面向大模型预训练的多模态行人轨迹预测隐私保护方案

魏建好[1],周淳森[2],李 闯[2*],文艳华[2],李克勤[3,4]
(1. 湖南工商大学前沿交叉学院,湖南长沙 410205;2. 湖南工商大学计算机学院,湖南长沙 410205;
3. 湖南大学信息科学与工程学院,湖南长沙 410082;4. 纽约州立大学计算机学院,美国纽约 12561)

摘 要: 在城市级交通大模型应用中,稀疏、异构及强时空关联的多模态行人轨迹数据面临大模型预训练的隐私安全问题. 然而,现有大模型隐私保护方法主要关注单一图像、文本或轨迹模态进行隐私保护,忽视了多模态之间在融合空间中的高维相关结构以及梯度中隐含的跨模态语义泄露风险,容易在模型反推或重构攻击下暴露用户真实轨迹模式和行为偏好,难以有效保护多模态融合数据和模型梯度关联性隐私. 此外,现有大模型注意力机制主要针对密集数据,难以高效处理稀疏的多模态交通数据,导致模型预测精度不高. 因此,本文提出了一种面向大模型预训练的多模态行人轨迹预测隐私保护方案(Privacy-preserving Multimodal Pedestrian Trajectory prediction scheme for Large model pre-training,PMPTL),实现了多模态数据和预训练模型的双重高效保护和高精度预测. 具体而言,创新的设计基于Transformer与Mamba相结合的多模态稀疏轨迹流融合方法(Multimodal Sparse trajectory flow fusion method based on a combination of Transformer and Mamba,MSTM),采用Transformer机制对行人轨迹序列进行全局依赖建模,引入Mamba机制降低长序列建模的复杂度,高效融合稀疏时空特征. 其次,提出基于分辨率网格划分的自适应加权差分隐私方法(Resolution-aware Grid partitioning-based Adaptive weighted Differential Privacy method,RGADP),根据网格分辨率和网格轨迹特征密度动态分配隐私预算,高可用保护融合特征隐私. 接着,提出基于双分支自适应稀疏自注意力机制的多模态特征增强算法(Dual-Branch Adaptive Sparse self-attention mechanism,DBAS),设计双分支自注意力机制,动态调整权重以增强稀疏数据特征表征,确保大模型在稀疏场景下高效表征稀疏轨迹的关键特征,提升预训练效率. 同时,采用自适应时空Top-K稀疏化的高效抖动量化隐私保护方法(Adaptive Spatiotemporal Top-K sparsification with Dithering Quantization method,ASDQ),减少梯度冗余,确保大模型预训练安全性. 最后,基于自适应加权聚合的多模态稀疏行人轨迹预测优化方法(Adaptive Weighted aggregation-based Multimodal sparse Trajectory prediction method,AWMT),对不同模型参数进行动态加权,平衡隐私保护强度与行人轨迹预测精度,以实现高精度轨迹预测. 通过理论分析论证了本文方案满足ϵ-差分隐私保护. 在真实数据集上的实验结果表明,本文方案的预测误差较现有先进方法降低10%,通信效率提升18.43%.

关键词: 大模型隐私保护;自适应差分隐私;行人轨迹预测;多模态数据高效融合;抖动量化;时空特征建模
基金项目: 国家自然科学基金(No.62402177);湖南省自然科学基金(No.2023JJ40237);湖南省教育厅优秀青年基金(No.22B0648);湘江实验室重大项目(No.23XJ01002,No.24XJJCYJ01003,No.22XJ01001)
中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112(2025)12-4376-18
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250638

Privacy-Preserving Multimodal Pedestrian Trajectory Predictio Scheme for Large Model Pre-Training

WEI Jian-hao[1], ZHOU Ting-sen[2], LI Chuang[2*], WEN Yan-hua[2], LI Ke-qin[3,4]
(1. *School of Advanced Interdisciplinary Studies*, *Hunan University of Technology and Business*, *Changsha*, *Hunan* 410205, *China*;
2. *School of Computer Science*, *Hunan University of Technology and Business*, *Changsha*, *Hunan* 410205, *China*;
3. *College of Computer Science and Electronic Engineering*, *Hunan University*, *Changsha*, *Hunan* 410082, *China*;
4. *Department of Computer Science*, *State University of New York*, *New York* 12561, *USA*)

Abstract: Multimodal pedestrian trajectory prediction in city-scale traffic models faces critical challenges including

收稿日期:2025-07-21;录用日期:2025-12-01;责任编辑:覃怀银
*通讯作者:李闯

sparse heterogeneous data with strong spatiotemporal correlations and privacy risks during large model pre-training. However, existing privacy-preserving methods for large models predominantly focus on protecting a single modality, such as images, text, or trajectories, while neglecting the high-dimensional correlation structures among modalities in the fusion space and the risk of cross-modal semantic leakage embedded in the gradients. As a result, these methods are vulnerable to model inversion and reconstruction attacks that can expose users' real trajectory patterns and behavioral preferences, and they fail to effectively protect the privacy of both multimodal fused data and gradient correlations. Moreover, conventional attention mechanisms designed for dense data struggle to efficiently process sparse multimodal traffic features, resulting in suboptimal prediction accuracy. To address these issues, this paper proposes a privacy-preserving multimodal pedestrian trajectory prediction scheme for large model pre-training (PMPTL), achieving dual-efficient protection for both multimodal data and pre-trained models, along with high-accuracy prediction. Specifically, we design an innovative multimodal sparse trajectory flow fusion method based on a combination of Transformer and Mamba (MSTM), where the Transformer mechanism models global dependencies in pedestrian trajectory sequences and the Mamba mechanism is introduced to reduce the complexity of long-sequence modeling, thereby enabling efficient fusion of sparse spatiotemporal features. Secondly, we propose a resolution-aware grid partitioning-based adaptive weighted differential privacy (RGADP) method, which dynamically allocates privacy budgets according to grid resolution and the density of grid-level trajectory features, thereby achieving high-utility protection of fused feature privacy. Next, we propose a multimodal feature enhancement algorithm based on a dual-branch adaptive sparse self-attention mechanism (DBAS). By designing a dual-branch self-attention structure that dynamically adjusts weights to strengthen the representation of sparse data features, DBAS enables the large model to efficiently capture key characteristics of sparse trajectories in sparse scenarios and thereby improves pre-training efficiency. Additionally, an adaptive spatiotemporal Top-K sparsification with dithering quantization (ASDQ) method is introduced to reduce gradient redundancy and ensure secure model training. Finally, we propose an adaptive weighted aggregation-based multimodal sparse trajectory prediction framework (AWMT), which dynamically re-weights different model parameters to balance the strength of privacy protection and the accuracy of pedestrian trajectory prediction, thereby achieving high-precision trajectory forecasting. Theoretical analysis demonstrates that our scheme satisfies ϵ-DP protection. Experimental results on two real-world datasets show that our scheme reduces prediction error by 10% compared to state-of-the-art approaches and improves communication efficiency by 18.43%.

Key words: privacy-preserving large model; adaptive differential privacy; pedestrian trajectory prediction; efficient multimodal data fusion; dithering quantization; spatiotemporal feature modeling

Foundation Item(s): National Natural Science Foundation of China (No.62402177); Natural Science Foundation of Hunan Province (No.2023JJ40237); Outstanding Youth Project of Hunan Provincial Education Department (No.22B0648); Major Program Project of Xiangjiang Laboratory (No.23XJ01002, No.24XJJCYJ01003, No.22XJ01001)

1　引言

随着城市大模型(如GPT、DeepSeek)在智慧交通中深度应用,行人轨迹预测[1~3]依托大模型对稀疏交通数据的强表征能力,提升了交通决策效率. 传统单模态方法无法充分建模行人的时空动态特征,难以支撑大模型的高精度推理. 因此,结合视觉和传感器等多模态数据融合的大模型预训练框架,显著强化了预测鲁棒性. 由于端云计算的大模型具有高计算需求与隐私泄露风险,导致现有端云计算平台难以满足大模型实时性和隐私保护的要求. 因此,云边端协同的大模型预训练框架应运而生[4],通过终端多模态数据融合处理、边缘大模型执行低延迟推理和云边进行模型协同训练,确保智慧交通系统安全高效决策.

然而,云边端协同的大模型预训练框架在执行多模态行人轨迹预测时,仍然面临严重的隐私安全问题[5]. 首先,终端设备将多模态数据发送至边缘层时,不可信边缘大模型可能通过链接攻击手段推断出用户敏感信息,如行人轨迹、兴趣点. 其次,在云边进行大模型协同预训练时,攻击者可能利用大模型预训练参数进行反向工程. 例如,在处理个性化数据(如用户图像信息或位置数据)时,攻击者用反向工程攻击方法可以恢复出与原始训练数据高度相似的内容,导致隐私泄露[6,7]. 因此,如何确保多模态交通数据和预训练模型参数的安全性,同时提高行人轨迹预测精度,是智慧交通大模型领域亟待解决的关键问题.

大模型预训练的多模态交通数据具有高稀疏性、高维性和复杂的时空关联性,会反映交通用户隐私,这要求终端层设备能够在本地对融合的多模态数据进行隐私保护[8]. 现有隐私保护技术主要包括匿名方法[9]、加密方法[10]和差分隐私机制[5,6,11]. 然而,匿名方法不能抵御链接攻击,隐私保护水平有限. 加密方法在进行大模型数据处理时的开销较大,并限制了交通数据的

共享性. 而差分隐私方法虽能解决两者的缺陷,但在高稀疏数据上会引入较大噪声,降低融合数据的可用性,影响大模型训练精度. 此外,在智慧交通大模型中,基于Transformer的自注意力融合机制[12,13]虽然提高了融合精度,但计算复杂度较高,难以高效处理大规模交通数据. 因此,如何对高稀疏的多模态交通数据进行高效融合保护是本文的首要挑战.

为提升大模型预训练的精度,边缘大模型需要挖掘扰动的多模态稀疏融合数据的行为模式,增强稀疏数据的表征能力[14]. 稀疏交通数据特征通常集中在特定区域,而大部分为噪声数据[15]. 然而,现有密集型Transformer的大模型预训练方法考虑所有数据特征,会引入大量冗余计算和无关噪声,降低了整体效率. 现有少数稀疏注意力机制[14,15]难以适应稀疏特征的动态分布差异,影响模型性能. 此外,为确保大模型预训练参数的安全性,现有大模型保护方法[16]主要结合联邦学习和差分隐私,采用固定Top-K的模型梯度进行扰动保护,但未考虑模型梯度的高时空动态关联性,导致梯度冗余,增加通信开销. 因此,如何增强预训练模型的安全高效性是本文的第二个核心挑战.

由于大模型训练数据和模型梯度都经过扰动保护,这虽然增强了大模型安全性,但也降低了预测精度. 现有预测方法[2,3]主要依赖真实数据,预测精度易受噪声影响. 此外,不同边缘大模型梯度不同. 现有全局模型聚合机制[17]对扰动的局部模型梯度进行平均,难以挖掘不同梯度对全局模型的影响,导致预测模型鲁棒性不足. 因此,如何设计一种高可用性的多模态行人轨迹预测方法,平衡多模态稀疏数据隐私与大模型预测精度,是本文的核心挑战.

为解决上述挑战,本文首次提出一种云边端框架下面向大模型预训练的多模态行人轨迹预测隐私保护方案(Privacy-preserving Multimodal Pedestrian Trajectory prediction scheme for Large model pre-training,PMPTL)针对第一个挑战,设计基于Transformer与Mamba相结合的多模态稀疏轨迹流融合方法(Multimodal Sparse trajectory flow fusion method based on a combination of Transformer and Mamba,MSTM),高效融合大模型预训练数据特征并保持时空一致性. 同时,提出基于分辨率网格划分的自适应加权差分隐私方法(Resolution-aware Grid partitioning-based Adaptive weighted Differential Privacy method,RGADP),考虑网格轨迹特征分布粒度,自适应分配隐私预算,高可用保护融合特征隐私. 为解决第二个挑战,设计一种基于双分支自适应稀疏自注意力机制的多模态特征增强算法(Dual-Branch Adaptive Sparse self-attention mechanism,DBAS),高效表征关键特征,过滤低查询-键匹配分数的噪声特征,提升大模型预训练的精度. 同时,设计自适应时空Top-K稀疏化的高效抖动量化隐私保护方法(Adaptive Spatiotemporal Top-K sparsification with Dithering Quantization method,ASDQ),考虑梯度的时空相关性,减少梯度冗余和提高通信效率,对模型梯度进行高效动态保护. 面对第三个挑战,设计基于自适应加权聚合的多模态稀疏行人轨迹预测优化方法(Adaptive Weighted aggregation-based Multimodal sparse Trajectory prediction method,AWMT),挖掘不同梯度的贡献度进行自适应加权聚合,实现高准确度预测. 通过理论安全性和性能分析,验证了本文PMPTL方案满足高水平的差分隐私保护,并显著降低了计算和通信开销.

本文主要的工作贡献如下:

(1)创新地提出一种强隐私保护、自适应特征增强与高效预训练的行人轨迹预测方案PMPTL,有效平衡模型预测精度和隐私保护水平.

(2)面对稀疏非均衡的多模态行人轨迹数据,提出一种高可用的多粒度自适应隐私保护机制. 考虑稀疏轨迹网格的分辨率动态分配隐私预算,并设计时空感知的Top-K稀疏抖动量化方法,对关键模型梯度动态高效保护,增强模型安全性和降低通信开销.

(3)考虑稀疏轨迹融合特征的高动态性,提出一种面向稀疏时空特征的多模态高效处理方法. 构建融合Transformer与Mamba的多模态稀疏特征表征模型,设计双分支自适应稀疏注意力机制,增强关键信息表达,引入自适应加权聚合策略,提升模型预测精度.

(4)通过理论安全性分析证明本文方案满足差分隐私约束. 实验结果表明,在强隐私保护下,本文方案预测误差比现有方法低10%,通信效率提高18.43%.

2 相关工作

2.1 多模态行人轨迹预测研究现状

多模态数据融合. 交通数据具有高稀疏性、强时空关联性,终端设备需高效融合这些数据以挖掘行人轨迹特征. 为此,研究者提出了基于Transformer的大模型自注意力融合机制[12,13],融合了多模态数据. 然而,Transformer融合方法具有$O(N^2)$的高计算复杂度,在处理大规模稀疏数据时计算效率较低.

大模型预训练. 多模态数据融合后,将发布至边缘层以支持大模型的深度学习预训练. 文献[11]提出一种面向多模态的预训练模型,通过多模态分组对齐捕捉多模态预训练数据之间的相关性. Aydar等人[18]通过为预训练的Transformer增加记忆机制,处理长时序序列. Zeng等人[19]提出一种多模态预训练框架,提高图像和文本的对齐能力. 然而,现有预训练方法主要关注密集型多模态数据,会考虑所有区域的多模态特征,难以对关键特征进行识别,导致大模型预训练效率和精度较低.

行人轨迹预测. 面对大规模稀疏交通数据,研究者

通常采用深度学习进行空间建模以实现高效预测. 例如,文献[1]提出一种基于扩散模型的通用轨迹预测框架,实现不同轨迹数据的统一表示,但难以捕捉稀疏数据的相互依赖. 文献[2]通过采用双Transformer架构,提高行人轨迹预测的准确性. 文献[3]提出一种基于记忆增强神经网络的轨迹预测模型,预测未来轨迹. 然而,这些方法难以挖掘多模态交通数据的时空关联性,且计算复杂度高. 此外,文献[20]提出一种基于Transformer的非自回归行人轨迹预测模型,解决序列长期依赖问题. 然而,上述这些方法大多依赖真实交通数据,没有考虑大模型预训练的多模态数据和梯度被扰动保护时,对预测精度的影响.

2.2　大模型预训练的隐私保护研究现状

大模型预训练的多模态数据融合隐私保护. 差分隐私方法通过向交通数据添加噪声提供更安全的保护. 然而,现有差分隐私技术也面临挑战. 例如,文献[5]提出了一种SPEDP(Sensitivity-aware Personalized Edge Differential Privacy model)隐私保护方法,提供细致的隐私控制. 然而,该方法在处理稀疏数据时会产生较大噪声误差. Liu等人[6]提出了一种基于边关系的本地差分隐私方案,利用三角计数估计器添加噪声. 文献[11]提出了基于差分隐私的轨迹相关性隐私保护机制,保护多用户轨迹数据的相关性. 但上述方法的计算复杂度高且通信效率低. 此外,现有方法主要关注单模态数据隐私,直接扰动多模态稀疏交通数据,会破坏模态间的关联性,导致大模型预训练精度不高.

大模型预训练梯度隐私保护. 在云边协同大模型训练中,攻击者通过分析模型梯度可推断模型结构和训练数据隐私. 为此,研究者提出使用差分隐私技术对梯度进行保护. 例如,文献[21]提出基于差分隐私的优化联邦学习方法. 文献[22]提出了一种基于去中心化联邦学习的差分隐私算法. 文献[17]提出一种个性化联邦学习框架DP-PFL(Differential Privacy based Personalized Federated Learning),在保护用户隐私的同时,生成个性化模型. 然而,这些方法在处理稀疏交通数据时仍存在通信效率低、系统开销大的问题,难以满足大模型在实际智慧交通环境中的实时性需求. 为解决这一问题,Wang等人[16]提出了一种面向稀疏抖动量化的差分隐私算法. 然而,该方法难以根据交通数据的变化来动态调整稀疏化程度,且没有考虑梯度间的时空相关性,导致梯度冗余的产生.

3　问题定义

3.1　系统模型

如图1所示,本文系统模型由三个层次组成:终端层、边缘层和云层. 在终端层,用户结合Transformer与

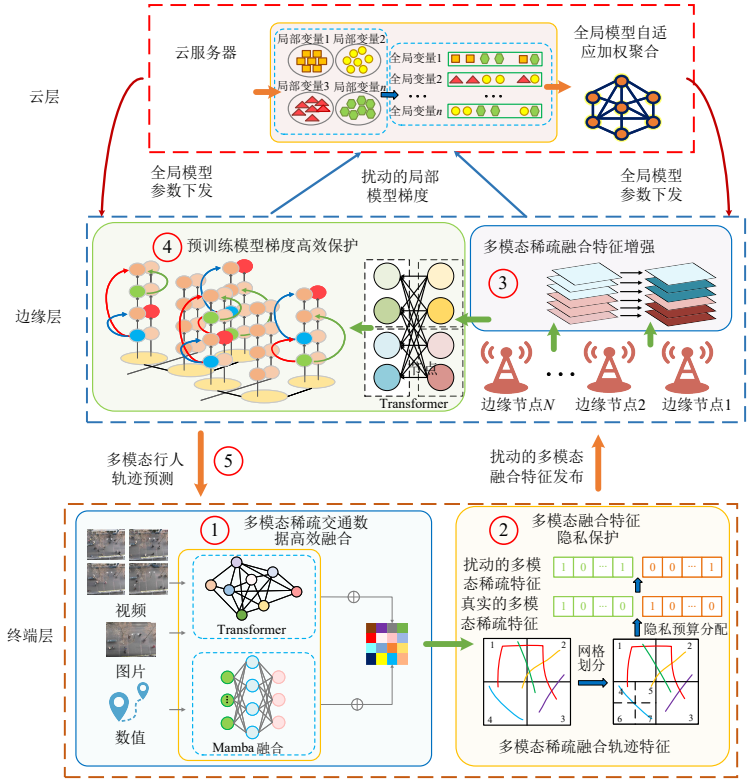


图1　系统模型图

Mamba机制提取多模态特征，实现多模态稀疏交通数据高效融合（步骤1）. 然后，基于RGADP方法，对多模态稀疏融合特征进行自适应加权差分隐私保护，并发布给边缘层（步骤2）. 在边缘层，大模型采用DBAS机制，对扰动的多模态稀疏融合特征进行表征增强，过滤噪声特征并减少冗余（步骤3）. 接着，基于增强的多模态稀疏特征，边缘层采用ASDQ方法对大模型预训练梯度进行高效保护，并上传扰动的局部梯度给云层（步骤4）. 最后，云服务器对局部梯度进行自适应加权聚合，下发给边缘层进行模型更新，执行高准确度的多模态行人轨迹预测（步骤5）.

3.2 设计目标

在系统模型中，终端用户是可信的，边缘大模型和云服务器是不可信的，其会窥探用户数据隐私. 为保护大模型预训练的多模态稀疏数据和梯度隐私，需对融合特征和模型梯度进行差分隐私扰动. 然而，双重噪声会增加轨迹预测的复杂性并降低准确性. 因此，本文的设计目标为：(1)多模态稀疏数据和模型梯度的高效保护；(2)多模态行人轨迹预测的高准确性.

目标1：多模态稀疏交通融合数据隐保护. 定义多模态交通数据的本地差分隐私机制如下.

定义1 ϵ^1-本地差分隐私[5,6]. 对任意两个多模态融合嵌入$\boldsymbol{x}_1$和$\boldsymbol{x}_2$，给定隐私预算ϵ^1，对于输出$\boldsymbol{y}$，隐私机制$\xi(\cdot)$实现ϵ^1-本地差分隐私，当且仅当满足：

$$\Pr\left[\xi\left(\boldsymbol{x}_1\right)=\boldsymbol{y}\right]\leqslant \mathrm{e}^{\epsilon^1}\Pr\left[\xi\left(\boldsymbol{x}_2\right)=\boldsymbol{y}\right] \tag{1}$$

其中，隐私预算ϵ^1决定隐私保护水平. 较小的值提供更强的隐私保护，但会引入更多的噪声，降低可用性.

目标2：预训练模型梯度隐私保护. 为提升模型保护效率，设计轻量级的差分隐私保护机制如下.

引理1 高斯机制-差分隐私[16]. 对于查询敏感度Δf，对于$\forall\delta\in(0,1)$，如果$\sigma\geqslant\sqrt{2\ln(1.25/\delta)}\,\Delta f/\epsilon^2$且噪声$Y\sim N\left(0,\boldsymbol{I}_m\sigma^2\right)$，其中$\boldsymbol{I}_m$表示$m\times m$的单位矩阵，则高斯机制$M(D)=f(D)+Y$满足$\left(\epsilon^2,\delta\right)$-差分隐私.

目标3：多模态稀疏行人轨迹预测的高准确性. 设计多模态行人轨迹预测的最小化损失函数$\mathcal{L}$为

$$\mathcal{L}=\lambda_1\mathcal{L}_{\mathrm{lt}}+\lambda_2\mathcal{L}_{\mathrm{clf}} \tag{2}$$

其中，λ_1和λ_2是损失函数权重；$\mathcal{L}_{\mathrm{lt}}=1/n\sum_{i=1}^{n}(\boldsymbol{y}_i-\hat{\boldsymbol{y}}_i)^2$是$L_2$损失函数；$\mathcal{L}_{\mathrm{clt}}=-\left[\boldsymbol{y}\cdot\log(\hat{\boldsymbol{y}})+(1-\boldsymbol{y})\cdot\log(1-\hat{\boldsymbol{y}})\right]$是交叉熵损失函数.

4 本文PMPTL方案

4.1 基于Transformer与Mamba相结合的多模态稀疏轨迹流融合方法

大模型预训练的交通数据呈现高稀疏性、强关联性和高维性，现有方法[12,13]难以表征交通数据的时空一致性且计算开销较大，影响模型的鲁棒性和效率. 因此，本文提出MSTM，不同于传统直接使用Transformer或Mamba的方法，本文创新地将Transformer的全局建模与Mamba的高效局部时空表征相结合，针对稀疏轨迹数据设计模态对齐重用模块与时空一致性建模机制，提升关键时间步的敏感性与预测精度，实现高效融合，如图2所示.

时空一致性建模. 为解决现有编码器在时空联合建模方面的不足，本文创新性地提出一种视频-图像编码器，在Transformer中加入卷积层（Convolutional，Conv），卷积核大小为3×3，通过局部时间（Local Time，LT）建模模块建模，捕捉时间动态变化. 在多头注意力层（Multi-Head Self-Attention，MHSA）和前馈层（Feed-Forward Network，FFN）进行空间建模. 该设计能同时捕获局部细粒度特征与全局语义依赖，而循环结构（如Long Short-Term Memory，LSTM）难以实现. 具体过程如图2所示.

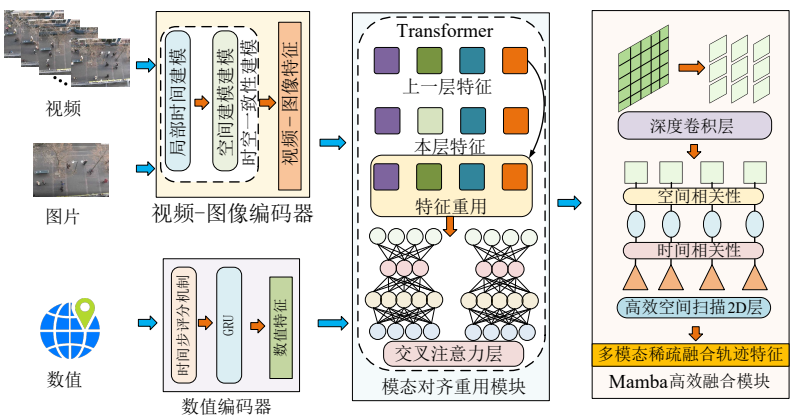


图2 本文MSTM方法的多模态高效融合流程图

$$\mathrm{VI}_{\mathrm{conv}}^{n-1}=\mathrm{Conv}\left(\mathrm{VI}^{n-1}\right) \tag{3}$$

$$\mathrm{VI}_{\mathrm{MHSA}}^{n}=\mathrm{LN}\left(\mathrm{MHSA}\left(\mathrm{VI}_{\mathrm{LT}}^{n}\right)+\mathrm{VI}_{\mathrm{LT}}^{n}\right) \tag{4}$$

$$\mathrm{VI}^{n}=\mathrm{LN}\left(\mathrm{FFN}\left(\mathrm{VI}_{\mathrm{MHSA}}^{n}\right)+\mathrm{VI}_{\mathrm{MHSA}}^{n}\right) \tag{5}$$

其中,LN是层归一化函数;VI表示视频-图像特征编码器函数;y表示层数.当$n=1$时,编码器的初始输入$\mathrm{VI}^{0}=\mathrm{Embed(Input)}\in\mathbb{R}^{B\times 256\times D_{\mathrm{model}}}$.其中,Embed是初始嵌入层;Input为嵌入输入;B为批次大小;D_{model}为隐藏层维度;$B\times 256\times D_{\mathrm{model}}$为每一层的向量维度.当$n\geqslant 2$时,$\mathrm{VI}^{n}=F^{n}(\mathrm{VI}^{n};\ \varphi^{n})\in\mathbb{R}^{B\times 256\times D_{\mathrm{model}}}$.其中,$F^{n}$是第$n$层的变换函数;$\varphi^{n}$是第$n$层的参数.局部时间建模模块基于多组融合策略提取特征,将视频帧按时间分组,每组处理同一空间位置在不同时间步的特征.

针对高维稀疏的行人轨迹时序数据,现有门控循环单元(Gated Recurrent Unit,GRU)编码器难以充分捕捉其中关键时间步的特征.本文在GRU中创新性地设计时间步评分机制,用于动态评估每个时间步的输入-状态对的重要性.该机制被嵌入到GRU的门控过程中,通过调节更新门、重置门及候选隐藏状态的权重分配,增强模型对轨迹关键变化点的感知能力.时间步评分机制的具体形式如下:

$$\boldsymbol{\alpha}_{t^{e}}=\mathrm{Soft\,max}\left(\boldsymbol{W}_{a}\cdot[\boldsymbol{h}_{t^{e}-1},\boldsymbol{x}_{t^{e}}]/\sqrt{d}\right) \tag{6}$$

其中,$\boldsymbol{x}_{t^{e}}$、$\boldsymbol{h}_{t^{e}-1}$分别是GRU网络的输入和上一时刻隐藏状态;d是输入数值的维度;$\boldsymbol{W}_{a}$是评分权重矩阵.将时间步评分$\boldsymbol{\alpha}_{t^{e}}$嵌入GRU中生成数值特征$\boldsymbol{H}$.

模态对齐重用模块.为突破现有Transformer计算开销大的问题,本文创新地设计了一种新的模态对齐重用模块,添加到自注意力(Self-Attention,SA)层,进行高效特征重用.随后,视频-图像特征$\mathrm{VI}_{\mathrm{SA}}^{i}$和数值特征$\boldsymbol{H}_{\mathrm{SA}}^{i}$输入到交叉注意力(Cross-Attention,CA)层中以对齐语义.具体过程如下:

$$\mathrm{VI}_{\mathrm{SA}}^{i}=\mathrm{LN}\left(\mathrm{SA}(\mathrm{VI}^{i-1})+\mathrm{VI}^{i-1}\right) \tag{7}$$

$$\boldsymbol{H}_{\mathrm{SA}}^{i}=\mathrm{LN}\left(\mathrm{SA}(\boldsymbol{H}^{i-1})+\boldsymbol{H}^{i-1}\right) \tag{8}$$

$$S_{\mathrm{VI}}=\mathrm{VI}_{\mathrm{SA}}^{i}\cdot\mathrm{VI}^{i-1}/\left(\|\mathrm{VI}_{\mathrm{SA}}^{i}\|_{2}\|\mathrm{VI}^{i-1}\|_{2}\right) \tag{9}$$

$$S_{H}=H_{\mathrm{SA}}^{i}\cdot\boldsymbol{H}^{i-1}/\left(\|\boldsymbol{H}_{\mathrm{SA}}^{i}\|_{2}\|\boldsymbol{H}^{i-1}\|_{2}\right) \tag{10}$$

$$\mathrm{VI}_{\mathrm{CA}}^{i}=\mathrm{LN}\left(\mathrm{CA}(\mathrm{VI}_{\mathrm{SA}}^{i},\mathrm{VI}^{n})+\mathrm{VI}_{\mathrm{SA}}^{i}\right) \tag{11}$$

其中,S_{VI}和S_{H}分别是该层与上一层视频-图像特征以及数值特征间的相似度.给定阈值θ,如果$|S_{\mathrm{VI}}|\geqslant\theta$或$|S_{H}|\geqslant\theta$,则表示该层特征和上一层特征相似,上一层特征会被复用,以降低计算复杂度.随后,将模态对齐重用模块的输出与原始特征通过交叉注意力层,生成视图特征$\mathrm{VI}^{\mathrm{CV}}$和数值特征$\boldsymbol{H}^{\mathrm{CM}}$.

Mamba高效融合模块.为有效捕捉行人轨迹数据中的时空特征,本文创新性地设计了一种具有时空关联性的Mamba模块,通过深度卷积(Depthwise convolution,Dwc)与高效空间扫描2D层(Efficient Spatial scanning 2D,ES2D)提取局部时空特征,以弥补Transformer细粒度局部建模的不足.首先,将不同模态的特征$(\mathrm{VI}^{\mathrm{CV}},\boldsymbol{H}^{\mathrm{CM}})$进行混合,生成混合特征VIH:

$$\begin{aligned}\mathrm{VIH}=&\mathrm{Dwc}\left(\mathrm{Linear}(\mathrm{VI}^{\mathrm{CV}})\right)\\&\otimes\mathrm{Dwc}\left(\mathrm{Linear}(\boldsymbol{H}^{\mathrm{CM}})\right)\oplus\mathrm{VI}^{\mathrm{CV}}\oplus\boldsymbol{H}^{\mathrm{CM}}\end{aligned} \tag{12}$$

其中,Dwc(·)为深度卷积操作;Linear为线性变换层.LN(·)为层归一化,将混合特征输入到ES2D中,捕捉混合特征的时空相关性:

$$\overline{\mathrm{VIH}_{1}}=\mathrm{LN}\left(\mathrm{ES2D(VIH)}\right)\otimes\mathrm{Linear}(\mathrm{VI}^{\mathrm{CV}}) \tag{13}$$

$$\overline{\mathrm{VIH}_{2}}=\mathrm{LN}\left(\mathrm{ES2D(VIH)}\right)\otimes\mathrm{Linear}(\boldsymbol{H}^{\mathrm{CM}}) \tag{14}$$

$$\overline{\mathrm{VIH}_{a}}=\overline{\mathrm{VIH}_{1}}\oplus\overline{\mathrm{VIH}_{2}} \tag{15}$$

将时空特征$\overline{\mathrm{VIH}_{a}}$输入通道注意力(Efficient Channel Attention,ECA)模块中,得到多模态稀疏融合轨迹特征:

$$\boldsymbol{X}=\mathrm{ECA}\left(\mathrm{Linear}\left(\overline{\mathrm{VIH}_{a}}\right)\right)\oplus\mathrm{ECA}\left(\mathrm{VI}^{\mathrm{CV}}\oplus\boldsymbol{H}^{\mathrm{CM}}\right) \tag{16}$$

4.2　基于分辨率网格划分的自适应加权差分隐私方法

大模型预训练的多模态融合特征需要隐私保护.面对高稀疏、分布不均衡的多模态行人轨迹数据,现有差分隐私技术[8,11]采用固定网格结构和相同隐私预算,没有考虑网格分辨率与轨迹特征密度,难以提供自适应的隐私保护,导致可用性较低.因此,本文提出RGADP,考虑网格轨迹特征密度和网格分辨率特性,自适应分配隐私预算,实现多模态稀疏数据隐私保护的高可用性,如图3所示.

隐私预算自适应加权分配.影响隐私预算的因素

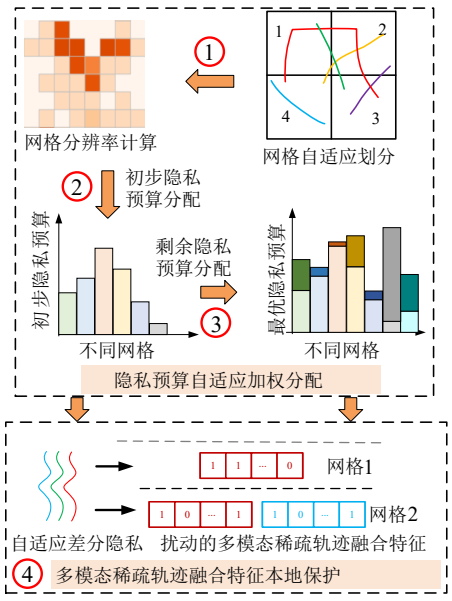


图3　本文RGADP方法的多模态稀疏融合数据保护流程图

主要包括网格轨迹特征密度和网格分辨率. 假设网格结构表示为 G,它包含一组网格单元 G_i,i 表示网格编号,计算网格轨迹特征密度 n_i 为

$$n_i = x_i^n \Big/ \sum_j x_j^n \tag{17}$$

其中,x_i^n 表示第 i 个网格单元的特征数量;$\sum_j x_j^n$ 表示整个网格区域的特征数量总和,通过归一化处理,每个网格单元的特征占比被统一,避免长尾分布问题.

其次,计算网格分辨率. 给定表示编号 i 的网格大小 s_i,得到网格分辨率为

$$r_i = n_i / s_i \tag{18}$$

由于 s_i 通常较小,最终网格分辨率 r_i 的相对值仍然受控,不会无限放大. 接着,为了平衡隐私保护强度与计算开销,在高密度区域执行更多迭代,在稀疏区域限制不必要迭代,则网格划分的最大迭代次数为

$$i_{\max} = \left\lceil \log_d \left(\|\boldsymbol{X}\| \Big/ \sum_i \gamma n_i \right) \Big/ 2 \right\rceil \tag{19}$$

其中,d 是子单元的划分大小;$\boldsymbol{X}$ 是多模态融合轨迹特征;γ 是一个调节因子. 每次迭代的隐私预算为 $\epsilon_c^1 = \epsilon^1 / i_{\max}$,其中 ϵ^1 是预定义的隐私预算;c 为迭代次数. 因此,每个网格 G_i 的初步分配隐私预算为

$$\epsilon_i^1 = r_i \times \epsilon_c^1 \tag{20}$$

给定网格总数 $|G|$,则剩余隐私预算为

$$\bar{\epsilon}^1 = \epsilon^1 - \sum_{i=1}^{|G|} \epsilon_i^1 \tag{21}$$

为充分利用剩余隐私预算,本文进一步考虑结合网格轨迹特征密度与网格分辨率. 采用加权向量 $\boldsymbol{w}$ 将剩余隐私预算自适应分配给网格,计算权重值 w_i 为

$$w_i = (r_i \times n_i) \Big/ \sum_{j=1}^{|G|} r_j \times n_j \tag{22}$$

多模态稀疏轨迹融合特征本地保护. 对划分网格的多模态稀疏融合特征 $\boldsymbol{x}_i$ 添加拉普拉斯噪声如下:

$$\widetilde{\epsilon_i^1} = |w_i| \times \bar{\epsilon}^1 \tag{23}$$

$$\widetilde{\boldsymbol{x}_i} = \boldsymbol{x}_i + \alpha \times \mathrm{Lap}\left(1/\widetilde{\epsilon_i^1}\right) + \beta \times \mathrm{Lap}\left(1/\epsilon_i^1\right) \tag{24}$$

其中,α 和 β 为权重参数. 可应用于大模型中的区域隐私差异化和多区域隐私保护. 最后,终端交通用户向边缘大模型上传扰动的多模态稀疏轨迹融合特征 $\tilde{\boldsymbol{X}}$.

4.3 基于双分支自适应稀疏自注意力机制的多模态特征增强算法

大模型预训练的轨迹数据特征具有高稀疏性和动态性,传统密集型 Transformer 会增加计算复杂度. 然而,现有稀疏注意力机制采用固定稀疏策略,如基于 Softmax 和修正线性单元(Rectified Linear Unit, ReLU)策略,难以适应稀疏行人轨迹特征的动态变化,存在稀疏区信息稀释、密集区信息冗余,导致模型训练性能不佳[14,15]. 因此,本文提出基于双分支自适应稀疏自注意力机制的多模态特征增强算法(DBAS),设计新的基于平方 LeakyReLU 的自注意力机制,动态增强特征表征,确保大模型在稀疏场景精准捕捉关键特征,如图 4 所示.

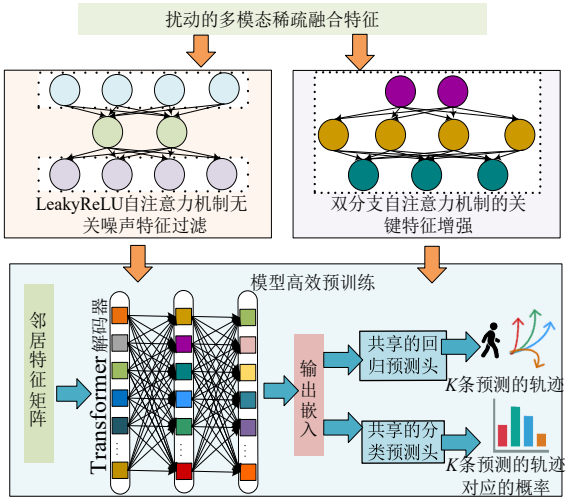


图 4 本文 DBAS 方法的多模态特征增强流程图

LeakyReLU 自注意力机制无关噪声特征过滤. 本文提出基于平方 LeakyReLU 的自注意力机制,过滤低查询-键匹配分数的无关噪声特征,缓解基于平方 ReLU 的激活函数的神经元死亡问题[23]. 首先,基于扰动的稀疏融合特征生成查询 $\boldsymbol{Q}$、键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$ 矩阵:

$$\boldsymbol{Q} = \boldsymbol{X}\boldsymbol{W}_Q, \boldsymbol{K} = \boldsymbol{X}\boldsymbol{W}_K, \boldsymbol{V} = \boldsymbol{X}\boldsymbol{W}_V \tag{25}$$

随后,计算注意力分数为

$$\boldsymbol{A} = f(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}} / \sqrt{d} + b)\boldsymbol{V} \tag{26}$$

其中,b 表示相对网格位置偏差;$f(\cdot)$ 为评分函数. 然后,计算稀疏自注意力(Learned Sparse Self-Attention, LSSA)分数如下:

$$\mathrm{LSSA} = \mathrm{Leaky\,ReLU}^2(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}} / \sqrt{d} + b) \tag{27}$$

其中,当 LSSA 值较低时,表示该特征与目标特征的相似性较低(约等于 0),则过滤掉该无关噪声特征.

双分支自注意力机制的关键特征增强. 考虑到单纯使用 LeakyReLU 可能导致特征过度稀疏,本文引入了密集自注意力以获得密集自注意力(Dense Self-Attention, DSA)分数:

$$\mathrm{DSA} = \mathrm{SoftMax}\left(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}} / \sqrt{d} + b\right) \tag{28}$$

采用双分支自注意力机制,计算最优的自适应注意力分数为

$$\boldsymbol{A} = (p_1 \times \mathrm{LSSA} + p_2 \times \mathrm{DSA})\boldsymbol{V} \tag{29}$$

其中,p_1和p_2是用于自适应调节双分支的两个归一化权重. 计算权重p_1和p_2为

$$p_n=\mathrm{e}^{a_n+\|r_i\|}\Big/\sum_{i=1}^{|G|}\mathrm{e}^{a_n+\|r_i\|},n=\{1,2\} \tag{30}$$

其中,a_1和a_2是可学习的参数. 该设计在过滤无关噪声特征和充分挖掘信息特征之间实现了更好的权衡.

大模型高效预训练. 将第n个邻居的观察轨迹表示为$\boldsymbol{X}_n$. 提取$\boldsymbol{X}_n$中多模态稀疏特征矩阵$\hat{\boldsymbol{X}}_n$,通过线性变换$\phi(\cdot,\cdot)$获得Transformer解码器的嵌入$\boldsymbol{E}_n$为

$$\boldsymbol{E}_n=\phi\left(\hat{\boldsymbol{X}}_n,\boldsymbol{W}_n\right) \tag{31}$$

其中,$\boldsymbol{W}_n$是可学习的参数矩阵. 随后,采用回归预测头和分类预测头分别预测多模态的未来轨迹和对应的概率. 计算真实值Y与L个聚类中心$C=\{c_1,c_2,\cdots,c_L\}$之间的距离,获得距离最近的聚类中心c_i为

$$i=\underset{i\in\{1,2,\cdots,L\}}{\operatorname{argmin}}\left(\|Y-c_i\|_2^2\right) \tag{32}$$

接着,采用最近邻假设,通过最近聚类中心c_i,得到预测轨迹的实际概率p_{traj}为归一化的负距离:

$$p_{\mathrm{traj}}=\mathrm{SoftMax}\left(\left\{-\|Y-c_i\|_2^2,i\in\{1,2,\cdots,L\}\right\}\right) \tag{33}$$

最后,预测未来轨迹$\hat{Y}$和对应的预测概率$\hat{p}_{\mathrm{traj}}$. 计算大模型训练的损失函数如下:

$$\mathcal{L}=\lambda_1\mathcal{L}_{\mathrm{lt}}(Y,\hat{Y})+\lambda_2\mathcal{L}_{\mathrm{clf}}(p_{\mathrm{traj}},\hat{p}_{\mathrm{traj}}) \tag{34}$$

其中,λ_1和λ_2被用来平衡损失函数;$\mathcal{L}_{\mathrm{lt}}$是$L_2$损失函数;$\mathcal{L}_{\mathrm{clt}}$为交叉熵损失函数,用于大模型自动驾驶系统.

4.4　自适应时空Top-K稀疏化的高效抖动量化隐私保护方法

面对高冗余、强关联的预训练模型梯度,需要进行梯度稀疏化和隐私保护. 然而,现有Top-K剪枝方法[16,17]难以动态调整稀疏化程度且忽略梯度的时空相关性,存在模型通信开销较大和预测精度不高的问题. 因此,本文提出ASDQ,动态筛选关键梯度并自适应添加噪声,实现大模型梯度的高效安全保护并降低通信开销,如图5所示.

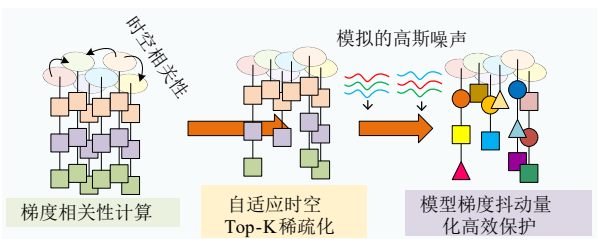


图5　本文ASDQ方法的预训练模型保护流程图

自适应时空Top-K稀疏化. 对于边缘大模型m,计算本轮平均模型梯度$\bar{\boldsymbol{g}}_m^t$和上一轮平均梯度$\bar{\boldsymbol{g}}_m^{t-1}$的时空相关性,动态计算每轮迭代中Top-K大小$K^t$. 给定常数$\mu$,通过Top-K算子计算关键稀疏梯度$\boldsymbol{g}_m^t$,具体如下所示:

$$\mathrm{sim}(\bar{\boldsymbol{g}}_m^t,\bar{\boldsymbol{g}}_m^{t-1})=\bar{\boldsymbol{g}}_m^t\cdot\bar{\boldsymbol{g}}_m^{t-1}/(\|\bar{\boldsymbol{g}}_m^t\|_2\|\bar{\boldsymbol{g}}_m^{t-1}\|_2) \tag{35}$$

$$K^t=\mathrm{TopK}\left(\|\bar{\boldsymbol{g}}_m^t\|,\mathrm{sim}(\bar{\boldsymbol{g}}_m^t,\bar{\boldsymbol{g}}_m^{t-1})\right) \tag{36}$$

$$\tilde{\boldsymbol{g}}_m^t=\left(\mathrm{sim}(\bar{\boldsymbol{g}}_m^t,\bar{\boldsymbol{g}}_m^{t-1})+\mu\right)\cdot\bar{\boldsymbol{g}}_m^t \tag{37}$$

$$\boldsymbol{g}_m^t=\mathrm{TopK}(\tilde{\boldsymbol{g}}_m^t) \tag{38}$$

模型梯度抖动量化高效保护. 对于关键稀疏梯度$\boldsymbol{g}_m^t$的每个分量l,边缘大模型m采样伽马变量$v_{m,l}\sim\Gamma(3/2,1/2)$,其参数选择依据是其对梯度抖动量化和隐私保护的影响[16],计算量化步长$\Delta_{m,l}$:

$$\Delta_{m,l}=2\sigma\sqrt{v_{m,l}}\Big/\max(\|\boldsymbol{g}_{m,l}^t\|_2,1) \tag{39}$$

其中,$\Delta_{m,l}$是模拟的高斯噪声的标准差. 边缘大模型m采样另一个均匀随机变量$u_{m,l}\sim\mathcal{U}(-\Delta_{m,l}/2,\Delta_{m,l}/2)$,得到扰动的量化梯度$\boldsymbol{q}_{m,l}^t$:

$$\boldsymbol{q}_{m,l}^t=Q(\boldsymbol{g}_{m,l}^t+u_{m,l}) \tag{40}$$

$$Q(x)=\left\lfloor(x-\Delta_{m,l})/2\Delta_{m,l}\right\rfloor\Delta_{m,l}+\Delta_{m,l}/2 \tag{41}$$

其中,$Q(\cdot)$为量化函数. 最后,边缘大模型将扰动的量化梯度$\boldsymbol{q}_{m,l}^t$发送至云服务器.

4.5　基于自适应加权聚合的多模态行人轨迹预测优化方法

云端在聚合扰动梯度后向边缘大模型下发全局参数,执行多模态稀疏行人轨迹预测. 针对平均聚合导致的梯度质量差异问题[17],本文提出AWMT,如图6所示.

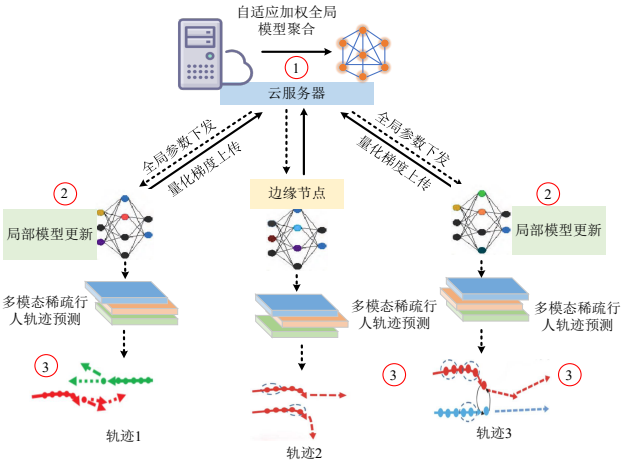


图6　本文AWMT方法的行人轨迹预测流程图

自适应加权全局模型聚合. 基于扰动的量化梯度$\boldsymbol{q}_{m,l}^t$,根据随机变量$u_{m,l}$,对扰动的量化梯度解码:

$$\hat{\boldsymbol{g}}_{m,l}^t=Q^{-1}(\boldsymbol{q}_{m,l}^t)-u_{m,l} \tag{42}$$

基于自适应加权方法,动态调整每个边缘大模型的贡献度. 聚合所有的解码梯度以更新全局模型参数:

$$\boldsymbol{g}^t=\sum_{m=1}^{M^t}\boldsymbol{w}_m^t\hat{\boldsymbol{g}}_m^t\Big/\sum_{m=1}^{M^t}\boldsymbol{w}_m^t \tag{43}$$

$$\boldsymbol{w}^{t+1}=\boldsymbol{w}^t-\left(\alpha^t/\sqrt{\boldsymbol{v}_t+\tau}\right)\boldsymbol{g}^t \tag{44}$$

$$\boldsymbol{v}_t=\beta^t\boldsymbol{v}_{t-1}+(1-\beta^t)\boldsymbol{g}_t^2 \tag{45}$$

其中，α^t表示第t次迭代时的学习率；β^t表示第t次迭代时梯度平方的指数衰减率；M^t表示参与第t轮训练的边缘大模型数量；$\boldsymbol{v}_t$表示梯度平方的滑动平均；τ是一个小常数，用来防止除零错误.

多模态稀疏行人轨迹预测. 云服务器将全局模型参数下发至各边缘大模型，进行高准确度行人轨迹预测. 为增强轨迹多样性，本文不仅选择Top-K个概率高的预测轨迹，而是排除高相似轨迹，具体如下：

$$\mathrm{Max}\sum_{i=1}^{K}P(\boldsymbol{T}_i)-\lambda\sum_{i,j=1}^{K}\mathrm{Sim}(\boldsymbol{T}_i,\boldsymbol{T}_j) \tag{46}$$

$$\mathrm{Sim}(\boldsymbol{T}_i,\boldsymbol{T}_j)=\sum_{t^e=1}^{T}\left\|\boldsymbol{T}_i(t^e)-\boldsymbol{T}_j(t^e)\right\| \tag{47}$$

其中，$\boldsymbol{T}_i$和$\boldsymbol{T}_j$分别为编号i与j的轨迹；$P(\cdot)$表示概率函数；$\mathrm{Sim}(\boldsymbol{T}_i,\boldsymbol{T}_j)$为两个轨迹的相似度；$\lambda\in[0,1]$是可调节参数[24]；$t^e$为当前时间步；$T$为总时间步.

5 安全性与效率分析

5.1 本文PMPTL方案的安全性分析

定理1 RGADP方法满足ϵ^1-本地差分隐私.

证明 首先，给定隐私预算为ϵ^1，每次划分网格的隐私预算为$\epsilon_c^1=\epsilon^1/i_{\max}$. 基于网格分辨率$r_i$，为每个网格$i$分配隐私预算：$\epsilon_i^1=r_i\times\epsilon_c^1$. 计算剩余预算为

$$\bar{\epsilon}^1=\epsilon^1-\sum_{i=1}^{|G|}\epsilon_i^1 \tag{48}$$

基于网格轨迹特征密度，自适应地将剩余隐私预算再次分配给网格，并向网格内添加拉普拉斯噪声：

$$\widetilde{\epsilon_i^1}=w_i\cdot\bar{\epsilon}^1 \tag{49}$$

则计算整个网格区域分配的总隐私预算为

$$\sum_{i=1}^{|G|}\epsilon_i^1+\sum_{i=1}^{|G|}\tilde{\epsilon}_i^1=\epsilon^1 \tag{50}$$

综上，RGADP方法满足ϵ^1-本地差分隐私得证. 证毕.

引理2[16] 如果一个隐私机制M在数据集D上满足(ϵ,δ)-差分隐私，且$|D|=n$. D_z是D的一个子集，大小$z\leqslant n$，且从D中均匀采样获得，则M满足(ϵ',δ')-差分隐私，其中$\epsilon'=\log\left[1+z(\mathrm{e}^{\epsilon}-1)/n\right]$且$\delta'=z\delta/n$.

引理3[16] 给定一个服从$\Gamma(3/2,1/2)$分布的伽马随机变量V，如果条件随机变量服从$(X|V=v)\sim\mathcal{U}(\mu-\sigma\sqrt{v},\mu+\sigma\sqrt{v})$，则随机变量满足$X\sim N\left(\mu,\sigma^2\right)$的分布.

引理4[16] 对于一个量化函数$Q(\cdot)$，其步长为$Q(\cdot)$，给定一个均匀随机变量$u\sim\mathcal{U}(-\Delta/2,\Delta/2)$和一个标量$Z$，采用$Q(\cdot)$量化，则$\hat{Z}=Q(Z+u)-u$，$\hat{Z}=Z+u'$. 其中$u$与$u'$是相同均匀分布，且与$u$独立.

定理2 在任意轮次t中，边缘大模型的估计梯度$\hat{\boldsymbol{g}}_m^t$是扰动稀疏梯度的有偏估计，其误差服从高斯分布$\hat{\boldsymbol{g}}_m^t$.

证明 考虑稀疏梯度$\boldsymbol{q}_{t,m,l}=Q\left(\boldsymbol{g}_{t,m,l}+u_{m,l}\right)$，其中$u_{m,l}\sim\mathcal{U}(-\Delta_{m,l}/2,\Delta_{m,l}/2)$. 而$\boldsymbol{q}_{t,m,l}$仅是通过函数$Q(\cdot)$和步长$\Delta_{m,l}$量化的梯度. 根据引理4，则得到$\hat{\boldsymbol{g}}_{t,m,l}=Q\left(\boldsymbol{g}_{t,m,l}+u_{m,l}\right)-u_{m,l}=\boldsymbol{g}_{t,m,l}+u'_{m,l}$，其中抖动误差源于均匀随机变量$u'_{m,l}\sim\mathcal{U}\left(-\Delta_{m,l}/2,\Delta_{m,l}/2\right)$. 由于$\Delta_{m,l}=2\sigma\sqrt{v_{m,l}}$依赖于伽马变量$v_{m,l}\sim\Gamma(3/2,1/2)$，根据引理3将伽马控制的均匀扰动转换为高斯误差，则得到$u'_{m,l}\sim N\left(0,\sigma^2\right)$. 由于梯度的每个维度相互独立，则$\hat{\boldsymbol{g}}_m=\boldsymbol{g}_m+\mathcal{U}_m$，其误差服从$\mathcal{U}_m\sim N\left(0,\boldsymbol{I}_m\sigma^2\right)$，其误差服从高斯分布，满足高斯机制. 证毕.

定理3 如果模拟的高斯噪声的标准差$\sigma\geqslant\sqrt{2\ln(1.25/\delta')}/|D|\epsilon^2$'，则自适应时空Top-K稀疏化的抖动量化隐私保护方法满足$\left(\epsilon^2,\delta\right)$-差分隐私，其中隐私预算$\epsilon^2=\log\left[1+B(\mathrm{e}^{\epsilon^2'}-1)/|D|\right]$且$\delta=B\delta'/|D|$在任何轮次$t$中成立，$|D|$是边缘大模型数据集的大小，$B$是本地训练的批量大小.

证明 首先，计算$\boldsymbol{g}_m^t(\cdot)$的查询敏感度：

$$\begin{aligned}\boldsymbol{g}_m^t(\cdot)&=\max_{D_m,D'_m}\|\mathrm{TopK}(\bar{\boldsymbol{g}}_m^t)-\mathrm{TopK}(\bar{\boldsymbol{g}}_m^t)\|_2\\&=\max_{D_m,D'_m}1/|D_m|\|\mathrm{TopK}\left(\sum_{x\in D_m}\tilde{\boldsymbol{g}}_{m,l}^t\right)\\&\quad-\mathrm{TopK}\left(\sum_{x\in D'_m}\tilde{\boldsymbol{g}}_{m,l}^t\right)\|_2\end{aligned} \tag{51}$$

其中，D'_m和D'_m是相邻的数据集. 对于D'_m中任何两个样本r_1和r_2，如果D'_m去掉r_1，D'_m去掉r_2，则模型在D'_m和D'_m平均梯度是相等的. 假设r_1和r_2在模型训练中的K维度上均匀分布且符号相反，则$\boldsymbol{g}_m^t(\cdot)$的查询敏感度为$1/|D|$.

基于引理1和定理2，因为本文模拟的高斯噪声标准差$\sigma\geqslant\sqrt{2\ln(1.25/\delta)}/|D|\epsilon^2$，则证明了$\boldsymbol{g}_m^t(\cdot)$满足$\left(\epsilon^2,\delta\right)$-差分隐私. 证毕.

定理4 本文PMPTL方案满足ϵ-差分隐私.

证明 由于多模态稀疏轨迹融合、稀疏特征高效增强和行人轨迹预测均未采用差分隐私，则满足0-差分隐私. 基于定理1和定理3，RGADP算法满足ϵ^1-差分隐私. 而ASDQ算法也遵循ϵ^2-差分隐私. 因此，基于差

分隐私的串行组合原理,本文方案满足$\epsilon=(\epsilon^1+\epsilon^2)$-差分隐私. 证毕.

5.2 算法敏感性分析

5.2.1 RGADP算法中调节因子的敏感性分析

定理5 给定子网格划分大小d,调节因子γ的最大变化量$\Delta\gamma_{\max}$和平均值$\bar{\gamma}$,则RGADP算法的最大迭代次数变化量$\Delta i_{\max}$(即敏感性)是有界的,$|\Delta i_{\max}|\leqslant\Delta\gamma_{\max}/2\ln(d)\cdot\bar{\gamma}$,具有高效性.

证明 RGADP算法关注的是当调节因子γ发生微小变化$\Delta\gamma$时,网格划分的最大迭代次数$i_{\max}$的变化$\Delta i_{\max}$有多大. 令$S(\gamma)=\sum_i\gamma n_i$,其中$\gamma=(\gamma_1,\gamma_2,\cdots,\gamma_M)$,则有

$$i_{\max}=f(S)=\left\lceil 1/2\log_d(\|\boldsymbol{X}\|/S)\right\rceil \tag{52}$$

$$\frac{\mathrm{d}f}{\mathrm{d}S}=\frac{1}{2\ln(d)}\cdot\frac{\mathrm{d}}{\mathrm{d}S}\left[\ln(\|\boldsymbol{X}\|)-\ln(S)\right]=-\frac{1}{2\ln(d)\cdot S} \tag{53}$$

变化量$|\Delta i_{\max}|\approx\left|\mathrm{d}f/\mathrm{d}S\right|\cdot|\Delta S|=|\Delta S|/2\ln(d)\cdot S$,则其$L_1$灵敏度为

$$|\Delta S|\leqslant\sum_i n_i|\Delta\gamma_i| \tag{54}$$

考虑所有γ_i的单位变化上限,即$\Delta\gamma_i\leqslant\Delta\gamma_{\max}$,则有$|\Delta S|\leqslant\Delta\gamma_{\max}\sum_i n_i=\Delta\gamma_{\max}\cdot N$. 其中$N$是输入序列长度. 设所有$\gamma_i$均值为$\bar{\gamma}$,代入得:

$$|\Delta i_{\max}|\leqslant\frac{1}{2\ln(d)\cdot\bar{\gamma}N}\cdot\Delta\gamma_{\max}\cdot N=\frac{\Delta\gamma_{\max}}{2\ln(d)\cdot\bar{\gamma}} \tag{55}$$

这表明$i_{\max}$对调节因子γ的变化量的上界是一个常数,且与输入序列长度N无关. 因此,证明了RGADP算法中调节因子γ的敏感性是有界且高效的. 证毕.

5.2.2 ASDQ算法中参数μ的敏感性分析

定理6 对于任意两个相邻数据集D和D'在一个查询函数f上函数输出的最大变化量Δf是有界的.

证明 本文考虑任意两个相邻数据集D和D'对于在一个查询函数f上函数输出的最大变化量Δf:

$$\Delta f=\max\ \|f(D)-f(D')\|_2 \tag{56}$$

令$f(\bar{\boldsymbol{g}}_m^t)=\left(\mathrm{sim}(\bar{\boldsymbol{g}}_m^t,\bar{\boldsymbol{g}}_m^{t-1})+\mu\right)\cdot\bar{\boldsymbol{g}}_m^t$,则平均梯度$\bar{\boldsymbol{g}}_m^t$变化满足:$\|\bar{\boldsymbol{g}}_m^t(D)-\bar{\boldsymbol{g}}_m^t(D')\|_2\leqslant 1/|D|$,则有

$$\begin{aligned}\Delta f&=\max\left\|f\left(\bar{\boldsymbol{g}}_m^t(D)\right)-f\left(\bar{\boldsymbol{g}}_m^t(D')\right)\right\|_2\\&=\max\left\|\left(s(D)+\mu\right)\cdot\bar{\boldsymbol{g}}_m^t-\left(s(D')+\mu\right)\cdot\bar{\boldsymbol{g}}_m^t\right\|_2\end{aligned} \tag{57}$$

其中,$s(D)=\mathrm{sim}(\bar{\boldsymbol{g}}_m^t,\bar{\boldsymbol{g}}_m^{t-1})\in[-1,1]$,且$\left\|\bar{\boldsymbol{g}}_m^t\right\|_2$有界;而$\mu$是常数,不影响梯度差异,因此:

$$\Delta f\leqslant\max\left\|s(D)\cdot\bar{\boldsymbol{g}}_m^t-s(D')\cdot\bar{\boldsymbol{g}}_m^t\right\|_2+\mu\cdot\left\|\bar{\boldsymbol{g}}_m^t-\bar{\boldsymbol{g}}_m^t\right\|_2 \tag{58}$$

这表明Δf对常数μ的变化量的上界是一个常数,且与输入序列长度无关. 因此,证明了ASDQ算法中常数μ的敏感性是有界且高效的. 证毕.

5.3 计算复杂度分析

5.3.1 多模态稀疏轨迹高效融合的计算复杂度分析

首先,模态对齐重用模块计算复杂度为$O(k^2d)$,交叉注意力计算复杂度为$O(kNd)$,其中k是token数量,d是特征维度,N是输入序列长度. 而Mamba融合模块计算复杂度为$O(Nd)$. 则本文方法的总计算复杂度为$O(k^2d+kNd+Nd)$,而$k=\sqrt{N}$,维度d的计算复杂度为$O(1)$,则总计算复杂度为$O(N^{3/2})$.

然而,现有Transformer融合机制的计算复杂度为$O(N^2)$. 显然,当$N\to\infty$时,$O(N^{3/2})\ll O(N^2)$,因此,本文方法的复杂度更低,有效提升了融合效率.

5.3.2 大模型预训练隐私保护算法的计算复杂度分析

设全局迭代总次数为T_t,模型参数的数量为s,稀疏比率为r,则Top-K稀疏化和抖动量化的计算复杂度分别为$O\left(s\log(rs)\right)$和$O(s)$. 随机梯度下降方法的计算复杂度为$O(m)$. 扩展到T_t轮全局迭代,得到本文模型预训练隐私保护算法的总计算复杂度为$O\left[T_t\left(s\log(rs)+m\right)\right]$因此,本文模型预训练隐私保护方法具有较高的效率.

5.4 大模型预训练通信效率分析

首先,考虑减法抖动过程,量化梯度向量$\boldsymbol{g}_j$为

$$\boldsymbol{g}_j=\left\lfloor(\boldsymbol{g}_j+u)/\Delta_{m,l}-1/2\right\rfloor\Delta_{m,l}+\Delta_{m,l}/2 \tag{59}$$

由于$\Delta_{m,l}$可以通过共享的随机种子被云服务器获得,则编码的信息为$\boldsymbol{x}=\left\lfloor(\boldsymbol{g}_j+u)/\Delta_{m,l}-1/2\right\rfloor$.

给定采样变量$u\sim\mathcal{U}\left(-\Delta_{m,l}/2,\Delta_{m,l}/2\right)$,$\boldsymbol{x}$的范围为$\left[\left\lfloor-1/\Delta_{m,l}-1\right\rfloor,\left\lfloor 1/\Delta_{m,l}\right\rfloor\right]$. 因此,本文预训练模型需要$n\log_2\left(2\left\lfloor 1/\Delta_{m,l}+1\right\rfloor\right)$位编码所有梯度向量$\boldsymbol{g}$,$n$表示梯度向量$\boldsymbol{g}$的维度.

稀疏化过程减少了传输的比特数. 设定$r=k/n$为稀疏化率,表示通过Top-K操作保留的元素比例,则保留的元素仅需$nr\log_2\left(2\left\lceil 1/\Delta_{m,l}\right\rceil+1\right)$比特进行编码,则比特数为$n\left[r\log_2\left(2\left\lceil 1/\Delta_{m,l}\right\rceil+1\right)+1-r\right]$. 因此,本文通信比特数为$n\left[r\log_2\left(2\left\lceil 1/\Delta_{m,l}\right\rceil+1\right)+1-r\right]$,有效提升了通信效率.

6 实验

6.1 实验设置

数据集. 采用两个公开的真实行人轨迹数据集ETH-UCY[24]和斯坦福无人机数据集(Stanford Drone Da-

taset,SDD)[24],以评估本文方法性能.ETH-UCY数据集包含了5个场景:ETH、HOTEL、UNIV、ZARA1和ZARA2.这些场景记录了行人在街道、酒店、校园、商场等大范围环境中的运动轨迹[25].该数据集包含大规模行人运动轨迹及图像和视频数据,图像大小均为720×576,图像格式为JPG格式,视频时长分别约为9、13、4、6和7 min,视频格式均为AVI格式[26].SDD数据集是由斯坦福大学计算机视觉与几何实验室在斯坦福大学校园、街道拥堵时段,收集了约60 GB的行人轨迹图像和视频数据,图像大小为960×540,图像格式为JPG格式,视频时长约为240 min,视频格式为MP4格式[27].

评价指标.通过两个指标比较本文方法和现有先进方法的性能:平均位移误差(Average Displacement Error,ADE)和最终位移误差(Final Displacement Error,FDE).给定真实的未来轨迹和预测的K条轨迹,ADE和FDE用于衡量它们之间的L_2距离:

$$\mathrm{ADE}=\sum_{t=T_o}^{T}\sqrt{(\boldsymbol{x}_t-\hat{\boldsymbol{x}}_t)^2+(\boldsymbol{y}_t-\hat{\boldsymbol{y}}_t)^2}\Big/T_p \tag{60}$$

$$\mathrm{FDE}=\sqrt{(\boldsymbol{x}_T-\hat{\boldsymbol{x}}_T)^2+(\boldsymbol{y}_T-\hat{\boldsymbol{y}}_T)^2} \tag{61}$$

其中,T_p表示预测的时间步长;T_o表示观测的起始时间步;$\boldsymbol{x}_T$和$\boldsymbol{y}_T$表示真实轨迹在最终时刻T的二维坐标;$\hat{\boldsymbol{x}}_T$和$\hat{\boldsymbol{y}}_T$表示预测轨迹在最终时刻T的二维坐标.

实验设置.在多模态融合阶段,相似度阈值θ设置在[0.5,0.9]范围内.在融合特征隐私保护中,将隐私预算ϵ^1设置在[0.2,1.0]内.在模型梯度隐私保护过程中,设定标准差σ在[0.01,0.1]内.将ETH-UCY数据集训练轮数设置为1 000,将SDD数据集训练轮数设置为2 000,将ETH-UCY和SDD数据集观察半径设置为2,观测时间步长设置为8,预测时间步长设置为12,采用Adam优化器的初始学习率为0.001,将Eth、Hotel、Univ、Zara2数据集模型隐藏层维度设置为128,Zara1和SDD数据集隐藏层维度设置为64.受文献[24]的参数实验调优启发,通过大量预处理实验,设定可调节参数λ为[0,1].为保持模型的稳定性和泛化能力,每个数据集被随机划分为80%用于训练,20%用于测试.所有实验均在单个RTX 4090 GPU上进行.

与现有方法对比.在多模态特征融合性能评估中,采用现有Transformer融合机制[13]、Mamba融合机制[28]与本文结合两者的MSTM方法进行预测精度性能对比.在多模态融合特征保护的可用性评估中,采用现有的SPEDP[5]方法、PRIVET(PRIVacy-preserved federated Estimator for Triangle count)[6]方法、TCPP(Trajectory Correlation Privacy-Preserving)[11]方法与本文的RGADP方法进行对比.在多模态扰动特征增强的时间开销评估中,采用现有密集型自注意力方法Drone-HAT(Drone-Hybrid Attention vision Transformer)[29]、Beyond Attention[18]、DiTFastAttn(Diffusion Transformers Fast Attention)[30]与本文的自适应稀疏自注意力机制DBAS进行性能对比.在预训练模型隐私保护的性能评估中,采用现有SPEDP[5]方法、AdapLDP-FL(Adaptive Local Differential Privacy-Federated Learning)[21]方法、DP-Norm(Differential Privacy-Normalization)[22]方法和DP-PFL[17]与本文的ASDQ方法进行对比.在行人轨迹预测评估中,将本文的PMPTL方案与现有先进方法SingularTrajectory[1]、ST-motion(SpatioTemporal-motion)[2]、SMEMO(Social MEmory MOdule)[3]、TUTR(Trajectory Unified TRansformer)[24]、EqMotion(Equivariant Motion)[25]、LED(LEapfrog Diffusion)[26]、MemoNet[31]、EigenTrajectory[27]进行对比.

6.2 多模态稀疏特征融合的预测精度评估

在SDD、ETH、HOTEL、UNIV、ZARA1、ZARA2等6个真实数据集中,评估本文MSTM方法和现有Transformer融合机制[13]、Mamba融合机制[28]的ADE,实验结果如图7所示.在数据集样本数量较小的情况下,三种多模态融合方法均表现出较高的ADE.随着数据集样本数的增加,三种方法的ADE均出现明显下降趋势,这是因为更大的训练数据集提供了更加丰富和多样的行人轨迹模式,使得模型在学习过程中能够更全面地捕捉轨迹的空间和时间分布特征,从而提升了预测精度.此外,本文方法在6个数据集上ADE均小于现有Transformer融合机制以及现有Mamba融合机制,预测精度提升近8%.这是因为Transformer和Mamba的融合不仅加强了模型的全局和局部学习能力,还通过提供多样化的特征表示和优化方法提高了模型的预测精度.因此,本文MSTM方法可以提高多模态稀疏特征融合的预测精度.

6.3 多模态稀疏融合特征保护的可用性评估

在6个数据集中,对于不同的隐私预算ϵ^1评估本文RGADP方法的数据可用性,且和现有的SPEDP[5]方法、PRIVET[6]方法、TCPP[11]方法进行对比,实验结果如图8所示.随着隐私预算的增加,所有方法的可用性逐渐提升.这是因为较大的隐私预算会对多模态融合特征添加较小的噪声扰动,有较高的数据可用性,这正好验证了第5.1节定理1的差分隐私性质的理论推导,形成了理论和实验的闭环.此外,在6个数据集上,本文RGADP方法的可用性均高于现有方法.这是因为本文考虑了轨迹特征的网格分辨率和密度,采用自适应加权隐私预算分配机制,动态添加噪声,提升可用性.而现有方法采用均匀的隐私预算,没有考虑融合特征的稀疏性,添加的噪声过大,导致可用性较低.在所有数据集中,本文RGADP方法的数据可用性较SPEDP方法、PRIVET方法和TCPP方法分别提升4.65%、8.40%、11.18%.因此,本文RGADP方法有效提升了可用性,为

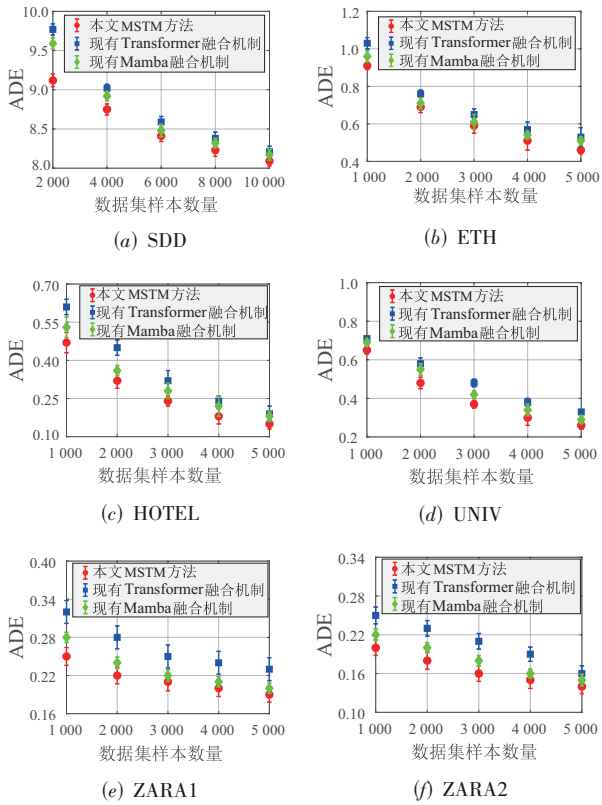


图7　不同多模态轨迹融合方法的预测精度评估

实现高准确的行人轨迹预测奠定基础.

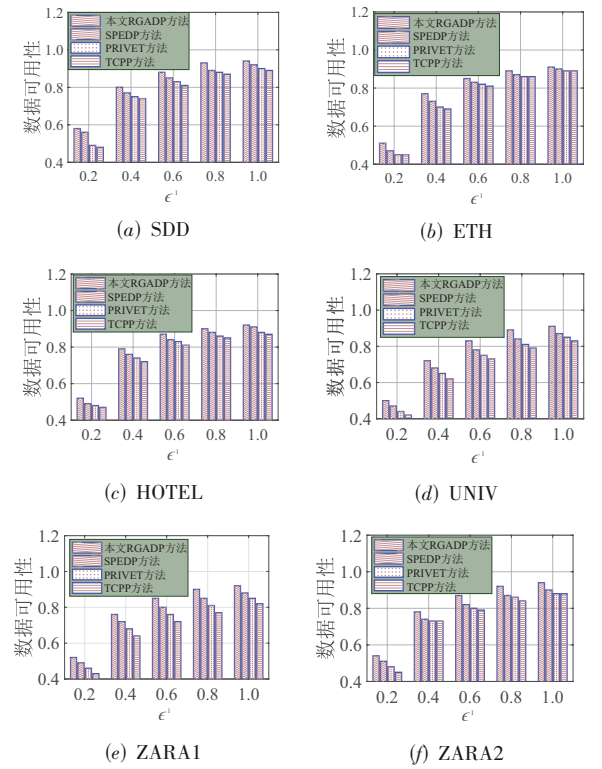


图8　不同多模态稀疏融合特征隐私保护方法的可用性评估

6.4　多模态融合扰动特征增强的效率性能评估

在6个数据集中,评估本文DBAS方法与现有密集型自注意力方法Drone-HAT[29]、Beyond Attention[18]、DiTFastAttn[30]的运行时间开销性能和特征增强预测效果,分别如图9和图10所示. 在图9中,所有算法的运行时间随着多模态稀疏数据样本数量的增大而上升,且本文DBAS方法的运行时间较短. 主要原因是当多模态稀疏样本数量较大时,稀疏融合特征的噪声表现更明显. 本文设计的稀疏自注意力机制,会过滤掉低查询-键匹配分数的无关噪声特征,增强关键特征表征. 而现有密集自注意力机制会计算所有特征,导致计算开销过大. 此外,本文DBAS方法的时间开销较现有密集自注意力机制整体提升24.1%. 在图10中,所有方法的预测性能(如ADE)均随样本数量增大而下降,且本文DBAS算法在6个数据集上的预测性能优于现有先进方法. 这是因为更大规模的训练数据有助于模型学习更丰富的时空特征,而本文DBAS算法增强了稀疏特征的表征效果,整体预测误差降低7.1%,验证了其在稀疏特征增强方面的有效性. 因此,本文DBAS方法可以实现高效的特征表征增强.

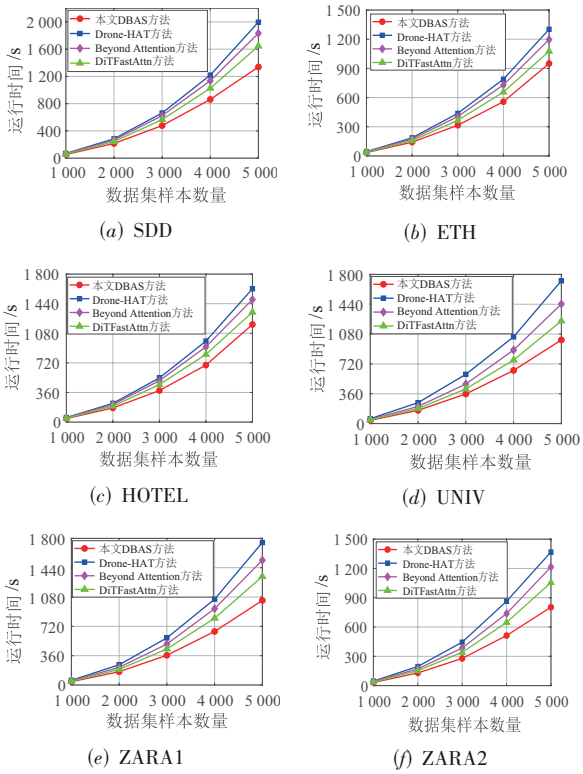
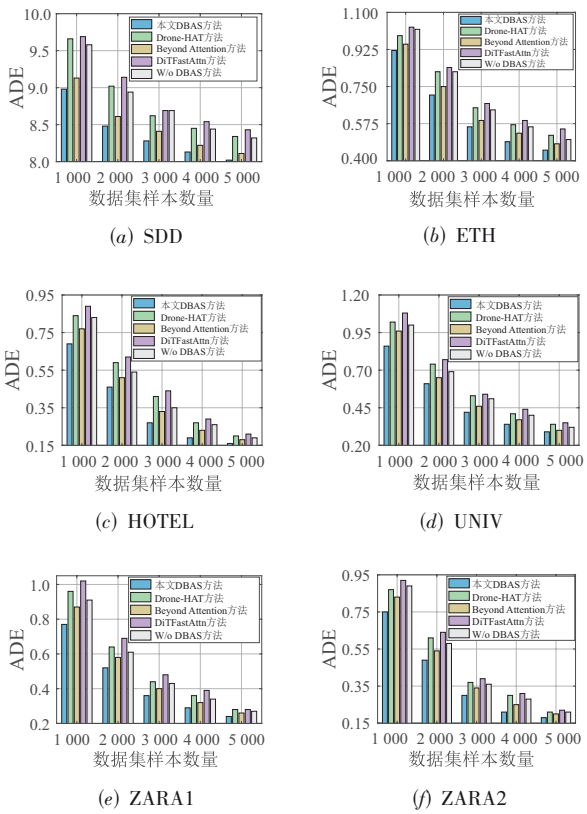


图9　不同多模态融合扰动特征表征增强方法的时间开销评估

6.5　预训练模型隐私保护的预测性能评估

在6个数据集中,评估不同隐私预算ϵ^1与不同标准差σ对本文ASDQ方法的预测性能影响,实验结果如图11所



(a) SDD (b) ETH

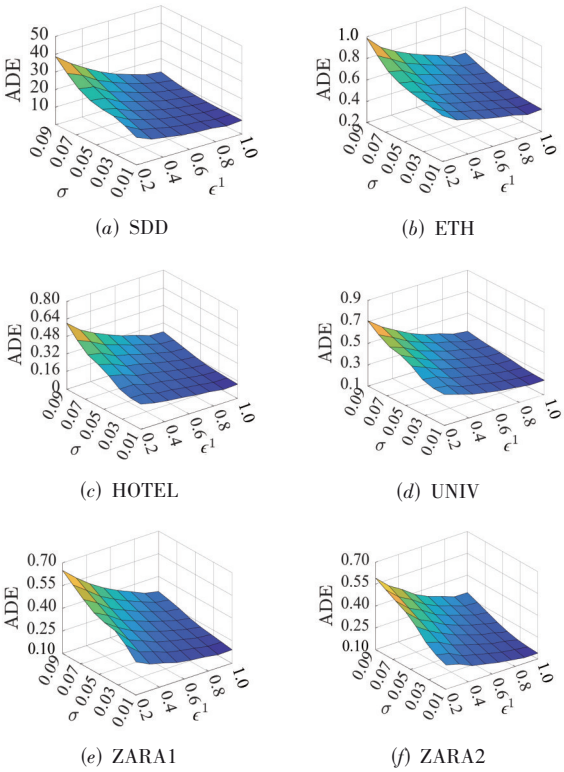
(c) HOTEL (d) UNIV

(e) ZARA1 (f) ZARA2

图10 不同数据集样本对不同注意力机制方法的预测性能评估

示. 对于较大的隐私预算较大ϵ^1和较小的标准差σ,预训练模型表现出更低的ADE值. 尤其在ETH数据集中.

当ϵ^1=1.0且σ=0.01时,本文ASDQ方法获得最优ADE值为0.41. 这归因于差分隐私性质,当隐私预算ϵ^1越大,标准差σ越小时,预训练模型梯度被添加的噪声越少,从而提升模型的预测性能. 同时,将本文ASDQ方法和现有SPEDP[5]方法、AdapLDP-FL[21]方法、DP-Norm[22]方法和DP-PFL[17]方法的预测性能进行对比,实验结果如图12和图13所示. 在6个数据集上,本文ASDQ方法的预测性能均优于现有方法的性能. 这是因为本文采用隐私预算自适应加权分配和模型梯度抖动量化高效保护多模态训练数据和模型梯度,而现有方法仅支持静态噪声添加,引入噪声过大,导致预训练模型的预测精度不高,这些实验结果佐证了第5.1节的定理2和定理3. 本文方法的ADE指标比SPEDP方法和AdapLDP-FL方法分别提升了10.65%和5.90%,整体可用性提升了6.05%和6.21%. 在FDE指标上较DP-Norm方法、AdapLDP-FL方法和DP-PFL方法分别提升17.98%、10.28%和25.03%. 因此,本文ASDQ方法在实现隐私保护的情况下具有更好的预训练模型性能.

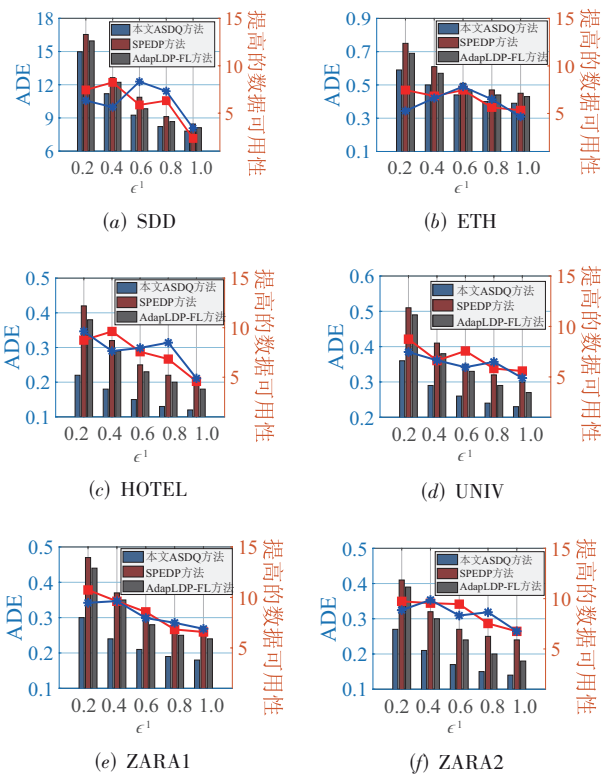


(a) SDD (b) ETH

(c) HOTEL (d) UNIV

(e) ZARA1 (f) ZARA2

图11 不同ϵ^1与σ对预训练模型隐私保护的预测性能影响



(a) SDD (b) ETH

(c) HOTEL (d) UNIV

(e) ZARA1 (f) ZARA2

图12 不同ϵ^1对不同预训练模型保护方法的预测性能评估

6.6 预训练模型隐私保护的通信效率评估

针对6个不同数据集场景，对本文ASDQ方法的通信效率（即每轮通信比特数）进行评估，与现有先进方法进行性能对比，实验结果如表1所示．选取的三类对比算法如下：

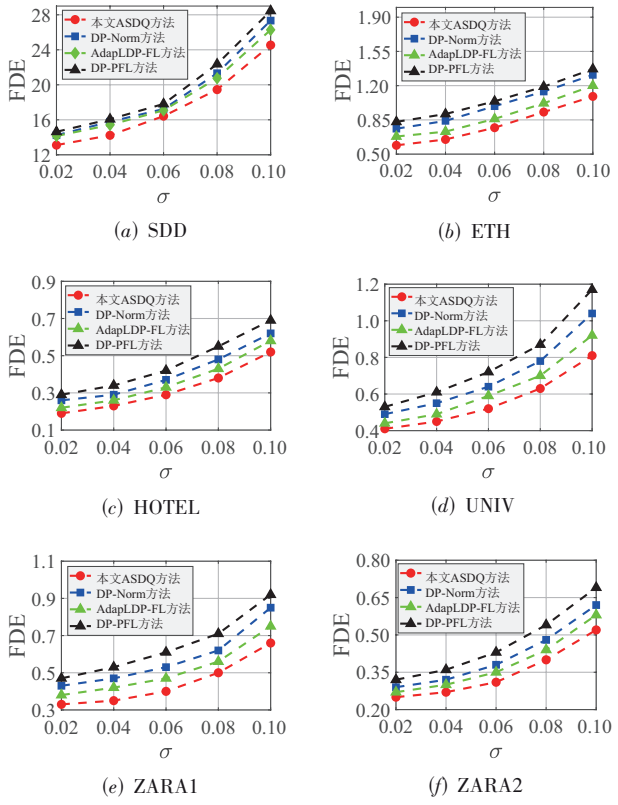


图13　不同σ对不同预训练模型保护方法的预测性能影响

（1）TopK-DP（TopK-Differential Privacy）[16]：预训练模型添加传统高斯噪声，结合Top-K梯度稀疏化降低通信量．

（2）DP-FL（Differential Privacy-Federated Learning）[16]：基于标准差分隐私机制的联邦学习基准方法对模型进行保护，未引入梯度稀疏化处理．

（3）FLDQ（Federated Learning Dithering Quantization）[32]：预训练模型采用固定抖动量化降低通信，缺乏动态稀疏化机制．

本文ASDQ方法在6个数据集中均显著优于现有算法．这主要是因为现有Top-K固定选择模型梯度，导致梯度冗余和通信增加．而本文ASDQ方法考虑了梯度间的时空相关性，动态计算K值，实现自适应Top-K稀疏化，提高通信效率．特别地，综合6个数据集结果，本文ASDQ方法整体通信量较TopK-DP、DP-FL、FLDQ等方法分别降低18.43%、73.87%和31.83%．综上所述，本文ASDQ方法在通信方面有显著优势，提高了预训练模型效率．

6.7 行人轨迹预测性能评估

消融实验．本文依次移除各核心模块，在6个数据集上设计了针对所提出的MSTM、RGADP、DBAS、ASDQ和AWMT方法的消融实验，如表2所示．当移除MSTM时，本文方案性能下降，表明结合Transformer与Mamba机制能有效捕捉行人轨迹预测场景的时空特征．而在移除RGADP和ASDQ后，本文方法性能有所提升．这是因为移除这两个方法后，没有向轨迹特征和模型梯度添加随机噪声，提升了预测精度，但存在隐私泄露风险．移除DBAS后，本文方法性能降低，表明其能有效抑制稀疏数据中的噪声．尽管本文PMPTL方案较w/oRGADP和w/oASDQ消融实验的性能有略微下降，但本文方案双重保护了训练数据和模型梯度，具有较高的安全性和鲁棒性．

固定权重与本文自适应权重方法对比．为验证DBAS方法中动态权重调节的有效性，本文对比了固定权重与自适应权重策略在行人轨迹预测中的性能差异，结果如表3所示．随着固定权重中稀疏分过滤过度，导致关键特征丢失，从而预测性能下降．本文PMPTL方案通过动态调节稀疏分支与密集分支的权重在6类场景中均取得最优性能，验证了其查询-键匹配分数动态过滤噪声特征的能力．

现有先进方法与本文预测方法性能对比．本文针对6个数据集场景，评估本文方案的预测性能，并且与现有先进方法的性能进行对比分析，实验结果如表4和表5所示．其中，受文献[33]启发，计算整体预测提升百分比为

$$\begin{aligned}&0.8\times(\mathrm{ADE}_{\mathrm{base}}-\mathrm{ADE}_{\mathrm{our}})/\mathrm{ADE}_{\mathrm{base}}\times100\%\\&+0.2\times(\mathrm{FDE}_{\mathrm{base}}-\mathrm{FDE}_{\mathrm{our}})/\mathrm{FDE}_{\mathrm{base}}\times100\%\end{aligned}\tag{62}$$

表1　不同预训练模型隐私保护方法在6种数据集中的通信效率对比结果

算法	数据集					
	ETH	HOTEL	UNIV	ZARA1	ZARA2	SDD
TopK-DP[16]	168 480 402	152 202 344	162 314 432	55 304 234	175 238 342	3 423 318
DP-FL[16]	562 462 340	423 450 752	452 124 348	421 348 562	430 460 212	10 123 824
FLDQ[32]	195 676 566	192 467 676	185 210 324	82 230 538	185 644 028	4 163 318
PMPTL(本文)	159 182 144	148 872 736	103 811 598	37 027 013	162 677 941	2 615 062

表2 行人轨迹预测的消融实验结果(ADE/FDE)

算法	数据集					
	ETH	HOTEL	UNIV	ZARA1	ZARA2	SDD
w/o MSTM	0.48/0.78	0.18/0.28	0.28/0.48	0.24/0.45	0.18/0.31	8.50/14.20
w/o RGADP	0.40/0.61	0.12/0.19	0.23/0.41	0.18/0.34	0.14/0.25	7.77/12.90
w/o DBAS	0.44/0.67	0.15/0.23	0.26/0.45	0.21/0.40	0.16/0.28	8.10/13.50
w/o ASDQ	0.39/0.59	0.11/0.18	0.22/0.40	0.17/0.34	0.13/0.24	7.75/12.73
w/o AWMT	0.43/0.64	0.14/0.21	0.25/0.44	0.20/0.38	0.15/0.27	7.95/13.30
本文PMPTL	0.41/0.62	0.13/0.20	0.24/0.43	0.19/0.37	0.14/0.26	7.83/13.11

表3 行人轨迹预测在固定权重与自适应权重下的结果(ADE/FDE)

算法	数据集					
	ETH	HOTEL	UNIV	ZARA1	ZARA2	SDD
固定权重p_1=0.3, p_2=0.7	0.44/0.68	0.15/0.23	0.27/0.49	0.22/0.41	0.17/0.31	8.11/13.23
固定权重p_1=0.5, p_2=0.5	0.45/0.70	0.17/0.26	0.30/0.54	0.23/0.43	0.19/0.35	8.52/13.82
固定权重p_1=0.7, p_2=0.3	0.48/0.77	0.18/0.28	0.36/0.66	0.27/0.50	0.22/0.40	9.07/14.54
本文PMPTL	0.41/0.62	0.13/0.20	0.24/0.43	0.19/0.37	0.14/0.26	7.74/12.58

表4 在ETH-UCY数据集上本文方案与现有先进方法的预测性能对比(ADE/FDE)

方法	数据集						整体提升百分比/%
	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG	
EqMotion* [25]	0.40/0.61	0.12/0.18	0.23/0.43	0.18/0.32	0.13/0.23	0.21/0.35	4.38
LED* [26]	0.39/0.58	0.11/0.17	0.26/0.43	0.18/0.26	0.13/0.22	0.21/0.33	3.20
SingularTrajectory*[1]	0.35/0.49	0.13/0.24	0.25/0.46	0.19/0.34	0.15/0.28	0.21/0.36	4.92
ST-motion* [2]	0.93/1.81	0.32/0.60	0.54/1.16	0.42/0.90	0.32/0.70	0.51/1.03	62.03
SMEMO* [3]	0.39/0.59	0.14/0.20	0.23/0.41	0.19/0.32	0.15/0.25	0.22/0.35	7.84
TUTR*[24]	0.40/0.61	0.11/0.18	0.23/0.42	0.18/0.34	0.13/0.25	0.21/0.36	4.92
TUTR[24]	0.44/0.71	0.15/0.24	0.28/0.55	0.23/0.44	0.16/0.28	0.25/0.44	12.78
PMPTL* (本文)	0.38/0.58	0.10/0.18	0.22/0.40	0.17/0.31	0.13/0.24	0.20/0.34	—
PMPTL(本文)	0.41/0.62	0.13/0.20	0.24/0.43	0.19/0.37	0.14/0.26	0.22/0.37	—

注:*表示模型没有添加噪声保护.

表5 在SDD数据集上本文方案与现有先进方法的预测性能对比

方法	ADE/FDE	整体提升百分比/%
MemoNet*[31]	8.56/12.66	7.79
EigenTrajectory*[27]	8.10/13.10	4.35
LED*[26]	8.48/11.66	5.40
EqMotion*[25]	7.90/11.90	0.48
SMEMO*[3]	8.11/13.06	4.38
TUTR*[24]	7.76/12.69	0.38
TUTR[24]	8.01/13.73	2.70
PMPTL* (本文)	7.74/12.58	—
PMPTL(本文)	7.83/13.11	—

注:*表示模型没有添加噪声保护.

从综合平均指标(AVeraGe overall metric,AVG)上看,在没有噪声扰动的原始模型性能方面,尽管本文PMPTL*方案在所有数据集中的FDE值较现有先进方法如EqMotion*与LED*提升不明显,但是ADE值为0.20是最低的.这主要是因为ADE值更能反映整条轨迹预测的精度,而FDE值仅关注预测终点误差,不能完全评估行人轨迹预测的结果,表明本文方案具有更高的预测精度.此外,本文PMPTL*方案较EqMotion*与LED*,在ETH-UCY数据集上的整体预测提升百分比分别为4.38%和3.20%.同时,本文方案的计算复杂度为$O(N^{3/2})$,优于二者的计算复杂度$O(N^2)$,具有明显的时间效率优势.在多模态数据和模型都被噪声保护下,本文PMPTL方法具有更好的噪声鲁棒性.其在ETH、HOTEL等场景的ADE/FDE误差较低,优于最先进方法TUTR的性能.这是因为本文对多模态数据添加自适应拉普拉斯噪声以及向模型梯度添加模拟可控的高斯噪声,避免噪声过大导致的预测精度降低.

6.8 行人轨迹预测可视化

在6个数据集场景下,本文方案的行人轨迹预测可视化结果如图14所示.每个子图的x轴和y轴表示二维空间中的坐标,用于展示行人的运动轨迹.本文方案可以有效地预测出行人的多种可能未来轨迹以及对应

的概率. 通过与地面真值轨迹的对比,可以观察到本文方案的预测结果与实际运动路径的比较接近,验证了本文方案在行人轨迹预测任务中的有效性. 此外,在不同的数据集中,模型的预测结果均展现出较好的多样性和准确性,表明本文PMPTL方案具有良好的泛化能力和鲁棒性.

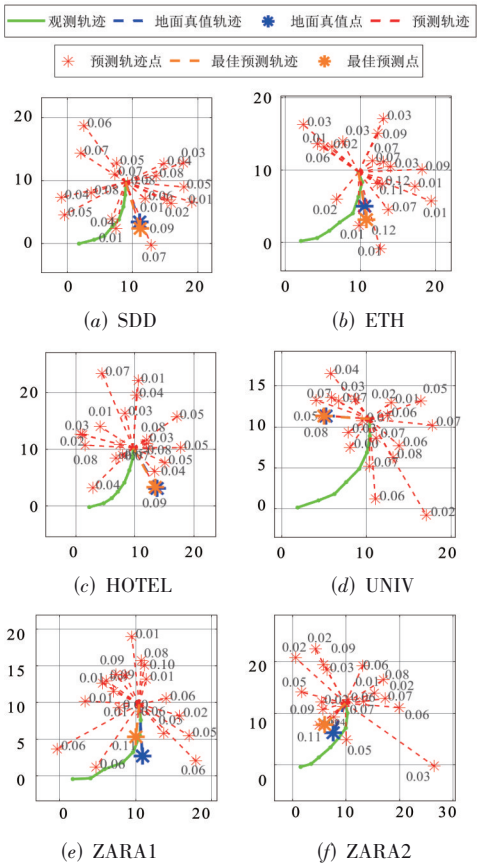


图14　本文方案的行人轨迹预测可视化图

7　结束语

针对大模型预训练中多模态稀疏行人轨迹预测存在隐私安全和计算复杂度高的问题,本文首次提出一种PMPTL方案. 首先,设计MSTM方法高效融合多模态稀疏时空特征. 结合RGADP算法,在动态分配隐私预算的同时,确保数据的可用性. 其次,为解决预训练模型隐私和通信效率问题,提出DBAS方法,高效表征关键特征,提升模型预训练效率. 引入ASDQ方法,在降低通信开销的同时确保大模型安全性. 最后,基于AWMT方法,提升模型预测精度. 通过安全性和效率分析及丰富实验结果表明,本文方案实现了安全高效的多模态行人轨迹预测.

未来,本文将深入研究:

(1)动态场景驱动的隐私预算分配. 结合大模型时空风险感知,动态分配隐私.

(2)边缘算力的动态资源调度. 设计面向边缘异构设备的混合稀疏计算资源调度策略,实现大模型的低延迟协同计算.

参考文献

[1] BAE I, PARK Y J, JEON H G. SingularTrajectory: Universal trajectory predictor using diffusion model[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 17890-17901.

[2] SAADATNEJAD S, GAO Y, MESSAOUD K, et al. Social-transmotion: Promptable human trajectory prediction[EB/OL]. (2024-12-04)[2025-11-10]. https://arXiv.org/abs/2312.16168.

[3] MARCHETTI F, BECATTINI F, SEIDENARI L, et al. SMEMO: Social memory for trajectory forecasting[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(6): 4410-4425.

[4] 崔双双, 吴限, 王宏志, 等. 面向云边端协同的多模态数据建模技术及其应用[J]. 软件学报, 2024, 35(3): 1154-1172.
CUI S S, WU X, WANG H Z, et al. Multimodal data modeling technology and its application for cloud-edge-device collaboration[J]. Journal of Software, 2024, 35(3): 1154-1172. (in Chinese)

[5] CHEN J J, HU C Q, SHENG W H, et al. Sensitivity-aware personalized differential privacy guarantees for online social networks[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 3116-3130.

[6] LIU Y H, WANG T H, LIU Y X, et al. Edge-protected triangle count estimation under relationship local differential privacy[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(10): 5138-5152.

[7] ZILBERMAN A, DVIR A, STULMAN A. SPRINKLER: A multi-RPL man-in-the-middle identification scheme in IoT networks[J]. IEEE Transactions on Mobile Computing, 2024, 23(10): 9971-9988.

[8] 康海燕, 王骁识. 基于数据特征相关性和自适应差分隐私的深度学习方法研究[J]. 电子学报, 2024, 52(6): 1963-1976.
KANG H Y, WANG X S. Research on the deep learning method based on data feature relevance and adaptive differential privacy[J]. Acta Electronica Sinica, 2024, 52(6): 1963-1976. (in Chinese)

[9] 李森森, 刘燕江, 郁滨, 等. 边缘计算环境下基于PUF的多接收者匿名签密方案[J]. 电子学报, 2024, 52(12): 4087-4100.

LI S S, LIU Y J, YU B, et al. PUF-based multi-receiver anonymous signcryption scheme in edge computing[J]. Acta Electronica Sinica, 2024, 52(12): 4087-4100. (in Chinese)

[10] 赵琪, 樊婷, 韦永壮. 基于MILP对轻量级密码算法FBC-128的差分分析[J]. 电子学报, 2024, 52(6): 1896-1902.
ZHAO Q, FAN T, WEI Y Z. MILP-based differential cryptanalysis of the FBC-128 lightweight cipher[J]. Acta Electronica Sinica, 2024, 52(6): 1896-1902. (in Chinese)

[11] WU L, QIN C Y, XU Z H, et al. TCPP: Achieving privacy-preserving trajectory correlation with differential privacy[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 4006-4020.

[12] LUO J, REN W Q, GAO X W, et al. Multi-exposure image fusion via deformable self-attention[J]. IEEE Transactions on Image Processing, 2023, 32: 1529-1540.

[13] LIU J, CHEN S H, HE X J, et al. VALOR: Vision-audio-language omni-perception pretraining model and dataset[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(2): 708-724.

[14] ZHOU S H, CHEN D S, PAN J S, et al. Adapt or perish: Adaptive sparse transformer with attentive feature refinement for image restoration[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2024: 2952-2963.

[15] CHEN X Y, CHENG Z H, CAI H Q, et al. Laplacian convolutional representation for traffic time series imputation[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(11): 6490-6502.

[16] WANG G, QI Q, HAN R, et al. P2CEFL: Privacy-preserving and communication efficient federated learning with sparse gradient and dithering quantization[J]. IEEE Transactions on Mobile Computing, 2024, 23(12): 14722-14736.

[17] WEI K, LI J, MA C, et al. Personalized federated learning with differential privacy and convergence guarantee[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 4488-4503.

[18] BULATOV A, KURATOV Y, KAPUSHEV Y, et al. Beyond attention: Breaking the limits of transformer context length with recurrent memory[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 38(16): 17700-17708.

[19] ZENG Y, ZHANG X S, LI H, et al. X^2-VLM: All-in-one pre-trained model for vision-language tasks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(5): 3156-3168.

[20] 桑海峰, 王金玉, 陈旺兴, 等. 基于第一视角的非自回归行人轨迹预测模型[J]. 电子学报, 2023, 51(5): 1266-1272.
SANG H F, WANG J Y, CHEN W X, et al. Non-autoregressive pedestrian trajectory prediction model based on the first perspective[J]. Acta Electronica Sinica, 2023, 51(5): 1266-1272. (in Chinese)

[21] YUE G F, YAN L, KANG L W, et al. AdapLDP-FL: An adaptive local differential privacy for federated learning[J]. IEEE Transactions on Mobile Computing, 2025, 24(6): 5569-5583.

[22] FUKAMI T, MURATA T, NIWA K T, et al. DP-norm: Differential privacy primal-dual algorithm for decentralized federated learning[J]. IEEE Transactions on Information Forensics and Security, 2024, 19: 5783-5797.

[23] DINH V, HO L, NGUYEN C. Hamiltonian Monte Carlo on ReLU neural networks is inefficient[C]//Advances in Neural Information Processing Systems 37. San Diego: NeurIPS Inc., 2024: 134107-134126.

[24] SHI L S, WANG L, ZHOU S P, et al. Trajectory unified transformer for pedestrian trajectory prediction[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2024: 9641-9650.

[25] XU C X, TAN R T, TAN Y H, et al. EqMotion: Equivariant multi-agent motion prediction with invariant interaction reasoning[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 1410-1420.

[26] MAO W B, XU C X, ZHU Q, et al. Leapfrog diffusion model for stochastic trajectory prediction[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2023: 5517-5526.

[27] BAE I, OH J, JEON H G. EigenTrajectory: Low-rank descriptors for multi-modal trajectory forecasting[C]//2023 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2024: 9983-9995.

[28] DONG W H, ZHU H D, LIN S H, et al. Fusion-mamba for cross-modality object detection[EB/OL]. (2024-04-14)[2025-11-10]. https://arXiv.org/abs/2404.09146.

[29] KHAN M, AHMAD J, EL SADDIK A, et al. Drone-HAT: Hybrid attention transformer for complex action recognition in drone surveillance videos[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2024: 4713-4722.

[30] DAI G H, LU P, NING X F, et al. DiTFastAttn: Attention compression for diffusion transformer models[C]//Ad-

vances in Neural Information Processing Systems 37. San Diego: eurIPS Inc., 2024: 1196-1219.

[31] XU C X, MAO W B, ZHANG W J, et al. Remember intentions: Retrospective-memory-based trajectory prediction[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2022: 6478-6487.

[32] HASıRCıOĞLU B, GÜNDÜZ D. Communication efficient private federated learning using dithering[C]//2024 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway: IEEE, 2024: 7575-7579.

[33] TRAN P, WU H R, YU C J, et al. What truly matters in trajectory prediction for autonomous driving[EB/OL]. (2023-11-6)[2025-11-10]. https://arXiv.org/abs/2306.15136.

作者简介

魏建好　男,1989年8月出生于河南省信阳市. 现为湖南工商大学副教授. 主要研究方向为人工智能安全.
E-mail: jianhao@hutb.edu.cn

周淳森　男,2000年11月出生于广西壮族自治区梧州市. 现为湖南工商大学在读研究生. 主要研究方向为智慧交通安全预测.
E-mail: zhoutingsen666@163.com

李　闯　男,1990年11月出生于湖南省湘乡市. 现为湖南工商大学副教授. 主要研究方向为高性能计算、人工智能.
E-mail: chuangli@hutb.edu.cn

文艳华　女,1985年9月出生于湖南省益阳市. 现为湖南工商大学副教授. 主要研究方向为联邦学习.
E-mail: yanhua-wen@hutb.edu.cn

李克勤　男,1963年5月出生于上海市. 现为湖南大学教授. 主要研究方向为并行计算、边缘计算、云计算.
E-mail: likq@hnu.edu.cn