

doi: 10.19562/j.chinasae.qcgc.2024.10.013

基于图卷积交互网络的车辆轨迹预测方法*

王梦茜¹, 蔡英凤¹, 王海², 饶中钰¹, 陈龙¹, 李祎承¹

(1. 江苏大学汽车工程研究院, 镇江 212013; 2. 江苏大学汽车与交通工程学院, 镇江 212013)

[摘要] 精确预测邻近车辆的未来轨迹对自动驾驶汽车的决策和运动规划至关重要, 现有研究倾向于利用递归神经网络(RNN)对车辆的时间交互进行建模, 但其对车辆交互建模的可解释性差, 忽略了实际的车道结构, 在捕捉车辆与其环境的交互方面存在不足。为解决这一问题, 本文提出了一种基于图卷积交互网络的考虑车道拓扑约束的车辆轨迹预测模型。其中车辆交互关系提取模块在构建车辆的空间关系时增加了边缘权重, 以考虑车辆的邻近交互, 使交互更具可解释性; 行驶场景表征模块旨在通过从高精地图中提取车道拓扑来提高车辆轨迹预测的准确性; 轨迹预测模块将上述两个模块的输出进行集成, 并输出预测的未来轨迹。这种集成允许对道路结构和车辆行驶轨迹之间的相互作用进行更精确的建模。实验结果表明, 与主流方法相比, 该模型在Argoverse数据集上取得了良好的性能, 提高了复杂道路结构下车辆轨迹规划的准确性和合理性。

关键词: 自动驾驶汽车; 轨迹预测; 图卷积网络; 交互建模

Vehicle Trajectory Prediction Method Based on Graph Convolutional Interaction Network

Wang Mengxi¹, Cai Yingfeng¹, Wang Hai², Rao Zhongyu¹, Chen Long¹ & Li Yicheng¹

1. Automotive Engineering Research Institute, Jiangsu University, Zhenjiang 212013;

2. School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang 212013

[Abstract] Accurate prediction of the future trajectory of surrounding vehicles is crucial to the decision-making and motion planning of autonomous vehicle. Existing research tends to use Recurrent Neural Networks (RNN) to model the time interaction of vehicles, but its interpretability of vehicle interaction modeling is poor, ignoring the actual lane structure, and there are deficiencies in capturing the interaction between vehicles and the environment. To address this problem, in this paper, a vehicle trajectory prediction model based on graph convolutional interactive networks that considers lane topology constraints is proposed. The vehicle interaction relationship extraction module adds edge weights when constructing the spatial relationship of vehicles to consider their neighboring interaction, making the interaction more interpretable. The driving scene representation module aims to improve the accuracy of vehicle trajectory prediction by extracting lane topology from high-precision maps. The trajectory prediction module integrates the output of the above two modules and outputs the predicted future trajectory. This integration allows for more precise modeling of the interaction between road structures and vehicle driving trajectories. The experimental results show that compared with mainstream methods, this model has achieved good performance on the Argoverse dataset, improving the accuracy and rationality of vehicle trajectory planning under complex road structures.

Keywords: autonomous vehicle; trajectory prediction; graph convolution network; interaction behavior

* 国家自然科学基金(52225212, U20A20333)资助。

原稿收到日期为2024年03月16日, 修改稿收到日期为2024年04月23日。

通信作者: 蔡英凤, 教授, 博士生导师, E-mail: caicaixiao0304@126.com。

前言

近年来,随着人们对提升交通安全、缓解交通拥堵、降低油耗以及减少废气排放等问题的日益关注,自动驾驶汽车作为一种有效的解决方案,逐渐赢得了公众的广泛认可^[1-5]。与传统的人工驾驶相比,自动驾驶汽车须能够准确地预测周围车辆的意图和行驶路径,以便在复杂多变的交通环境中实现安全和高效的行驶。然而,在实际的交通环境中,车辆的运动轨迹不单是由其与其他交通参与者之间的互动所决定,还受到车道环境的影响,包括道路的几何形状、交通标志以及交通规则等因素。因此,要想提高轨迹预测的准确性,就必须在预测过程中综合考虑车道环境的约束以及车辆在驾驶场景中与周围交通参与者的交互作用。

现有的车辆轨迹预测方法主要分为两类:一是基于物理模型的非交互式预测方法,二是融合深度学习技术的动态交互式预测方法。其中基于物理模型的预测方法主要通过分析车辆的运动学和动力学特性来预测其未来轨迹。这类方法的代表性模型包括蒙特卡洛模拟^[6]、贝叶斯网络^[7]和隐马尔可夫模型(HMM)^[8]。然而,这些方法主要依赖于对车辆过去运动状态的分析,却忽视了各个交通主体明确的意图与机动性,这导致它无法适应复杂的道路环境。为解决这些问题,学者们进一步提出了基于机动的预测方法,如支持向量机(SVM)^[9]、隐马尔可夫模型^[10]和高斯混合模型^[11]等。这些方法尝试对交通参与者的行为进行独立建模,虽然大幅提升了预测的灵活性,但由于这种方式需要对每个参与者单独计算,计算成本较高,且未能有效捕捉交通参与者之间的相互作用。

近年来,随着深度学习技术的迅猛发展,基于深度学习的交互式预测模型在车辆轨迹预测领域显示出了卓越的性能。这类模型有效应对了车辆轨迹预测面临的多种挑战,包括车辆行为之间的依赖性、交通规则与驾驶环境的限制,以及车辆行为的多样性特征。其中,递归神经网络(RNN)擅长提取时间特征,适合处理时间序列信息;卷积神经网络(CNN)则擅长提取包括交通参与者之间相互作用相关因素在内的空间特征。因此许多研究致力于结合RNN与CNN来实现轨迹预测。Chandra等^[12]提出了一种基于CNN-LSTM混合网络的TraPhic模型来预测交通参与者的轨迹,该模型将自车的状态和周围物体作

为输入,通过CNN-LSTM网络提取特征,并将这些特征与LSTM解码器连接,得到自车的预测轨迹。随着图论的发展,交通场景作为不规则图的特性引起了学者们的关注。因此,一些研究开始探索基于图结构的模型,如图神经网络(GNN),它们在处理车辆轨迹预测中的相互作用问题方面展现出独特的优势。Diehl等^[13]结合图卷积网络(GCN)和图注意力网络(GAT)来进行轨迹预测,验证了基于图结构的模型在捕捉交互相关因素方面的有效性。尽管基于图的方法在建模交通参与者之间的配对交互方面取得了良好的性能,但这些方法只关注动态参与者之间的交互,在构建交互时没有考虑车道结构特征和不同个体权重的影响。考虑到现实世界交通流的特点,这些方法不利于自动驾驶汽车做出安全可靠的运动决策。

随着高精地图技术的快速发展,近期有研究开始关注如何将静态车道结构的约束融入车辆轨迹预测中^[14],从而提升预测的准确性和效率。在这方面,光栅化渲染高精地图成为了一种常见的方法,它通过将地图元素(如道路、人行横道等)渲染成像素层,并对车道线进行编码。Chai等^[15]将高精地图进行光栅化,并对道路类型、交通信号灯和限速等交通信息进行了编码。Hong等^[16]则对自上而下的空间网格中的静态实体、动态实体和语义地图信息进行编码。尽管光栅化方法因其直观性而受到广泛应用,但该方法在渲染过程中易于丢失地图细节,且效率较低。为解决光栅化方法的局限性,最近的研究趋势是向构建环境的矢量化拓扑结构转变。这种方法巧妙地运用矢量化的方式对周围的静态环境进行表示,以捕获场景的上下文信息。Gao等^[17]提出的VectorNet是一个分级图神经网络,首先将车辆和行人的轨迹及地图信息进行编码向量表示,并使用折线子图从向量中抽取实例的特征,生成的特征通过完全连接的全局网络进行聚合,以实现不同实例间的信息交换。Liang等^[18]提出的LaneGCN是将高精地图矢量化,提取结构化的地图信息,并使用图卷积网络计算车道特征,这些特征与融合网络中的车辆特征以及其他车道特征相结合,从而更准确地捕捉车辆在复杂交通环境中的动态行为。这种基于向量图的方法不仅能够有效捕捉车道的拓扑结构和边界特性,且相比基于光栅化的方法,它能够以更少的参数实现更优的性能。

可以看出,大多数现有的轨迹预测方法都有两个局限性:首先,对于车辆间的交互关系提取不充

分,没有考虑影响权重的时变性质来预测动态物体的相互作用,即真实交通环境中车辆是运动的,交互图的影响权重应该不断更新,而不仅仅是在初始状态的固定权重下;其次,对于车道信息利用不到位,没有有效结合驾驶环境中车道结构约束和动态交互对象的影响,例如,感兴趣的车道可能在车道方向上延伸很长一段距离,而不感兴趣的车道则只须考虑附近一小段的影响。

为解决这些问题,本文提出了一种新的轨迹预测模型,该模型同时考虑了车道结构和周边车辆的交互作用。首先,改进了基于图的交互感知预测模型,将图的构建扩展到时间和空间结构相结合的层面,并提取时间和空间层面的图形关系变化,以适应实际场景中时刻变化的交互效果。同时,为更好地捕捉周围不同车辆对目标车辆的影响,本文改进了图的空间边缘,根据人工阈值设置边缘,以根据车辆的空间距离分配边缘权重。通过对不同空间位置的车辆边缘分配不同的注意力权重,可以突出影响交互的关键信息,而抑制不重要的信息,显著提高了模型操作的效率。此外,将静态车道拓扑约束^[18]编码到向量图网络中,有效地捕捉了车道图的复杂拓扑和长距离依赖关系。

本研究的贡献总结如下:

(1)将车辆交互的影响和车道拓扑结构相结合,通过独立提取车道拓扑网络和车辆交互关系,使得模型在考虑周边车辆的交互作用时还能兼顾车道性质,保证输出可行驶区域内的轨迹。

(2)所提模型改进了空间连边的定义,赋予距离

权重,改善了人工定义阈值所构建的连边没有考虑近距离车辆不同影响力的缺陷,能够更好地反映车辆空间距离和交互影响力的关系。

(3)所提模型在 Argoverse1 数据集上进行了实验验证,结果表明该模型取得了良好的效果。

1 方法

1.1 问题定义

车辆轨迹预测问题通常可被定义为:已知目标车辆的历史轨迹,对其在未来场景中的行驶轨迹进行预测。本文将目标车辆的历史轨迹表示为 $X = [p^{(1)}, p^{(2)}, \dots, p^{(t_h)}]$, 预测轨迹表示为 $Y = [p^{(t_h+1)}, p^{(t_h+2)}, \dots, p^{(t_h+t_f)}] = f(X)$, 其中 $p^{(t)} = [x^{(t)}, y^{(t)}]$ 为目标车辆在 t 时刻的位置坐标, t_h 为历史轨迹时间段, t_f 为未来预测时间段, $f(\cdot)$ 为所提出的预测模型。

本文将进一步扩展,在预测时综合考虑了周边车辆和静态交通道路信息对目标车辆轨迹的影响。这意味着影响目标车辆未来轨迹的因素不仅限于其历史轨迹信息,还包括周边车辆的历史位置信息 $S = [s^{(1)}, s^{(2)}, \dots, s^{(t_h)}]$ 以及目标车辆所处交通场景的道路信息 L 。则经过模型训练之后,输出 $t_h + 1$ 到 $t_h + t_f$ 时刻的目标车辆未来坐标序列 Y 为

$$Y = [p^{(t_h+1)}, p^{(t_h+2)}, \dots, p^{(t_h+t_f)}] = f(X, S, L) \quad (1)$$

1.2 基于图卷积交互网络的车辆轨迹预测模型

本节详细介绍所提的基于图卷积交互网络的车辆轨迹预测模型的具体架构。图1显示了模型的整

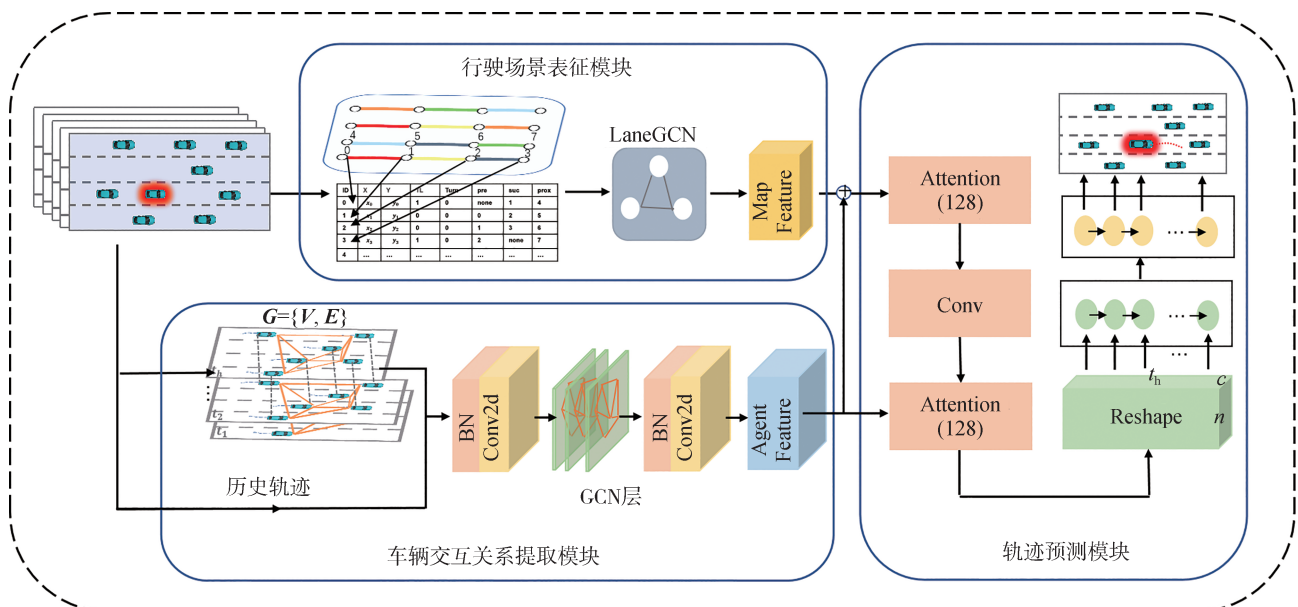


图1 基于图卷积交互网络的车辆轨迹预测模型总体框架图

体框架图。该模型由3个模块组成,分别是车辆交互关系提取模块、驾驶场景表征模块和轨迹预测模块。

1.2.1 车辆交互关系提取模块

车辆交互关系提取模块包括两部分:预测的目标车辆和周围车辆的历史轨迹交互图的构建以及用于提取轨迹信息的车辆GCN网络。

(1)交互图构建

考虑到在自动驾驶行驶场景中,车辆行驶要受到其周围对象的运动影响,这与人们在社交网络上的行为非常相似(一个人通常会受到他/她的朋友的影响)^[19]。因此本文将交互行为定义为 $G = (V, E)$,其中节点 V 表示场景中每一个车辆对象, E 表示节点间的交互影响连边。在构建交互图前,首先需要原始数据进行特征表示,即将原始数据构造能够高效输入图形的形式。具体做法是,针对观察到的最后一个帧(t_{obs})中的 N 个车辆对象,将原始数据构建成一个三维矩阵 (N, T_h, C) ,其中 T_h 是历史轨迹中的帧数, C 是输入特征数,本文 $C = 2$ 表示物体的 x 和 y 坐标。

基于上述原始数据的特征表示,构建了交互图 $G = (V, E)$ 。考虑到节点集中每个节点在不同的时间步可能具有不同的状态,定义节点集中的每个节点为 $V_i = \{v_i^t | t = 1, 2, \dots, t_h\}$,是时间 t 处节点 i 的特征向量,整个节点集 V 即输入的原始数据矩阵 $N \times T_h \times C$ 。

对于交互图的连边 E ,考虑到时空上轨迹点之间彼此影响,因此将交互图连边表示为时间和空间交互连边两种类型。首先,本文通过一个时间连边 e_i^t 将每个车辆 v_i 与另一个时间步中的自身连接起来。因此,时间连边集 $E_t = \{e_i^t | (v_i^t, v_i^{t+1})\}$ 表示每个节点在 t_h 上的历史轨迹。其次,对于空间连边 E_s ,本文从最后观察到的帧 t_{obs} 获得一个邻接矩阵来表示。在数据处理时,将 t_{obs} 所有轨迹考虑在内,假设影响目标车辆轨迹的周边车辆存在一个阈值范围(本文取经验值10 m)。以要预测车辆为中心,划定阈值大小为影响半径 r ,计算周边车辆与目标车辆之间的距离值 l ,将 l 与经验半径距离 r 相比, $l \leq r$ 时认为车辆之间有交互影响,并设置邻接矩阵值为1,反之为0,如图2右上,目标车与阈值范围内3辆车有连边,这些周边车辆形成一个交互节点集 D 。因此,空间连边集被定义为 $E_s = \{e_{ij} | (v_i, v_j \in D)\}$,用连边的二元邻接矩阵 A 表示成大小为 (N, N) 的对称矩阵:

$$A = \begin{cases} 1, & (v_i^t, v_j^t) \in D \\ 0, & \text{其他} \end{cases} \quad (2)$$

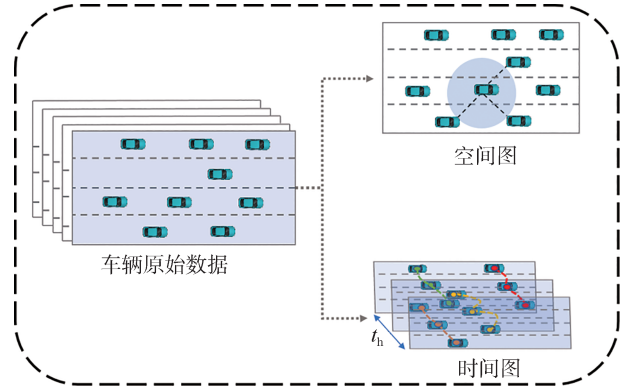


图2 车辆时空交互图的构建

考虑到交通参与者更容易受到近邻的影响,本文做一个新的细化连接边缘代理对象的阈值范围,即赋予了边缘权重作为初始固定注意分数,该权重值由每个边的核函数计算,定义为边的逆:

$$A = \begin{cases} \frac{1}{\|v_i^t - v_j^t\|_2}, & (v_i^t, v_j^t) \in D \\ 0, & \text{其他} \end{cases} \quad (3)$$

(2)车辆GCN提取轨迹信息

数据处理完成后,将历史轨迹和交互图 G 传入车辆交互关系表征模块。该模块首先将其通过一层二维卷积层映射到高维卷积层,来聚合节点之间的特征信息;然后通过两层图卷积层进行时空交互。图卷积层由CNN和GCN组成,分别提取时间信息和空间信息。空间交互的时间变化由两部分之和来形成空间交互的卷积内核,即当前观测帧的交互图 G_{obs} 与和 G 大小一致但参与训练的可训练图 G_{train} 之和,其中 G_{obs} 为所构建的空间邻接矩阵 A 。为保证空间信息交互后特征图维度前后一致,对 A 进行图卷积的归一化操作:

$$G_{\text{obs}} = \tilde{D}^{-\frac{1}{2}}(A + I)\tilde{D}^{-\frac{1}{2}} \quad (4)$$

式中度矩阵 \tilde{D} 中的值表示每个车辆节点的相邻车辆数量。节点的相互作用是一种非定向图结构,每个节点的变化将导致其相邻节点的函数更新,最终将驱动整个图结构中所有节点的函数变化。这种变化被认为是一种图形信号的传输和扩散。每个节点将接收由周围节点通过边传递到它的信息,同时将自己的特征信息传递给周围的节点。本文考虑到观察帧设置的固定图阈值的局限性,除在空间图设置时

增添距离的核函数的逆作为边缘权重外,也增添了可训练图 $\mathbf{G}_{\text{train}}$,因此空间交互信息的更新计算过程如下:

$$f_{\text{graph}} = \sum_{j=0}^1 (G_{\text{obs}}^j + G_{\text{train}}^j) f_{\text{conv}} \quad (5)$$

这样,通过由 \mathbf{G} 和 $\mathbf{G}_{\text{train}}$ 做卷积核的GCN提取空间交互信息后,再对时间层面上通过卷积核大小固定为 (1×3) 的时间卷积层,使得 $n \times t_h \times c$ 维度的数据沿着时间维度(第二维度)处理数据。空间层和时间层交替处理后输出维度保持 $n \times t_h \times c$ 不变的数据,该数据后续用于与行驶场景表征模块输出数据融合。

1.2.2 行驶场景特征描述模块

为构建一个能够充分描述各种车道变换(如环形车道)的车道图,本文通过设置一个中心线转向阈值来确定车道是否相邻。具体来说,如果车道段之间的转向角超过阈值,则视双车道段为相邻车道,否则视为同一车道,这可以避免预测的轨迹与车道图不匹配的情况。对于场景中的每个参与者 i ,首先检索该参与者在预测视野中可能到达的所有相关车道,以及来自历史视野的所有相关车道,之后将车道转换为有向图 $\mathbf{G}(\mathbf{V}, \mathbf{E})$ 的矢量化数据形式。

节点:车道中心线网络同时捕获车道方向和每个驾驶员可以遵循的合法路线,并试图用这两种道路属性来描述静态交通环境的约束,因此使用车道中心线作为车道图的节点。由于车道线的长度都不相同,且为将其矢量化表示,本文按照长度对其均匀划分,将较长的车道线离散为一组 N 个固定长度的小向量。其中,每个小向量的中点对应车道图中的一个节点,每个节点包含着丰富的几何和语义信息,如图3所示。即车道图中的节点 \mathbf{V} 可表示为一系列特征向量:

$$\mathbf{v}_{1:N} = [f_1^v, \dots, f_N^v] \quad (6)$$

$$\mathbf{f}_n^v = [x_n^v, y_n^v, \theta_n^v, K_n^v] \quad (7)$$

式中: x_n^v, y_n^v 和 θ_n^v 为几何特征,表示第 n 个向量的位置和曲率; K 为语义特征,指示该车道段是否为转弯车道、当前是否受交通信号灯控制等。因此,该节点函数捕获了车道中心线沿线的几何图形和交通控制元素。

边缘:通过定义车道图中的边缘 \mathbf{E} ,使通过该图的任何连通路程都对应于车辆可以在场景中行驶的合法路线。本文考虑以下几种类型的边来表示车道段间的拓扑关系。前导边(E_{pre})表示沿目标车辆行

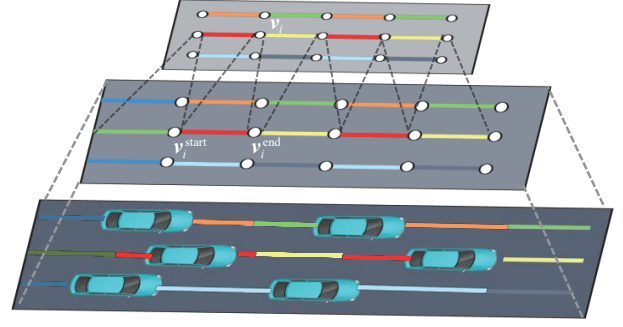


图3 车道矢量拓扑关系提取示意图

驶方向的未来将经过的车道线,后继边(E_{suc})表示逆着目标车辆行驶方向的已经过的车道线,为防止变道情况的出现,还将引入左邻边(E_{left})和右邻边(E_{right})。这4种类型的边全面包含了车道线之间的拓扑关系,则车道图的边缘 \mathbf{E} 可使用邻接矩阵进行概括:

$$\mathbf{E} = \{ \mathbf{A}_i \}_{i \in (\text{suc}, \text{pre}, \text{left}, \text{right})} \quad (8)$$

式中 $\mathbf{A}_i \in R^{N \times N}$, $\mathbf{A}_{i,jk} = 1$ 表示节点 j 和节点 k 之间存在着第 i 种拓扑关系。

行驶场景表征模块依据上述构图方法对原始高精地图提取特征,首先对车道段的局部特征进行编码,之后通过全连接网络对单条车道信息进行有效聚合:

$$x_i = FC(v_i^{\text{end}} - v_i^{\text{start}}) + FC(v_i) \quad (9)$$

式中: \mathbf{v}_i 为第 i 个节点的特征向量, v_i^{start} 和 v_i^{end} 分别为该向量的起始点,即 $v_i^{\text{end}} - v_i^{\text{start}}$ 为该条车道段的向量表示。 x_i 为第 i 个节点的输入特征,所有 x_i 的集合即为特征矩阵 \mathbf{X} 。之后参考 LaneGCN 中的地图编码方法,使用 LonvConv 算子来获取车道间的关系,并将所得的车道图 \mathbf{G} 输入到图卷积网络中,即

$$\mathbf{Y} = \text{ReLU}(\text{BN}(\mathbf{P}_0 + \mathbf{P}_1)) \quad (10)$$

$$\mathbf{P}_0 = \mathbf{X}\mathbf{W}_0 + \sum_{i \in \{E_{\text{adj}}\}} \mathbf{A}_i \mathbf{X}\mathbf{W}_i \quad (11)$$

$$\mathbf{P}_1 = \sum_{c=1}^C (\mathbf{A}_{\text{pre}}^{k_c} \mathbf{X}\mathbf{W}_{\text{pre},k_c} + \mathbf{A}_{\text{suc}}^{k_c} \mathbf{X}\mathbf{W}_{\text{suc},k_c}) \quad (12)$$

式中: \mathbf{A}_i 为第 i 条车道的邻接矩阵; \mathbf{W}_i 为第 i 条车道的权重矩阵; \mathbf{X} 为特征矩阵; k_c 表示卷积层中第 c 个卷积核的膨胀大小。考虑到预测过程中存在车辆在固定历史时间段内由于速度过快,容易在直线车道段产生长距离历史轨迹的情况,在直线路段可采取膨胀卷积的方式,增大邻接矩阵来扩大视野域。然后在卷积网络之后通过一层 $n \times 128$ 维的全连接层输出车道特征。

1.2.3 轨迹预测模块

轨迹预测模块由融合网络和预测头组成。前者聚合上述两个模块的输出(即车辆信息和车道信息),后者根据融合信息对轨迹进行解码。

融合网络主要集成了车辆交互提取模块和驾驶场景表示模块的输出特征。其更新过程如下:(a)将车辆信息传送到车道节点,以获得车道占用率或其他使用条件;(b)更新车道节点之间的信息,实现车道线段之间的实时互联;(c)整合更新后的地图特征和实时交通信息,然后将反馈信息传递给车辆。具体来说,车道节点之间的信息更新仍然采用 LaneGCN,通过由车道信息构造的邻接矩阵进行卷积核实现。对于车辆和车道之间信息的相互传输,通过注意机制实现步骤(a)和步骤(c)的更新过程。由于这两个步骤具有相同的结构,此处以步骤(a)为例,详细描述它的结构。

为将车辆信息传递给车道节点,以车道节点 x_i 为中心节点,聚合周边的车辆信息 x_j ,首先得到它们之间的注意力权重系数 e_{ij} :

$$e_{ij} = \phi(\text{concat}(x_i, \Delta_{i,j}, x_j)W_1)W_2 \quad (13)$$

式中: $\Delta_{i,j}$ 表示 x_i 和 x_j 之间的位置距离权重,使用 MLP 聚合这两个节点之间的权重信息,即 $\Delta_{i,j} = \text{MLP}(x_i - x_j)$; ϕ 表示归一化层和激活层的组合; W_1 和 W_2 表示节点间的更新权重。获得的注意力权重用于更新图形每一层每个节点的特征,如图4所示。最终更新后的车道节点信息为

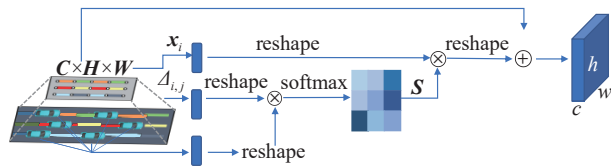


图4 融合网络的更新过程

$$y_i = x_i W_0 + \sum_j e_{ij} \quad (14)$$

其中上下文节点被定义为车道节点与车辆节点的 l_2 距离小于阈值的通道节点,此处阈值可取经验值 6 m。具体网络设置为:行驶场景表征模块提取出的 $n \times 128$ 的二维车道位置信息和 $n \times 4$ 维的车道性质特征(是否转向、是否有交通控制以及是否是交叉口)组成新的地图特征信息和车辆的二维特征数据作为该单元输入数据,经过两层图注意力机制的堆栈和一层全连接后,输出带有车辆信息的车道特征,

且保持维度为 $n \times 128$ 。第(c)部分与第(a)部分网络结构一致,最后提取出含有车道信息以及车道交互信息的车辆特征信息,维度输出同样保持为 $n \times 128$ 。

预测头由 GRU 解码器和全连接网络组成。轨迹预测模块将融合后的车辆特征信息作为输入,通过编码器解码器处理上述模块中提取的上下文向量信息,并由多模态预测头输出最终的运动轨迹预测。具体过程是:首先将融合后的车辆特征扩维为 $n \times t_h \times c$,输入 Seq2Seq 结构网络,与 LSTM 相比,GRU 内部结构更为简洁,参数较少,但功能却能与 LSTM 相当。考虑到时间成本以及硬件的计算能力,本文选择效益更明显的 GRU 进行实现。将表示车辆在每个时间维度上的特征向量传入编码器 GRU 的对应输入单元。接着,编码器 GRU 的隐藏特征以及车辆在先前时间步长的坐标信息被共同送入解码器 GRU,以预测当前时间步长的位置坐标。在解码过程中,首先输入的是最后观测帧中车辆的坐标信息,之后当前步骤的输出将被送入下一个 GRU 单元作为输入。这样的解码过程将重复进行多次,直到模型输出未来所有预期时间步长的位置坐标。

1.3 实施细节

本文主要采用 Python 编程语言,并利用 PyTorch 库来构建和训练模型。为加速模型的学习速度,选用了 Adam 优化器,并将其学习率设定为 0.001。

训练过程中将模型训练为回归任务。由于多模态轨迹的输出,因此将损失函数设置为车道分类误差和轨迹回归误差的加和,计算如下:

$$L = L_{\text{cls}} + \alpha L_{\text{reg}} \quad (15)$$

式中: L_{cls} 是预测节点特征多模态分类的交叉熵损失; L_{reg} 是地面真相未来轨迹的负高斯对数似然; $\alpha = 1.0$,是平衡两个损失项的标量。

2 实验

本节在 Argoverse 数据集^[20]上进行了广泛的实验,以证明本文模型的有效性。

2.1 实验设置

数据集:Argoverse 是一个公共数据集,可用于自动驾驶汽车 3D 追踪和运动预测,同时也是首个包含高精地图信息的数据集。它共收集了 327 793 个场景的 5 s 序列,每个序列采样率为 10 Hz,即下一个时间戳中出现的相同对象的间隔约为 0.1 s。整个数

据集序列被划分为训练集、验证集和测试集,分别包含 205 942 个、39 472 个和 78 143 个序列。在训练集和验证集中,每个目标对象的数据都提供了完整的 5 s 轨迹信息。对于测试集,仅提供了前 2 s 的轨迹数据,以评估模型在不完整信息下的预测能力。

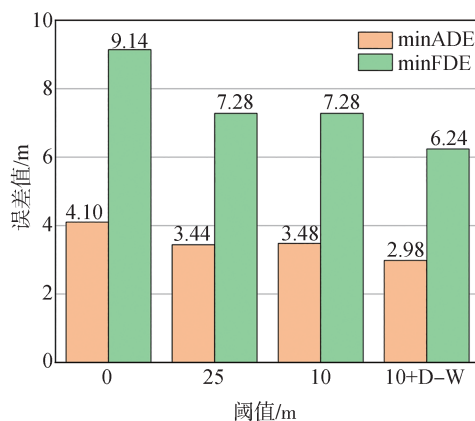
评价指标:与其他轨迹预测方法类似,本文将采用最为广泛使用的平均位移误差(ADE)和最终位移误差(FDE)作为本研究的评价指标。其中 ADE 是指整个时间步长上预测轨迹与地面真实值之间的平均 L2 距离,FDE 为预测轨迹端点与地面真值之间的位移误差。因为本文将对 Argoverse 进行多模态预测评估,因此使用 minADE 和 minFDE 作为评价指标。本预测任务还考虑了失误率(MR),失误率指根据端点误差,在地面真实度的某个阈值(2 m)内预测轨迹的百分比。

2.2 成分消融研究

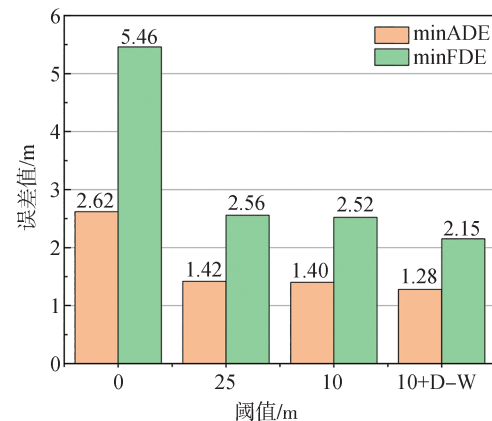
文中所提出的模型被划分为独立的模块。消融实验在 Argoverse 验证集上进行,以评估每个模块对预测误差的影响。实验以没有任何交互的 GRU 网络为基线,依次添加其他模块来比较结果。其中包括车辆交互提取模块 AgentGCN、驾驶场景表征模块 LaneGCN 和轨迹预测模块(MLP 和 GRU),如表 1 所示。

表 1 消融实验结果

模型模块			K=1		K=6		
AgentGCN 10 m + D-W	LaneGCN	预测模块		min ADE	min FDE	min ADE	min FDE
		MLP	GRU				
			√	4.10	9.14	2.62	5.46
√			√	2.98	6.24	1.28	2.15
√	√	√		2.77	5.74	1.20	1.97
√	√		√	2.09	4.67	0.82	1.25



(a) K=1 验证集上阈值影响对比



(b) K=6 验证集上阈值影响对比

图 5 不同阈值对位移误差影响的比较

表 1 中的结果表明,基线 GRU 在 $K=1$ 和 $K=6$ 时误差最大,因为它无法捕捉周围车辆和车道拓扑的相互作用。在添加车辆交互信息提取模块 AgentGCN 后, $K=1$ 和 $K=6$ 的 minADE 分别降低了 27% 和 51%。结果表明,车辆之间的相互作用对未来的轨迹有很大影响。然后考虑了 GRU 和 AgentGCN 静态道路环境库的影响,将 LaneGCN 加入到车道拓扑提取模型中。结果表明,当 $K=1$ 和 $K=6$ 时,minADE 分别降低了 7% 和 6%,预测效果较好。

此外,还对预测头(GRU 或 MLP)的选择进行比较实验。结果表明,使用 GRU 的序列到序列结构比 MLP 更有效。最后,与不考虑任何相互作用的模型相比,本文模型将 $K=1$ 的 minADE 降低了 49%,将 $K=6$ 的 minADE 降低了 68%,这表明本文模型更容易捕捉交互作用,并在复杂的道路环境中预测更准确的驾驶轨迹。

2.3 交互阈值研究

在本节中评估了交互阈值的大小对最终结果的影响,如图 5 所示。横坐标分别显示 0、25、10 m 和 10+D-W 的阈值。0 m 表示不考虑车辆之间的相互作用;25 和 10 m 意味着在相应的阈值内车辆之间存在相互作用;10+D-W 表示 10 m 以内的车辆之间存在相互作用,不同车辆之间的相互作用与车辆的实际空间距离有关。同时用不同数量的预测轨迹($K=1$ 和 $K=6$)进行了实验。结果表明,无论输出轨迹 K 如何,当不考虑车辆相互作用时(横坐标值为 0),minADE 和 minFDE 都高于其他列。这意味着,当车辆被视为一个独立的实体,且不考虑相互作用来预测轨迹时误差将很大。相反,考虑周围车辆的相互作用确实有助于更准确的预测。

2.4 预测模态数的影响

由于目标的未来轨迹具有多种可能性,因此轨迹预测也具有很大的不确定性。为验证本文模型能够更好地反映这种多可能性,设计了实验使模型能够输出不同数量的轨迹。其中,输出设置为 num=1、num=3、num=6 共 3 种可能性。

从图 6 的结果可以看出,与确定性的单个轨迹 (num=1) 相比,预测少量多峰轨迹 (num=3) 对结果几乎没有改善。本文模型 (num=6) 对 ADE 和 FDE 都具有最好的预测结果,即它对最终位移误差和平均位移误差都保留了较小的值。这表明所提出的模型能够很好地解释现实中车辆轨迹的不确定性。

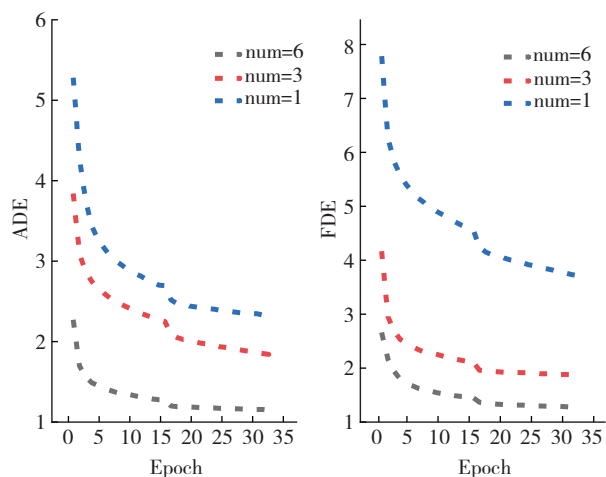


图 6 不同预测轨迹条件下的预测误差的比较

2.5 模型预测精度的比较

如表 2 所示,将本文所提模型与官方基线和其他先进的模型在 Argoverse 运动预测的测试集数据上进行了比较。

表 2 本文模型与其他模型在 Argoverse 测试集上的比较 ($K=6$)

模型	minADE	minFDE	MR
GRIP++	1.40	2.52	0.31
TNT	0.94	1.54	0.13
LaneGCN	0.87	1.36	0.16
HOME	0.92	1.36	0.11
mmTransformer	0.84	1.34	0.15
SSL-Lanes	0.84	1.25	0.13
FRM	0.82	1.27	0.14
Our Model	0.82	1.25	0.14

GRIP++^[19]使用 GCN 网络来同时提取代理之间的交互特征,但仅纯粹构建代理之间的图网络来考虑交互效应,不考虑静态车道环境的影响。

TNT^[21]是一种基于历史数据生成目标的轨迹状态序列方法,并基于似然估计得到紧凑的轨迹预测集。

LaneGCN^[18]利用了一种新颖的结构化地图表示和演员-地图交互。它使用多个邻接矩阵和车道展开图卷积,利用融合网络捕捉参与者和地图之间的复杂互动。

HOME^[22]提出了一种解决运动预测问题的框架,该模型结合了经典 CNN 与注意力机制,用于智能体与环境特征提取及交互。

mmTransformer^[23]使用堆叠 Transformer 作为主干,用来聚合固定轨迹建议的上下文信息;特征解码器通过轨迹生成器和选择器分别对每个特征进行解码。

SSL-Lanes^[24]将自监督学习方法应用于轨迹预测任务,证明了自监督学习在轨迹预测中的有效性,显著提高了轨迹预测性能。

FRM^[25]提出了使用车道信息来预测对象之间的随机未来关系的方法。通过预测车辆占用车道级路点的概率,并使用概率分布对交互进行建模。

与不考虑车道特征的 GRIP++ 模型相比,本文模型将平均位移误差降低了 41.4%,最终位移误差降低了 50.4%,显著提高了预测结果。与 LaneGCN 模型相比,本文模型平均位移误差降低了 5.7%,最终位移误差降低 8.1%。与最新的 FRM 相比,本文模型具有与 FRM 相当的平均位移误差,且最终位移误差略好于 FRM,失误率也与 FRM 相当。总而言之,本文模型在 Argoverse 上取得了良好的结果。

2.6 可视化分析

在 Argoverse 数据集上可视化了所提出的模型预测结果,如图 7 所示。分别对不考虑车道图的 GRIP++ 模型和本文所提模型在相同场景下的轨迹进行可视化,其中目标车辆的历史轨迹为黄色,真实地面轨迹为红色,预测的多模态轨迹为绿色。

通过分析不同场景下车辆的预测轨迹和真实轨迹,发现本文所提模型可以生成可行且平滑的多模态轨迹,且这些轨迹相较于 GRIP++ 模型在相同环境下更符合地图约束。

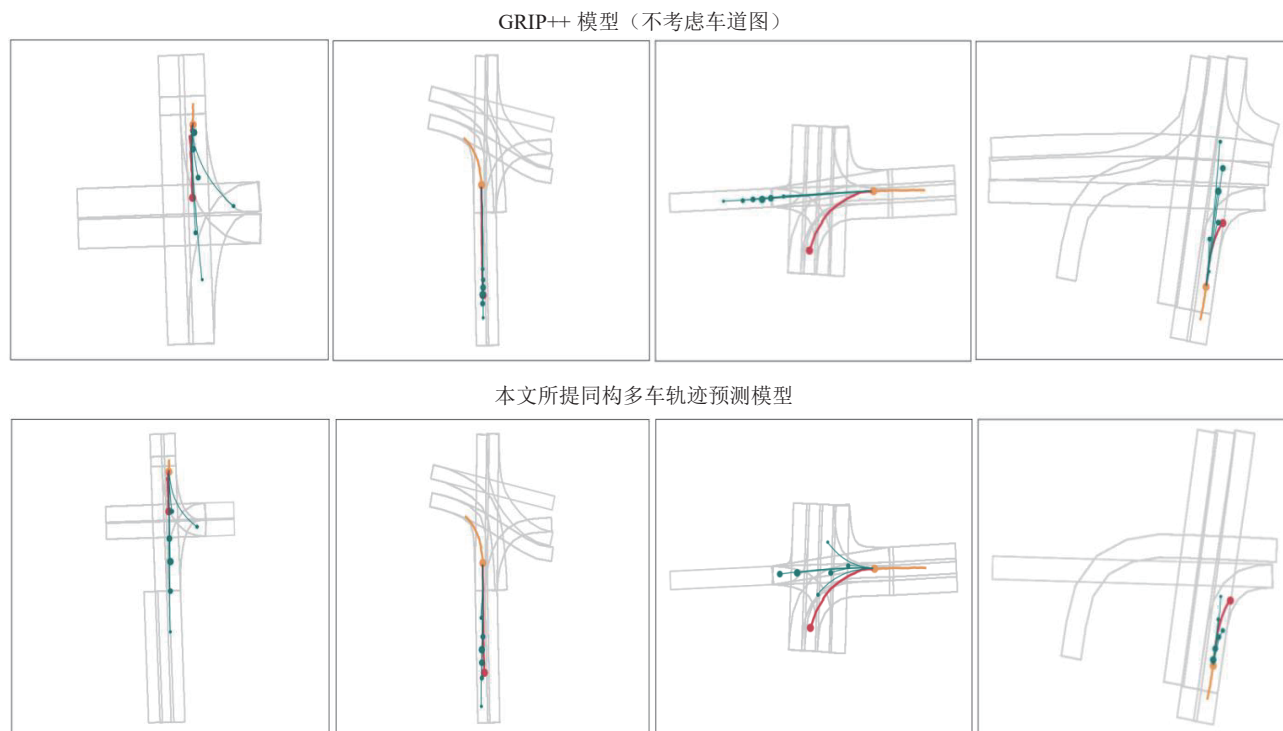


图7 Argoverse 验证集上的运动预测结果可视化

3 结论

提出一种新的轨迹预测模型,该模型扩展了从周围车辆预测车辆轨迹时考虑的相互作用,包括车道拓扑的预测,并独立提取车道拓扑网络和车辆相互作用关系,这确保了该模型在考虑周围车辆的相互作用以及可驾驶区域内轨迹的输出的同时考虑了车道。在构建交互关系时,所提出的模型将距离权重分配给空间关系的边缘,这改进了基于阈值构建的边缘不考虑邻近车辆的不同影响。这可以更好地反映车辆的空间距离与交互影响之间的关系。并在 Argoverse 数据集上取得了良好的效果,特别是与不考虑车道环境的预测模型相比。这表明在复杂的城市环境中,周围车辆的动态影响和静态车道拓扑的约束对于预测至关重要。

然而,由于本文中使用的数据集没有考虑行人、骑行者等对象,因此所提出的模型在具有多参与者交互的复杂交通场景中的效果仍有待测试。在未来的工作中,计划考虑方向和速度等限制,对密切互动进行进一步分析。如何提取高清矢量地图的更多局部信息,如是否变道和限速等,也是未来分析的一部分。

参考文献

- [1] CAI Y, DAI L, WANG H, et al. DLnet with training task conversion stream for precise semantic segmentation in actual traffic scene[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(11): 6443-6457.
- [2] ZHOU J, OLOFSSON B, FRISK E. Interaction-aware motion planning for autonomous vehicles with multi-modal obstacle uncertainty predictions[J]. IEEE Transactions on Intelligent Vehicles, 2023.
- [3] 胡启慧,蔡英凤,王海,等. 基于层次图注意的异构多目标轨迹预测方法[J]. 汽车工程, 2023, 45(8): 1448-1456.
HU Qihui, CAI Yingfeng, WANG Hai, et al. Heterogeneous multi-object trajectory prediction method based on hierarchical graph attention [J]. Automotive Engineering, 2023, 45 (8) : 1448-1456.
- [4] 汪梓豪,蔡英凤,王海,等. 基于单目视觉运动估计的周边多目标轨迹预测方法[J]. 汽车工程, 2022, 44(9): 1318-1326.
WANG Zihao, CAI Yingfeng, WANG Hai, et al. Surrounding multi-target trajectory prediction method based on monocular visual motion estimation [J]. Automotive Engineering, 2022, 44 (9): 1318-1326.
- [5] 高镇海,鲍明喜,高菲,等. 基于LSTM概率多模态预期轨迹预测方法[J]. 汽车工程, 2023, 45(7): 1145-1152.
GAO Zhenhai, BAO Mingxi, GAO Fei, et al. The method of probabilistic multi-modal expected trajectory prediction based on LSTM[J]. Automotive Engineering, 2023, 45(7): 1145-1152.

- [6] DANIELSSON S, PETERSSON L, EIDEHALL A. Monte Carlo based threat assessment: analysis and improvements [C]. 2007 IEEE Intelligent Vehicles Symposium. IEEE, 2007: 233–238.
- [7] LEFÈVRE S, LAUGIER C, IBAÑEZ-GUZMÁN J. Exploiting map information for driver intention estimation at road intersections [C]. 2011 IEEE Intelligent Vehicles Symposium (iv). IEEE, 2011: 583–588.
- [8] FIRL J, STÜBING H, HUSS S A, et al. Predictive maneuver evaluation for enhancement of car-to-x mobility data [C]. 2012 IEEE Intelligent Vehicles Symposium. IEEE, 2012: 558–564.
- [9] AOUDE G S, DESARAJU V R, STEPHENS L H, et al. Behavior classification algorithms at intersections and validation using naturalistic data [C]. 2011 IEEE Intelligent Vehicles Symposium (iv). IEEE, 2011: 601–606.
- [10] MORRIS B T, TRIVEDI M M. Trajectory learning for activity understanding: unsupervised, multilevel, and long-term adaptive approach [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(11): 2287–2301.
- [11] LI Z, WANG B, GONG J, et al. Development and evaluation of two learning-based personalized driver models for pure pursuit path-tracking behaviors [C]. 2018 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2018: 79–84.
- [12] CHANDRA R, BHATTACHARYA U, BERA A, et al. Traffic trajectory prediction in dense and heterogeneous traffic using weighted interactions [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 8483–8492.
- [13] DIEHL F, BRUNNER T, LE M T, et al. Graph neural networks for modelling traffic participant interaction [C]. 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019: 695–701.
- [14] 连静, 李硕贤, 刘一获, 等. 基于车道目标引导的车辆轨迹预测 [J]. 汽车工程, 2023, 45(8): 1353–1361.
LIAN Jing, LI Shuoxian, LIU Yidi, et al. Goal supervised attention network for vehicle trajectory prediction [J]. Automotive Engineering, 2023, 45(8): 1353–1361.
- [15] CHAI Y, SAPP B, BANSAL M, et al. Multipath: multiple probabilistic anchor trajectory hypotheses for behavior prediction [J]. arXiv preprint arXiv:1910.05449, 2019.
- [16] HONG J, SAPP B, PHILBIN J. Rules of the road: predicting driving behavior with a convolutional model of semantic interactions [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 8454–8462.
- [17] GAO J, SUN C, ZHAO H, et al. VectorNet: encoding HD maps and agent dynamics from vectorized representation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11525–11533.
- [18] LIANG M, YANG B, HU R, et al. Learning lane graph representations for motion forecasting [C]. Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020: 541–556.
- [19] LI X, YING X, CHUAH M C. Grip++: enhanced graph-based interaction-aware trajectory prediction for autonomous driving [J]. arXiv preprint arXiv:1907.07792, 2019.
- [20] CHANG M F, LAMBERT J, SANGKLOY P, et al. Argoverse: 3D tracking and forecasting with rich maps [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 8748–8757.
- [21] ZHAO H, GAO J, LAN T, et al. TNT: target-driven trajectory prediction [J]. arXiv preprint arXiv:2008.08294, 2020.
- [22] GILLES T, SABATINI S, TSISHKOU D, et al. HOME: heatmap output for future motion estimation [J]. arXiv preprint arXiv:2105.10968, 2021.
- [23] LIU Y, ZHANG J, FANG L, et al. Multimodal motion prediction with stacked transformers [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021: 7573–7582.
- [24] BHATTACHARYYA P, HUANG C, CZARNECKI K. SSL-Lanes: self-supervised learning for motion forecasting in autonomous driving [C]. Conference on Robot Learning. PMLR, 2023: 1793–1805.
- [25] PARK D, RYU H, YANG Y, et al. Leveraging future relationship reasoning for vehicle trajectory prediction [J]. arXiv preprint arXiv:2305.14715, 2023.

~~~~~  
(上接第1862页)

- [19] DONG J, YANG G H.  $H_\infty$  controller synthesis via switched PDC scheme for discrete-time T-S fuzzy systems [J]. IEEE Transactions on Fuzzy Systems, 2009, 17(3): 544–555.
- [20] SALEH L, CHEVREL P, CLAVEAU F, et al. Shared steering

control between a driver and an automation: stability in the presence of driver behavior uncertainty [J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(2): 974–983.