集组成。数据集包括室外监控摄像头拍摄的5个视频,共包含2206条行人轨迹,可以反映轨迹在直线运动和曲线运动样条之间变化的不同特征。在大学入口处,视频捕捉人们的动作场景,在ETH数据集中,这些动作场景包含了更多笔直的轨迹,几乎没有社交互动,而UCY数据集展示了更多与人、空间互动有关的场景。例如,UCY-Zara1数据集中包括在商店入口处弯曲的行人轨迹,而UCY-Univ具有更多的社交互动。此外,除非考虑社会和空间环境,否则这些社交互动的情况会增加单个路径的不可预测

性。目前绝大部分轨迹预测模型将ETH和UCY公开数据集作为训练集和测试集,并在此数据集上测试模型性能。

DUT数据集包含通过航拍采集的大连理工大学主校区具有典型混杂交通特征的路段,适用于人车交互运动的轨迹预测模型,如图4所示。图4(a)为无交通信号的交叉路口。由于没有交通信号灯,当行人与车辆交互时,可以体现人车交互的“社会性”。图4(b)为环岛,行人和车辆在较大的活动空间内自由移动,而不受制于路面,可以得到多个方向上的预测结果。本课题组采用一架型号为DJI-Mavic-Pro的无人机在行人和车辆很难察觉的高度拍摄,以达到不会因为无人机的出现影响行人运动的目的。视频分辨率为1920×1080,每秒帧数为25。数据集共有17个交叉路口场景片段和11个环岛场景片段,共包含1793条轨迹。



(a) 无交通信号的交叉路口



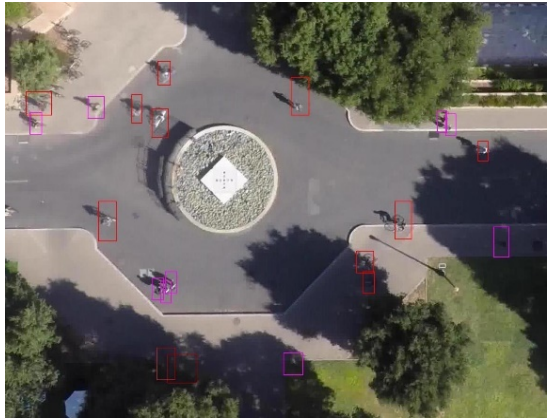
(b) 环岛

图4 DUT数据集

斯坦福无人机数据集 (Stanford drone dataset) 是由 Robicquet A 等人^[49]提出的一个大型数据集,旨在解决目标跟踪或轨迹预测类的任务。斯坦福无人机数据集是一个大型且先进的数据集,以图像和视频的形式收集了各种类型的智能体(不仅是行人,还包括自行车、滑板、汽车、公共汽车和高尔夫球车)(如图5所示)在现实户外环境(例如大学校园)中的行驶信息。



(a) 方庭场景



(b) 转盘场景

图5 斯坦福无人机数据集

4.2 评价指标

轨迹预测问题作为预测问题的子问题,可以被看作一个序列生成问题,该问题的输入序列对应观测到的所有行人的位置,输出序列表示行人在未来不同时刻所处的位置。即假设在任意一个时刻 T ,存在 n 个行人,在场景中行人 i 所处的实际位置被表示为 (x_i^T, y_i^T) ,从时间 $T=1$ 到 $T=T_{\text{obs}}$ (obs表示观测序列长度)所有行人所处的位置序列,然后预测时间 $T=T_{\text{obs}+1}$ 到 $T=T_{\text{pred+obs}}$ (pred表示预测序列长度)所有行人所处的预测位置 $(\hat{x}_i^T, \hat{y}_i^T)$ 。在评估轨迹预

测方法的性能时,通常使用平均位移误差(average displacement error, ADE)以及最终位移误差(final displacement error, FDE)两个评价指标对每个场景中智能体的轨迹进行预测评估。

(1) ADE

ADE 是指每个行人的真实轨迹与预测轨迹的位置序列在每一个时间步长内的平均欧氏距离差值,被用来评估预测过程的总体性能,如式(1)所示:

$$\text{ADE} = \frac{1}{n} \sum_{i=1}^n \frac{1}{\text{pred}} \sum_{T=T_{\text{obs}+1}}^{T_{\text{pred+obs}}} \sqrt{(x_i^T - \hat{x}_i^T)^2 + (y_i^T - \hat{y}_i^T)^2} \quad (1)$$

(2) FDE

FDE 是指每个行人的真实轨迹与预测轨迹的位置序列在终点位置的平均欧氏距离差值,如式(2)所示:

$$\text{FDE} = \frac{1}{n} \sum_{i=1}^n \sqrt{(x_i^{T_{\text{pred+obs}}} - \hat{x}_i^{T_{\text{pred+obs}}})^2 + (y_i^{T_{\text{pred+obs}}} - \hat{y}_i^{T_{\text{pred+obs}}})^2} \quad (2)$$

4.3 性能比较

目前领域内常被用于性能测试的数据集是 ETH 和 UCY 数据集。虽然在后期有许多针对轨迹预测方法的大型数据集相继被提出,但是由于这些数据集出现的时间较为靠后,一些早期发布的轨迹预测方法在这些数据集上的验证缺乏充足的实验数据,很难准确地测试这些方法的效果,故本文只比较了基于 ETH 和 UCY 数据集的模型性能。表1对比了本文提到的一些轨迹预测方法在公开数据集上的性能。一些方法要用到场景或行人骨架等语义信息,故实验是在模型只使用轨迹信息时得出的预测结果。

从表1中的数据可知,基于深度学习的轨迹预测方法(S-LSTM、O-LSTM、S-GAN-P、Sophie、Social-ways、GAT、Social-BiGAT、NEXT、TPNet-20、Social-STGCNN)效果更佳,其识别精度远远高于基于传统的浅层学习方法(Linear),且在 ETH、UCY 两个主流公开数据集上的不同场景中性能表现都十分优异。随着基于生成式网络与图神经网络技术的发展以及这两种网络与各种语义场景信息的结合,诸多轨迹预测方法在公开数据集上的测试性能愈发优异,其中以 Social-ways 为代表的基于生成对抗网络的学习方法性能显著,识别率相对较

表1 各轨迹预测方法在 ETH、UCY 数据集上的性能

方法	参考文献序号	评价指标 (ADE 值/FDE 值)					
		ETH	Hotel	Univ	Zara1	Zara2	平均值
Linear	[24]	1.33/2.94	0.39/0.72	0.82/1.59	0.62/1.21	0.77/1.48	0.79/1.59
S-LSTM	[24]	1.09/2.35	0.79/1.76	0.67/1.40	0.47/1.00	0.56/1.17	0.72/1.54
O-LSTM	[24]	1.05/2.21	0.81/1.68	0.71/1.45	0.47/1.02	0.64/1.25	0.74/1.52
S-RNN	[45]	2.72/4.60	0.85/1.35	1.05/2.20	1.60/3.50	1.45/3.00	1.53/2.93
Social-attention	[50]	3.60/4.70	0.79/1.44	1.30/2.66	0.95/2.05	1.00/2.14	1.53/3.52
S-GAN	[29]	0.81/1.52	0.72/1.61	0.60/1.26	0.34/0.69	0.42/0.84	0.58/1.18
S-GAN-P	[29]	0.87/1.62	0.67/1.37	0.76/1.52	0.35/0.68	0.42/0.84	0.61/1.21
Sophie	[32]	0.70/1.43	0.76/1.67	0.54/1.24	0.30/0.63	0.38/0.78	0.54/1.15
Social-ways	[30]	0.39/0.64	0.39/0.66	0.55/1.31	0.44/0.64	0.51/0.92	0.46/0.82
GAT	[33]	0.68/1.29	0.68/1.40	0.57/1.29	0.29/0.60	0.37/0.75	0.52/1.07
STGAT	[51]	0.70/1.35	0.37/0.67	0.59/1.23	0.35/0.69	0.31/0.64	0.47/0.92
Social-BiGAT	[33]	0.69/1.29	0.49/1.01	0.55/1.32	0.30/0.62	0.36/0.75	0.48/1.00
NEXT	[36]	0.73/1.65	0.30/0.59	0.60/1.27	0.38/0.81	0.31/0.68	0.46/1.00
TPNet-20	[37]	0.84/1.73	0.24/0.46	0.42/0.94	0.33/0.75	0.26/0.60	0.42/0.90
Social-STGCNN	[46]	0.64/1.11	0.79/0.85	0.44/0.79	0.34/0.53	0.30/0.48	0.44/0.75

高, Social-ways 方法的 ADE 值与 FDE 值达到了 0.46 和 0.82, 在性能上较之前提出的方法有了大幅度的提高。而且可以发现基于图网络的轨迹预测方法 (GAT、STGAT、Social-BiGAT、TPNet-20、Social-STGCNN) 进展迅速, 在 S-RNN 提出之时, S-RNN 的表现性能仅为基于 LSTM 的方法的 50%, 但随着图模型的改进以及 GCN 和 GAT 模型的应用, 基于图网络的轨迹预测方法的预测精度也获得了提升, 尤其是 2020 年提出的模型 (如 TPNet-20、Social-STGCNN), 其预测效果已经超过了基于生成式网络的轨迹预测方法, 且预测准确率较 S-LSTM 提高了 50% 以上。总而言之, 通过合理地设计网络结构和损失函数, 基于深度学习的轨迹预测方法可以相对精确地预测出行人接下来的轨迹, 可以较好地建立轨迹预测任务的基础框架。

5 结束语

本文对近年来轨迹预测领域的工作进行了研究, 以模型的结构设计与优化为出发点, 对目前

轨迹预测方法进行了分类, 并对不同算法的优缺点加以总结。结合轨迹预测的发展趋势可以发现, 基于浅层学习的轨迹预测方法已成为历史, 未来随着具有相当规模的大型数据集相继被提出, 无论从精度还是效率上, 基于深度学习的轨迹预测方法都比浅层学习方法更有效。目前基于深度学习的轨迹预测方法已有向多任务学习以及模型融合发展的趋势, 且对于模型结构的使用有显著的模块化特点, 即先明确问题, 后从备选模块中选用合适的模块进行拼接。未来, 神经网络结构将递归构造为更为复杂的深度网络, 并且随着数据集规模的发展以及特征信息的丰富, 基于深度学习的轨迹预测方法的模型的预测精度会不断提高, 但是随之而来的是模型预测效率的降低。而在实际场景的应用过程中, 需要同时保证轨迹预测算法运行的高效性及识别的高精度性, 当前领域内轨迹预测技术在实际场景应用中还存在不足, 需要结合实际考虑诸多问题, 因此未来的研究工作还可从下面几个方面展开。

- 当前大部分行人轨迹预测数据集为俯瞰视

角, 基于该数据集的训练测试数据不足以支撑行人轨迹预测的实际应用, 由于行人轨迹预测应用场景大部分为道路交通场景, 创建一个各种信息完备(场景信息、轨迹信息以及人体骨架关键点信息等)的第一人称视角的大型数据集是十分有必要的。

- 在实际场景中对行人轨迹预测算法的效率及识别精度有较高的要求, 但当前存在两者不可兼顾的问题。如何设计出能够被应用在嵌入式或移动设备上高效且轻量化的网络结构, 并在此基础上保证高精度的识别效果, 也是未来领域内的一大研究热点。

- 目前的模型对计算得到的交互信息缺乏可解释性。行人之间的交互是复杂而抽象的, 很难在算法中被精确建模, 仍然依赖于数据驱动, 即使使用基于图的算法, 图算法的各种缺点也会随之而来。如何通过交互的可解释建模以及采用更为先进的图模型来提升模型预测的准确性和可解释性, 也是行人轨迹预测领域的研究热点之一。

- 目前行人轨迹预测框架强烈依赖于实际场景中的应用领域或者应用领域中的特定场景(例如车辆汇集以及智能车领域中的行人过马路场景), 因此达到任何形式的绝对要求几乎是一件不可能的事情。对于具体的场景用例、行业驱动领域而言, 诸如智能车领域内提出的需求拟定、解决方案以及行业标准都是较为成熟的(例如国际标准化组织对城市驾驶场景下智能车的紧急制动的预测范围和度量精度的要求进行了描述, 以最高车速、交通规则以及行人速度和角速度分布为条件, 定义了车辆舒适加速/减速率的规范等)。智能车紧急制动这一具体场景用例的解决方案已经达到了工业级消费品的标准, 对于其他具体实际场景用例而言, 更加标准化及具体化的行业标准的提出对于加快领域内预测技术的发展也至关重要。

轨迹预测技术在近5年的时间里已经取得了飞速的发展, 随着思路的扩展, 基于多任务学习和模块化思想的轨迹预测算法必将成为轨迹预测的一大趋势。随着技术的更新和进步、算法性能的提高, 相信轨迹预测技术走进实际生活的距离不会很远。

参考文献:

[1] RUDENKO A, PALMIERI L, HERMAN M, et al. Human motion

trajectory prediction: a survey[J]. *The International Journal of Robotics Research*, 2020(1): 027836492091744.

- [2] DAI M M, WANG J X, YIN G D, et al. Dynamic output-feedback robust control for vehicle path tracking considering different human drivers' characteristics[C]//*Proceedings of the 36th Chinese Control Conference*. Piscataway: IEEE Press, 2017: 9407-9412.
- [3] MOUSSAID M, PEROZO N, GARNIER S, et al. The walking behaviour of pedestrian social groups and its impact on crowd dynamics[J]. *PLoS ONE*, 2010, 5(4): e10047.
- [4] LEFEVRE S, VASQUEZ D, LAUGIER C. A survey on motion prediction and risk assessment for intelligent vehicles[J]. *Robomech Journal*, 2014, 1(1): 1-14.
- [5] KELLER C G, GAVRILA D M. Will the pedestrian cross? A study on pedestrian path prediction[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2014, 15(2): 494-506.
- [6] SCHNEIDER N, GAVRILA D M. Pedestrian path prediction with recursive Bayesian filters: a comparative study[C]//*Proceedings of the 35th German Conference on Pattern Recognition*. Berlin: Springer Press, 2013: 174-183.
- [7] PAVLOVIC V, REHG J M, MACCORMICK J. Learning switching linear models of human motion[C]//*Proceeding of the Conference and Workshop on Neural Information Processing Systems*. New York: Curran Associates Press, 2000: 981-987.
- [8] FOX E B, SUDDERTH E B, JORDAN M, et al. Bayesian nonparametric inference of switching dynamic linear models[J]. *IEEE Transactions on Signal Processing*, 2011, 59(4): 1569-1585.
- [9] KOOLIJ F P, SCHNEIDER N, FLOHR F, et al. Context-based pedestrian path prediction[C]//*Proceedings of the 13th European Conference on Computer Vision*. Berlin: Springer Press, 2014: 618-633.
- [10] HELBING D, MOLNAR P. Social force model for pedestrian dynamics[J]. *Physical Review E*, 1995, 51(5): 4282-4286.
- [11] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. *Computer Science*, 2014.
- [12] ALAHI A, RAMANATHAN V, LI F F. Socially-aware large-scale crowd forecasting[C]//*Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2014: 2211-2218.
- [13] YI S, LI H S, WANG X G. Understanding pedestrian behaviors from stationary crowd groups[C]//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2015: 3488-3496.
- [14] GOLI S A, FAR B H, FAPOJUWO A O. Vehicle trajectory prediction with Gaussian process regression in connected vehicle environment[C]//*Proceedings of 2018 IEEE Intelligent Vehicles Symposium*. Piscataway: IEEE Press, 2018: 550-555.
- [15] ELLIS D, SOMMERLADE E, REID I. Modelling pedestrian trajectory patterns with Gaussian processes[C]//*Proceedings of 2009 IEEE International Conference on Computer Vision*. Piscataway: IEEE Press, 2009: 1229-1234.
- [16] RASMUSSEN C E, WILLIAMS C K I. *Gaussian processes for machine learning*[M]. Cambridge: MIT Press, 2005.
- [17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. *Neural Computation*, 1997, 9(8): 1735-1780.
- [18] CHUNG J Y, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. *Eprint Arxiv*, 2014.

- [19] GRAVES A, JAITLEY N. Towards end-to-end speech recognition with recurrent neural networks[C]//Proceedings of the 31st International Conference on Machine Learning. New York: ACM Press, 2014: 1764-1772.
- [20] CHOROWSKI J, BAHDANAU D, CHO K, et al. End-to-end continuous speech recognition using attention-based Recurrent NN: first results[J]. arXiv preprint, 2014, arXiv: 1412.1602.
- [21] DONAHUE J, HENDRICKS L A, GUADARRAMA S, et al. Long-term recurrent convolutional networks for visual recognition and description[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 2625-2634.
- [22] WU H, CHEN Z Y, SUN W W, et al. Modeling trajectories with recurrent neural networks[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2017: 3083-3090.
- [23] KARATZOLOU A, JABLONSKI A, BEIGL M. A Seq2Seq learning approach for modeling semantic trajectories and predicting the next location[C]//Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2018: 528-531.
- [24] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: human trajectory prediction in crowded spaces[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 961-971.
- [25] KITANI K M, ZIEBART B D, BAGNELL J A, et al. Activity forecasting[C]//Proceedings of the 12th European Conference on Computer Vision. Berlin: Springer Press, 2012: 201-214.
- [26] XUE H, HUYNH D Q, REYNOLDS M. SS-LSTM: a hierarchical LSTM model for pedestrian trajectory prediction[C]//Proceedings of 2018 IEEE Workshop on Applications of Computer Vision. Piscataway: IEEE Press, 2018: 1186-1194.
- [27] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the Conference and Workshop on Neural Information Processing Systems. New York: Curran Associates Press, 2014: 2672-2680.
- [28] KIPF T N, WELING M. Variational graph auto-encoders[J]. arXiv preprint, 2016, arXiv:1611.07308.
- [29] GUPTA A, JOHNSON J, LI F F, et al. Social GAN: socially acceptable trajectories with generative adversarial networks[C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2255-2264.
- [30] AMIRIAN J, HAYET J B, PETTRE J. Social ways: learning multi-modal distributions of pedestrian trajectories with GANs[J]. arXiv preprint, 2019, arXiv:1904.09507.
- [31] VARSHNEYA D, SRINIVASARAGHAVAN G. Human trajectory prediction using spatially aware deep attention models[J]. arXiv preprint, 2017, arXiv:1705.09436.
- [32] SADEGHIAN A, KOSARAJU V, SADEGHIAN A, et al. Sophie: an attentive GAN for predicting paths compliant to social and physical constraints[C]//Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 1349-1358.
- [33] KOSARAJU V, SADEGHIAN A, MARTIN-MARTIN R, et al. Social-BiGAT: multimodal trajectory forecasting using bicycle-GAN and graph attention networks[C]//Proceedings of the Conference and Workshop on Neural Information Processing Systems. New York: Curran Associates Press, 2019: 137-146.
- [34] CHENG H, YANG W L M Y, SESTER M, et al. Context conditional variational autoencoder for predicting multi-path trajectories in mixed traffic[J]. arXiv preprint, 2020, arXiv:2002.05966.
- [35] YANG B, YAN G, WANG P, et al. TPPO: a novel trajectory predictor with pseudo oracle[J]. arXiv preprint, 2020.
- [36] LIANG J W, JIANG L, NIEBLES J C, et al. Peeking into the future: predicting future person activities and locations in videos[C]//Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 5725-5734.
- [37] SUN J H, JIANG Q H, LU C W. Recursive social behavior graph for trajectory prediction[J]. arXiv preprint, 2020, arXiv:2004.10402.
- [38] LI G H, MULLER M, QIAN G C, et al. DeepGCNs: can GCNs go as deep as CNNs?[C]//Proceedings of 2019 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 9267-9276.
- [39] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint, 2016, arXiv:1609.02907.
- [40] SCHOLKOPF B, TSUDA K, VERT J, et al. Kernel methods in computational biology[M]. Cambridge: MIT Press, 2004.
- [41] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]//Proceedings of the European Semantic Web Conference. Berlin: Springer Press, 2018: 593-607.
- [42] BERG R V D, KIPF T N, WELING M. Graph convolutional matrix completion[J]. arXiv preprint, 2017, arXiv:1706.02263.
- [43] ZHANG L D, SHE Q, GUO P. Stochastic trajectory prediction with social graph network[J]. arXiv preprint, 2019, arXiv:1907.10233.
- [44] YAN S J, XIONG Y J, LIN D H, et al. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]//Proceedings of the National Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2018: 7444-7452.
- [45] HADDAD S, WU M Q, WEI H, et al. Situation-aware pedestrian trajectory prediction with spatio-temporal attention model[J]. arXiv preprint, 2019, arXiv:1902.05437.
- [46] MOHAMED A, QIAN K, ELHOSEINY M, et al. Social-STGCNN: a social spatio-temporal graph convolutional neural network for human trajectory prediction[J]. arXiv preprint, 2020, arXiv:2002.11927.
- [47] PELLEGRINI S, ESS A, GOOL L V. Improving data association by joint modeling of pedestrian trajectories and groupings[C]//Proceedings of the 11th European Conference on Computer Vision. Berlin: Springer Press, 2010: 452-465.
- [48] LERNER A, CHRYSANTHOU Y, LISCHINSKI D. Crowds by example[J]. Computer Graphics Forum, 2007, 26(3): 655-664.
- [49] ROBICQUET A, SADEGHIAN A, ALAHI A, et al. Learning social etiquette: human trajectory prediction in crowded scenes[C]//Proceedings of the 14th European Conference on Computer Vision. Berlin: Springer Press, 2016: 549-565.
- [50] VEMULA A, MUELLING K, OH J. Social attention: modeling attention in human crowds[C]//Proceedings of 2018 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2018: 4601-4607.
- [51] HUANG Y F, BI H K, LI Z X, et al. STGAT: Modeling spatial-temporal interactions for human trajectory prediction[C]//Proceedings of 2019 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 6272-6281.

[作者简介]



李琳辉（1981- ），男，博士，大连理工大学汽车工程学院副教授，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。



任威威（1996- ），男，大连理工大学汽车工程学院硕士生，主要研究方向为智能车辆规划决策、轨迹预测等。



周彬（1997- ），男，大连理工大学汽车工程学院硕士生，主要研究方向为智能车辆规划决策、轨迹预测等。



连静（1980- ），女，博士，大连理工大学汽车工程学院副教授，主要研究方向为新能源汽车智能化、轨迹预测等。