

基于目标锚点驱动的多模态轨迹预测方法

高镇海, 鲍明喜, 赵 睿, 唐明弘, 高 菲

(吉林大学 汽车底盘集成与仿生全国重点实验室, 长春 130022)

摘 要: 针对现有轨迹预测方法在车辆与地图交互方面的不足, 导致轨迹预测结果不符合道路拓扑结构这一缺陷, 提出了一种耦合自车运动趋势的目标锚点驱动的多模态轨迹预测方法 (TMTP)。该模型通过图模型高效地将交通场景的先验知识引入算法中, 以便能精准地描述交通场景中的异构互动关系。同时, 该模型充分考虑了动态场景图的车辆历史轨迹、自车未来轨迹和静态场景图中矢量化地图的拓扑信息的交互作用, 并通过注意力网络聚合不同节点之间的特征, 实现了更好的局部-全局之间的特征融合。此外, TMTP 将驾驶意图表征为目标锚点, 简化了意图空间的复杂性。本文在大规模 Argoverse 运动预测基准上对本文方法进行了评估, 结果表明: 本文模型相比于官方基准模型在 $\min FDE_1$ 和 $\min FDE_6$ 上分别提升 56.2%、56.6%, 可出色地完成轨迹预测任务。

关键词: 车辆工程; 轨迹预测; 目标锚点; 图神经网络

中图分类号: U463 **文献标志码:** A **文章编号:** 1671-5497(2026)01-0021-10

DOI: 10.13229/j.cnki.jdxbgxb.20240682

Multimodal trajectory prediction based on target anchor-driven

GAO Zhen-hai, BAO Ming-xi, ZHAO Rui, TANG Ming-hong, GAO Fei

(National Key Laboratory of Automotive Chassis Integration and Bionics, Jilin University, Changchun 130022, China)

Abstract: Existing trajectory prediction methods often overlook the interaction between vehicles and the map, resulting in trajectory predictions that do not conform to road topologies. To address this issue, this paper proposes a target-anchor-driven multimodal trajectory prediction method (TMTP) that couples the vehicle's motion trend. The proposed model efficiently incorporates prior knowledge of traffic scenarios into the algorithm through a graph model, allowing for precise description of heterogeneous interactions within traffic scenes. The model thoroughly considers the interaction between the vehicle's historical trajectories in the dynamic scene graph, the future trajectories of the ego vehicle, and the topological information of the vectorized map in the static scene graph. By utilizing an attention network, the model aggregates features from different nodes, achieving enhanced local-global feature fusion. Furthermore, TMTP represents driving intentions as target anchors, simplifying the complexity of the intention space. The proposed method was evaluated on the large-scale Argoverse motion forecasting benchmark. The results demonstrate that the model introduced in this paper outperforms the official benchmark model by 56.2% and

收稿日期: 2024-06-19.

基金项目: 国家自然科学基金项目 (52394261, 52202494, 52202495); 吉林省科技发展计划项目 (202302013).

作者简介: 高镇海 (1973-), 男, 教授, 博士. 研究方向: 汽车智能安全与自动驾驶. E-mail: gaozh@jlu.edu.cn

通信作者: 高菲 (1989-), 女, 副教授, 博士. 研究方向: 座舱人机交互安全. E-mail: gaofei123284123@jlu.edu.cn

56.6% in metrics $\min \text{FDE}_1$ and $\min \text{FDE}_6$, respectively, exhibiting an exemplary capability in accomplishing the task of trajectory prediction.

Key words: vehicle engineering; trajectory prediction; target anchor; graph neural networks

0 引言

轨迹预测是自动驾驶和机器人导航等领域的关键技术之一,它通过分析历史数据和实时信息来预测交通参与者在未来一段时间内的移动路径,这对于提高道路安全性和行驶效率、改善用户体验至关重要;准确的轨迹预测能够帮助系统提前识别并避免潜在的碰撞风险,同时在复杂的交通环境中做出更加合理和高效的决策。然而,轨迹预测具有固有的多模态性以及不确定性^[1-3]。尤其在高动态性和复杂性的城市工况下,单模态模型往往难以准确预测未来的轨迹,其输出结果可能趋近于各种意图的平均值,导致预测结果不准确。因此,多模态轨迹预测模型随之被广泛提出用于解决预测未来的不确定性问题,充分表示未来的预测空间^[4,5]。

近年来,受益于计算机视觉技术的发展,基于卷积神经网络 CNN^[6-8]和长短期记忆网络(LSTM)^[9-11]的深度学习模型已经在处理序列数据和捕捉时间依赖性方面取得了显著的成果。然而,这些方法主要在欧氏空间内进行操作,往往忽略了交通环境中复杂的非欧几里得结构,即交通参与者之间的相互作用和道路拓扑结构。这种简化的空间假设限制了模型对交通场景中复杂关系和动态行为的理解和表达能力。

为了克服这些限制,图神经网络^[12,13]被专门设计用于处理图结构数据,它能够直接在非欧几里得空间中建模和学习复杂的拓扑关系。然而,大多数基于图的方法的缺点是它们依赖仅基于邻域和车辆距离构建的邻接矩阵来表示交互。这种方法可能无法充分反映车辆与地图之间动态交互的真实影响。

在轨迹预测领域,如何构建周围环境的精确上下文表示是至关重要的。一些深度学习方法尝试通过将场景和代理信息栅格化为图像风格来进行环境表示,其中不同的颜色用于区分车道线、障碍物和代理等环境元素^[14-16]。然而,栅格化图像的方法存在一些固有的缺陷,最主要的是有损编码问题。为了解决栅格化表示的局限性,最

近的工作^[17-19]通过采用矢量化方法进行更紧凑的场景表示,矢量化表示使用数学模型来精确描述环境元素的几何形状和位置,如车道线、障碍物边缘和代理的轨迹点。

然而,现有基于图的方法在处理车辆交互时,往往解耦了预测和决策模块^[20,21],忽略了未来可能发生的交互对当前轨迹预测的影响。再者,大多数研究都是通过预定义的驾驶意图来建模表征未来的驾驶行为,在多变和动态的交通环境中无法对所有的驾驶行为进行表示。因此,通过引入目标锚点驱动^[22,23]的轨迹预测方法,可以显著提高模型对交通场景的理解能力,目标锚点驱动通过识别和预测每个交通参与者的未来目标位置,使模型能够更加精准地捕捉到个体的行为动机和未来动向。

综上所述,现有轨迹预测方法还存在一些问题亟待解决:①规划和预测模块的解耦导致预测轨迹未能充分考虑自车未来规划对相邻车辆的潜在约束,这可能引发不安全的驾驶行为。②现有模型未能充分模拟车辆与环境因素(如地图)之间的复杂交互,限制了对交通场景的深入理解。③预定义的驾驶意图空间维度过高,降低了模型对驾驶异常行为的敏感性,难以适应多变的道路条件。为解决上述问题,本文提出了一种耦合自车运动趋势的目标锚点驱动的多模态轨迹预测方法,简称 TM-TP。首先,该模型将静态、动态交互场景图表征为复杂的、动态变化的图结构数据,用以捕捉和建模非欧空间中节点的复杂交互关系。其次, TM-TP 综合考虑历史时域和未来时域的车辆之间以及车辆和地图之间的交互。模型利用注意力聚合节点的特征信息,并生成潜在目标锚点候选集,进而生成不同模式下的轨迹。

1 运动预测理论阐述

本文模型架构如图 1 所示。该模型遵循目标锚点驱动这一特性,通过信息融合,使模型能更全面地捕捉和建模静态场景图和动态场景图的交互作用信息。

本文的目标是给定车辆的过去运动状态、自

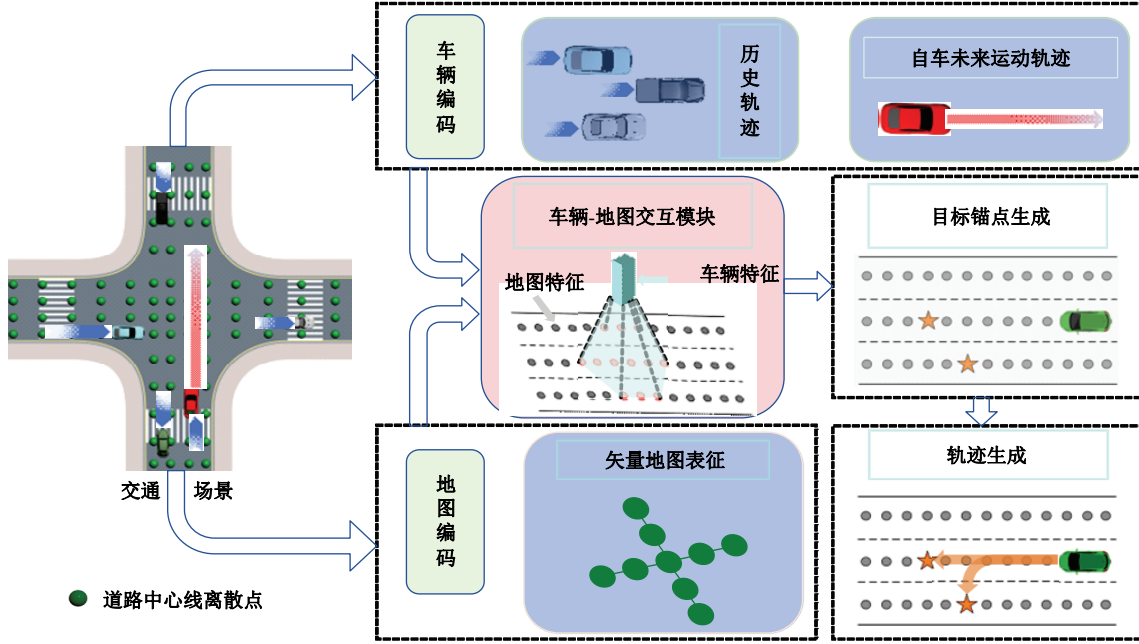


图 1 TMTP 架构

Fig. 1 TMTP architecture

车的运动趋势以及交通场景图来预测场景中目标车辆的未来轨迹以及每条轨迹相对应的概率。与现有的工作不同,本文聚焦于聚合动态场景中车辆的特征信息以及静态场景图中的地图特征信息去完成不同的任务。基于这种图结构表示方法,本文中图的节点信息不仅考虑了车辆的运动信息,还考虑了拓扑地图的远程特征信息,使得特征向量更具结构化 and 表现力,能够出色完成轨迹预测任务。

交互场景图是由静态道路和动态车辆两种不同类型的实体(即节点)和关系(即边)组成的异质图。异质图被定义为 $\mathcal{G} = \{\mathcal{U}, \mathcal{E}\}$, 其中 $\mathcal{U}$ 和 $\mathcal{E}$ 分别代表节点集合和边集合。每个节点 $u \in \mathcal{U}$ 和每条边 $e \in \mathcal{E}$ 都与它们的映射函数 $\phi(u): \mathcal{U} \rightarrow \mathcal{U}$ 和 $\phi(e): \mathcal{E} \rightarrow \mathcal{E}$ 相关联。 $\mathcal{U}$ 、 $\mathcal{E}$ 分别表示节点类型集合和边类型集合,其中 $|\mathcal{U}| + |\mathcal{E}| > 2$ 。边类型 $e: u \rightarrow v$ 表示源节点 u 指向目标节点 v 的关系。

动态交互场景图包括感兴趣区域中的所有车辆。车辆信息集 $\mathcal{X}$ 包括自车 X_0 、目标车辆 X_1 和其他所有感知的车辆 $\{X_2, X_3, \dots, X_{NS}\}$, 其中 NS 是其他感知智能体的数量。需要注意的是,本文一致编码局部视图中元素之间的相对关系,而不是处理绝对坐标。局部坐标的原点是目标车辆处于 T_{obs} 的位置。针对目标车辆,本文设定其历史轨迹信息:

$$X_i^{tar} = \{(\Delta x_i^{tar}, \Delta y_i^{tar}) \in R^2 | i = 1, 2, \dots, T_{obs}\} \quad (1)$$

式中: X_i^{tar} 为目标车辆在可观测时域内 T_{obs} 的历史轨迹; Δx_i^{tar} 和 Δy_i^{tar} 分别为鸟瞰视角下的 2D 坐标(纵向和横向坐标)。x 轴的正方向定义为自车的驱动方向, y 轴为 x 轴逆时针旋转 90° 的方向。

设定对接近于目标车辆的其他障碍车辆的运动信息:

$$X_i^{nbr} = \{(\Delta x_i^{nbr}, \Delta y_i^{nbr}) \in R^2 | i = 1, 2, \dots, T_{obs}\} \quad (2)$$

式中: X_i^{nbr} 为临近车辆的历史轨迹信息,其组成部分和目标车辆一样。

此外,本文考虑了自车运动趋势是影响目标车辆未来驾驶行为的主要因素之一。

$$X_{T_{obs}+i}^{ego} = \{(\Delta x_{T_{obs}+i}^{ego}, \Delta y_{T_{obs}+i}^{ego}) \in R^2 | i = 1, 2, \dots, T_{pred}\} \quad (3)$$

式中: $X_{T_{obs}+i}^{ego}$ 为自车在预测时域 T_{pred} 的轨迹坐标信息。本文采用数据集中自车后 3 s 的轨迹数据作为自车的未来轨迹。

场景信息集 $\mathcal{M}$ 包含矢量地图的节点信息 $\mathcal{M}_c$ 以及邻接矩阵集 $\mathcal{S}_c$ 。地图的节点信息是对道路中心线均匀采样而来。 $\mathcal{M}_c$ 、 $\mathcal{S}_c$ 分别表示了车道的几何特征以及道路之间的连接关系:

$$\mathcal{M}_c = \{(\Delta x_l, \Delta y_l, u_l^{turn}, u_l^{control}, u_l^{intersection}) \in \mathcal{U} | l = 1, 2, \dots, N_l^c\} \quad (4)$$

$$\mathcal{S}_c = \{(\mathcal{S}_l, \mathcal{S}_r, \mathcal{S}_p^{1:6}, \mathcal{S}_s^{1:6}) \in \mathcal{E}\} \quad (5)$$

式中: $\Delta x_l, \Delta y_l$ 为地图中第 l 个节点相对于目标车辆的相对位置; $u_l^{\text{turn}}, u_l^{\text{control}}, u_l^{\text{intersection}}$ 分别为第 l 个节点左转、右转时交通灯的信号, 以及是否在十字路口; N_l 为矢量地图节点的数量; $\mathcal{S}_l, \mathcal{S}_r$ 分别表明了当前节点和左侧节点、右侧节点的连接关系; $\mathcal{S}_p^{1:6}, \mathcal{S}_s^{1:6}$ 分别表明了当前节点和前后 6 阶邻居节点的连接关系。

预测模型的最终任务是输出预测时域中目标车辆的未来轨迹:

$$\begin{cases} p(y|\mathcal{X}, \mathcal{M}) \approx \sum_{g_i \in \mathcal{G}} p(y|g_i, \mathcal{X}, \mathcal{M}) p(g_i|\mathcal{X}, \mathcal{M}) \\ p(y|\mathcal{X}, \mathcal{M}) = \{(\Delta x_{T_{\text{obs}+i}}^{\text{tar}}, \Delta y_{T_{\text{obs}+i}}^{\text{tar}}, \sigma_{x, T_{\text{obs}+i}}^{\text{tar}}, \sigma_{y, T_{\text{obs}+i}}^{\text{tar}}, \delta_{T_{\text{obs}+i}}^{\text{tar}}) \in R^2 | i=1, 2, \dots, T_{\text{pred}}\} \end{cases} \quad (6)$$

式中: 模型的输出是预测时域 T_{pred} 内的二元高斯分布; $\Delta x_{T_{\text{obs}+i}}^{\text{tar}}, \Delta y_{T_{\text{obs}+i}}^{\text{tar}}$ 分别为 x 轴、 y 轴的均值; $\sigma_{x, T_{\text{obs}+i}}^{\text{tar}}, \sigma_{y, T_{\text{obs}+i}}^{\text{tar}}, \delta_{T_{\text{obs}+i}}^{\text{tar}}$ 分别为方差和相关系数; $\mathcal{G}$ 为一组有限的候选目标; $p(g_i|\mathcal{X}, \mathcal{M})$ 为目标锚点预测; $p(y|g_i, \mathcal{X}, \mathcal{M})$ 为以目标锚点、静态、动态交互场景图为条件的轨迹生成。上述所有轨迹信息均从被预测车辆的视角出发进行构建。

2 TMTP 模型

本节将详细介绍 TMTP 模型, 该模型通过整合自车的运动趋势以及对静态和动态交互场景的深入理解, 实现对车辆未来运动的精准预测。

2.1 车辆编码模块

本节旨在介绍一种多尺度特征融合的车辆编码模块, 该模块由卷积残差网络 (Residual neural network, ResNet) 和循环特征金字塔网络 (Recurrent feature pyramid networks, RFPN) 组成, 该模块的目的是捕获目标车辆在密集交通中与其他因素之间的复杂时间和空间相关性, 以实现密集交通中精确的轨迹预测。卷积残差网络和循环金字塔网络分别如图 2、图 3 所示。ResNet 有效解决了深层网络训练过程中的梯度消失和梯度爆炸问题。RFPN 聚合不同尺度的特征, 即从粗到细地捕捉图像的层次结构信息。该模块同时作用于其他车辆的历史轨迹信息和自车的未来轨迹信息。在卷积残差网络中, 每一次卷积过后均进行组归一化和 ReLU, 共同增强网络的稳定性和表达能力。

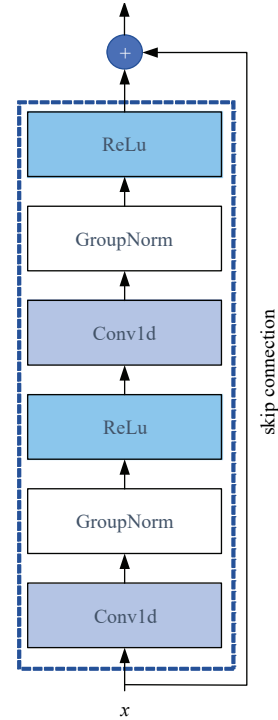


图 2 卷积残差网络

Fig. 2 Convolutional residual network

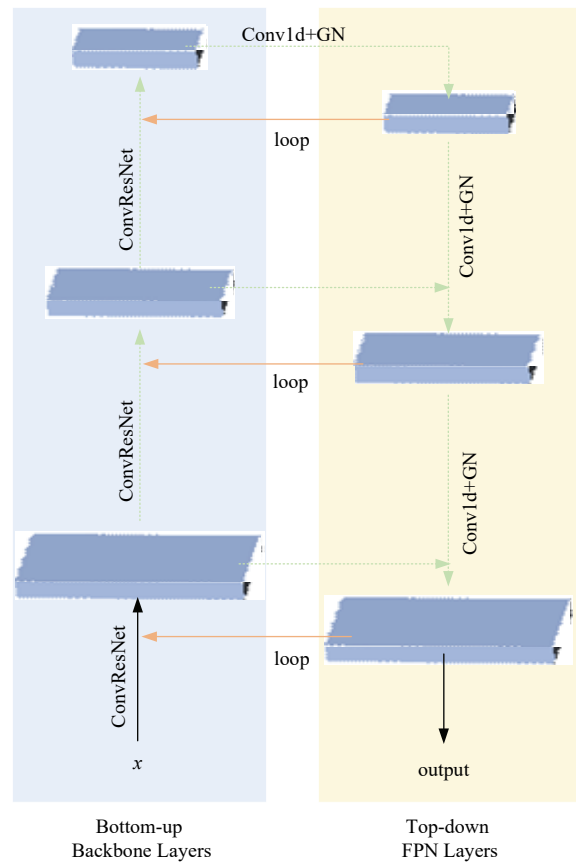


图 3 循环特征金字塔网络

Fig. 3 Recurrent feature pyramid networks

2.2 矢量地图特征编码

本文通过矢量地图数据获取交通环境的几何信息和语义信息。具体来说,道路几何信息是由地图数据中车道段的一系列中心点组成。然而,车辆除了按照车道中心线行驶外,还需要将交通信息反馈到地图的节点中。本文将车道节点定义为中心线任意两个连续点形成的直线段。车道节点的位置是前后端点的平均值,这和 LaneGCN^[19]采取一样的操作。本文对地图节点进行特征编码,对每一时间步需要考虑当前节点的位置以及前后节点所形成的方向。地图节点特征如下:

$$\hat{u}_i \leftarrow \sigma(W_{\text{node}}(\oplus [W_{\text{shape_node}}(\oplus [(x_i^{\text{end}} - x_i^{\text{start}}), (y_i^{\text{end}} - y_i^{\text{start}}), \tan(\theta_i)]) + \sigma(W_{\text{loc_node}}(\oplus [(x_i^{\text{end}} - x_i^{\text{start}}), (y_i^{\text{end}} - y_i^{\text{start}})])], u_i^{\text{turn}}, u_i^{\text{control}}, u_i^{\text{intersection}}]) \quad (7)$$

式中: $\sigma(\cdot)$ 为 LeakyReLU 非线性激活函数和 GroupNorm 归一化函数; $W_{\text{shape_node}}$ 、 $W_{\text{loc_node}}$ 分别为处理节点形状和位置对应的可学习权重参数; $(x_i^{\text{start}}, y_i^{\text{start}})$ 、 $(x_i^{\text{end}}, y_i^{\text{end}})$ 分别为第 i 个节点的开始点和结束点; θ_i 为第 i 个节点的角度; $\hat{u}_i \in U$ 为地图节点特征矩阵的第 i 行元素,表示其第 i 个节点。

本文通过感兴趣区域的节点位置索引信息,捕捉当前节点的一阶邻居信息,包括前后左右,以及 n 阶前、后邻居信息,并通过多尺度聚合模块集成邻居信息,使其能够捕捉远程的特征信息。

$$\hat{U} \leftarrow U W_0 + \sum_{i \in \text{left, right}} \sigma(A_i U W_i) + \sum_{c=1}^C \sigma(A_{\text{pre}}^c U W_{\text{pre}, c}) + \sigma(A_{\text{suc}}^c U W_{\text{suc}, c}) \quad (8)$$

式中: A_i 、 W_i 分别对应于第 i 个连接类型的邻接矩阵和权重矩阵; A_{left} 、 A_{right} 分别为相邻车道和本车道之间的信息流动,由于本文构建了多个尺度下源节点到目标节点的关系,可使得高阶邻居信息聚合到本节点上; C 为扩张多尺度层的大小,表示依次从前、后扩张 6 次的邻居节点信息聚合到当前节点; U 、 $\hat{U}$ 分别为原节点特征矩阵和经过线性残差层后该节点的输出特征,该特征既保留了原始特征信息,又包含了多阶邻居特征信息。

2.3 车辆-地图交互模块

2.2 节获取了新的节点表示,每个节点都包含了自身信息和相关节点的特征信息。然而,并不是所有的地图节点信息都与车辆相关,不相关

的特征会使神经网络学习变得更加困难。因此,本文需要了解车辆和地图节点之间的连接关系。由于本文在历史轨迹信息基础上还加入了自车未来轨迹的约束作用,此连接关系包含每一时间步下车辆的历史轨迹信息和自车未来轨迹信息,对应于一定阈值内地图的节点关系。因此,需要融合地图的先验信息和多维度下的车辆运动信息。本文使用空间注意力机制有目的地选择需要关注的元素,进而聚合智能体和地图的节点信息。本文通过欧几里得距离来选择车辆和地图节点的感兴趣特征集,这表明距离较近的节点对会有更大的影响,而距离较远的节点对的影响会减小。

$$S_D = \{(x_i, x_j, d_i - d_j) \mid \|d_i - d_j\|_2 < \epsilon_{\text{th}}, \forall i \in \{1, 2, \dots, |U_{\text{agent}}|\}, \forall j \in \{1, 2, \dots, |U_{\text{CGLL}}|\}\} \quad (9)$$

式中: ϵ_{th} 为感兴趣区域范围的阈值; d_i 、 d_j 分别为 x_i 、 x_j 在局部坐标系中的坐标。

$$U^{\text{att}} = \phi(U_{\text{agent}} + (\phi(W_{\text{ROI}} U_{\text{ROI}}) + \phi(W_{\gamma} (\oplus [U_{\text{agent}}, U_{\text{CGLL}}, \phi(W_{\text{dist}}(d_i - d_j))])))) \quad (10)$$

式中: U_{agent} 为智能体原始特征; U^{att} 为经过空间注意力的输出特征; ϕ 为 ReLU 激活函数。

2.4 目标锚点集成模块

本节旨在预测未来轨迹分布的目标锚点,未来轨迹的运动模式可以由不同的目标锚点表征。由于车辆总是按照道路中心线行驶,因此,本文从地图节点信息(即道路中心线)中选择未来的目标锚点。目标锚点示意图如图 4 所示,首先生成锚点,即预测车辆在未来可能到达的关键位置,然后将这些关键位置作为重要的先验信息。利用这些锚点,该模型能够更精确地拟合车辆到达每个潜在位置的轨迹。与传统方法相比,这种方法除了车道保持、左转、右转等简单的预定义意图外,还

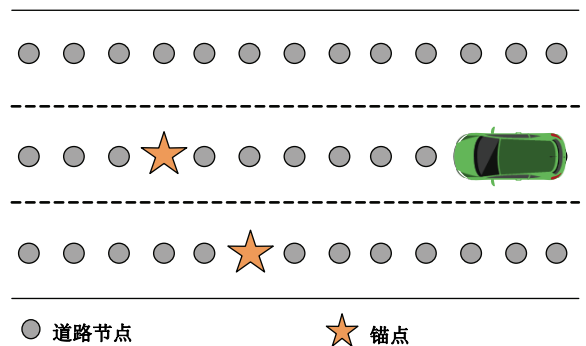


图 4 目标锚点示意图

Fig. 4 Anchor diagram

能更深入地捕捉到个体的驾驶行为动机和未来动向,从而提供更为精确和细致的预测结果。

目标锚点集成模块主要分为:目标锚点分类模块以及目标锚点生成模块。目标锚点分类模块目的是对地图的节点信息进行从低到高的重要性排序,该模块主要是通过MLP网络和Sigmoid将地图节点输出分数映射为0-1。目标锚点生成模块是基于分类模块获得较高分数的节点特征、位置以及预测分数通过多头注意力模块进一步捕捉到局部特征和长距离依赖关系。

多头注意力模块用于捕捉不同子空间下的信息,以增强模型对目标车辆未来轨迹的预测能力。多头注意力模块如图5所示。针对多头注意力层,本文将目标锚点的分数、坐标以及特征组合在一起表示为一组查询 Q 、一组键 K 、一组值 V ,并通过缩放点积注意力层去挖掘目标锚点集之间的依赖关系。单个缩放点积注意力如式(11)所示:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\underbrace{\frac{QK^T}{\sqrt{d_k}}}_{\text{attention matrix}}\right)V \quad (11)$$

式中: QK^T 为目标锚点之间的相似度得分,衡量它们之间的匹配程度; d_k 为维度缩放因子。

$\text{Softmax}(\cdot)$ 将 $\frac{QK^T}{\sqrt{d_k}}$ 转换为概率分布,来获取目标

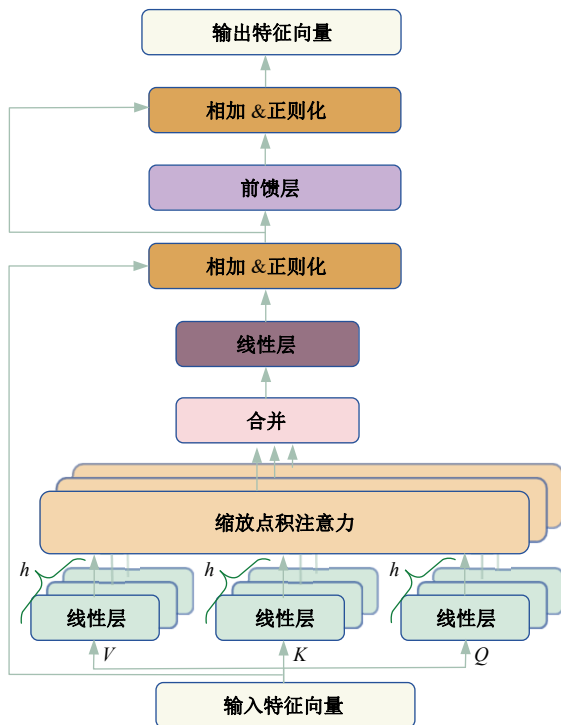


图5 多头注意力模块

Fig. 5 Multi-head attention module

锚点之间的注意力权重。多头注意力并行执行缩放的点积注意力函数 h 次,每个注意力层均采用残差连接缓解梯度消失和提高模型的性能。

多头注意力公式如(12)所示:

$$\text{MultiHead}(Q, K, V) = (\oplus(\text{head}_1, \dots, \text{head}_h))W^o$$

$$\text{where head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \quad (12)$$

3 基于Argoverse数据集的实验分析

3.1 数据集

本文在大规模Argoverse^[24]运动预测基准评估本文算法,该基准提供了车辆和矢量地图的相关信息,数据集以10 Hz进行采样,并包含了327 790个场景,每个场景包含自动驾驶车辆5 s的行驶轨迹,并分为208 272条训练集、40 127条验证集、79 391条测试集。本文算法采用 $T_{\text{obs}} = 2 \text{ s}$ 、 $T_{\text{pred}} = 3 \text{ s}$ 来预测多模态轨迹。Argoverse数据集场景图如图6所示。



图6 Argoverse数据集场景图

Fig. 6 Argoverse dataset scene diagram

3.2 实验细节

为了提高轨迹预测算法在未知场景中的有效性,本研究采用了数据增强技术。具体而言,通过对所有数据集进行随机旋转等操作,模拟了更多样化的交通场景,从而提升了模型对不同情境的适应性。所有实验均在Intel Core i5-13600KF CPU, NVIDIA GeForce RTX 4070 12 GB GPU, Ubuntu 20.04 LTS, Python 3.8, Pytorch 1.8, CUDA 11.7完成。为了适配大规模参数模型以及收敛快速,模型训练的具体超参数设置详见表1。

3.3 评价指标

本文使用最小平均位移误差(Average displacement error, minADE)、最小最终位移误差(Final displacement error, minFDE)和漏检率

表 1 超参数设置

Table 1 The specific hyperparameter settings

超参数	Argoverse
优化器	Adam
学习率	0.01
训练次数	30
Dropout	0.1
批量尺寸	12
权重衰减	3×10^{-4}

(Miss rate, MR)作为算法的评价指标。ADE为预测轨迹的每个坐标点与其对应的真值点之间的平均位移。 $\min ADE_K$ 为预测的 K 个轨迹中最小ADE, $\min FDE_K$ 是 K 个预测轨迹和真值最终位置之间的最小位移。 $MR_{2,K}$ 为地面真值端点与最佳预测端点之间的距离大于2.0 m的比率。 $K=1$ 表示预测轨迹中最可能的一条轨迹,代表模型的最佳性能。 $K=k(k>1)$ 表示前 k 个最可能的轨迹,可以更全面地评估模型的多样性。

$$\begin{cases} \min ADE_K = \min_{k \in 1, 2, \dots, K} \left(\frac{1}{NT_{\text{pred}}} \sum_{t=1}^{T_{\text{pred}}} \sum_{i=1}^N \|p_{i,t} - (\hat{p}_{i,t})_k\|_2 \right) \\ \min FDE_K = \min_{k \in 1, 2, \dots, K} \frac{1}{N} \sum_{i=1}^N \|p_{i,T_{\text{pred}}} - (\hat{p}_{i,T_{\text{pred}}})_k\|_2 \\ MR_{2,K} = \frac{1}{N} I[\|p_{i,T_{\text{pred}}} - (\hat{p}_{i,T_{\text{pred}}})_k\|_2 > 2] \end{cases} \quad (13)$$

式中: $I(\cdot)$ 为指示函数,仅 $\|p_{i,T_{\text{pred}}} - (\hat{p}_{i,T_{\text{pred}}})_k\|_2$ 大于2时等于1。

3.4 基础模型

为了验证 TMTP 算法的有效性,本文将 TMTP 与近几年的先进算法模型进行了比较。

(1) Nearest-neighbor (NN) regression (Argoverse Baseline)^[24]:该模型是 Argoverse 官方提供的基准模型。

(2) Target-driven trajectory (TNT)^[23]:通过预测目标状态、条件轨迹生成和轨迹评分选择3个阶段来实现对移动代理未来轨迹的多模态预测模型。

(3) THOMAS^[25]:通过层次化的热图解码和可学习的轨迹重组模块构成的一种联合多智能体轨迹预测框架。

(4) GoHome^[14]:一种基于图表示和稀疏投影的轨迹预测方法。

3.5 定性分析

为了评估本文模型能够捕捉轨迹的多模态关

键特征,同时让模型在场景下的表现更易被理解,本文基于 Argoverse 数据集进行了定性实验,定性实验如图7所示。图7(a)~(c)、(d)~(f)、(g)~(i)中本文分别给出了T型路口、十字路口以及其他复杂道路场景下 EVMT-GAT 生成的合理未来运动结果。绿线标记是模型给出的多个预测轨迹,实线红线表示被预测车辆未来3 s的真实轨迹,蓝色实线代表被预测车辆2 s的历史轨迹,淡粉色和深粉色分别代表自车2 s的历史轨迹和3 s的未来轨迹。每个潜在轨迹均给出对应的概率,且所有轨迹总概率和为1。从图7中可以看出, TMTP 能很好地捕捉到车道内的不同行为模式。

3.6 定量分析

为了量化模型预测的轨迹和实际轨迹之间的差异,通过一些指标对本文模型进行了定量实验。在 Argoverse 运动预测数据集中,本文进行了一项深入的比较研究,旨在评估本文模型与其他先进模型在 $\min ADE$ 和 $\min FDE$ 两个关键性能指标上的表现。为了全面评估模型的性能,本文评估当 $K=1$ 和 $K=6$ 时模型在未来3 s的 $\min ADE_K$ 、 $\min FDE_K$ 、 $MR_{2,K}$ 。基于 Argoverse 的定量实验结果如表2所示,所有评估指标数值越低,代表性能越好。相比于官方基准模型-Nearest-neighbor,当 $K=1$ 时,本文模型在 $\min ADE_1$ 、 $\min FDE_1$ 、 $MR_{2,1}$ 上分别提升了51.9%、56.2%、35.9%。当 $K=6$ 时,本文模型在 $\min ADE_6$ 、 $\min FDE_6$ 、 $MR_{2,6}$ 分别上提升了44.4%、56.6%、80.6%,这主要归功于本文融合了地图和车辆信息。与 TNT 相比,当 $K=1$ 时,本文模型在 $\min ADE_1$ 、 $\min FDE_1$ 、 $MR_{2,1}$ 上分别提升了23.69%、30.4%、21.2%。当 $K=6$ 时,本文模型在 $\min ADE_6$ 、 $\min FDE_6$ 、 $MR_{2,6}$ 上分别提升了3.5%、1.8%、6.2%,这是因为本文模型考虑了自车的未来规划信息。与 GoHome 相比,当 $K=1$ 时,本文模型在 $\min ADE_1$ 、 $\min FDE_1$ 、 $MR_{2,1}$ 上分别提升了1.7%、5.4%、2.3%。当 $K=6$ 时,本文模型在 $\min ADE_6$ 、 $\min FDE_6$ 、 $MR_{2,6}$ 上分别提升了2.8%、1.5%、0.9%,这是因为在解码器中本文采用目标锚点代替了热力图。与 THOMAS 相比,当 $K=1$ 时,本文模型在 $\min ADE_1$ 、 $\min FDE_1$ 、 $MR_{2,1}$ 上分别提升了0.5%、4.0%、0.4%。当 $K=6$ 时,本文模型在 $\min ADE_6$ 、 $\min FDE_6$ 、 $MR_{2,6}$ 上分别提升了

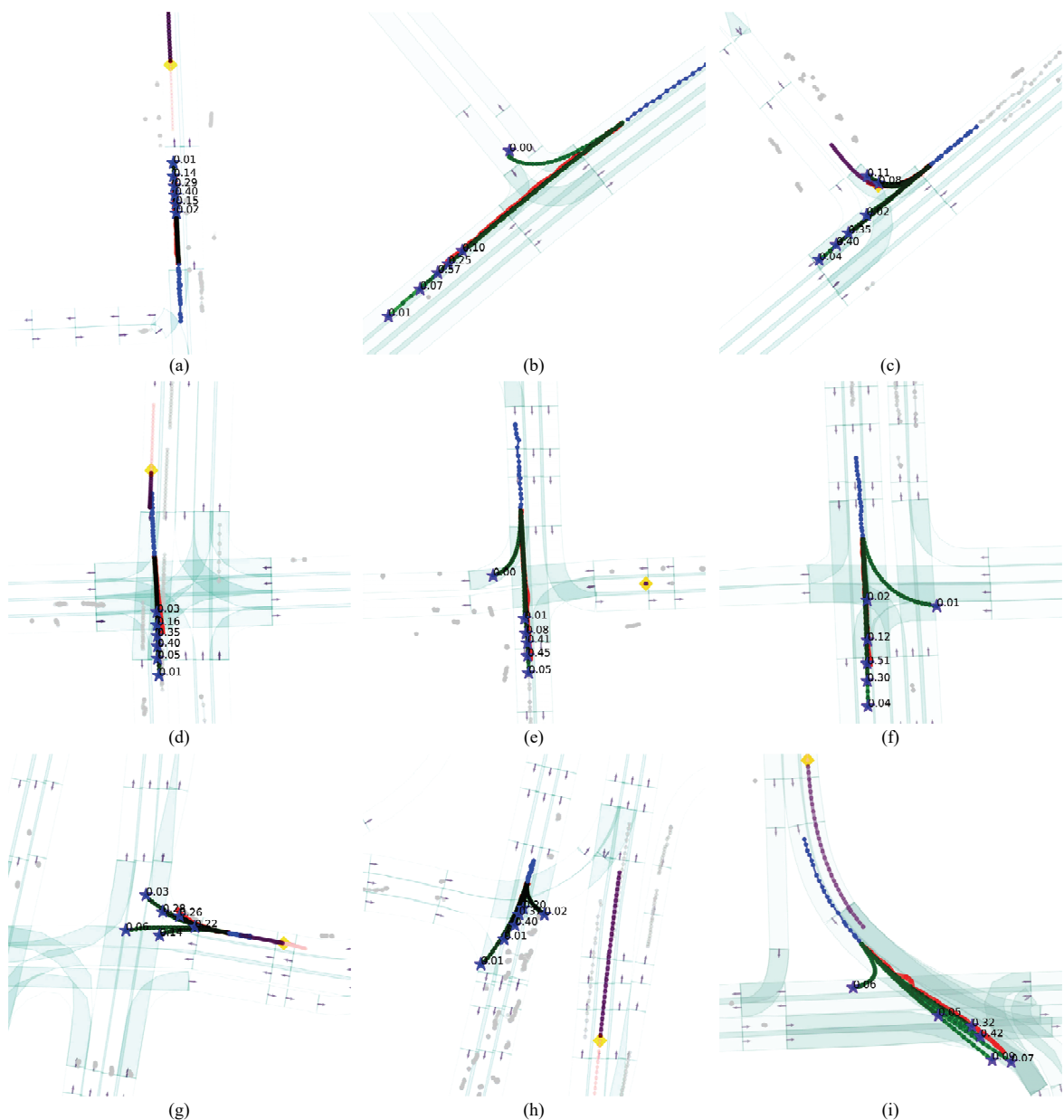


图 7 TMTP在Argoverse上的定性实验结果

Fig. 7 Qualitative test results of TMTP on Argoverse

表 2 TMTP在Argoverse数据集预测时域 3 s 的定量实验结果

模型	K = 1			K = 6		
	min ADE ₁	min FDE ₁	MR _{2,1}	min ADE ₆	min FDE ₆	MR _{2,6}
NN(baseline)	3.455	7.883	0.872	1.713	3.287	0.537
TNT	2.174	4.959	0.709	0.987	1.446	0.166
GOHome	1.689	3.647	0.572	0.979	1.450	0.105
THOMAS	1.669	3.593	0.561	0.942	1.439	0.105
本文	1.661	3.450	0.559	0.952	1.428	0.104

−0.7%、0.8%、0.9%。由于 THOMAS 对多个智能体进行联合预测,导致本文模型的 min ADE₆ 高于 THOMAS 方法,但是在其他各项指标下, TMTP 均低于其余各种方法。数值越低,表明该

模型的预测精度和多模态特性越好,越符合人类的驾驶行为。

定量分析实验结果证明,本文模型在诸多指标下的实验性能均优于列出的先进模型,展示了

巨大的未来行为预测空间,表现出了强大的多模态轨迹预测能力。

4 结束语

本文提出了一种耦合自车运动趋势目标锚点驱动的注意力网络多模态轨迹预测方法。该模型遵循编码器-解码器架构,能更全面地捕捉和建模静态场景图和动态场景图的交互作用信息。该方法将目标锚点作为先验知识,并结合地图上下文信息的车辆特征解码出目标车辆未来多模态轨迹分布。相比于官方基准模型,TMTP在 $\min ADE_1$ 和 $\min ADE_6$ 上分别提升了56.2%、56.6%。本文证明了TMTP在理解高动态性和复杂性的交通环境方面表现出色,显著提升了轨迹预测的准确性,并增强了模型的解释能力。此外,TMTP在Argoverse运动预测基准测试中表现出了先进的性能,验证了该模型在实际应用中的潜力。尽管本文的模型在多模态轨迹预测方面取得了显著的成果,但是在新的或未见过的场景中泛化能力不足。因此,未来轨迹预测的重心将集中在深化上下文信息的利用上,从而更准确地捕捉和推理社会属性和交互行为,研究意图驱动的预测模型,以提高长期预测的准确性,增强模型的泛化能力和自适应学习能力。

参考文献:

- [1] Huang Y, Du J, Yang Z, et al. A survey on trajectory-prediction methods for autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2022, 7(3): 652-674.
- [2] Gao Z, Bao M, Cui T, et al. Collision risk assessment for intelligent vehicles considering multi-dimensional uncertainties[J]. IEEE Access, 2024, 12: 57780-57795.
- [3] 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, 1162.
Gao Zhen-hai, Bao Ming-xi, Gao Fei, et al. The method of probabilistic multi-modal expected trajectory prediction based on LSTM[J]. Automotive Engineering, 2023, 45(7): 1145-1152, 1162.
- [4] Gao Z H, Bao M X, Gao F, et al. Probabilistic multi-modal expected trajectory prediction based on LSTM for autonomous driving[J]. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2023, 238(9): 2817-2828.
- [5] Jia X, Wu P, Chen L, et al. HDGT: heterogeneous driving graph transformer for multi-agent trajectory prediction via scene encoding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(11): 13860-13875.
- [6] Deo N, Trivedi M M. Convolutional social pooling for vehicle trajectory prediction[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops(CVPRW), Salt Lake City, USA, 2018: 1581-1589.
- [7] Cai Y, Wang Z, Wang H, et al. Environment-attention network for vehicle trajectory prediction[J]. IEEE Transactions on Vehicular Technology, 2021, 70(11): 11216-11227.
- [8] Zhong Z, Luo Y, Liang W. STGM: vehicle trajectory prediction based on generative model for spatial-temporal features[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 18785-18793.
- [9] Deo N, Trivedi M M. Multi-modal trajectory prediction of surrounding vehicles with maneuver based LSTMs[C]//IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 2018: 1179-1184.
- [10] Wang Y, Wang J, Jiang J, et al. SA-LSTM: a trajectory prediction model for complex off-road multi-agent systems considering situation awareness based on risk field[J]. IEEE Transactions on Vehicular Technology, 2023, 72(11): 14016-14027.
- [11] Lin L, Li W, Bi H, et al. Vehicle trajectory prediction using LSTMs with spatial-temporal attention mechanisms[J]. IEEE Intelligent Transportation Systems Magazine, 2022, 14(2): 197-208.
- [12] Wu Z, Pan S, Chen F, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.
- [13] Zhou J, Cui G, Hu S, et al. Graph neural networks: a review of methods and applications[J]. AI Open, 2020, 1(1): 57-81.
- [14] Gilles T, Sabatini S, Tsishkou D, et al. GOHOME: graph-oriented heatmap output for future motion estimation[C]//International Conference on Robotics and Automation(ICRA), Philadelphia, USA, 2022: 9107-9114.
- [15] Gilles T, Sabatini S, Tsishkou D, et al. HOME: heatmap output for future motion estimation[C]//IEEE International Intelligent Transportation Sys-

- tems Conference(ITSC), Indianapolis, USA, 2021: 500-507.
- [16] Hong J, Sapp B, Phibin 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(CVPR), Long Beach, USA, 2019: 8446-8454.
- [17] Zhang L, Li P, Chen J, et al. Trajectory prediction with graph-based dual-scale context fusion[C] // IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS), Kyoto, Japan, 2022: 11374-11381.
- [18] Gu J, Sun C, Zhao H. DenseTNT: end-to-end trajectory prediction from dense goal sets[C] //IEEE/CVF International Conference on Computer Vision (ICCV), Online, 2021: 15283-15292.
- [19] Liang M, Yang B, Hu R, et al. LaneGCN: learning lane graph representations for motion forecasting[C] // The 16th European Conference on Computer Vision, Glasgow, UK, 2020: 541-556.
- [20] Song H, Ding W, Chen Y, et al. PiP: planning-informed trajectory prediction for autonomous driving [C] // The 16th European Conference on Computer Vision, Glasgow, UK, 2020: 598-614.
- [21] Guo H, Meng Q, Cao D, et al. Vehicle trajectory prediction method coupled with ego vehicle motion trend under dual attention mechanism[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-16.
- [22] Zhang L, Su P H, Hoang J, et al. Map-adaptive goal-based trajectory prediction[J/OL]. [2024-05-26]. <https://arxiv.org/abs/2009.04450v2>.
- [23] Zhao H, Gao J, Lan T, et al. TNT: Target-driveN trajectory prediction[C] //IEEE Conference on Robot Learning, London, UK, 2021: 895-904.
- [24] Chang M F, Ramanan D, Hays J, et al. Argoverse: 3D tracking and forecasting with rich maps[C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), Long Beach, USA, 2019: 8740-8749.
- [25] Gilles T, Sabatini S, Tsishkou D, et al. THOMAS: trajectory heatmap output with learned multi-agent sampling[J/OL]. [2024-05-26]. <https://arxiv.org/abs/2110.06607>.