

时空图注意力网络在交叉口车辆轨迹预测的应用

曾伟良 陈漪皓 姚若愚 廖睿翔 孙为军

广东工业大学自动化学院 广州 510006

(weiliangzeng@gdut.edu.cn)

摘要 随着人工智能和大数据技术的快速发展,以深度学习为代表的自动驾驶轨迹预测是未来的热点研究方向。在混合交通场景下,如何准确地预测机动车与非机动车的轨迹,是实现自动驾驶技术中安全行驶和高效轨迹规划等问题的前提。针对交叉路口中不同运动对象之间发生交互时的轨迹预测问题,提出了基于图注意力网络的建模方案。所采用的模型结合了时间与空间上研究对象之间的相互作用,对机动车与非机动车的未来轨迹做出了更准确的预测,可应用于自动驾驶的轨迹规划方案,确保在复杂交通场景下,机动车与非机动车能够安全且高效地通过交叉路口。该模型在简单交互情况下,可取得较小的轨迹平均位移误差和最终位移误差,而在复杂交互情况下,可提供更为合理的规划路径。

关键词: 自动驾驶;轨迹预测;深度学习;图注意力网络

中图分类号 TP391

Application of Spatial-Temporal Graph Attention Networks in Trajectory Prediction for Vehicles at Intersections

ZENG Wei-liang, CHEN Yi-hao, YAO Ruo-yu, LIAO Rui-xiang and SUN Wei-jun

School of Automation, Guangdong University of Technology, Guangzhou 510006, China

Abstract With the rapid development of artificial intelligence and big data technology, the application of deep learning to trajectory prediction on autonomous driving has become a hot topic in recent years. The premise of keeping the safety of navigation and efficient path planning in autonomous driving is to make trajectory predictions of motor and non-motor vehicles accurate, especially in mixed traffic scenes. Aiming at tackling the problem related to path planning when interactions happen among different research objects at an intersection, a modelling scheme based on graph attention networks is proposed. The model applied combines spatial and temporal interactions among traffic-agents to improve the accuracy of trajectory predictions for motor and non-motor vehicles. In addition, the proposed model can be applied to path planning of autonomous driving, ensuring that motor and non-motor vehicles are capable of passing the interactions safely and efficiently in complex traffic scenes. In the case of simple interactions, the average displacement error and final displacement error of the trajectories derived from the model reach relatively small. And in the case of complex interactions, the future paths provided by the model are more reasonable than the future ground truths.

Keywords Autonomous driving, Trajectory prediction, Deep learning, Graph Attention Networks

1 引言

人工智能逐渐走进了人们的视野,例如机器翻译^[1-2]、语音识别^[3-4]、图像识别^[5]等,给人们的生活带来了空前的便利。在交通领域,人工智能也带来了许多崭新的应用,如地图匹配、车牌识别、警用机器人等^[6]。其中,最令人瞩目和期待的应用之一是自动驾驶^[7]。自动驾驶所要实现的最终目标是车

辆对周围环境进行感知后,高速运算轨迹规划并做出决策,从而能够在复杂的道路环境下自主安全地行驶。传统车辆的驾驶员能够通过自身对周边环境信息的收集,在短时间内进行判断,从而采取进一步的行驶决策。但对于在道路上行驶的自动驾驶车辆来说,像人类一样在极短时间内对周边车辆的运动行为和未来行驶轨迹趋势进行预知是非常困难的。

城市道路交叉口是城市道路的重要连接点,但在交叉口

基金项目:国家自然科学基金(61803100,U1911401);广东省智能交通系统重点实验室开放基金(202005003);广东省科技计划(2019B010121001,2019B010118001,2019B01019001);工信部工业互联网创新发展工程(TC190A3X9-2-2);国家重点研发计划(2018YFB1802400)

This work was supported by the National Natural Science Foundation of China(61803100,U1911401),Guangdong Provincial Key Laboratory of Intelligent Transportation System(202005003),Science and Technology Planning Project of Guangdong Province,China(2019B010121001,2019B010118001,2019B01019001),Industrial Internet Innovation and Development Project of MIIT(TC190A3X9-2-2)and National Key Research and Development Project(2018YFB1802400).

通信作者:孙为军(14341569@qq.com)

处存在着严重的混合交通现象。各种机动车交通流与非机动车交通流特性各异,这不但影响了交叉口的运行效率,而且极易诱发交通事故。状况复杂的城市道路,特别是车流量较大的交叉路口环境,为自动驾驶的实现带来了众多挑战。车辆周边运动对象的种类繁多、交通对象形状及运动方式的异质性都会对机动车与非机动车的行驶轨迹规划造成严重干扰。基于这个问题,本研究将通过采集城市道路环境下机动车与非机动车的真实轨迹坐标信息,对所研究车辆的历史运动信息及运动方式进行分析,并结合周边环境,改进现有轨迹预测算法,对机动车与非机动车未来的行驶轨迹提供合理的规划方案。

2 文献回顾

2.1 手工构造函数

针对交通领域中的轨迹预测问题,科学家们已提出了许多的方法,较为经典的方法有:1)社会力(Social Force),将运动对象与环境之间的相互作用量化为吸引力和斥力,然后利用这些力的合力驱动研究对象运动^[8];2)高斯过程(Gaussian Process),通过连续域(时间或空间)上随机变量的联合,预测研究对象的速度、加速度、角偏量等参数^[9];3)卡尔曼滤波(Kalman Filter),卡尔曼滤波器在预测点附近对历史轨迹进行线性化,并使用最小方差估计^[10]。

上述经典方法对于分析交通对象的轨迹预测问题做出了巨大贡献,也为自动驾驶领域奠定了良好的基础。但是此类方法主要存在以下不足之处:1)模型需要手工构造函数以刻画交互,而不是通过数据驱动的方式推断并生成未来轨迹。这导致了模型的适应性较差,只能适用于简单的交互情景,在较为复杂的场景中表现欠佳,对于分析长期轨迹预测问题稍有欠缺。2)模型通常针对的是在研究对象周围较短距离内发生的交互情景,而没有考虑更远距离运动对象与其他个体可能发生的交互,对周边环境信息的考虑并不周全^[11]。

2.2 序列-序列预测模型

针对传统方法的不足,一些学者开始提出利用深度学习技术,基于序列-序列(Sequence-Sequence, Seq2Seq)预测结构,对车辆的运动轨迹进行预测^[12]。1)社会长短期记忆模型(Social LSTM):采用了在序列预测上表现突出的长短期记忆网络,模型在LSTMs的每一步加入了池化层,整合了当前预测对象的“邻居”的隐藏状态,并将其作为隐藏状态信息的一部分传递下去,解决了多个研究对象的LSTMs之间无法捕捉空间中的交互的问题^[11]。2)社会生成对抗网络(Social-GAN):结合了序列预测和以生成器-判别器为框架的生成对抗网络,将全局的对象信息纳入考虑并进行池化,采用博弈训练去对抗一个递归判别器,从而生成可行的未来轨迹。此外,利用新的损失函数 Variety Loss 可生成多条可行的路径^[13-14]。3)SoPhie:结合了社会注意力机制和物理注意力机制,从周围对象中提取出重要的轨迹信息,建模捕获未来路径的不确定性,并利用生成对抗网络生成更真实的样本^[15]。4)Multi-Agent Tensor Fusion:将多个智能体的过去轨迹和场景信息编码为张量,应用卷积组合捕获运动对象的交互,同时保留了运动对象和场景的空间区域结构,最后采用对抗性损失学习随机预测,递归地解码生成未来轨迹^[16]。

2.3 注意力机制

注意力模型最初被用于机器翻译,现在已成为神经网络领域的一个重要概念。在人工智能领域,注意力机制已成为神经网络结构的重要组成部分,并在自然语言处理、统计学习、语音和计算机等领域有着大量的应用^[17]。注意力机制在交通领域中的应用可以利用人类视觉机制进行直观解释。例如,当人们处于一个拥挤场景中,相比较于建筑物本身或人们身后的人与事物,人们会更加关注眼前的障碍物以及即将向他们走来的人^[18]。因此,注意力机制根据信息的重要性,为历史信息分配不同的注意力权重,有助于克服循环神经网络(Recurrent Neural Network, RNN)中的一些缺点和限制,如随着输入长度的增加性能下降,以及输入顺序不合理导致的计算效率低下。

2.4 图网络

图神经网络(Graph Neural Network, GNN)是一种直接运行在图结构上的神经网络,在与图网络相关的分析中表现良好^[19]。基于GNN的图卷积网络(Graph Convolutional Network, GCN)有着广泛的应用,可以应用于任何在图上有卷积的网络拓扑图^[20-22]。近些年来,基于图神经网络的方法在轨迹预测领域取得优秀表现。基于这一类网络,图注意力网络(Graph Attention Network, GAT)被提出,其对邻域内不同节点分配不同的权重,而不需要进行任何代价高昂且复杂的矩阵运算,也不需要预先知道图形结构^[23-24]。

因此,受上述网络和方法论的启发,本研究提出了一种结合了图注意力网络的模型,基于Seq2Seq结构,通过挖掘时间图和空间图中的隐藏信息,获取机动车和非机动车之间关键的交互和冲突信息。本研究所采用的模型能够结合所有有用信息,产生较为理想的预测结果。通过对轨迹信息进行计算,本方法预期在混合交通场景下,将预测轨迹与未来真实轨迹之间的误差最小化。而在冲突行为较为复杂的交互场景下,本方法可辅助所研究的机动车和非机动车,为其提供更安全、高效的轨迹规划方案。

3 基于时空图注意力网络的轨迹预测

3.1 问题陈述

在第 t 个时间步下,第 i 辆车的横纵坐标记为 (x_i^t, y_i^t) 。在观察时间段1至 t_{obs} 内每辆车的历史运动轨迹及其与周围车辆的交互后,本研究旨在对时间段 t_{obs+1} 至 $t_{obs+pred}$ 内车辆的未来运动轨迹进行预测。其中, obs 和 $pred$ 表示观察点和预测点的数量。

3.2 模型框架

模型总体框架如图1所示。车辆的轨迹预测问题本质上是一个序列预测问题,考虑到Encoder-Decoder(编码-解码)框架在处理序列预测问题上取得了优秀表现^[25],本文方法基于该框架,将轨迹预测问题转化为序列预测问题。对于每个车辆个体,其未来的运动状态、行驶轨迹不仅与驾驶者本身、车辆的历史运动信息等因素有关,还与其周围的车辆对其产生的交互影响有关。因此,本文方法引入图注意力网络GAT对时间及空间上的交互信息进行捕捉并量化,作为模型计算过程的一部分。

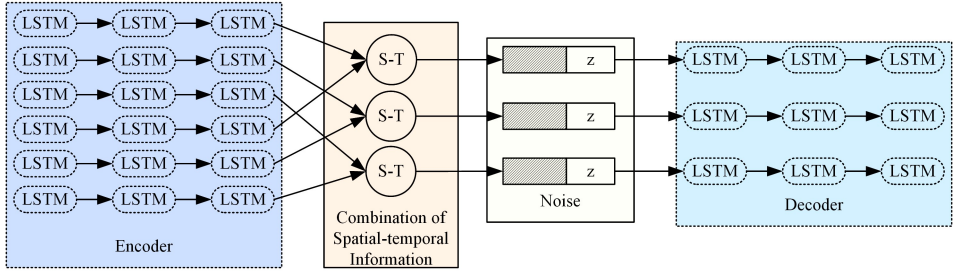


图1 模型框架

Fig. 1 Framework of model

3.3 相对位置编码

每个机动车或非机动车个体具有其独特的运动方式。而 LSTM 网络在捕捉、收集研究对象个体的历史运动轨迹表现优秀^[26], 因此, 我们使用 LSTM 单元代表每一个所研究的车辆个体, 从而获得其运动状态的序列化表示。

首先, 需要对第 i 辆车对于前一个时间步的相对位置信息作出定义:

$$\Delta x_i^t = x_i^t - x_i^{t-1} \quad (1)$$

$$\Delta y_i^t = y_i^t - y_i^{t-1} \quad (2)$$

将相对位置信息嵌入到一个固定长度的向量 e_i^t 中:

$$e_i^t = \Phi(\Delta x_i^t, \Delta y_i^t; \mathbf{W}_{rel}) \quad (3)$$

其中, $\Phi(\cdot)$ 为嵌入函数, \mathbf{W}_{rel} 为嵌入权重。

将该嵌入层向量处理为 LSTM 单元进行空间信息传递:

$$p_i^t = LSTM(p_i^{t-1}, e_i^t; \mathbf{W}_p) \quad (4)$$

其中, p_i^t 为空间 LSTM 隐藏状态, \mathbf{W}_p 为空间 LSTM 单元嵌入权重。

3.4 图注意力网络

图是一种很灵活的数据结构, 它囊括了很多其他的数据结构。例如, 如果没有边, 那么图就会变成一个集合; 如果只有相互垂直的边, 其中任意两个节点都相连, 那么就会得到一个数据树。

神经网络本身可以看作是一个图, 其中节点是神经元, 边缘是权重, 或者说节点表示层, 边缘表示向前或向后传递的流程。在没有对 LSTM 进行其他模块的添加和改进的情况下, 仅仅依靠为每个机动车和非机动车分配 LSTM 单元, 无法很好地捕捉到车辆个体之间的交互与冲突行为。因此, 为了使得多个机动车与非机动车个体之间的运动信息能够被收集, 进而被纳入考量, 图神经网络 GNN 将每个车辆个体视为计算图中的每个节点, 从而对机动车与非机动车的轨迹信息进行与图有关的操作。

此外, 图注意力网络 GAT 是一种基于空间的图卷积网络, 注意力机制在其中的表现在于, 在聚合特征信息时, 将注意力机制用于确定节点邻域的权重, 通过为不同的节点分配不同的注意力权重, 从而聚合来自周围车辆个体的信息。因此, 采用 GAT 用于交互及冲突信息的共享。

如图 2 所示, 以 5 个车辆个体为例, 用图中 5 个节点表示不同个体, 它们之间的连线为边缘, 表示机动车与非机动车等车辆个体之间的交互影响信息。

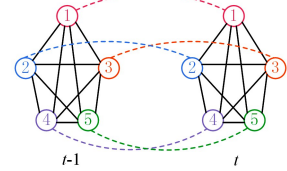


图2 图、节点与边缘

Fig. 2 Graph, node and edge

在自然语言处理中, 自注意力 self attention 利用了注意力机制, 计算每个单词与其他所有单词之间的关联^[17]。类似地, 在所研究的车辆轨迹预测问题中, 通过自注意力, 对机动车与非机动车之间的交互信息进行计算与量化。基于自注意力机制, GAT 对以图形为结构的数据进行计算, 通过关注每个节点周围的其他节点, 得到每个图上的所有节点的特征。最终, 利用层层叠加的自注意力机制, 使车辆轨迹信息得到新的考虑了车辆之间交互信息的表征。

对于注意力层, 注意力系数通过以下公式进行计算:

$$a_{ij}^t = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T [\text{concat}(\mathbf{W}_{p_i}^t, \mathbf{W}_{p_j}^t)]))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(\mathbf{a}^T [\text{concat}(\mathbf{W}_{p_i}^t, \mathbf{W}_{p_k}^t)]))} \quad (5)$$

其中, a_{ij}^t 表示在 t 时间步下, 节点 j 对 i 的注意力权重系数, \mathbf{a} 表示单层神经网络前向传播的权值向量, \mathbf{a}^T 表示 \mathbf{a} 矩阵的转置, N_i 表示节点 i 周围的车辆个体数量, \mathbf{W} 表示一个用于共享线性变换权重的矩阵, $\text{concat}(\cdot)$ 为拼接函数, 并通过 LeakyReLU 非线性函数进行激活。

接下来, 利用权重系数 a_{ij}^t 计算所有空间隐藏层状态加权之和, 生成新的状态向量:

$$\hat{p}_i^t = \tanh\left(\sum_{j \in N_i} a_{ij}^t \mathbf{W}_{p_j}^t\right) \quad (6)$$

其中, \hat{p}_i^t 表示车辆 i 在 t 时刻的聚合空间隐藏状态, 包含了其他车辆对象在空间图上的影响, \tanh 为非线性激活函数。

3.5 时空图信息结合

为了结合时间与空间上的隐藏状态信息, 对于每个机动车或非机动车车辆个体, 在已经分配一个 LSTM 单元用于传递空间图上的信息之余, 为其分配另一个 LSTM 单元传递时间图上的轨迹信息:

$$m_i^t = LSTM(m_i^{t-1}, \hat{p}_i^t; \mathbf{W}_m) \quad (7)$$

其中, \mathbf{W}_m 为时间 LSTM 单元嵌入权重。

在编码阶段, 时间 LSTM 用于捕捉当前车辆个体在时间图上的运动状态、方式和信息, 而空间 LSTM 用于捕捉当前车辆个体在空间图上的运动状态、方式和信息。

为了将两者进行结合, 构成时空图注意力网络的隐藏信息, 需要将这两部分信息通过多层感知机 MLP 结构进行线

性映射,之后进行时空图信息的结合。

$$p_i = \delta_p(p_i^{T_{obs}}) \quad (8)$$

$$m_i = \delta_m(m_i^{T_{obs}}) \quad (9)$$

$$h_i = \text{concat}(p_i, m_i) \quad (10)$$

其中, T_{obs} 为任意观测时刻, $\delta(\cdot)$ 为多层感知机结构,用于进行数据映射。

3.6 解码生成预测轨迹

为了提高模型的鲁棒性,以及缓解训练过程中的过拟合现象,引入噪声,与编码器的输出信息 h_i 进行连接:

$$d_i^{T_{obs}} = \text{concat}(h_i, z) \quad (11)$$

其中, z 为所添加的噪声。

对于下一时刻 $T_{obs} + 1$,通过解码器对隐藏信息进行解码,并经过多层感知机 MLP 线性层结构,转换到坐标空间生成所预测的相对轨迹信息:

$$d_i^{T_{obs}+1} = \text{LSTM}(d_i^{T_{obs}}, e_i^{T_{obs}}; \mathbf{W}_{dec}) \quad (12)$$

$$(\Delta x_i^{T_{obs}+1}, \Delta y_i^{T_{obs}+1}) = \delta(d_i^{T_{obs}+1}) \quad (13)$$

其中, \mathbf{W}_{dec} 为 LSTM 解码权重。

此后,将相对位置信息结合坐标,转化为绝对坐标。

3.7 损失函数

在训练过程中,采用基于 L2 损失改进的 *VarietyLoss* 函数作为损失函数^[13,24]:

$$L_{variety} = \min(k) \| Y_i - \hat{Y}_i^k \|_2 \quad (14)$$

其中, Y_i 是机动车与非机动车的未来真实轨迹, \hat{Y}_i^k 是机动车与非机动车的未来预测轨迹,超参数 k 指代生成器中在生成解码器的初始隐藏状态时噪声的随机取样次数。

4 模型仿真

4.1 环境配置

本研究的相关环境如下。CPU: AMD Ryzen 7 3750H, 2132MHz(21.333×100)。NVIDIA 通用并行计算架构 CUDA 版本:10.0。内存:16GB。显卡:NVIDIA GTX 1650。显卡内存:4GB。操作系统:Ubuntu 18.04 LTS。编程语言:Python,版本为 3.7。深度学习框架:Pytorch,版本为 1.0。

4.2 数据集

如图 3 所示,本研究采用中国某城市某交叉口在 2016 年 11 月 10 日上午 7 时 52 分至 8 时 23 分的监控视频进行分析。以该城市交叉口的车辆为研究对象,共采集 1140 辆右转机动车和 460 辆非机动车的行驶轨迹。此外,将每个斑马线及其相邻右转车道作为一个整体,从而将整个道路场景划分为 4 个不同区域,并按顺时针顺序分别标记为 1 至 4。

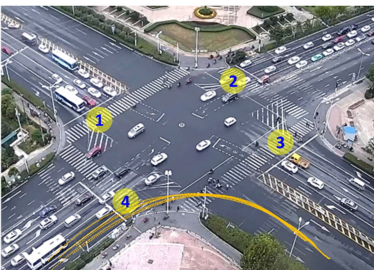


图 3 数据采集场景

Fig. 3 Scene for data collection

如图 4 所示,使用由 Suzuki 开发的基于图像处理技术的软件^[27],对机动车和非机动车轨迹进行手工提取。在下一阶段,将提取的轨迹及其对应的坐标信息进行导出,并经过预处理得到 4 列数据,每列分别为采集时刻、车辆编号、轨迹的 x 坐标和轨迹的 y 坐标。最后,以 4 个不同区域的车辆轨迹数据作为 4 个不同场景下的数据集,分别进行仿真。

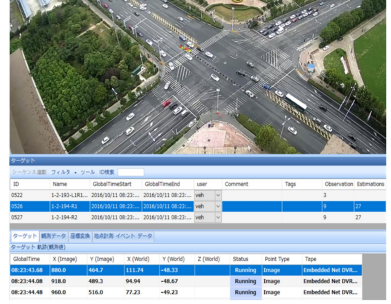


图 4 轨迹提取软件

Fig. 4 Software for data extraction

4.3 参数选择

目前为止,轨迹预测算法大多针对行人进行研究,并采用 0.4 s 作为采样周期,即下一个采样点到当前采样点的时间间隔为 0.4 s。而非机动车通过道路交叉口斑马线和机动车在右转车道右转弯时,其速度通常保持在 10~15 km/h 之间,大约是行人行走速度(4~5 km/h)的 2~3 倍。因此,在采样周期的选择上,本研究选择 0.16 s 作为采样的时间间隔,在保证确保采样数据充足的情况下,使模型仿真过程更为合理。

在训练阶段,模型共经历 400 次迭代,即 Number of Epochs 为 400;一次训练所选取的样本数 Batch Size 为 64;采用维数为 16 的高斯噪声;采用带泄漏修正线性单元 LeakyReLU 函数作为非线性激活函数 Activation Function;为了优化神经网络的训练,采用批量归一化 Batch Normalization 与 Adam 优化器;考虑到 Dropout 方法与批量归一化 Batch Normalization 之间的不兼容性^[28],将为缓解过拟合所设计的 Dropout 设置为 0;将学习率 Learning Rate 设置为 0.003,并在 250 次迭代后的训练中引入验证集验证模型性能;超参数 k 设置为 20;观测轨迹点个数 Observed Length 为 8,预测轨迹点个数 Predicted Length 为 12;对于每个数据集,训练集、测试集和验证集的比例接近 6:2:2。

4.4 评估标准

模型所生成预测轨迹的准确度由平均位移误差(Average Displacement Error, ADE)及最终位移误差(Final Displacement Error, FDE)进行评估。

ADE 具体表示为所有车辆的预测轨迹与真实轨迹的平均欧几里得距离误差,计算公式如下:

$$ADE = \frac{1}{n} \sum_{i=1}^n \frac{1}{t_{pred} - t_{obs} + 1} \sum_{t=t_{obs}+1}^{t_{obs}+t_{pred}} \sqrt{(\hat{x}_i^t - x_i^t)^2 + (\hat{y}_i^t - y_i^t)^2} \quad (15)$$

其中, n 为车辆总数, \hat{x}_i^t, \hat{y}_i^t 分别为模型所预测的车辆轨迹的横、纵坐标。

FDE 具体表示为车辆在最终的预测位置与相应的真实位置之间的平均欧几里得距离误差,计算公式如下:

$$FDE = \frac{1}{n} \sum_{i=1}^n \sqrt{(\hat{x}_{i_{pred}}^i - x_{i_{pred}}^i)^2 + (\hat{y}_{i_{pred}}^i - y_{i_{pred}}^i)^2} \quad (16)$$

其中, n 为车辆总数, $\hat{x}_{i_{pred}}^i$, $\hat{y}_{i_{pred}}^i$ 分别车辆在最后一个预测时刻下的轨迹的横、纵坐标。

4.5 实验结果

4.5.1 激活函数与优化器

神经网络中的每个神经元接收上一层神经元的输出值作为本神经元的输入值, 并将输入值传递给下一层, 输入层神经元节点会将输入属性值直接传递给下一层(隐藏层或输出层)。在多层神经网络中, 上层节点的输出和下层节点的输入之间具有一个函数关系, 这个函数称为激活函数^[29]。若不使用激活函数, 会导致神经网络的每层仅完成线性变换, 多层

输入叠加后仍属于线性变换。由于通常线性模型的表达能力不够, 因此, 需要通过非线性激活函数以引入非线性因素。常见的激活函数有 sigmoid, tanh, ReLU, LeakyReLU 等。

优化器是神经网络的眼睛, 是神经网络蓬勃发展的基础。不同的优化器试图通过不同的优化方法寻找模型的最优解, 进行神经网络众多参数的学习^[30]。常见的优化器有 SGD, Adagrad, RMSProp, Adam 等。

为了初步探究本方法所适合的激活函数, 首先固定 Adagrad 作为优化器, 分别使 sigmoid, tanh, ReLU, LeakyReLU 作为激活函数, 对 ADE 及 FDE 进行计算, 结果如表 1 所列。

表 1 激活函数与误差

Table 1 Activation functions and errors

Activation Function	Dataset 1		Dataset 2		Dataset 3		Dataset 4		Average	
	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE
sigmoid	0.595	1.304	0.548	1.226	0.708	1.507	0.654	1.275	0.627	1.328
tanh	0.574	1.288	0.524	1.077	0.610	1.452	0.552	1.212	0.565	1.257
ReLU	0.568	1.263	0.513	1.134	0.621	1.409	0.533	1.172	0.559	1.245
LeakyReLU	0.547	1.241	0.493	1.001	0.592	1.382	0.511	1.158	0.536	1.196

同理, 为了初步探究本文方法所适合的优化器, 首先固定 tanh 作为激活函数, 分别使 SGD, Adagrad, RMSProp 以及

Adam 作为优化器, 对 ADE 及 FDE 进行计算, 结果如表 2 所列。

表 2 优化器与误差

Table 2 Optimizers and errors

Optimizer	Dataset 1		Dataset 2		Dataset 3		Dataset 4		Average	
	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE	ADE	FDE
SGD	0.591	1.311	0.542	1.101	0.623	1.473	0.569	1.241	0.581	1.282
Adagrad	0.574	1.288	0.524	1.077	0.610	1.452	0.552	1.212	0.565	1.257
RMSProp	0.547	1.246	0.492	1.034	0.588	1.401	0.511	1.160	0.535	1.210
Adam	0.538	1.238	0.483	0.991	0.572	1.382	0.499	1.149	0.523	1.190

探究激活函数对模型预测精度的影响时, 分别对 4 个数据集进行仿真, 由表 1 可知, 相比较于 sigmoid, tanh 以及 ReLU 激活函数, Leaky ReLU 激活函数在 4 个数据集上均取得最小 ADE 值和 FDE 值。因此, 选择 LeakyReLU 函数作为本方法的最优激活函数。

同理, 探究优化器对模型预测精度的影响时, 分别对 4 个数据集进行仿真, 由表 2 可知, 相比较于 SGD, Adagrad 以及

RMSProp 优化器, Adam 优化器在 4 个数据集上均取得最小 ADE 值和 FDE 值。因此, 选择 Adam 优化器作为本方法的最优优化器。

4.5.2 误差分析

训练过程中, 在经过 250 次迭代, 即 epochs 达到 250 之后, 引入验证集对模型进行验证。因此, 得到 4 个数据集下在验证过程中的 ADE 曲线及 FDE 曲线, 分别如图 5 及图 6 所示。

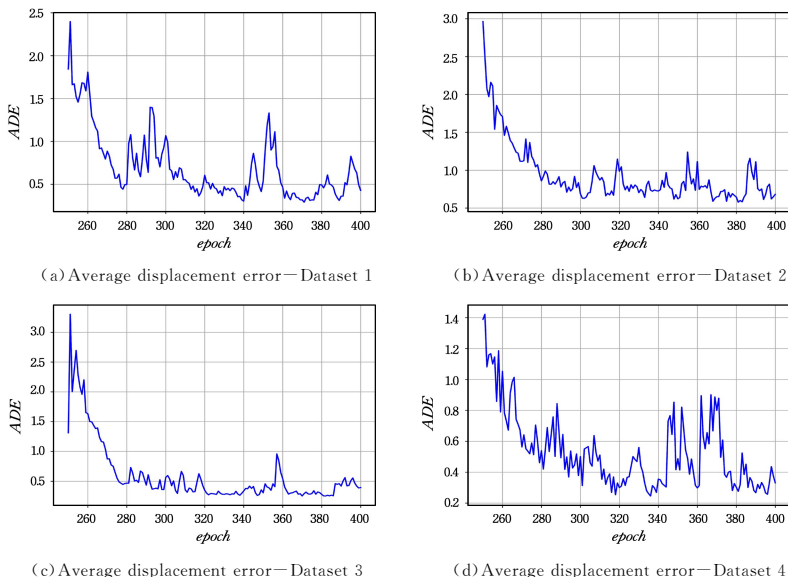


图 5 验证集 ADE 曲线

Fig. 5 ADE curves for validation set

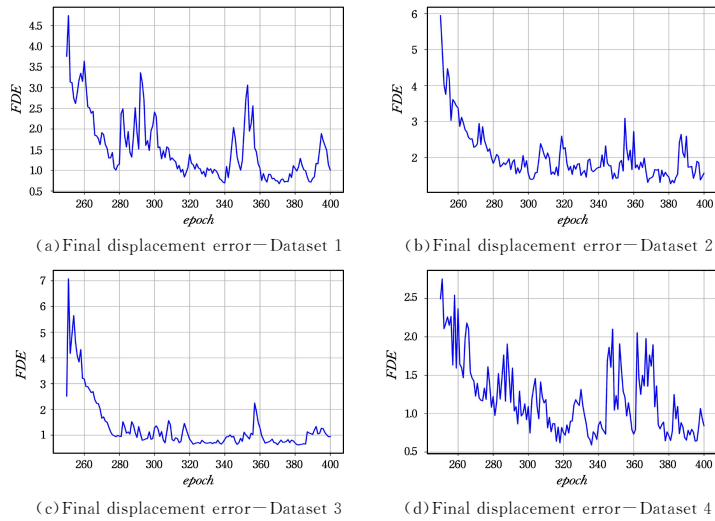


图6 验证集 FDE 曲线

Fig. 6 FDE curves for validation set

为了与本方法进行对比,采用 Social LSTM 方法和 Social GAN 方法对本实验所采用数据集进行仿真。经过调优后,3 种方法的参数配置如表 3 所列。

表 3 3 种方法参数配置

Table 3 Parameter configurations of three methods

参数	S-LSTM	S-GAN	本方法
Num_epochs	400	400	400
Batch_size	64	64	64
Noise_type	Gaussian	Gaussian	Gaussian
Noise_dim	16	16	16
Activation	ReLU	ReLU	Leaky ReLU
Dropout	0	0	0
Optimizer	Adagrad	Adam	Adam
Batch_norm	True	True	True
Learning_rate	0.003	0.003	0.003
k	—	20	20
Obs_len	8	8	8
Pred_len	12	12	12

3 种方法仿真所得结果如表 4 所列,在本实验所采用的数据集上,3 种方法均取得较好表现。将 3 种方法分别在 4 个数据集上仿真得到的 ADE, FDE 进行算术平均后,可知本方法取得了最好的效果。具体表现在:对比于 Social LSTM 算法,本方法所求得平均位移误差 ADE 低于其运算结果

51.77%,最终位移误差 FDE 低于其运算结果 42.14%;对比于 SocialGAN 算法,本方法所求得平均位移误差 ADE 低于其运算结果 15.64%,最终位移误差 FDE 低于其运算结果 12.82%。因此,通过误差分析可知,本研究所采用的模型更为适合在复杂道路交叉口环境下为车辆提供轨迹规划方案。

表 4 仿真结果

Table 4 Simulation results

误差评估	数据集	S-LSTM	S-GAN	本方法
ADE	1	0.967	0.593	0.507
ADE	2	0.998	0.614	0.452
ADE	3	1.032	0.582	0.541
ADE	4	1.074	0.538	0.463
AVG	—	1.018	0.582	0.491
FDE	1	1.981	1.401	1.201
FDE	2	1.971	1.456	0.959
FDE	3	1.947	1.256	1.349
FDE	4	2.093	1.191	1.115
AVG	—	1.998	1.326	1.156

4.6 损失函数曲线

图 7 给出了根据 4 个数据集训练过程中损失函数值进行绘制得到的损失函数曲线。

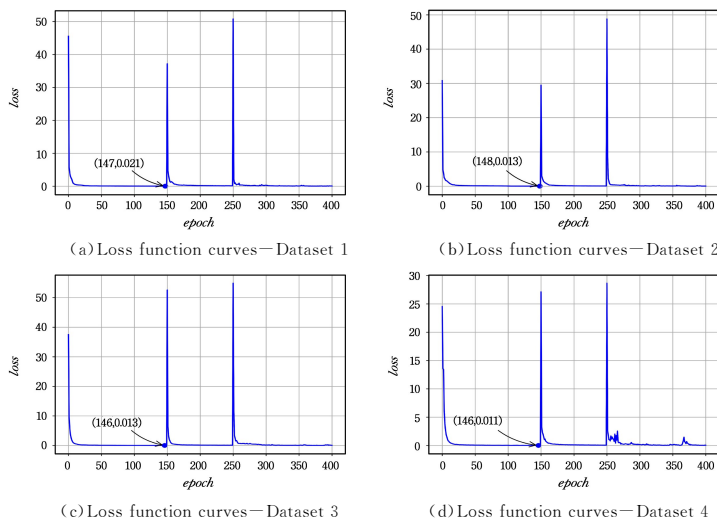


图 7 损失函数曲线

Fig. 7 Loss function curves

观察可知,在4个不同场景的数据集上,本方法分别在第149,148,384,147个epoch上取得loss最小值。而训练过程中出现的函数值骤增的情况,本质上是由循环神经网络(Recurrent Neural Network, RNN)所带来的梯度爆炸问题导致的^[31-32]。即便采用了在传统RNN上做出改进的长短期记忆网络LSTM,也仅能起到缓解梯度爆炸的作用,并不能完全解决这一问题^[33]。

4.7 案例分析

通过将生成的预测轨迹与交叉口场景相结合进行可视化,可以将研究时间段内每辆车的观察轨迹、预测轨迹和未来真实轨迹进行绘制。为了直观地反映每一辆车行驶的速度,在可视化时根据车辆速度的不同对所生成的轨迹点个数进行插值处理。如果车辆以较快速度行驶,会产生较多轨迹点;反之,生成轨迹点较少。

本研究选取了4个交互情况复杂,且生成的预测轨迹与未来真实轨迹差异较大的场景进行分析,以说明本方法所提供的轨迹规划方案的优势,如图8所示。

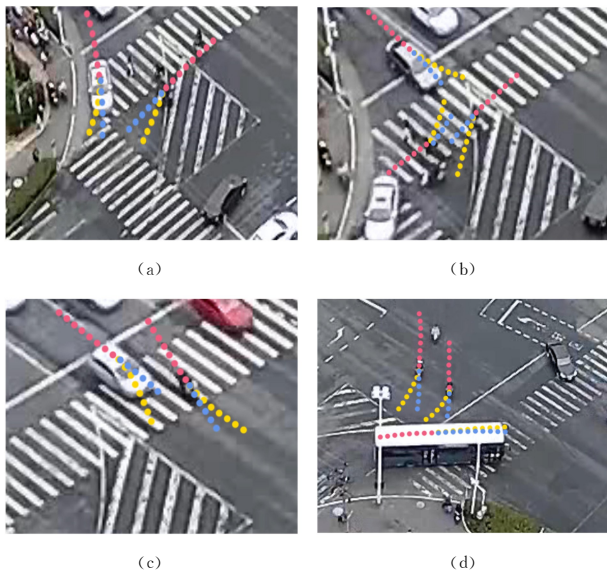


图8 交互案例(电子版为彩色)

Fig. 8 Interaction cases

4.7.1 交互场景1

如图8(a)所示,处于区域1最右车道的白色轿车,与其左前方的电动车发生交互行为。此时,本方法经仿真所得的轨迹规划方案为:白色轿车向右做小角度转向,电动车向左做小角度转向,实现两车相互背离。此规划方案可使得白色轿车与电动车在不需要做大幅度减速的情况下,安全驶离路口。而未来真实轨迹显示,电动车选择按原方向直行,而没有在原来直行的情况下向左转向,进而达到两车相互背离的目的。代表未来真实轨迹的蓝色点数目较代表预测轨迹的黄色点数目明显减少,表明两车在未来短时间内均做出大幅减速。因此,相较于真实轨迹,通过本方法得到的规划方案能够使两车更为安全、高效地完成交互,从而在降低风险的情况下更快地驶离交叉路口。

4.7.2 交互场景2

如图8(b)所示,处于区域1最右车道的银色轿车前方有一自行车和一电动车与其发生交互。未来真实轨迹显示,三者共同选择大幅度减速以在确保安全的情况下完成交互。而

在此类交互情况较为复杂的情况下,本方法考虑到三者之间的相互影响,做出如下规划:银色轿车向左做小角度转向,自行车向左缓慢行驶,电动车向左转向,从而使三者相互背离。与未来真实轨迹相比,该轨迹规划方案能够明显提高三者发生交互冲突时的交通安全性,为车辆在行驶过程中提供更优选择。

4.7.3 交互场景3

如图8(c)所示,区域1中一电动车在经过道路交叉口时,并未按照我国交通规则在斑马线上行驶,而采取了与机动车一同在机动车道上行驶的做法,属于违反交通规则的行为。注意到电动车有向右前方行驶的趋势,如果此时轿车不采取向右做小角度转向的方式,则不得不大幅度降低自身速度以做出避让。因此,本方法为白色轿车和电动车提供的规划路径方案为:白色轿车向右做小角度转向,电动车向左做小角度转向,从而在提高两者加速阶段安全性的同时,使两者保持一定的初速度,不对后方来车的行驶通畅性造成影响。故在此类发生违反交通规则的现象的情况下,合理的轨迹规划更加保证了道路安全性和行驶高效性。

4.7.4 交互场景4

如图8(d)所示,两电动车沿对角线通过交叉路口时,遭遇前方处于右转状态的公交车。由于公交车体积较大,转弯时间较长,因此,两电动车在距公交车一定距离时主动降低速度。但为了减少两电动车在交叉路口处的停留时间,电动车若能采用向右前方行驶的方案,则能够以更短时间完成与公交车的交互,从而更为高效地通过交叉路口。

结束语 本研究采取基于深度学习技术的方法,将城市道路交叉口的车辆轨迹进行序列化,转化为序列-序列问题,并结合循环神经网络、长短期记忆网络、注意力机制等方法对车辆进行轨迹预测并做出合理的轨迹规划。结果表明,在交叉路口的4个区域中,对于简单的交互场景,本方法能够取得较小的车辆轨迹预测误差,而对于多车辆同时产生交互、冲突等行为的复杂场景,本方法能够充分考虑到研究对象周围车辆与其发生的交互行为,从而为每个车辆个体提供更优的未来行驶路径选择。

与此同时,本研究在考虑道路交叉口环境限制条件上仍存在欠缺之处,例如在规划非机动车在斑马线上的路径时,未能将人行道这一影响非机动车行驶轨迹的因素纳入考量,因此产生的规划路径仍不够完善。故下一步研究可以通过图像处理技术,提取整个道路交叉口视频、图片的信息,采取添加物理注意力机制等模块的方式,对人行道、障碍物等客观环境条件制定规则,提高该模型的泛化能力,使其适用于更多复杂的场景。

参考文献

- [1] WU Y, SCHUSTER M, CHEN Z, et al. Google's neural machine translation system: bridging the gap between human and machine translation[J]. arXiv:1609.08144, 2016.
- [2] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[J]. arXiv:1409.0473, 2014.
- [3] GRAVES A, JAITLY N. Towards end-to-end speech recognition with recurrent neural networks[C]// International Conference on Machine Learning. 2014:1764-1772.
- [4] CHOROWSKI J, BAHDANAU D, CHO K, et al. End-to-end

- continuous speech recognition using attention-based recurrent nn:First results[J]. arXiv:1412.1602,2014.
- [5] XU K,BAJ,KIROS R,et al. Show,attend and tell:Neural image caption generation with visual attention[C]//International Conference on Machine Learning. 2015:2048-2057.
- [6] LAROCA R,SEVERO E,ZANLORENSI L A,et al. A robust real-time automatic license plate recognition based on the yolo-detector[C]//International JointConference on Neural Networks. 2018:1-10.
- [7] BERGER C. From a competition for self-driving miniature cars to a standardized experimental platform;concept,models,architecture,and evaluation[J]. arXiv:1406.7768,2014.
- [8] HELBING D,MOLNAR P. Social force model for pedestrian dynamics [J]. Physical Review E Statistical Physics Plasmas Fluids & Related Interdisciplinary Topics,1998,51(5):4282.
- [9] TAY C,LAUGIER C. Modelling smooth paths using gaussian processes[C]//Springer Tracts in Advanced Robotics,2007,42:381-390.
- [10] CHAN Y T,HU A G C,PLANT J B. A kalmanfilter based tracking scheme with input estimation[J]. IEEE Transactions on Aerospace and Electronic Systems,1979,AES-15(2):237-244.
- [11] ALAHI A,GOEL K,RAMANATHAN V,et al. Social lstm: human trajectory prediction in crowded spaces[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2016.
- [12] ILYA S,ORIOLO V,QUOC V L. Sequence tosequence learning with neural networks [C]//Advances in Neural Information Processing Systems. 2014:3104-3112.
- [13] AGRIM G,JUSTIN J,LI F,et al. Social gan:socially acceptable trajectories with generative adversarial networks [C]//IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- [14] JAVAD A,JEAN-BERNARD H,JULIAN P. Social ways: learning multi-modal distributions of pedestrian trajectories with gans[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.
- [15] AMIR S,VINEET K,ALI S,et al. SoPhie:an attentive gan for predicting paths compliant to social and physical constraints [C]//IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- [16] ZHAO T,XU Y,MONFORT M,et al. Multi-agent tensor fusion for contextual trajectory prediction[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2019.
- [17] ASHISH V,NOAM S,NIKI P,et al. Attention is all you need [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017,30.
- [18] VEMULA A,MUELLING K,OH J. Social attention: modeling attention in human crowds[C]//2018 IEEE international Conference on Robotics and Automation (ICRA). IEEE,2018.
- [19] XU K,HU W,LESKOVEC J,et al. How powerful are graph neural networks? [J]. arXiv:1810.00826.
- [20] SIJIE Y,YUANJUN X,DAHUA L. Spatial temporal graph convolutional networks for skeleton-based action recognition [C]//Thirty-Second AAAI Conference on ArtificialIntelligence. 2018.
- [21] YAN S J,XIONG Y J,LIN D H. Spatial temporal graph convolutional networks for skeleton-based action recognition [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018(1).
- [22] CHENYANG S,WENTAO C,WEI W,et al. An attention enhanced graph convolutional lstm network for skeleton-basedaction recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019:1227-1236.
- [23] VELIKOVI P,CUCURULL G,CASANOVA A,et al. Graph attention networks[J]. arXiv:1710.10903.
- [24] HUANG Y,BI H,LI Z,et al. STGAT:modeling spatial-temporal interactions for human trajectory prediction[C]//IEEE International Conference on Computer Vision. 2019:6272-6281.
- [25] CHO K,VAN M B,GULCEHRE C,et al. Learning phrase representations usingrnnencoder-decoder for statistical machine translation[J]. arXiv:1406.1078.
- [26] HASIM S,ANDREW S,FRANCOISE B. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition[J]. arXiv:1402.1128.
- [27] SUZUKI K,NAKAMURA H. Traffic Analyzer-the integrated video image processing system for traffic flow analysis [C]//Proceedings of the 13th World Congress on Intelligent Transportation Systems. 2006.
- [28] LI X,CHEN S,HU X,YANG J. Understanding the disharmony between dropout and batch normalization by variance shift[C]//IEEE Conference on Computer Vision and Pattern Recognition. 2019:2677-2685.
- [29] NWANKPA C,IJOMAH W,GACHAGAN A,et al. Activation functions:comparison of trends in practice and research for deep learning[J]. arXiv:1811.03378.
- [30] BELLO I,ZOPH B,VASUDEVAN V,et al. Neural optimizer search with reinforcement Learning[C]//International Conference on Machine Learning. 2016:459-468.
- [31] HOJJATS.SHARN S,JOSEPH B. Recent advances in recurrent neural networks[J]. arXiv:1801.01078,2017.
- [32] WOJCIECH Z,ILYA S,ORIOLO V. Recurrent neural network regularization[J]. arXiv:1409.2329.
- [33] SEPP H,JURGEN S. Long short-term memory [J]. Neural Computation,1997,9(8):1735-1780.



ZENG Wei-liang, born in 1986, Ph.D associate professor. His main research interests include routing problem in complex network, traffic simulation, and big data visualization for smart city.



SUN Wei-jun, born in 1975, Ph.D, is a member of CCF YOCSEF AC. His main research interests include internet of thing and machine learning.