DOI:10.11918/202006051

图卷积神经网络行人轨迹预测算法

王天保[1],刘 昱[1],郭继昌[2],晋玮佩[2]

(1.天津大学 微电子学院,天津 300072;2.天津大学 电气自动化与信息工程学院,天津 300072)

摘 要: 针对行人轨迹预测任务中行人间的交互模式难以被有效构建的问题,提出了一种基于图卷积神经网络的算法TP-GCN来建立行人交互模型并进行轨迹预测.首先对行人的轨迹序列使用长短期记忆网络提取轨迹运动特征;随后将行人视为图结构中的顶点,创建表示相互关系的邻接矩阵,并根据视觉盲区范围筛除无关顶点间的连接权重;然后对于轨迹运动特征,使用图卷积神经网络提取不同轨迹间的交互信息,同时增加顶点对自身所隐含交互信息的提取,并使用长短期记忆网络将交互信息编码为轨迹交互特征;之后通过深度图信息最大化方法,对图卷积神经网络的权重进行优化,使得个人的运动模式符合场景内所有行人共有的运动模式;最后将轨迹运动特征和轨迹交互特征使用长短期记忆网络进行解码,完成轨迹预测.在公开数据集ETH和UCY上的实验结果表明,所提算法能够根据行人间的交互模式做出与真实行为接近的符合行人习惯的预测,整体预测精度高.同时,消融实验和预测轨迹的可视化也显示了算法的有效性及良好的可解释性.

关键词: 轨迹预测;交互模式;图卷积神经网络;长短期记忆网络;互信息

中图分类号: TP391 **文献标志码:** A **文章编号:** 0367-6234(2021)02-0053-08

Pedestrian trajectory prediction algorithm based on graph convolutional network

WANG Tianbao[1], LIU Yu[1], GUO Jichang[2], JIN Weipei[2]

(1. School of Microelectronics, Tianjin University, Tianjin 300072, China;
2. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China)

Abstract: To solve the problem that the pedestrian interaction model is difficult to be effectively constructed in the pedestrian trajectory prediction task, a trajectory prediction algorithm based on graph convolutional network (TP-GCN) was proposed to establish pedestrian interaction and predict future trajectories of pedestrians. First, the long short-term memory was used to extract the trajectory motion features of the trajectory sequences of pedestrians. Then, the pedestrians were considered as the nodes on the graph, and adjacency matrix was built to represent the created interactions. Next, the connection weights between unrelated nodes were screened out according to the blind zone. For trajectory motion features, the graph convolutional network was applied to extract the interactions between the trajectories and increase the extraction of the interaction in each trajectory, and the interaction was then encoded as trajectory interaction features by using long short-term memory. Furthermore, the weights of the graph convolutional network were optimized by the Deep Graph Info method to ensure that the motion pattern of individual accords with those of all the pedestrians in the scene. Finally, the trajectory motion features and trajectory interaction features were decoded using long short-term memory to complete the trajectory prediction. According to the experiment on the public datasets ETH and UCY, the proposed algorithm could make the predictions of pedestrian habits close to the real trajectories based on the interaction model between pedestrians, and the overall prediction accuracy was high. In addition, the ablation experiment and the visualization of the predicted trajectory also verified the effectiveness and interpretability of the algorithm.

Keywords: trajectory prediction; interaction model; graph convolutional network; long short-term memory; mutual information

在城市街道等密集行人场景中,自动驾驶车辆、机器人等运动主体需要根据其他行人的位置规划自身路径,通过对目标的位置预测得以保持安全距离并排除风险因素,行人未来位置预测的准确性对于运动主体的决策系统至关重要[1].行人轨迹预测是一项复杂任务,由于每个行人自身的运动习惯有着天然差异,并且群体环境中存在人与人的交互,个人的运动模式会受到周围行人隐含的影响,人们会遵循社会规则方面的常识来调整自己的路线,运动主

收稿日期: 2020-06-09
基金项目: 云南省重大科技专项计划项目(202002AD080001)
作者简介: 王天保(1994—),男,硕士研究生;
刘 昱(1976—),男,教授,博士生导师
通信作者: 郭继昌,jcguo@tju.edu.cn

体需要预测他人的动作和社会行为[2]. 构建具有较高可解释性和泛化能力的行人交互模式是轨迹预测问题的重点.

早期的行人轨迹预测使用手工设计特征的方法构建社会力量(social force, SF)[3-4]模型,由此表示行人在运动过程中相互吸引和排斥的情况,然而完全依靠手工设计特征难以表示复杂场景中隐含的交互行为. 近年来以数据驱动为主导的循环神经网络(recurrent neural network, RNN)编解码结构广泛应用于轨迹预测任务,具有代表性的是 Alahi 等[5]使用长短期记忆网络(long short-term memory, LSTM)[6]编码器 - 解码器结构,通过社交池化(social-pooling)获取不同距离行人间的依赖关系,从而表现个体间隐含的交互信息;Gupta 等[7]将轨迹预测看作序列生成问题,使用生成对抗网络(generative adversarial networks, GAN)体现轨迹的多模态性质,并且对历史轨迹编码进行最大池化(max-pooling),生成社交可接受的轨迹;在考虑多种物理特征的方面,Hasan 等[8]将获取的行人头部朝向特征纳入编码过程,其结果证实对周围行人的关注程度与自身视线方向具有高度相关性;张志远等[9]使用行人间的距离及方向信息构建注意力模型,并使用生成对抗方法训练轨迹生成; Amirian 等[10]使用 infoGAN 结构,通过优化输入隐含变量与输出轨迹分布的互信息来提升轨迹生成效果,并根据行人间的位置、方向、可接近的最小距离等物理特征进行注意力池化.

图神经网络(graph neural network, GNN)将深度学习应用在非欧几里得结构上,构建顶点和边表示对象间的关系,展现出良好的鲁棒性和可解释性,因此通过图拓扑结构建模行人之间的交互模式是一种有效的方式. Vemula 等[11]在轨迹预测问题上使用时空图网络构建交互模型,使预测目标对周围行人分配不同软注意力权重,获取时间和空间上的轨迹交互信息;由于行人运动具有时间连续性,空间上的行人交互模式不仅与当前位置有关,还应考虑历史影响, Huang 等[12]基于图注意力网络(graph attention network, GAT)[13]对周围行人分配注意力以进行运动 LSTM 编码; Kosaraju 等[14]使用图注意力网络表示空间交互关系,通过 Bicycle-GAN 生成多模态预测; Mohamed 等[15]根据行人位置构建邻接矩阵,通过图卷积神经网络(graph convolutional network, GCN)[16]构建交互模式,并使用时间外推卷积神经网络进行轨迹预测. 然而,使用图注意力网络进行注意力分配由于依赖于高维特征间的相关性,其过程并不直观,且没有考虑图的结构关系;另一方面,由于正常人眼关注度高的区域主要分布在视野中部,并且双眼水平视场角约为 188°,在行走状态下人眼存在较大盲区,现有图网络所分配得到的交互注意力往往会错误地将盲区中的行人纳入其中.

考虑到图网络在建立交互模型中所具有的优势及存在的问题,本文提出一种新的基于图卷积神经网络的轨迹预测模型(trajectory prediction graph convolutional network, TP - GCN)用于构建行人间的交互模式并进行轨迹预测. 算法使用图卷积神经网络处理编码过的高维行人轨迹特征,从而构建行人间的交互模式,根据盲区信息优化图卷积神经网络的邻接矩阵,并加强了对自身隐含交互模式的获取,同时使用深度图信息最大化方法将图结构的局部特征和整体特征间的互信息最大化,优化图网络的特征提取效果. 在公开数据集上进行实验,结果表明本文算法可以取得较精确的预测效果,同时具有较强泛化效果及可解释性.

1 图卷积神经网络

卷积神经网络(convolutional neural network, CNN)利用固定尺寸的卷积核在图像上进行卷积操作并平移,从而提取图像中的所需特征. 图卷积神经网络的原理与 CNN 类似,并将欧氏空间的卷积操作推广到非欧空间,对图结构中顶点的特征进行提取,以完成后续的顶点分类等任务. 具体地,若无向图 $G=(V,E)$ 中有 n 个顶点,顶点为 $V=\{V_i \mid \forall i \in \{1,2,\cdots,n\}\}$,连接顶点的边为 $E=\{e_{ij} \mid \forall i,j \in \{1,2,\cdots,n\}\}$,每个顶点包含 d 维特征,则根据各顶点 V 之间的边 E 构成的 $n \times n$ 维的邻接矩阵 $\boldsymbol{A}$,通过训练卷积核系数,计算中心顶点的邻接顶点与卷积核的卷积结果,从而实现特征提取. 单层 GCN 结构如式(1)所示

$$H^{l+1}=\sigma(\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}H^{l}W_{\mathrm{g}}^{l}), \tag{1}$$

式中: $\tilde{\boldsymbol{A}}=\boldsymbol{A}+\boldsymbol{I}$, $\boldsymbol{A}$ 为图的邻接矩阵, $\boldsymbol{I}$ 为单位矩阵; $\tilde{\boldsymbol{D}}^{-\frac{1}{2}}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-\frac{1}{2}}$ 为规范化对称邻接矩阵; H^{l} 为 l 层顶点的特征; W_{g}^{l} 为 l 层的图卷积权重,其将对应图顶点的特征维数 d 进行变换; σ 为激活函数.

两层图卷积网络示意图如图 1 所示,其中每层顶点的连接关系均由共享的邻接矩阵 $\tilde{\boldsymbol{A}}$ 表示,输入特征为 V,通过两层 GCN 得到输出 Z,如式(2)所示

$$Z=f(V,\tilde{\boldsymbol{A}})=\sigma(\tilde{\boldsymbol{A}}\sigma(\tilde{\boldsymbol{A}}VW_{\mathrm{g}}^{0})W_{\mathrm{g}}^{1}), \tag{2}$$

式中 f 为两层 GCN 的特征传播公式.

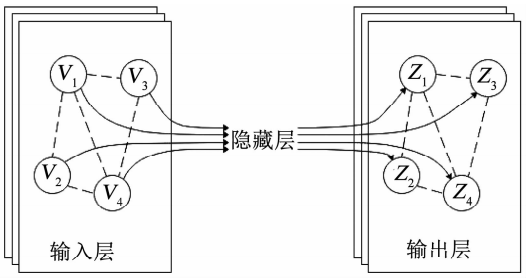

图 1　两层图卷积网络示意

Fig. 1　Two-layer graph convolutional network

2　本文算法

行人轨迹预测任务定义为，假设场景中有 n 个行人，通过给定 $t = 1,2,\cdots,T_{\text{obs}}$ 时刻场景内全部行人的观测轨迹坐标 $X = X_1, X_2, \cdots, X_n$，预测 $t = T_{\text{obs}}+1,\cdots,T_{\text{final}}$ 时刻全部行人的未来轨迹坐标 $\hat{X} = \hat{X}_1, \hat{X}_2, \cdots, \hat{X}_n$. 每个行人 $i = 1,2,\cdots,n$ 在 $t = 1,2,\cdots, T_{\text{obs}}, T_{\text{obs}}+1,\cdots,T_{\text{final}}$ 时刻的真实坐标为 $X_i^t = (x_i^t, y_i^t)$，行人 $i = 1,2,\cdots,n$ 在 $t = T_{\text{obs}}+1,\cdots,T_{\text{final}}$ 时刻的预测坐标为 $\hat{X}_i^t = (\hat{x}_i^t, \hat{y}_i^t)$，预测时长为 $T_{\text{pred}} = T_{\text{final}} - T_{\text{obs}}$.

轨迹预测任务中任意时刻的每个行人 i 都与不同数量和运动状态的其他行人存在交互关系，行人间内在的影响方式复杂且随时间而变化，若以向量表示每个行人的运动状态，那么同一时刻相关联的所有行人构成了一组典型的图结构数据.

本文提出的轨迹预测模型 TP－GCN 中，将行人作为图结构中的顶点，利用 GCN 在图结构中良好的特征提取能力来获取行人间的交互关系，并通过最大互信息优化方法进一步提升 GCN 的运算效果，从而完成轨迹预测. 算法框图见图 2.

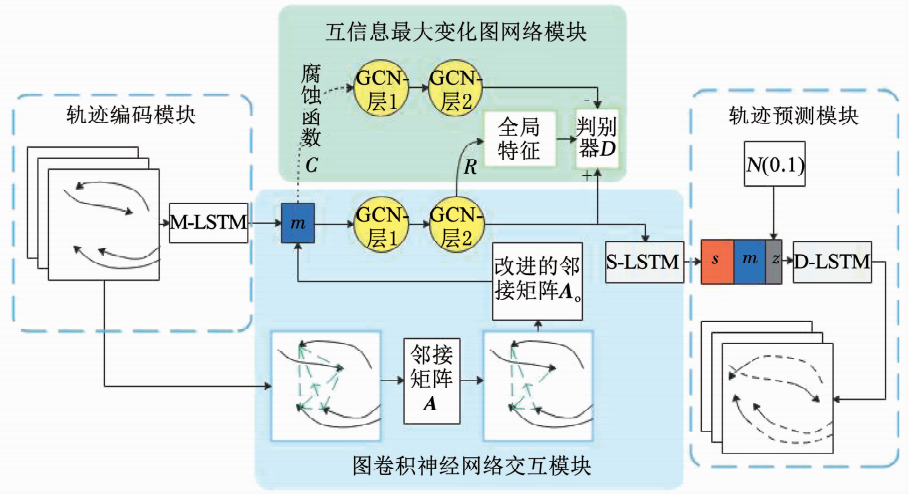

图 2　图卷积神经网络轨迹预测模型(TP－GCN)

Fig. 2　Trajectory prediction based on graph convolutional network (TP-GCN)

如图 2 所示，TP－GCN 由 4 个模块构成，分别为：1）轨迹编码模块：将原始轨迹使用 LSTM 编码得到轨迹运动特征；2）图卷积神经网络交互模块：通过原始轨迹计算改进的邻接矩阵，将轨迹运动特征输入 GCN 计算轨迹交互特征；3）互信息最大化图网络模块：最大化 GCN 输出中局部特征与全局特征间的互信息，从而优化 GCN 的特征提取效果；4）轨迹预测模块：将提取的轨迹运动特征与轨迹交互特征进行 LSTM 解码，得到轨迹预测结果.

2.1　轨迹编码模块

该模块的输入为多帧历史轨迹图像，每帧历史轨迹图像包括若干条轨迹，各帧间轨迹的数目相同，轨迹编码模块对每条轨迹的相对位置变化进行 M－LSTM 编码，输出轨迹运动特征. 为了获取单一轨迹中具有较强表达能力的特征[5,7-8]，首先将前后帧之间单个轨迹的相对位置变化 $\Delta X_i^t = (x_i^t - x_i^{t-1}, y_i^t - y_i^{t-1})$ 编码为固定长度的运动向量 q_i^t，如式(3)所示

$$q_i^t = \phi(\Delta X_i^t, \boldsymbol{W}_{\text{q}}), \tag{3}$$

式中，ϕ 为相对位置编码函数，$\boldsymbol{W}_{\text{q}}$ 为相对位置编码权重. 随后使用 LSTM 结构进行 M－LSTM 编码以获取轨迹运动特征 m_i^t，如式(4)所示

$$m_i^t = f_{\text{M-LSTM}}(m_i^{t-1}, q_i^t, \boldsymbol{W}_{\text{m}}), \tag{4}$$

式中：$f_{\text{M-LSTM}}$ 为编码器 M－LSTM；m_i^{t-1} 为 M－LSTM 在 $t-1$ 时刻的隐藏状态，隐藏层单元个数为 32；$\boldsymbol{W}_{\text{m}}$ 为 M－LSTM 的编码权重.

2.2　图卷积神经网络交互模块

由于行人的轨迹受到周围行人运动模式隐含的影响，仅对每个轨迹分别进行编码难以完整表达场景内多个轨迹的复杂运动模式，需要构建合理的模型表达行人间交互模式. 使用图结构 $G^t = (V^t, E^t)$

建立 t 时刻行人间的交互模型,将行人作为图结构中顶点的集合 V^t,行人间的交互关系为边的集合 E^t,其表达式为

$$V^t = \{V_i^t \mid \forall i \in \{1,2,\cdots,n\}\}, \tag{5}$$

$$E^t = \{e_{ij}^t \mid \forall i,j \in \{1,2,\cdots,n\}\}. \tag{6}$$

式中:V_i^t 为第 i 个顶点所具有的特征;e_{ij}^t 为顶点之间值为 0 或 1 的边,若 i 与 j 间存在交互关系则为 1,否则为 0.

将每一个时间点中顶点 V^t 的连接关系 E^t 表示为邻接矩阵 $\boldsymbol{A}^t$,所有时刻邻接矩阵的集合为 A. 行人间相互的影响程度随着距离的增加而减小[15],本文将邻接矩阵 $\boldsymbol{A}^t$ 中的边 e_{ij}^t 根据距离不同分配的权重 a_{ij}^t,如式(7)所示

$$a_{ij}^t = \begin{cases} 1/\| X_i^t - X_j^t \|_2, \| X_i^t - X_j^t \|_2 \neq 0; \\ 0, \text{其他}. \end{cases} \tag{7}$$

由于个体在运动中存在较大盲区,为了避免盲区中行人的干扰,故应主动筛除为盲区中的其他行人分配交互权重的情况,本文中假设行人的盲区范围为速度方向两侧各 >90°的范围. 如图 3 所示,在 t 时刻,$X_1 = (x_1, y_1)$,$X_2 = (x_2, y_2)$,$\Delta X_1 = (x_1', y_1')$,$\Delta X_2 = (x_2', y_2')$,背向而行的两个行人均处于对方的盲区中.

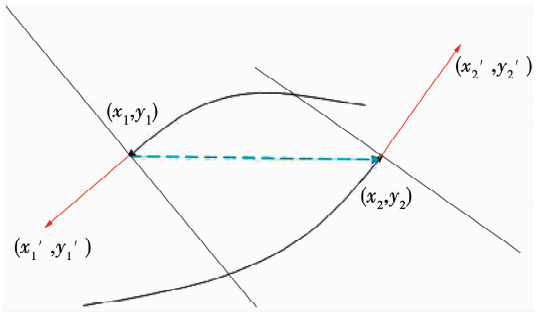

图 3 盲区中行人示意

Fig. 3 Pedestrians in blind zone

此时两个行人的速度向量与相对位置向量满足式(8)

$$[\Delta X_1 (X_1 - X_2)][\Delta X_2 (X_1 - X_2)] < 0. \tag{8}$$

由以上情况推广,根据位置和速度方向信息,将行人背离情况下的边权重 a_{ij}^t 设置为 0,此时 a_{ij}^t 如式(9)所示

$$a_{ij}^t = \begin{cases} 1/\| X_i^t - X_j^t \|_2, \\ \quad [\Delta X_i^t (X_i^t - X_j^t)][\Delta X_j^t (X_i^t - X_j^t)] > 0; \\ 0, \text{其他}. \end{cases} \tag{9}$$

将经过 M - LSTM 编码的轨迹运动特征 m_i^t 作为图卷积网络中顶点的输入特征 V_i^t,因为 LSTM 的记忆门使得若干时刻前 m_i 的影响依然保留,而行人在此段时间中的轨迹隐式地表现了其他轨迹的交互影响,间接体现了交互模式,故在图网络中对中心顶点分配额外权重,增加对行人自身所含的隐式交互信息的提取,得到 t 时刻改进的邻接矩阵 $\boldsymbol{A}_o^t$,如式(10)所示

$$\boldsymbol{A}_o^t = \tilde{\boldsymbol{D}}^{-\frac{1}{2}} \tilde{\boldsymbol{A}}^t \tilde{\boldsymbol{D}}^{-\frac{1}{2}} + k\boldsymbol{I}. \tag{10}$$

式中,k 为中心顶点额外权重系数,本文使用 $k = 2$;$\boldsymbol{I}$ 为单位矩阵.

本文将两层图卷积网络相叠加,通过两层 GCN 结构得到第 i 条轨迹的输出特征

$$Z_i^t = f'(V_i^t, \boldsymbol{A}_o^t) = \sigma(\boldsymbol{A}_{oi}^t \sigma(\boldsymbol{A}_o^t V_i^t \boldsymbol{W}_g^0) \boldsymbol{W}_g^1). \tag{11}$$

式中,σ 为激活函数 PReLU;$\boldsymbol{W}_g^0$、$\boldsymbol{W}_g^1$ 为第 1、2 层 GCN 的权重. 场景内的整体输出 Z^t 为

$$Z^t = f(V^t, \boldsymbol{A}_o^t) = \sigma(\boldsymbol{A}_o^t \sigma(\boldsymbol{A}_o^t V^t \boldsymbol{W}_g^0) \boldsymbol{W}_g^1). \tag{12}$$

对图网络输出 Z_i^t 进行 S - LSTM 编码,从而得到轨迹交互特征

$$s_i^t = f_{\text{S-LSTM}}(s_i^{t-1}, Z_i^t, \boldsymbol{W}_s). \tag{13}$$

式中:$f_{\text{S-LSTM}}$ 为编码器 S - LSTM;s_i^{t-1} 为 S - LSTM 在 $t-1$ 时刻的隐藏状态,隐藏层单元个数为 32;$\boldsymbol{W}_s$ 为 S - LSTM 编码权重.

2.3 互信息最大化图网络模块

由于受到周围行人和潜在社交规则的影响,群体中个体的运动模式倾向于场景内所有个体的平均运动模式. 本文使用深度图信息最大化方法[17]最大化 GCN 输出局部特征与全局特征间的互信息,使得局部特征可以获得接近全局特征的向量表示,也就意味着在行人间的交互模型中,每个个体行人学习到了场景内全体行人所共有的运动模式.

本文的深度图信息最大化方法中,局部特征为第 2 层 GCN 的输出特征 Z_i^t,全局特征为所有局部特征 Z^t 的平均值 $\vec{S}^t$,表示为

$$\vec{S}^t = R(Z^t) = \text{softmax}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^t\right), \tag{14}$$

式中 R 为读取函数.

为了使局部特征 Z_i^t 与全局特征 $\vec{S}^t$ 间的互信息最大化,首先利用腐蚀函数 C 在保留原有图结构中顶点的连接方式的前提下添加扰动,对每个顶点的特征进行随机排序,从而制作负样本 $(\tilde{V}_i^t, \boldsymbol{A}_o^t) = C(V_i^t, \boldsymbol{A}_o^t)$;然后负样本通过两层 GCN,得到局部特征 $\tilde{Z}_i^t = f(\tilde{V}_i^t, \boldsymbol{A}_o^t)$;随后使用判别器 D 判别局部特征 Z_i^t、$\tilde{Z}_i^t$ 包含于全局特征 $\vec{S}^t$ 所在范围的概率分数,即判别 $D(Z_i^t, \vec{S}^t)$ 与 $D(\tilde{Z}_i^t, \vec{S}^t)$;最后通过训练使判别器尽可能给正样本打高分且给负样本打低分,完

成对 GCN 特征提取效果的优化，训练判别器的损失函数 L_{inf} 如式(15)所示

$$L_{\text{inf}} = \frac{1}{2n}\Big(\sum_{i=1}^{n} f(\boldsymbol{V},\boldsymbol{A}_{\text{o}})[\log D(Z_i,\vec{S})] + \sum_{j=1}^{n} f(\tilde{\boldsymbol{V}},\boldsymbol{A}_{\text{o}})[\log(1-D(\tilde{Z}_j,\vec{S}))]\Big). \tag{15}$$

2.4　轨迹预测模块

为了表示真实场景中的不确定性，本文从标准正态分布 $N(0,1)$ 中抽取噪声 z[7]，将 z 与轨迹运动特征和轨迹社交特征叠加，得到 T_{obs} 时刻的轨迹解码特征 $d_i^{T_{\text{obs}}} = m_i^{T_{\text{obs}}} \parallel s_i^{T_{\text{obs}}} \parallel z$. 使用 D-LSTM 将 $d_i^{T_{\text{obs}}}$ 解码得到下一时刻的轨迹解码特征 $d_i^{T_{\text{obs}}+1}$，如式(16)所示

$$d_i^{T_{\text{obs}}+1} = f_{\text{D-LSTM}}(d_i^{T_{\text{obs}}}, q_i^{T_{\text{obs}}}, \boldsymbol{W}_{\text{d}}). \tag{16}$$

式中：$f_{\text{D-LSTM}}$ 为解码器 D-LSTM；$q_i^{T_{\text{obs}}}$ 为式(4)中编码过的运动向量，$T_{\text{obs}}+1$ 时刻起所使用的位置为预测位置；$\boldsymbol{W}_{\text{d}}$ 为 D-LSTM 权重. 利用 $d_i^{T_{\text{obs}}+1}$ 计算每个时间点的预测未来轨迹

$$\hat{X}_i^{T_{\text{obs}}+1} = (\hat{x}_i^{T_{\text{obs}}+1}, \hat{y}_i^{T_{\text{obs}}+1}) = \delta(d_i^{T_{\text{obs}}+1}). \tag{17}$$

式中 δ 为线性层. 因为未来轨迹存在多种合理分布，本文使用多样损失函数[7,12]生成多个轨迹样本，进而选取轨迹样本中与真实轨迹间 L2 距离最小的预测轨迹，多样损失函数 L_{variety} 如式(18)所示

$$L_{\text{variety}} = \min_{p} \parallel X_i - \hat{X}_i^{(p)} \parallel_2. \tag{18}$$

式中：X_i 为 $t = T_{\text{obs}}+1,\cdots,T_{\text{final}}$ 时刻真实未来轨迹序列；$\hat{X}_i^p$ 为 $t = T_{\text{obs}}+1,\cdots,T_{\text{final}}$ 时刻的预测序列；p 为训练时生成的样本数，本文使用 $p=20$. 本文在训练神经网络模型的同时，使用深度图信息最大化方法对模型进行优化，最终损失函数如式(19)所示

$$L_{\text{total}} = L_{\text{variety}} + L_{\text{inf}}. \tag{19}$$

3　实验与分析

实验基于 PyTorch 1.1 建立网络模型，使用 Adam 优化器进行参数优化，LSTM 学习率为 0.01，GCN 学习率为 0.03，判别器 D 学习率为 0.001，批处理大小为 64，训练数据集训练轮数为 500，单个 RTX 2 080 Ti GPU 进行训练，生成测试样本数 $N=20$.

本文在公开轨迹预测数据集 ETH[18] 和 UCY[19] 上进行实验，ETH 包含 ETH 和 HOTEL 2 个子数据集，UCY 包含 UNIV、ZARA1 和 ZARA2 3 个子数据集，所有数据集均使用俯拍视角，包含了不同场景中 1 500 多名行人的运动轨迹. 使用世界坐标系，将行人表示为坐标点，获取时间间隔为 0.4 s 的坐标序列. 保留同时存在 n 个目标的序列，即每段序列中行人的数量保持不变. 采用留一法[5]，即在 4 个数据集上进行训练和验证，在剩下的一个数据集上进行测试.

本文使用两种基本评价指标：

1)平均偏移误差(ADE)：全部时间点的预测序列与真实序列间的均方误差，单位为 m.

2)最终偏移误差(FDE)：预测结束时刻的预测序列与真实序列间的误差，单位为 m.

3.1　定量分析

定量分析使用不同算法在相同数据集上进行对比实验，并使用除去了部分模块的本文算法进行消融实验，具体分析如下.

3.1.1　算法对比

为评估 TP-GCN 的准确性，选取了多种对比算法，分别是 S-LSTM[5]、S-Atten[11]、S-GAN[7]、SoPhie[2]、Next[20]、S-ways[10]、Social-BiGAT[14]、STGAT[12]，观测时长 $T_{\text{obs}}=8$(3.2 s)，预测时长 $T_{\text{pred}}=12$(4.8 s)，使用 ADE 和 FDE 进行评价，所有生成多样本轨迹的算法均产生 20 个预测样本，本文算法与其他对比算法在 5 个数据集上的预测精度比较结果见表 1，表中黑体为表现最好的预测结果.

由表 1 可以看出，TP-GCN 在 HOTEL 和 UNIV 数据集上两个指标均优于其他所有算法，并在 5 个数据集的平均 ADE 和 FDE 并列第一. 相较于 SoPhie 和 Next 使用环境信息和行人姿态信息，TP-GCN 仅使用坐标序列信息而没有使用环境信息，更利于在多种场景中泛化；TP-GCN 在 ETH 数据集上效果一般，原因在于 ETH 的测试集较小，各种算法普遍在 ETH 数据集上效果一般，但相较于使用图网络的 Social-BiGAT 和 STGAT，TP-GCN 在 HOTEL、UNIV、ZARA1 和 ZARA2 这 4 个数据集上表现更好，同时取得了良好的稳定性. 与对比算法相比，一方面，本算法使用图卷积神经网络建立交互模式，利用盲区信息筛除错误交互行为的干扰，并且加强了对行人自身运动习惯的挖掘，使算法具有较强的可解释性；另一方面，本算法通过深度图信息最大化方法，使得场景中个体行人与全体行人间的运动模式一致程度更高，从而在多种场景下依然具有较好的鲁棒性. 综上所述，本文算法的总体预测精度较高.

3.1.2　消融实验

为评估 TP-GCN 各个部分的作用，调整多个指定模块，其中算法 1 去掉最大互信息模块，算法 2 去掉邻接矩阵 $\boldsymbol{A}$，算法 3 没有使用盲区信息优化 A，算法 4 单位矩阵系数 $k=0$，算法 5 训练样本数 $p=1$，

测试样本数 $N=1$,算法 6 训练样本数 $p=1$,测试样本数 $N=20$,观测时长 $T_{obs}=8$(3.2 s),预测时长 $T_{pred}=12$(4.8 s),使用 ADE 和 FDE 进行评价,本文算法与调整指定模块后的算法在 5 个数据集上的预测精度比较结果见表 2. 表中黑体为表现最好的预测结果.

表 1 本文算法 TP – GCN 与对比算法的 ADE 和 FDE 比较结果

Tab. 1 Comparison of ADE and FDE results of TP – GCN and other algorithms m

使用算法	ADE						FDE					
	ETH	HOTEL	UNIV	ZARA1	ZARA2	平均值	ETH	HOTEL	UNIV	ZARA1	ZARA2	平均值
S – LSTM[5]	1.09	0.79	0.67	0.47	0.56	0.72	2.35	1.76	1.40	1.00	1.17	1.54
S – Atten[11]	1.39	2.51	1.25	1.01	0.88	1.41	2.39	2.91	2.54	2.17	1.75	2.35
S – GAN[7]	0.81	0.72	0.60	0.34	0.42	0.58	1.52	1.61	1.26	0.69	0.84	1.18
SoPhie[2]	0.70	0.76	0.54	**0.30**	0.38	0.54	1.43	1.67	1.24	0.63	0.78	1.15
Next[20]	0.73	0.30	0.60	0.38	0.31	0.46	1.65	0.59	1.27	0.81	0.68	1.00
S – ways[10]	**0.39**	0.39	0.55	0.44	0.51	0.46	**0.64**	0.66	1.31	0.64	0.92	**0.83**
Social – BiGAT[14]	0.69	0.49	0.55	**0.30**	0.36	0.48	1.29	1.01	1.32	**0.62**	0.75	1.00
STGAT[12]	0.65	0.35	0.52	0.34	0.29	**0.43**	1.12	0.66	1.10	0.69	**0.60**	**0.83**
TP – GCN	0.74	**0.28**	**0.50**	0.33	**0.28**	**0.43**	1.24	**0.51**	**1.07**	0.71	0.61	**0.83**

表 2 本文算法 TP – GCN 在调整指定模块情况下的 ADE 和 FDE 比较结果

Tab. 2 Comparison of ADE and FDE results of TP – GCN with different control settings m

使用算法	ADE						FDE					
	ETH	HOTEL	UNIV	ZARA1	ZARA2	平均值	ETH	HOTEL	UNIV	ZARA1	ZARA2	平均值
算法 1	0.76	0.29	0.50	0.34	0.29	0.44	1.29	0.52	1.07	**0.71**	0.63	0.84
算法 2	0.76	0.29	0.50	0.34	0.29	0.44	1.58	**0.51**	1.07	0.72	0.63	0.90
算法 3	**0.73**	0.29	0.50	0.34	0.29	**0.43**	1.26	0.52	**1.06**	**0.71**	0.63	0.84
算法 4	0.77	0.30	**0.49**	**0.33**	0.29	0.44	1.40	0.53	**1.06**	**0.71**	0.62	0.86
算法 5	0.91	0.38	0.52	0.42	0.34	0.51	1.95	0.73	1.15	0.91	0.76	1.10
算法 6	0.78	0.34	0.51	0.38	0.32	0.47	1.65	0.61	1.11	0.83	0.70	0.98
TP – GCN	0.74	**0.28**	0.50	**0.33**	**0.28**	**0.43**	**1.24**	**0.51**	1.07	**0.71**	**0.61**	**0.83**

由表 2 可以看出,与算法 1 做对比,由于最大互信息模块进行了图网络输出结果的局部特征和全局特征间的互信息最大化,使得受到周围行人交互影响的个体行人运动模式更趋近于周围所有人的平均运动模式,符合环境中集体所默认的潜在社交规则,TP – GCN 的预测结果全面优于对比算法 1. 与算法 2、3、4 做对比,TP – GCN 通过构建基于盲区信息的邻接矩阵并外加单位矩阵构建交互模式,既考虑了周围其他行人直接的交互影响,又提取了自身所受到的隐式交互影响. 3 种对比算法整体表现均不如 TP – GCN,而值得注意的是算法 3 和 4 在 UNIV 数据集中表现优于本文算法,本文理解为由于此数据集中行人远密集于其他数据集并且行人转头环顾四周情况明显增多,周围行人的直接交互影响更为明显,在此情景下本文算法单位矩阵权重过大且盲区范围过大,但另一方面,这也恰恰说明交互权重在密集场景中的重要性. 与算法 5、6 做对比,TP – GCN 考虑了轨迹的多样性和不确定性,预测效果明显优于算法 5 和算法 6,在同为产生 20 个预测样本的情况下比算法 6 的 ADE 提升了 8.5%,FDE 提升了 15.3%. 通过消融实验的对比结果可知,本文所使用算法的预测精度较高.

3.2 定性分析

通过对轨迹序列进行可视化,进一步分析本文所提出算法的可解释性. 从 ZARA2 测试数据集中提取本文算法所使用和生成的轨迹,实线轨迹为观察轨迹,时长为 3.2 s,点划线轨迹为真实未来轨迹,虚线为预测未来轨迹,时长为 4.8 s,轨迹可视化结果见图 4.

从图 4(a)、(b)中可以观察到,在密集行人场景中,处于图像右侧的个体行人自右向左运动,左侧的群体行人并排自左向右运动,此时右侧行人通过,

由于右侧行人经过了左侧行人原本朝向的方向，左侧群体的路径受到了轻微影响. 从图 4(c)、(d)中可观察到，处于相向行走的两组行人，相遇时两组人依照社会规则相互绕行，期间基本没有改变组内个体间的距离关系；另一方面，预测最终时刻行人的盲区范围如图中矩型阴影范围所示，由于此时背向而行的行人位于彼此的盲区之中，没有相互的交互影响，故而视觉盲区中的交互信息被筛除，行人保持原有方向运动. 从图 4(e)、(f)中可以看出，图 4(e)右侧的两名并行的行人和图 4(f)右侧同向而行的3 个行人，受到周围不经过自身路线的行人影响较小，即原本沿近乎直线行走的行人，能够察觉附近的行人不妨碍自身运动时，行人可以保持原有路线运动，这也与人的运动习惯相符. 通过分析预测结果，证明本文算法能够基于交互信息做出与真实行为接近的符合行人习惯的预测.

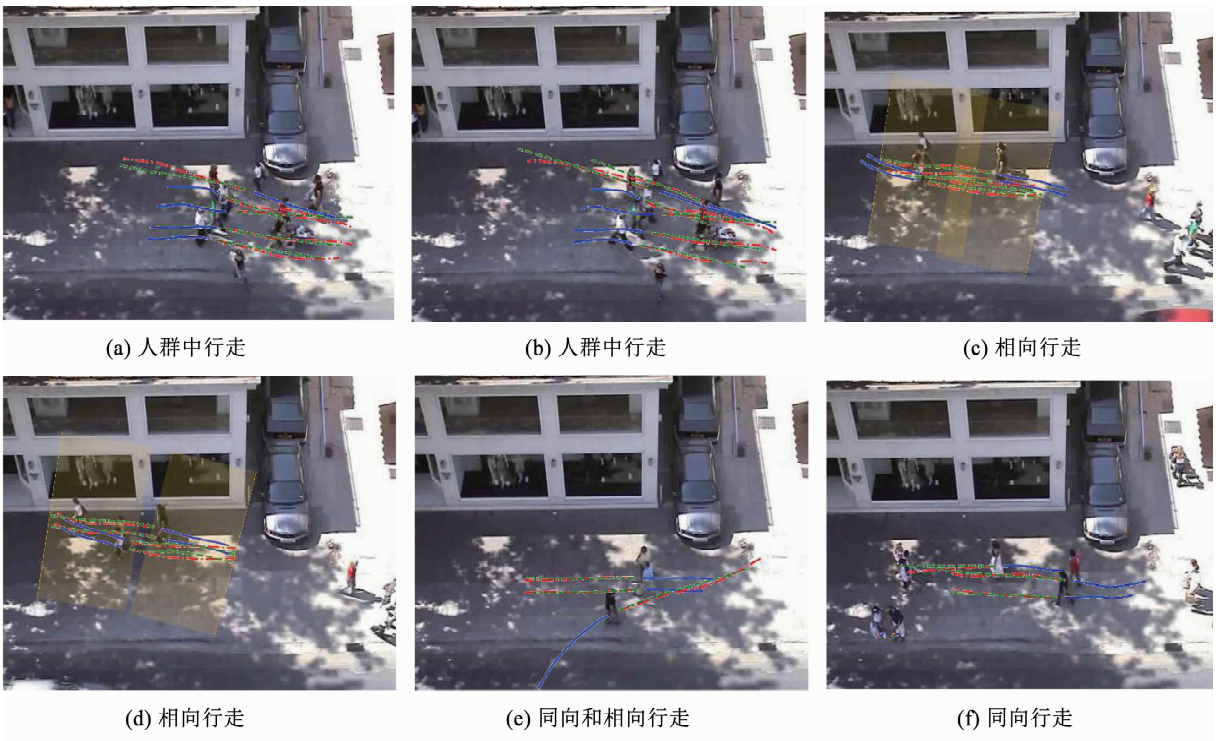

图 4　预测轨迹可视化结果

Fig. 4　Visualization of predicted trajectories

4　结　论

本文提出了一种基于视觉盲区信息和互信息最大化图卷积神经网络的算法 TP－GCN 来建立行人间的交互模式并进行轨迹预测. 该算法克服了图注意力网络构建交互模式不直观的问题，筛除了盲区中行人的交互影响，综合考虑了行人间直接的交互模式和隐含的交互信息，并使得个体运动符合群体运动的社交规则，具有良好的可解释性和泛化性能. 在公开数据集 ETH 和 UCY 上与目前先进的算法进行对比，本文算法的整体预测精度较高，同时消融实验和预测轨迹的可视化也显示了本文算法的有效性及良好的可解释性.

参考文献

[1] HUDNELL M, PRICE T, FRAHM J M. Robust aleatoric modeling for future vehicle localization[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Long Beach: IEEE, 2019: 2944. DOI: 10.1109/CVPRW.2019.00355

[2] SADEGHIAN A, KOSARAJU V, SADEGHIAN A, et al. SoPhie: an attentive GAN for predicting paths compliant to social and physical constraints[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019:1349. DOI: 10.1109/CVPR.2019.00144

[3] CHOI W, SAVARESE S. A unified framework for multi-target tracking and collective activity recognition [C]// European Conference on Computer Vision. Berlin: Springer, 2012: 215. DOI: 10.1007/978－3－642－33765－9_16

[4] PELLEGRINI S, ESS A, SCHINDLER K, et al. You'll never walk alone: modeling social behavior for multi-target tracking[C]// 2009 IEEE 12th International Conference on Computer Vision. Kyoto: IEEE, 2009: 261. DOI: 10.1109/ICCV.2009.5459260

[5] ALAHI A, GOEL K, RAMANATHAN V, et al. Social LSTM: human trajectory prediction in crowded spaces [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 961. DOI: 10.1109/CVPR.2016.110

[6] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735. DOI: 10.1162/neco.1997.9.8.1735

[7] GUPTA A, JOHNSON J, LI Feifei, et al. Social GAN: socially acceptable trajectories with generative adversarial networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 2255. DOI: 10.1109/CVPR.2018.00240

[8] HASAN I, SETTI F, TSESMELIS T, et al. MX-LSTM: mixing tracklets and vislets to jointly forecast trajectories and head poses [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 6067. DOI: 10.1109/CVPR.2018.00635

[9] 张志远,刁英华. 结合社会特征和注意力的行人轨迹预测模型[J]. 西安电子科技大学学报, 2020, 47(1): 10
ZHANG Zhiyuan, DIAO Yinghua. Pedestrian trajectory prediction model with social features and attention[J]. Journal of Xidian University, 2020, 47(1): 10. DOI: 10.19665/j.issn1001-2400.2020.01.002

[10] AMIRIAN J, HAYET J B, PETTRÉ J. Social ways: learning multi-modal distributions of pedestrian trajectories with GANs[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Long Beach: IEEE, 2019: 2964. DOI: 10.1109/CVPRW.2019.00359

[11] VEMULA A, MUELLING K, OH J. Social attention: modeling attention in human crowds [C]//2018 IEEE International Conference on Robotics and Automation (ICRA). Brisbane: IEEE, 2018: 4601. DOI: 10.1109/ICRA.2018.8460504

[12] HUANG Yingfan, BI Huikun, LI Zhaoxin, et al. STGAT: modeling spatial-temporal interactions for human trajectory prediction[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019:6271. DOI: 10.1109/ICCV.2019.00637

[13] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[Z]. arXiv:1710.10903

[14] KOSARAJU V, SADEGHIAN A, MARTÍN-MARTÍN R, et al. Social-BiGAT: multimodal trajectory forecasting using Bicycle-GAN and graph attention networks[Z]. arXiv: 1907.03395

[15] MOHAMED A, QIAN Kun, ELHOSEINY M, et al. Social-STGCNN: a social spatio-temporal graph convolutional neural network for human trajectory prediction[C]//IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 14412

[16] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[Z]. arXiv:1609.02907

[17] VELIČKOVIĆ P, FEDUS W, HAMILTON W L, et al. Deep Graph Infomax[Z]. arXiv:1809.10341

[18] PELLEGRINI S, ESS A, VAN GOOL L. Improving data association by joint modeling of pedestrian trajectories and groupings [C]//European Conference on Computer Vision. Berlin: Springer, 2010:452. DOI: 10.1007/978-3-642-15549-9_33

[19] LEAL-TAIXÉ L, FENZI M, KUZNETSOVA A, et al. Learning an image-based motion context for multiple people tracking[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus: IEEE, 2014: 3542. DOI: 10.1109/CVPR.2014.453

[20] LIANG Junwei, JIANG Lu, NIEBLES J C, et al. Peeking into the future: predicting future person activities and locations in videos [C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 5718. DOI: 10.1109/CVPR.2019.00587

(编辑 苗秀芝)