

文章编号: 2095-2163(2023)12-0182-05

中图分类号: TP391.41

文献标志码: A

融合坐标注意力与卷积社交池的车辆轨迹预测

刘芝妍, 闫建红, 王震

(太原师范学院 计算机科学与技术学院, 山西 晋中 030619)

摘要: 针对车辆行驶受相邻车辆的影响, 本文提出了一种特征增强的模型, 在 LSTM 编码器-解码器架构上, 使用卷积社交池提取一定范围内的车辆轨迹信息特征, 再添加坐标注意力进行特征增强。使用 NGSIM 数据集实验结果表明, 本文所提模型表现性能良好。相较于其他模型, 5 s 内在均方根误差指标上平均降低了 14.10%、15.76%、17.49%、17.81%、17.20%。

关键词: 车辆轨迹预测; 坐标注意力; 卷积社交池

Vehicle trajectory prediction integrating coordinate attention and convolutional social pool

LIU Zhiyan, YAN Jianhong, WANG Zhen

(School of Computer Science and Technology, Taiyuan Normal University, Jinzhong Shanxi 030619, China)

Abstract: Addressing the impact of neighboring vehicles on vehicle motion, this paper introduces a feature-enhanced model. Within the LSTM encoder-decoder architecture, it employs convolutional social pooling to extract vehicle trajectory information features within a specific range and further enhances these features with coordinate attention. Experimental results using the NGSIM dataset demonstrate the excellent performance of the proposed model. Compared to other models, it achieves an average reduction of 14.10%, 15.76%, 17.49%, 17.81%, and 17.20% in the root mean square error (RMSE) metric within a 5-second prediction horizon.

Key words: vehicle trajectory prediction; coordinate attention; convolutional social pool

0 引言

过去几年, 得益于计算机视觉、传感器信号处理、硬件设计等领域的技术进步, 自动驾驶已经从“可能成为可能”走向“商业化”。然而, 2018年 Tesla 和 Uber 自动驾驶汽车发生的两起交通事故, 引发了人们对自动驾驶汽车安全性的担忧。因此, 提高自动驾驶汽车上运行智能算法的性能至关重要, 如预测周围物体(车辆、行人、自行车等)的未来轨迹。作为自动驾驶的核心任务之一, 车辆轨迹预测的准确性为下游的规划控制提供保障, 如果自动驾驶汽车能够准确预测周围车辆的位置, 就可以避免此类交通事故的发生。机器学习、神经网络在各个领域的大量应用, 也使车辆轨迹预测从之前的基于物理、模型的方法向机器学习、深度学习转变。

目前现有的轨迹预测方法大致分为3类: 基于物理、基于机动和基于交互感知^[1]。S. Ammoun 等^[2]提出了一种基于物理的方法, 考虑对象的个体动力学来预测其运动, 但忽略了目标车辆可能受道路结构和邻近车辆的影响限制。M. Althoff 等^[3]提出了一种基于机动的方法, 考虑机动选项并预测以机动为条件轨迹, 但忽略了周围车辆对目标车辆的影响。

近年来, 基于交互感知的车辆轨迹预测引起了越来越多的研究者的兴趣, 基于交互感知的方法将驾驶视为一种交互活动, 与纯基于物理和基于机动的方法相比, 具有更好的性能, 且可以扩展为同时考虑物理和机动。Li 等^[4]根据预测目标建立坐标系, 使用图表示近距离对象的交互, 利用 LSTM 编码器-解码器进行轨迹预测。Mo 等^[5]提出了将车辆和基础设施信息的交互表示为异构图, 利用图神经网络

基金项目: 山西省重点研发计划(N202102010101008)。

作者简介: 刘芝妍(1998-), 女, 硕士研究生, 主要研究方向: 深度学习、人工智能; 王震(1997-), 男, 硕士研究生, 主要研究方向: 深度学习、图像处理。

通讯作者: 闫建红(1972-), 女, 博士, 教授, CCF 会员, 主要研究方向: 机器学习、网络安全。Email: yan_jian_hong@163.com

收稿日期: 2023-11-01

对轨迹的高级交互进行建模并预测。Li 等^[6]提出了一种用于多智能体轨迹预测的通用生成神经网络,通过 Wasserstein 图双注意力网络进行交互感知轨迹预测。现有的大多数交互感知的方法表示共享坐标系中所有智能体的运动,该坐标系对平移和旋转敏感,旨在预测单个智能体的轨迹。然而,自动驾驶应该同时预测多个周围主体(如车辆和行人)的未来状态,以便在复杂且高度动态的城市驾驶场景中进行导航。

随着循环神经网络在各个领域的应用,如自然语言处理、图像处理等,车辆轨迹预测也兴起了使用深度学习的热潮。Deo 等^[7-9]提出了对于高速公路上周围车辆的机动分类和轨迹预测的统一框架。首先使用 LSTM(长短时记忆网络)模型将所有观察到的汽车(目标车辆和其周围车辆)的历史轨迹和相对位置表示为上下文向量,使用上下文向量来进行机动分类,并使用 LSTM 来预测车辆的未来位置;之后,考虑到仅使用 LSTM 模型存在捕获车辆轨迹相互依赖性能很差的问题,因此在模型中添加卷积社交池化层来增强模型的预测能力,使模型可以访

问周围车辆的运动状态及空间关系。

如今,注意力机制也被广泛应用于各种模型中用来提升模型的性能。如:SEnet^[10]借助 2D 全局池化来计算通道注意力,并以相当低的计算成本提供显著的能力提升。Hou Qibin 等^[11]将位置信息嵌入到通道注意力中,提出了一种叫做坐标注意力的移动网络注意力机制,可以灵活的插入各种网络。连静等^[12]将注意力机制与图模型结合对车辆轨迹进行预测,但该模型依赖于高精度地图,且仅考虑车辆之间的交互。

针对以上问题,本文综合考虑目标车辆及周围相邻车辆的历史轨迹信息和相对空间位置等对目标车辆未来轨迹的影响,提出了融合坐标注意力的卷积社交池化网络(CSCA-LSTM),用来对车辆轨迹进行预测。

1 模型与框架

模型框架如图 1 所示,由基于 LSTM 的编码器、融合坐标注意力的卷积社交池、基于 LSTM 的解码器等部分组成。

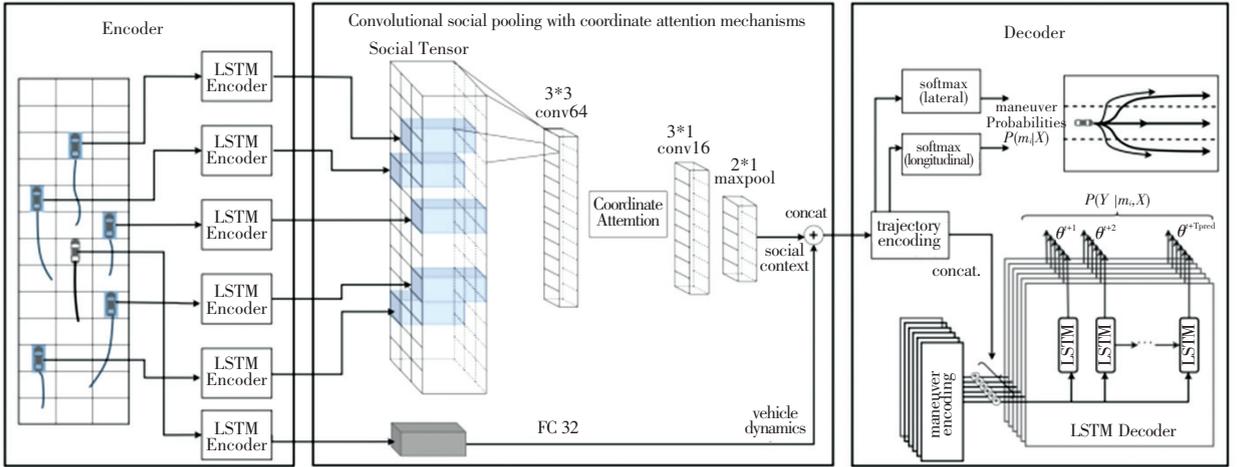


图 1 CSCA-LSTM 整体模型框架

Fig. 1 CSCA-LSTM overall model framework

1.1 LSTM 编码器

模型中使用 LSTM 编码器,是用来学习车辆运动。对于每个时刻,向 LSTM 编码器输入被预测车辆(v)及周围车辆最近 t 帧的历史轨迹和被预测车辆及周围车辆之间的相对空间位置,以及所处车道信息等。使用编码器进行编码,得到每辆车的状态张量,且该状态张量会逐帧更新。每辆车使用的 LSTM 具有共享权重,允许所有车辆的 LSTM 状态组件之间直接对应。

1.2 融合坐标注意力的卷积社交池

卷积社交池用来捕获场景中车辆之间轨迹运动的相互依赖性。

首先,通过围绕被预测的智能体定义一个 13×3 的空间网格,并根据场景中其他智能体的相对空间位置,使用其 LSTM 状态张量进行填充,形成社交张量,表示模型可以访问周围车辆的运动状态和其空间配置。在社交池化之后添加卷积层,卷积网络等方差可以帮助模型学习社交张量中的局部有用特

征,最大池化层可以增加局部平移不变性。两个卷积层和一个池化层应用于社交张量来获得社交上下文编码。被预测车辆的 LSTM 状态张量单独通过一个全连接层来获得该车的轨迹编码,将这两个编码连接起来形成完整的轨迹编码,传递给解码器。

如图 2 所示,为了使特征提取更为精准,在第一个卷积网络后添加了坐标注意力。坐标注意力 (Coordinate Attention, CA) 是一种高效注意力机制,能够获取通道间信息,而且考虑与方向相关的位置信息,具有灵活轻量的特点,能够简单的插入移动网络的核心结构中,有助于模型更好地定位和识别目标。可以把 CA 看作一个特征增强计算单元,对于输入 $X = [x_1, x_2, \dots, x_c] \in \mathbb{R}^{C \times H \times W}$, 经 CA 模块处理后输出与 X 维度相同的增强特征 $Y = [y_1, y_2, \dots, y_c]$ 。CA 模块由坐标信息嵌入和坐标信息生成两部分组成。

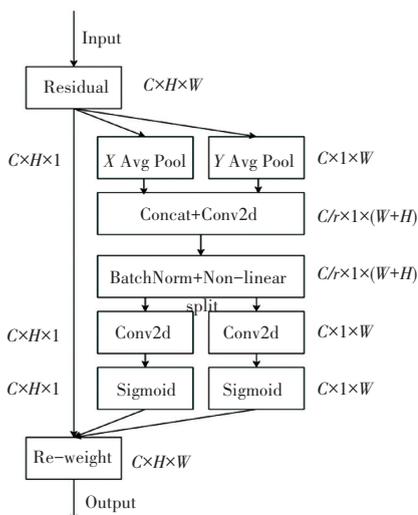


图 2 坐标注意力结构图

Fig. 2 Coordinate attention block

1.2.1 坐标信息嵌入

给定输入的特征图 X , 使用池化核 $(H, 1)$ 和 $(1, W)$ 分别沿着水平坐标和垂直坐标对每个通道进行编码。则高度为 h 和宽度为 w 的第 c 个通道的输出可表示为:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i) \quad (1)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq i \leq H} x_c(i, w) \quad (2)$$

式中: $z_c^h(h)$ 代表高度为 h 的第 c 个通道的输出, $z_c^w(w)$ 代表宽度为 w 的第 c 个通道的输出, $x_c(i, j)$ 代表输入特征图 X 的第 c 个通道在位置 (i, j) 上的值。

1.2.2 坐标信息生成

首先将公式(1)和公式(2)的输出拼接起来,使用 1×1 卷积和非线性激活函数进行特征转化,表达式如式(3)所示:

$$f = \delta(F_1([z^h, z^w])) \quad (3)$$

式中: $f \in \mathbb{R}^{C/r \times (H+W)}$ 为包含横向和纵向空间信息的中间特征, r 为缩减因子。

随后,将 f 分为两个独立的特征 $f^h \in \mathbb{R}^{C/r \times H}$ 和 $f^w \in \mathbb{R}^{C/r \times W}$, 使用另外两个 1×1 卷积 F_h 和 F_w 统一通道数,再用 Sigmoid 函数进行特征转化,使其维度与输入 X 一致,表达式如下:

$$g^h = \delta(F_h(f^h)) \quad (4)$$

$$g^w = \delta(F_w(f^w)) \quad (5)$$

最后,将输出 g^h 和 g^w 合并成权重矩阵与输入的特征图相乘,得到该注意力模块的输出,表达式如式(6)所示:

$$y_c(i, j) = x_c(i, j) \times g_c^h(j) \times g_c^w(i) \quad (6)$$

1.3 LSTM 解码器

使用 LSTM 解码器用来生成目标车辆未来轨迹的预测分布。该解码器具有两个 softmax 层,即输出车辆在横向上的换道概率和纵向上的机动概率。根据换道信息得到目标车辆未来轨迹的多模态预测分布。

2 实验结果与分析

2.1 数据集

本文使用公开的 NGSIM US-101 和 I-80 数据集进行实验。每个数据集包含在 45 min 的时间跨度内以 10 Hz 捕获的真实高速公路交通轨迹。每个数据集由 15 min 的轻度、中度和拥堵交通状况片段组成。将完整的数据集分为 7 : 1 : 2 的训练集、验证集和测试集;将轨迹分成 8 s 的片段,其中使用 3 s 历史轨迹和 5 s 的预测范围。在将每个片段下采样 2 倍后,再输入 LSTM 编码器,以降低模型复杂性。

2.2 评价指标

使用反映预测轨迹与真实轨迹之间均方根误差 (RMSE) 作为评价指标,值越小误差越小,值越大,误差越大。产生二元正态分布时,使用均值来计算 RMSE,生成多模态分布时,使用概率权值最高的模态来计算 RMSE。计算公式如式(7)所示。

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_t - \hat{x}_t)^2 + (y_t - \hat{y}_t)^2} \quad (7)$$

其中, (\hat{x}_t, \hat{y}_t) 和 (x_t, y_t) 分别表示 t 时刻目标

车辆 i 的真实位置和预测位置。

2.3 实验环境及参数设置

本文模型使用 Pytorch 实现, 具体硬件设施见表 1。使用 Adam 优化器训练模型, 学习率设置为 0.001, 编码器维度为 64, 解码器维度为 128, 对所有层使用 $\alpha = 0.1$ 的 leaky-ReLU 函数激活。

表 1 具体硬件信息

Table 1 Hardware information

软硬件名称	详细信息
CPU	13th Gen Intel(R) Core(TM) i9-13900HX 2.20 GHz
GPU	NVIDIA GeForce RTX4060
内存大小	32.0 GB
显存大小	16.0 GB
操作系统	Windows11
编程语言	Python 3.8.15
CUDA 版本	11.3
深度学习框架	Pytorch 1.11.0

2.4 对比实验

本文使用 3 s 的历史轨迹预测 5 s 的未来轨迹。为验证模型的有效性, 首先与原模型 CS-LSTM 进行比较, 数据集、参数等均与原文保持一致, 进行 5 轮预训练和 3 轮正式训练。损失值对比、纬度坐标预测准确率、经度坐标预测准确率如图 3、表 2、表 3 所示。

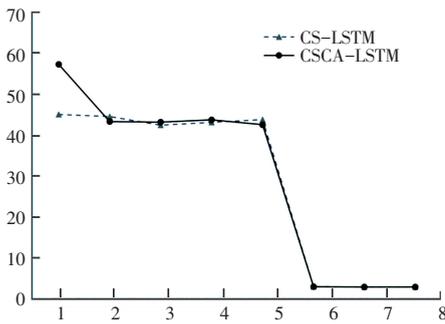


图 3 训练损失

Fig. 3 Training loss

图 3 表明, 添加了坐标注意力的模型收敛性更好、速度更快, 这也意味着 CSCA-LSTM 模型具有更好的鲁棒性。

表 2 纬度坐标预测准确率

Table 2 Prediction accuracy of latitude coordinates

模型	正式迭代次数		
	1	2	3
CS-LSTM	98.070 8	98.136 5	98.161 3
CSCA-LSTM	98.063 8	98.089 6	98.140 1
	-0.01	-0.05	-0.02

表 3 经度坐标预测准确率

Table 3 Prediction accuracy of longitude coordinate

模型	正式迭代次数		
	1	2	3
CS-LSTM	89.716 4	89.603 4	89.642 1
CSCA-LSTM	89.850 5	89.844 3	89.992 5
	0.15	0.27	0.39

表 2、表 3 显示相较于 CS-LSTM 和 CSCA-LSTM 模型纬度坐标预测准确率分别降低了 0.01、0.05、0.02; 经度坐标预测准确率分别提升了 0.15、0.27、0.39。虽然, 纬度预测准确率下降了, 但经度坐标预测准确率提升更多, 表明本文所提模型具有提升经度坐标预测准确率的能力。为使模型达到最优, 经调整正式训练迭代次数实验表明, 最优正式训练迭代次数为 11 次。

与 C-VGMM + VIM^[8]、CS-LSTM^[9]、LSTM、MATF Multi Agent^[13] 等多个轨迹预测模型做对比, 均方根误差 RMSE 结果见表 4 (标黑为最佳结果)。

其中, CS-LSTM 使用共享信息的卷积社交池; C-VGMM+VIM 结合变分高斯混合模型和隐马尔可夫模型预测车辆轨迹框架; LSTM 仅使用长短时记忆网路; MATF Multi Agent 是应用聚合函数的一种模型。

表 4 均方根误差对比

Table 4 Comparison of root mean square error

模型	1 s	2 s	3 s	4 s	5 s
C-VGMM+VIM	0.66	1.56	2.75	4.24	5.99
CS-LSTM	0.62	1.29	2.13	3.20	4.52
LSTM	0.66	1.62	2.94	4.63	6.63
MATF Multi Agent	0.67	1.51	2.51	3.71	5.12
CSCA-LSTM	0.56	1.25	2.10	3.18	4.51

从表 4 中可以看出, 本文所提模型在每一秒的预测位置都相较于其他模型更准确。相较于 C-VGMM+VIM 每一秒的均方根误差降低了 15.15%、19.87%、23.64%、25.00%、24.71%; 相较于 CS-LSTM 每一秒的均方根误差降低了 9.68%、3.10%、1.41%、0.63%、0.22%; 相较于 LSTM 每一秒的均方根误差降低了 15.15%、22.84%、28.57%、31.32%、31.98%; 相较于 MATF Multi Agent 每一秒的均方根误差降低了 16.42%、17.22%、16.33%、14.29%、11.91%。实验结果表明, 本文所提模型是有效的, 整体性能优于其他模型。