

面向智能网联车辆的轨迹预测模型

王 健, 贾晨威

(吉林大学 计算机科学与技术学院, 长春 130012)

摘要:相较于传统的单车智能自动驾驶系统只能根据自身对于环境感知的结果对未来进行预测,智能网联自动驾驶系统可以通过V2X技术获取额外的周围道路环境动态信息进行融合预测。本文在单车智能轨迹预测的基础上,使用特殊编码器使得轨迹预测模型可以融合自身的感知信息与来自V2X共享的动态道路信息。在CARLA仿真数据集上的实验结果证明,使用V2X技术获取周围道路环境的动态信息相较于未使用动态环境信息的轨迹预测算法能够更准确地预测车辆轨迹。

关键词:计算机应用技术;自动驾驶;车联网;轨迹预测

中图分类号:TP399 **文献标志码:**A **文章编号:**1671-5497(2025)06-1963-10

DOI:10.13229/j.cnki.jdxgxb.20231046

Trajectory prediction model for intelligent connected vehicle

WANG Jian, JIA Chen-wei

(College of Computer Science and Technology, Jilin University, Changchun 130012, China)

Abstract: In contrast to traditional single-vehicle intelligent autonomous driving systems, which can only make predictions about the future based on their own perception of the environment, intelligent connected autonomous driving systems have the capability to enhance predictions by incorporating additional dynamic information about the surrounding road environment through V2X technology. Building upon the foundation of single-vehicle intelligent trajectory prediction, a specialized encoder was employed to enable the trajectory prediction model to seamlessly fuse its own perceptual information with dynamic road data obtained via V2X communication. The experimental results on the CARLA simulation dataset demonstrate that using V2X technology to obtain dynamic information of the surrounding road environment can more accurately predict vehicle trajectories compared to trajectory prediction algorithms that do not use dynamic environment information.

Key words: computer application technology; autonomous driving; internet of vehicles; trajectory prediction

0 引 言

随着自动驾驶应用的日益发展,运动预测模

型也引起了广泛的关注。尽管单车智能在许多环境中表现出色,但受限于传感器的感知能力,因此

收稿日期:2023-10-03.

基金项目:国家自然科学基金项目(62272194).

作者简介:王健(1982-),男,教授,博士.研究方向:车联网与自动驾驶.E-mail:wangjian591@jlu.edu.cn

仍面临着许多与驾驶安全相关的长尾问题。预测模块作为自动驾驶系统中的重要组成部分。它根据感知到的物体的历史运动状态预测未来可能的轨迹。预测出的轨迹信息会被提供给下游的决策和规划模块,使自车做出合理和安全的驾驶行为。因此,预测模块的性能表现与自动驾驶车辆的行驶安全息息相关。

预测模型通常需要有效地整合多种输入,包括交通场景的静态、社交和时序方面的要素,以实现准确有效的预测。为了更好地捕捉交通场景中实体之间的复杂交互作用,目前研究人员探索了多种用于交通场景表示和建模的技术。

在建模交通场景的道路信息时,有 2 种常见的输入类型,分别是栅格地图输入和高清地图输入。栅格地图输入是指栅格化的周围环境俯视图。通常使用卷积神经网络编码栅格地图中带有各种特征通道的时空信息^[1-12]。但是,当环境负载过高导致输入的特征通道非常多时,模型的推理性能会受到非常严重的影响。作为替代方案,高清地图输入通过使用有限点的序列来表示道路的几何信息和代理轨迹,并且通过为不同点分配不同属性表示复杂的特征信息^[13-19]。由于使用了稀疏的点集表示复杂的交通场景,在相同的数据量下高精地图输入具有更高的信息容纳能力。

代理之间的社交互动也会对代理的运动产生影响。目前研究人员探索了多种方法建模代理间的互动关系。社交汇集(Social pooling)根据代理之间的社交关系建模互动特征^[20-23],但这种方法通常会忽略代理之间互动的时序特征。得益于图神经网络(Graph neural network, GNN)对关系数据的强大建模能力,越来越多的轨迹预测模型使用 GNN 更高效地捕捉代理之间的空间和时间交互关系^[13-16, 24, 25],但其在计算和存储复杂性方面仍存在劣势。由于交通场景的复杂性,模型通常需要对不同的场景信息做出不同的反应。注意力机制模拟了人类注意力的行为,允许模型动态地关注输入数据的不同部分,在社交互动中采用注意力机制,可以使代理捕捉输入序列不同部分之间的依赖关系,并有选择地关注最显著的特征以进行预测,同时也提高了可解释性^[17, 26-35]。

Transformer 模型^[36]已成功地应用于各个领域^[37-50],在众多的序列任务中都展示了其高效的

建模能力。轨迹预测任务也是一种序列预测任务,其可以很好地适配 Transformer 模型。目前,在一些关于运动预测的研究中都使用 Transformer 结构处理代理时序运动趋势、代理之间的时空依赖关系以及代理和环境元素之间的依赖关系^[26-30, 51, 52],并展现了 Transformer 模型在复杂场景中提高轨迹预测准确性和效率的潜力。同时,Transformer 模型可以高效地处理多模态的特征,对于轨迹预测任务可以在不同阶段将不同模态的预测关联信息作为键值,使得每个代理可以融合多种模态信息。

对于单车智能的自动驾驶系统,为了解决更多的长尾问题,不得不在传感器、数据和模型方面投入更多的成本和精力,但由于单车感知角度的限制,单车的感知能力是有上限的。然而,V2X 技术可以通过较少的传感器成本,利用多角度传感器的信息共享,大幅扩大车辆感知的时空范围^[53, 54]。在目前的自动驾驶领域,对于轨迹预测模型的研究多是建立在单车智能的条件下,而没有考虑 V2X 技术对轨迹预测模型的作用。

由于不同自动驾驶系统对预测任务的定位不同,许多自动驾驶公司和研究机构都基于自身对于预测任务的需求开源了自己的数据集以使研究人员可以训练和评估自己的方法。目前存在的轨迹预测数据集都是从单车智能的角度采集的,其仅包含了自车感知范围内的轨迹和道路信息,而缺乏 V2X 场景下的共享环境数据,如信号灯的时序信息这样的交通环境的动态数据^[55-58]。

综上所述,目前的轨迹预测方法和数据集都只关心单车智能场景下的问题,而对于智能网联场景下,自动驾驶车辆可以与周围的网联设备进行通信,获取周围环境的未来预测窗口内的动态道路变化信息,但自动驾驶车辆如何充分地利用这些额外的数据仍是个问题。所以本文从这一点出发,研究了如何在单车智能轨迹预测模型的基础上,增加特殊的动态环境编码器编码这些额外的动态道路变化信息,并将得到的动态道路变化特征与代理特征进行融合,最终实现更准确的轨迹预测结果。

1 建模方法

1.1 数据表示

1.1.1 轨迹信息表示

本文将交通场景数据分为代理特征(Agent

features)和车道特征(Lane features)^[15, 17, 52]2种类型。代理特征描述了每个目标在每个时刻和周围的空间关系。为了捕捉每个时刻的空间信息,本文构建了以每个代理为中心的子图,子图内的所有其他代理位置信息都被转换到中心代理的坐标系下。子图中每个节点的值是从代理在2个相邻时刻计算得到的运动矢量,即:

$$E'_{A-A,i}(j) = \mathbf{R}_i^T(p'_j - p'_i) \quad (1)$$

$$N'_{A-A,i}(j) = \mathbf{R}_i^T(p'_j - p'_{j-1}) \quad (2)$$

式中: $E'_{A-A,i}(j)$ 为以agent_i为中心的子图中的边信息; $N'_{A-A,i}(j)$ 为子图中的节点信息; \mathbf{R}_i^T 为agent_i的旋转矩阵; p'_t 为时刻 t 时agent_i的位置。

1.1.2 车道空间特征表示

本文所提模型需要处理交通环境的动态和静态信息。因此,本文进一步将车道特征分为车道空间特征(Lane spatial feature)、车道动态特征(Lane dynamic feature)和车道静态特征(Lane static feature)。

车道空间特征是由一个指向车道当前位置节点到下一个位置节点的矢量表示的,矢量的长度等于2个节点之间的空间距离。每个代理都需要构建以该代理为中心的道路信息子图。子图中的边信息包括每个车道节点与中心代理之间的相对几何关系。子图中的节点信息包括车道节点之间的位置矢量。这些车道空间特征描述了车道的几何结构和代理与车道之间的空间关系,即:

$$E_{A-L,i}(\xi) = \mathbf{R}_i^T(p_\xi - p'_i) \quad (3)$$

$$N_{A-L,i}(\xi) = \mathbf{R}_i^T(p_\xi^0 - p_\xi^1) \quad (4)$$

式中: p_ξ 为车道中节点 ξ 的几何位置; p_ξ^0 和 p_ξ^1 为车道中2个连续节点的几何位置。

1.1.3 车道动态特征表示

车道动态特征是由V2X网络共享得到的每个车道的未来预测时间窗口内连续的交通控制信息。这些控制信息影响着车辆和行人在未来的运动目标。具体的,智能网联车辆通过近场通信网络与周围路侧V2X交换动态交通信息,使得车辆可以获知未来预测时间窗口内的控制信号的时序变化数据。这些数据会在当前预测模型被调用时进行预处理和输入。本文使用one-hot编码来表示车道在每个时刻的动态信息,具体每个位置的编码含义如表1所示。

表1 动态特征的one-hot编码

Table 1 Dynamic features one-hot encoding

下标	含义
0	车道状态未知
1	车道允许通行(被绿灯控制)
2	车道被黄灯控制
3	车道不允许通行(被红灯控制)

1.1.4 车道静态特征表示

车道静态特征是由自车感知模块采集得到的车道固有信息,其不会随着时间的变化而变化,包括车道的转向方向、车道标线是否允许变道和车道线类型等。同样使用one-hot编码表示车道静态信息,具体的编码含义如表2所示。

表2 静态特征的one-hot编码

Table 2 Static features one-hot encoding

下标	含义
0~21	车道类型
22~25	车道方向
26~36	左车道类型
37~47	右车道类型
48	是否在交叉路口

1.2 模型结构

图1为本文动态智能网联预测模型(Dynamic intelligent connected prediction, DICP)模型的整体框架。算法的大体流程如下:

(1)数据预处理流程。将轨迹数据和道路信息向量化表示,对于每个需要预测的代理,都会构建一个以自身为中心的子图。

(2)使用Agent-Agent交互模块编码每个代理每个时刻的空间关系特征,然后使用时序Transformer获取每个代理空间特征的时序变化特征。

(3)使用Agent-Lane静态交互模块编码每个代理与周围静态道路特征的依赖关系。

(4)使用Agent-Lane动态交互模块编码每个代理与周围动态道路特征的依赖关系。

(5)使用全局交互模块使每个代理能够捕捉与场景中的其他代理在未来时刻中的依赖关系。

(6)使用多个多层感知机(Multi-layer perceptron, MLP)解码出每个代理的多模态轨迹和每个模态的置信度分数。

(7)算法的后处理流程。将每个代理输出的轨迹还原到自车坐标系下。

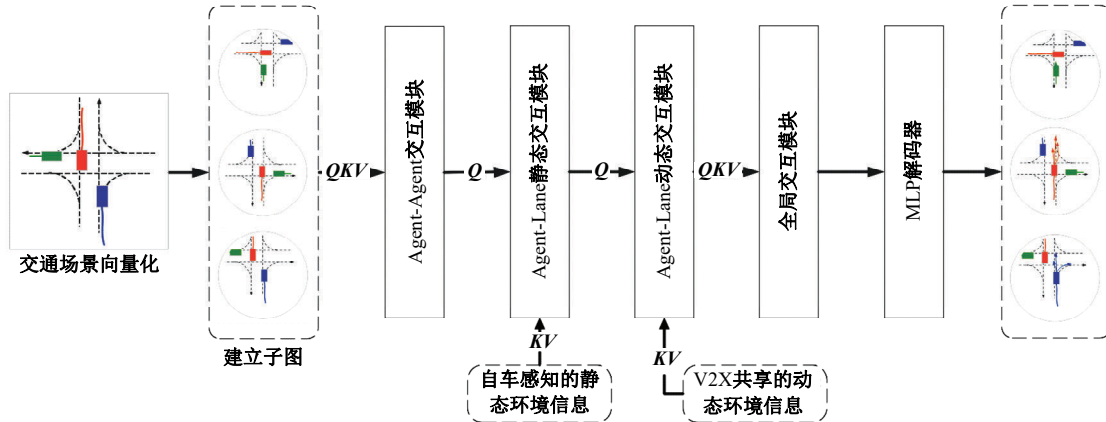


图 1 整体模型架构

Fig. 1 Overall model architecture

1.2.1 Agent-Agent 交互模块

Agent-Agent 交互模块用于捕捉场景中代理之间的时空关系。它以场景中每个时刻所构建的代理关系子图作为输入,中心代理的信息作为 Query,将连接的其他代理的信息作为 Key 和 Value。与 Transformer 的编码器计算方式相似,Agent-Agent 交互模块使用多头交叉注意力(Multi-head cross-attention)和一个前馈网络(Feed-forward network)获得融合的空间特征,并且使用门控单元更新中心节点的信息,然后获得每个代理在每个时间步的一系列空间关系特征 $F'_{A-A,i}$ 。这些连续的空间特征经过一个时序 Transformer 编码器来获取时间特征 $F_{A-A,i}$ 。与 Bert 类似^[39],Agent-Agent 交互模块在最后增加一个可学习的嵌入汇集整个时间段的时空变化特征。具体的结构如图 2 所示,其中各个模块的计算方式分别如下所示。

$$A'_{i,j} = \text{Softmax} \left(\frac{Q'_i K'_j{}^T}{\sqrt{(d)}} \right) \quad (5)$$

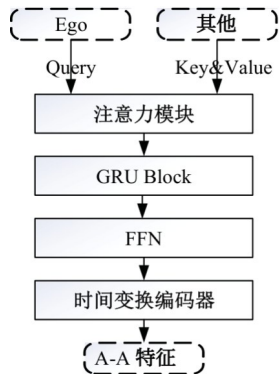


图 2 Agent-Agent 交互模块

Fig. 2 Agent-Agent interaction module

$$\text{Attn}'_i = \sum_j A'_{i,j} V'_j \quad (6)$$

$$F'_{A-A,i} = \text{GRU}(\text{FFN}(\text{Attn}'_i)) \quad (7)$$

$$F_{A-A,i} = \text{TfEnc}(\text{cat}(F'_{A-A,i}, \text{token}_{\text{ctx}}), \mathbf{M}_{\text{temp}}) \quad (8)$$

式中: Q'_i 为中心代理在 t 时刻的查询特征; K'_j 和 V'_j 分别为在 t 时刻其他代理的键和值; $A'_{i,j}$ 为注意力分数; GRU 为门控循环单元; FFN 为前馈神经网络; Attn'_i 为注意力特征; TfEnc 为标准的 Transformer 模型编码器模块; $\text{token}_{\text{ctx}}$ 为一个可学习的嵌入,用于汇聚全局的时序特征; \mathbf{M}_{temp} 为下三角矩阵,使得每个时刻只能注意到当前和过去时刻的信息。

1.2.2 车道特征编码器

对于车道的静态特征,本文使用了 1 个 MLP 将静态特征的 one-hot 编码转换成特定维度的特征。

对于车道的动态特征,本文使用 1 个时间注意力编码器(Temporal transformer encoder)获取时间维度上的特征。首先,把动态特征的每个状态转换为对应的可学习的嵌入,并在末尾添加一个单独的可学习嵌入以汇聚全局的时间信息。然后,为每个位置添加了一个可学习的位置编码标记时序信息。这些特征通过时间注意力编码器,最终转换成时间维度上的特征。

1.2.3 Agent-Lane 静态交互模块

本文使用了一个代理与车道信息的静态交互模块,使得每个代理能够收集到局部的静态车道信息,从而产生初步的只关心静态信息的预测结果。

具体的,Agent-Lane 静态交互模块使用了交叉注意力的结构。其中,每个代理的特征作为 Query,车道的静态信息则作为 Key 和 Value。然

后,使用一个前馈神经网络和门控结构更新每个代理的特征。具体的结构如图 3 所示。

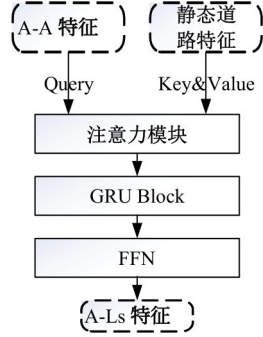


图 3 Agent-Lane 静态交互模块

Fig. 3 Agent-Lane static interaction module

1.2.4 Agent-Lane 动态交互模块

在经过 Agent-Lane 静态交互模块之后,只得到了静态环境信息的预测头。因此,还需要进一步地使每一个代理收集周围动态环境的信息,从而强化预测头,进一步得出更准确的预测结果。

具体的,将已经收集得到的静态环境信息的预测头作为 Query,动态环境信息作为 Key 和 Value。再经过前馈神经网络和门控结构使每个预测头融合动态环境信息。具体的结构如图 4 所示。

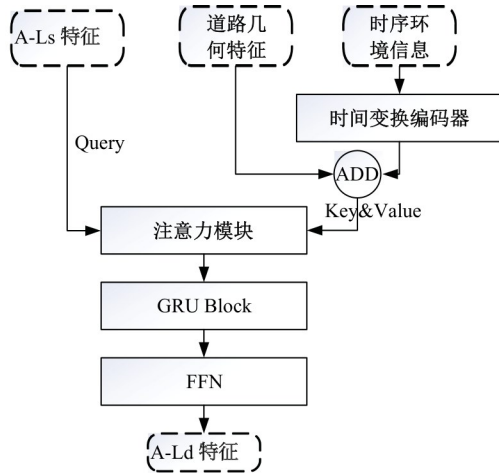


图 4 Agent-Lane 动态交互模块

Fig. 4 Agent-Lane dynamic interaction module

1.2.5 Agent-Agent 全局交互模块

为了捕获场景中代理之间的远程依赖关系,本文采用了一个 Agent-Agent 全局交互模块。该模块利用坐标系之间的几何关系,促进了局部区域之间的信息传递。具体而言,Agent-Agent 全局交互模块考虑了场景中不同区域的相对位置和方向,以实现远程区域之间的有效通信。通过融合坐标系之间的几何关系,Agent-Agent 全局交

互模块能够有效地建模场景不同部分之间的远程依赖关系和交互作用。对于同一场景中的所有代理,计算所有代理的多头自注意力,并且使用前馈神经网络和门控单元更新每个代理的特征。

这一模块的引入有助于更好地理解并捕获场景中的全局信息,特别是远程部分之间的相互作用,使模型能够考虑到整个场景的动态性和复杂性,从而提高预测的准确性和鲁棒性。

1.2.6 轨迹解码器模块

本文采用了 Laplace 混合模型表示未来轨迹的分布。具体的,使用 MLP 融合每个代理的局部和全局信息。然后,分别使用 3 个不同的 MLP 解码聚合后的信息,生成多模态分布中的不同参数。对于每个代理的每个模态,输出每个时刻位置分布的均值 $\mu'_{i,f} \in \mathbb{R}^2$ 和分布的尺度 $b'_{i,f} \in \mathbb{R}^2$ 信息。同时还输出了每个代理的每个模态的置信度 $\hat{\phi}$,其张量形状为 $[N, F]$,即:

$$f_{\text{Laplace}} = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right) \quad (9)$$

1.3 损失函数

在模型的训练过程中,采用 2 个损失函数鼓励模型准确预测未来位置并生成多个轨迹模态。首先,计算预测的每个模态轨迹与目标值最后时刻之间的 ℓ_2 距离,具有最小 ℓ_2 距离的模态被选为最佳预测轨迹。回归损失是最佳预测轨迹的所有时刻的平均 ℓ_2 距离。分类损失是将每个模态最后时刻的 ℓ_2 距离经过 Softmax 的输出作为软目标,然后计算与每个模态的置信度的交叉熵损失。最后,将回归损失 Loss_{reg} 和分类损失 Loss_{cls} 相加,得到总损失值 Loss ,即:

$$\text{Loss}_{\text{reg}} = -\frac{1}{NT} \sum_{i=1}^N \sum_{t=0}^T \log P(p'_i | \hat{\mu}'_i, \hat{b}'_i) \quad (10)$$

$$\text{Loss}_{\text{cls}} = \text{CE}(\hat{\phi}, \text{Softmax}(\|\hat{\mu}'_{i,f} - p'_i\|)) \quad (11)$$

$$\text{Loss} = \text{Loss}_{\text{reg}} + \text{Loss}_{\text{cls}} \quad (12)$$

式中: $P(\cdot|\cdot)$ 为 Laplace 分布的概率密度函数; $\hat{\phi}$ 为模型预测的每个轨迹模态的置信水平; $\hat{\mu}'_{i,f}$ 为 agent_i 最优预测中最后时刻的坐标均值; CE 为交叉熵; p'_i 为 agent_i 在时间 t 的实际坐标; $\hat{\mu}'_i$ 和 \hat{b}'_i 分别为最佳预测轨迹的位置均值和尺度。

2 实验

2.1 数据集

本文分别在静态环境条件和动态环境条件下

验证模型性能。在静态环境条件下的实验是为了验证传统单车智能的轨迹预测性能。在动态环境条件下的实验是为了验证车联网技术带来的动态环境信息对车辆轨迹预测性能的影响。因为目前缺少公开的动态环境轨迹预测的数据集,所以本文通过CARLA仿真器生成了一个包含100万条动态环境轨迹数据的数据集——DICP1M。同时,本文还采用了公开的Argoverse 1 Motion Forecasting^[55]数据集评估模型在静态环境下的轨迹预测性能。

2.1.1 Argoverse 数据集

Argoverse 1 Motion Forecasting 是一个精选的轨迹预测数据集,其中包含了324 557个交通场景,每个场景持续5 s,其中2 s用作历史轨迹,3 s用作未来轨迹。该数据集包含以鸟瞰图表示的每个被跟踪对象的2D中心点,采样频率为10 Hz。

2.1.2 DICP1M 数据集

DICP1M数据集是通过CARLA^[59]自动驾驶仿真软件采集得到的。CARLA是一个强大的开源自动驾驶仿真软件,目前已经在工业界和学术界得到了广泛的认可和应用。CARLA提供了包括车辆传感器、驾驶行为、天气以及交通场景等多种因素的仿真。本文通过在8个CARLA所提供的虚拟城市中,随机生成不同的交通场景,并以10 Hz的频率采集自动驾驶车辆周围5 s内的车辆运动轨迹和动态交通环境信息,最终得到DICP1M数据集。其中每个CARLA虚拟城市的描述和采样设置如表3所示。

DICP1M包含了100万个交通预测场景,每

表3 CARAL虚拟城市描述和采样设置

Table 3 CARLA virtual city description and sample setup

城市	描述	训练集采样数	验证集采样数
1	小型城市交通场景1	100 000	25 000
2	小型城市交通场景2	100 000	25 000
3	大型的城市地图,带有环岛和大型路口场景	100 000	25 000
4	小型城市交通场景,有一条“8字形”公路	100 000	25 000
5	方形网格城市交通场景	100 000	25 000
6	高速公路场景,有长的高速公路和许多出入口,包含密歇根式左转弯路口	100 000	25 000
7	乡村道路场景,道路狭窄,几乎没有红绿灯	100 000	25 000
8	大型城市的市中心场景	100 000	25 000

个预测场景持续5 s,其中2 s用作历史轨迹,3 s用作未来轨迹。DICP1M中800 000条数据用于训练,200 000条用于验证。与Argoverse数据集类似,DICP1M数据集每个场景以10 Hz频率采样的轨迹的二维鸟瞰图表示。与目前一些主流的轨迹预测数据集相比,DICP1M提供了更多的场景和轨迹数量,具体内容如表4所示。

表4 DICP1M与主流轨迹预测数据集对比

Table 4 Comparison between DICP1M and main stream trajectory prediction data sets

属性	nuScenes	Argoverse	DICP1M
预测时长/s	2	3	3
轨迹数量/条	4.3×10^3	324×10^3	1×10^6
城市数量/个	2	2	8
采样频率/Hz	2	10	10
动态交通环境信息	—	—	✓
总计时长/h	5.5	320	1 389

2.2 评价指标

最小最终位移误差 minFDE是指在预测时段T内,最佳预测的最终位置与实际最终位置之间的 ℓ_2 距离。其表达式为:

$$\text{minFDE} = \ell_2(\hat{y}_T^{\text{best}}, y_T) \quad (13)$$

式中: \hat{y}_T^{best} 为最佳预测轨迹的最后一个点的坐标; y_T 为实际轨迹的最后一个点的坐标。

最小平均距离 minADE是指最佳预测轨迹与实际轨迹之间所有时刻的平均 ℓ_2 距离:

$$\text{minADE} = \frac{1}{T_{\text{pred}}} \sum_{t=0}^{T_{\text{pred}}} \ell_2(\hat{y}_t^{\text{best}}, y_t) \quad (14)$$

式中: T_{pred} 为预测的步数; \hat{y}_t^{best} 为最佳预测轨迹t时刻的坐标; y_t 为实际轨迹t时刻的坐标。

最小失效率 minMR是指最佳预测轨迹所有时刻的位置与实际位置之间的 ℓ_2 距离超过2.0 m的比率:

$$\text{minMR} = \frac{\hat{s}^{\text{best}}}{T_{\text{pred}}} \quad (15)$$

式中: \hat{s}^{best} 为最佳预测轨迹在阈值范围之外的步数数量。

模型训练的具体超参数设置如表5所示。

2.3 实验结果

本文模型在DICP1M验证集和Argoverse验证集上的实验结果如表6所示。其中,“↓”表示越小越优,上标“1”表示在单车智能模型条件下的预测结果,上标“2”表示在智能网联模型条件下的预测结果。

表 5 超参数设置

Table 5 Specific hyperparameter settings

超参数	Argoverse	DICP1M
嵌入维度	128	128
学习率	5×10^{-4}	5×10^{-4}
训练轮次	64	64
随机失活概率	0.1	0.1
训练精度	bf16	bf16
批量大小	16	128
权重衰减率	3×10^{-4}	3×10^{-4}

表 6 DICP 模型在不同数据集上的表现

Table 6 Performance of DICP on different datasets

数据集	minADE ↓	minFDE ↓	minMR ↓
Argoverse ¹	0.669 5	0.985 4	0.096 1
DICP1M ¹	0.301 9	0.620 9	0.074 6
DICP1M ²	0.281 3	0.571 4	0.067 4

2.5 预测结果可视化

本文模型在智能网联场景下的轨迹预测可视化如图 5~图 8 所示,所使用的数据集均为 DICP1M。其中,箭头表示道路中心线,箭头的方向表示道路方向,红色的中心线表示该车道当前不允许通行,绿色表示可以通行,黄色表示红色和绿色变化的中间态,灰色表示车道状态不受控制。每个车辆的当前位置用彩色圆标出,红色实线表示车辆轨迹真值,蓝色虚线表示车辆轨迹预测值。

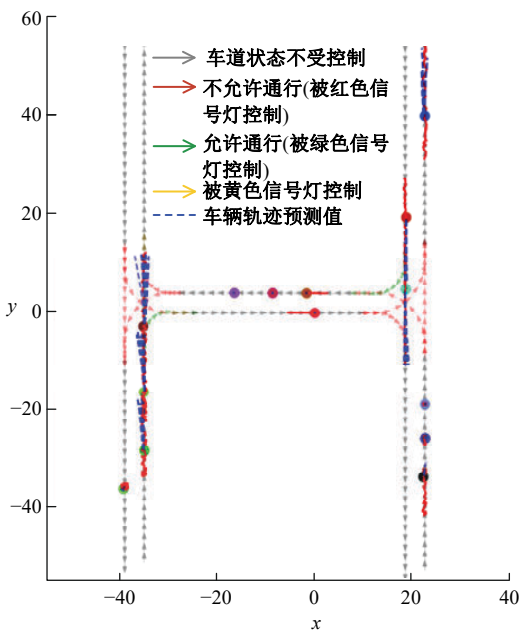


图 5 车辆受道路信号控制场景

Fig. 5 Scenario of vehicle is controlled by road signals

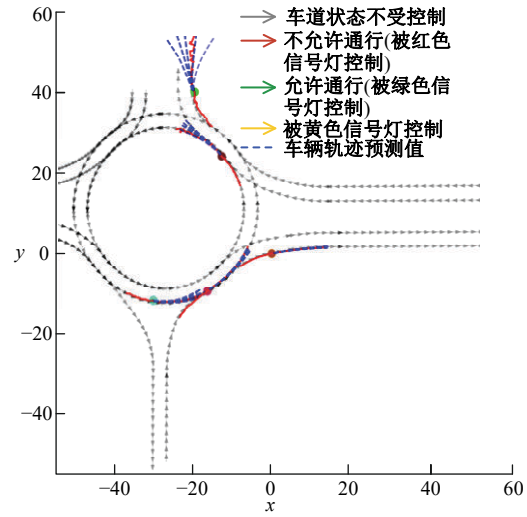


图 6 车辆通过环岛场景

Fig. 6 Scenario of vehicles passing through roundabout

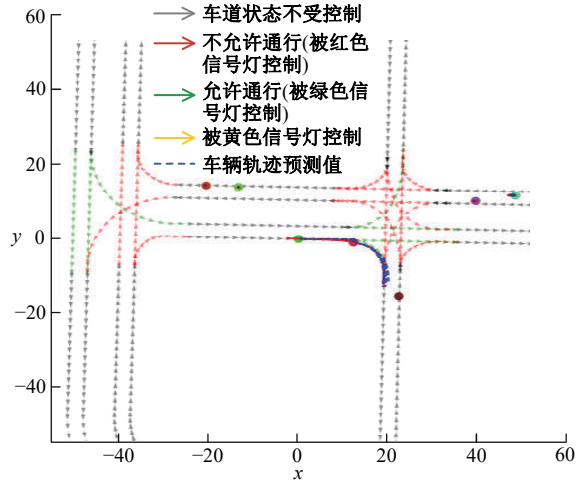


图 7 车辆受道路信号控制场景

Fig. 7 Scenario of vehicle is controlled by road signals

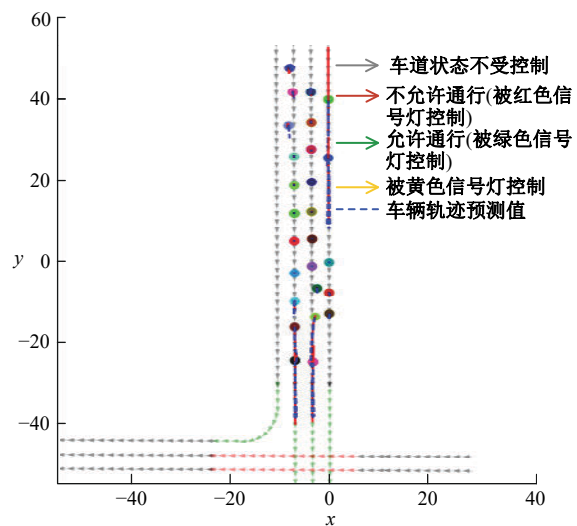


图 8 车辆获知道路将可以通行场景

Fig. 8 Scenario of vehicles are allowed to pass

3 结束语

本文首先构建了一个用于智能网联车辆轨迹预测的大规模数据集——DICP1M。其中包含了 100 万条轨迹数据样本,包括每个时刻的轨迹信息,以及每个时刻车辆周围的动态道路信息。其次,提出了一种适用于智能网联车辆的轨迹预测模型——DICP。在单车智能轨迹预测模型的基础上增加处理动态环境信息的结构,使得模型可以融合通过 V2X 技术共享而来的车道动态变化信息。最后,在 DICP1M 数据集上的实验结果表明,在 V2X 技术的支持下,智能网联车辆可以更准确地预测动态交通环境下的运动轨迹。

参考文献:

- [1] Chai Y, Sapp B, Bansal M, et al. MultiPath: multiple probabilistic anchor trajectory hypotheses for behavior prediction[J/OL]. [2023-07-13]. <https://doi.org/10.48550/arXiv.1910.05449>
- [2] Cui H G, Radosavljevic V, Chou F C, et al. Multimodal trajectory predictions for autonomous driving using deep convolutional networks[J/OL]. [2023-07-13]. <https://ar5iv.labs.arxiv.org/html/1809.10732>
- [3] Casas S, Luo W, Urtasun R. Intentnet: learning to predict intention from raw sensor data[C]//Conference on Robot Learning, Zürich, Switzerland, 2018: 947-956.
- [4] Lee, N, Choi W, Vernaza P, et al. Desire: Distant future prediction in dynamic scenes with interacting agents[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017: 336-345.
- [5] Hong J, Sapp B, Philbin J. Rules of the road: predicting driving behavior with a convolutional model of semantic interactions[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019: 8454-8462.
- [6] Jain A, Casas S, Liao R, et al. Discrete residual flow for probabilistic pedestrian behavior prediction[DB/OL]. [2023-07-13]. <https://doi.org/10.48550/arXiv.1910.08041>
- [7] Deo N, Trivedi M M. Trajectory forecasts in unknown environments conditioned on grid-based plans[DB/OL]. [2023-07-13]. <https://arxiv.org/abs/2001.00735>
- [8] Khandelwal S, Qi W, Singh J, et al. What-if motion prediction for autonomous driving[DB/OL]. [2023-07-13]. <https://arxiv.org/abs/2008.10587>
- [9] Mangalam K, An Y, Girase H, et al. From goals, waypoints & paths to long term human trajectory forecasting[C]//IEEE/CVF International Conference on Computer Vision, Nashville, TN, USA, 2021: 15233-15242.
- [10] Phan-Minh T, Grigore E C, Boulton F A, et al. Governet: multimodal behavior prediction using trajectory sets[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 14074-14083.
- [11] Rhinehart N, McAllister R, Kitani K, et al. Precog: prediction conditioned on goals in visual multiagent settings[C]//IEEE/CVF International Conference on Computer Vision, Long Beach, CA, USA, 2019: 2821-2830.
- [12] Ridel D, Deo N, Wolf D, et al. Scene compliant trajectory forecast with agent-centric spatiotemporal grids[J]. IEEE Robotics and Automation Letters, 2020, 2(5): 2816-2823.
- [13] Casas S, Gulino C, Liao R, et al. Spaggn: Spatially-aware graph neural networks for relational behavior forecasting from sensor data[C]//IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020: 9491-9497.
- [14] Huang Y, Bi H, Li Z, et al. Stgat: Modeling spatial-temporal interactions for human trajectory prediction[C]//IEEE/CVF International Conference on Computer Vision, Long Beach, CA, USA, 2019: 6272-6281.
- [15] Liang M, Yang B, Hu R, et al. Learning lane graph representations for motion forecasting[C]//Computer Vision-ECCV 2020, Glasgow, UK, 2020: 541-556.
- [16] Mohamed A, Qian K, Elhoseiny M, et al. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 14424-14432.
- [17] Gao J, Sun C, Zhao H, et al. Vectornet: Encoding hd maps and agent dynamics from vectorized representation[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 11525-11533.
- [18] Zhao H, Gao J, Lan T, et al. Tnt: Targetdriven trajectory prediction[C]//Conference on Robot Learning, London, UK, 2021: 895-904.
- [19] Zeng W, Liang M, Liao R, et al. Lanercnn: distrib-

- uted representations for graph-centric motion forecasting[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems, Prague, Czech Republic, 2021: 532-539.
- [20] Alahi A, Goel K, Ramanathan V, et al, Social lstm: Human trajectory prediction in crowded spaces[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Las Vegas, Nevada, USA, 2016: 961-971.
- [21] Deo N, Trivedi M M. Convolutional social pooling for vehicle trajectory prediction[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 2018: 1468-1476.
- [22] Gupta A, Johnson J, Li F F, et al. Social GAN: socially acceptable trajectories with generative adversarial networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018: 2255-2264.
- [23] Mangalam K, Girase H, Agarwal S, et al, It is not the journey but the destination: endpoint conditioned trajectory prediction[C] //Computer Vision-ECCV 2020, Glasgow, UK, 2020: 759-776.
- [24] Zhang L, Su P H, Hoang J, et al, Map-adaptive goal-based trajectory prediction[C]//Conference on Robot Learning, London, UK, 2021: 1371-1383.
- [25] Kosaraju V, Sadeghian A, Mart'in-Mart'in R, et al, Social-bigat: multimodal trajectory forecasting using bicycle-GAN and graph attention networks[C]//Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 2019, 32: 137-146.
- [26] Li L L, Yang B, Liang M, et al, End-to-end contextual perception and prediction with interaction transformer[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, USA, 2020: 5784-5791.
- [27] Liu Y, Zhang J, Fang L, et al, Multimodal motion prediction with stacked transformers[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 7577-7586.
- [28] Mercat J, Gilles T, Zoghby N E, et al, Multi-head attention for multimodal joint vehicle motion forecasting[C]//IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 2020: 9638-9644.
- [29] Ngiam J, Caine B, Vasudevan V, et al. Scene transformer: a unified architecture for predicting multiple agent trajectories [DB/OL]. [2023-07-13]. <https://arxiv.org/abs/2106.08417>
- [30] Yu C, Ma X, Ren J, et al. Spatio-temporal graph transformer networks for pedestrian trajectory prediction[DB/OL]. [2023-07-13]. <https://doi.org/10.48550/arXiv.2005.08514>
- [31] Salzmann T, Ivanovic B, Chakravarty P, et al. Trajectron++: multi-agent generative trajectory forecasting with heterogeneous data for control[DB/OL]. [2023-07-13]. https://www.ecva.net/papers/eccv_2020/papers_ECCV/papers/123630664.pdf
- [32] Ye L, Wang Z, Chen X, et al. GSAN: graph self-attention network for learning spatial-temporal interaction representation in autonomous driving[J]. IEEE Internet of Things Journal, 2021, 9(12): 9190-9204.
- [33] Luo C, Sun L, Dabiri D, et al. Probabilistic multimodal trajectory prediction with lane attention for autonomous vehicles[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 2020: 2370-2376.
- [34] Messaoud K, Deo N, Trivedi M M, et al. Multi-head attention with joint agent-map representation for trajectory prediction in autonomous driving[DB/OL]. [2023-07-13]. <https://arxiv.org/abs/2005.02545>
- [35] Gilles T, Sabatini S, Tsishkou D, et al. HOME: heatmap output for future motion estimation[J/OL]. [2023-07-13]. <https://arxiv.org/pdf/2109.01827>
- [36] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J/OL]. [2023-07-13]. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fd053c1c4a845aa-Paper.pdf
- [37] Brown T B, Mann B, Ryder N, et al. Language models are few-shot learners[DB/OL]. [2023-07-13]. <https://dl.acm.org/doi/pdf/10.5555/3495724.3495883>
- [38] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//Computer Vision-ECCV 2020: 16th European Conference, Berlin: Springer, 2020: 213-229.
- [39] Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[DB/OL]. [2023-07-13]. <https://aclanthology.org/N19-1423.pdf>
- [40] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [DB/OL]. [2023-07-13]. <https://arxiv.org/abs/2010.11929>
- [41] Arnab A, Dehghani M, Heigold G, et al. Vivit: a video vision transformer[C]//IEEE/CVF Conference

- on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 6836-6846.
- [42] Ho J, Kalchbrenner N, Weissenborn D, et al. Axial attention in multidimensional transformers [DB/OL]. [2023-07-13]. <https://arxiv.org/abs/1912.12180>
- [43] Lee J, Lee Y, Kim J, et al. Set transformer: a framework for attention-based permutation-invariant neural networks[C] //International Conference on Machine Learning, Long Beach, California, USA, 2019: 3744-3753.
- [44] Bello I, Zoph B, Vaswani A, et al. Attention augmented convolutional networks[C] //IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 2019: 3286-3295.
- [45] Srinivas A, Lin T Y, Parmar N, et al. Bottleneck transformers for visual recognition[C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 16519-16529.
- [46] Hung W C, Kretzschmar H, Lin T Y, et al. Soda: Multi-object tracking with soft data association[DB/OL]. [2023-07-13]. <https://arxiv.org/abs/2008.07725>
- [47] Ramachandran P, Parmar N, Vaswani A, et al. Stand-alone self-attention in vision models[DB/OL]. [2023-07-13]. <https://dl.acm.org/doi/pdf/10.5555/3454287.3454294>
- [48] Tay Y, Dehghani M, Abnar S, et al. Long range arena: a benchmark for efficient transformers [DB/OL]. [2023-07-13]. <https://arxiv.org/abs/2011.04006>
- [49] Tay Y, Dehghani M, Bahri D, et al. Efficient transformers: a survey[J]. ACM Computing Surveys, 2022, 55(6): 1-28.
- [50] He K, Chen X, Xie S, et al. Masked autoencoders are scalable vision learners[C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 2022: 16000-16009.
- [51] Giuliari F, Hasan I, Cristani M, et al. Transformer networks for trajectory forecasting[C] //International conference on pattern recognition (ICPR), Montréal, QC, Canada, 2021: 10335-10342.
- [52] Zhou Z, Ye L, Wang J, et al. Hivt: hierarchical vector transformer for multi-agent motion prediction[C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 2022: 8823-8833.
- [53] Jung C, Lee D, Lee S, et al. V2X communication-aided autonomous driving: system design and experimental validation[J]. Sensors, 2020, 20(10): No. 2903.
- [54] Deng R, Di B, Song L. Cooperative collision avoidance for overtaking maneuvers in cellular V2X-based autonomous driving[J]. IEEE Transactions on Vehicular Technology, 2019, 68(5): 4434-4446.
- [55] Chang M F, Lambert J, Sangkloy P, et al. Argoverse: 3D tracking and forecasting with rich maps[C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019: 8748-8757.
- [56] Houston J, Zuidhof G, Bergamini L, et al. One thousand and one hours: self-driving motion prediction dataset[C] //Conference on Robot Learning, London, UK, 2021: 409-418.
- [57] Ettinger S, Cheng S, Caine B, et al. Large scale interactive motion forecasting for autonomous driving: the waymo open motion dataset[C] //IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 2021: 9710-9719.
- [58] Caesar H, Bankiti V, Lang A H, et al. Nuscenes: a multimodal dataset for autonomous driving[C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 2020: 11621-11631.
- [59] Dosovitskiy A, Ros G, Codevilla F, et al. CARLA: an open urban driving simulator[C] //Conference on Robot Learning, Mountain View, California, 2017: 1-16.