移动群智感知中的因果增强深度时空真值发现方法

蒋伟进[1,2]，石顺究[1,2]，欧阳宇明[1,2]，许宇胜[3]，刘溪溪[1,2]，崔新雨[1,2]，刘志华[1,2]
（1.湖南工商大学计算机学院，湖南 长沙 410205；2.湘江实验室，湖南 长沙 410205；3.北方导航控制技术股份有限公司，北京 100176）

摘要：针对无真值移动群智感知（MCS）场景中预测误差传播与环境混淆导致的质量误归因问题，提出了一种因果增强深度时空真值发现（CE-DSTTD）框架。该框架由两部分组成：一是多尺度时空Transformer（MSTT）预测参照生成模块，用于生成稳定预测参照以降低误差传播；二是去混淆校正模块，通过构建面向质量评估的结构因果模型（SCM），并驱动因果增强的基于预测与声誉的真值发现（CE-PRBTD）算法对质量评分与声誉更新实施去混淆校正。实验结果表明，所提方法在低质量用户识别F1分数、声誉区分度与降噪效果方面均优于对比方法，并在跨域场景下表现出良好的稳健性。该方法可提升无真值MCS场景下质量评估的稳健性与合理性。

关键词：移动群智感知；真值发现；时空预测；去混淆校正；结构因果模型
中图分类号：TP181
文献标志码：A
doi：10.11959/j.issn.1000–436x.TXXB260206

Causal-enhanced deep spatio-temporal truth discovery method for mobile crowdsensing

Jiang Weijin[1,2], Shi Shunjiu[1,2], Ouyang Yuming[1,2], Xu Yusheng[3], Liu Xixi[1,2], Cui Xinyu[1,2], Liu Zhihua[1,2]
1. School of Computer Science, Hunan University of Technology and Business, Changsha 410205, China
2. Xiangjiang Laboratory, Changsha 410205, China
3. North Navigation Control Technology Co., Ltd., Beijing 100176, China

Abstract: To address the problem of quality misattribution caused by prediction error propagation and environmental confounding in no-ground-truth mobile crowdsensing (MCS) scenarios, a causal-enhanced deep spatio-temporal truth discovery (CE-DSTTD) framework was proposed. The framework consisted of two components. The first was a multi-scale spatio-temporal Transformer (MSTT) prediction reference generation module, which generated stable prediction references to mitigate error propagation. The second was a deconfounding correction module, which constructed a structural causal model (SCM) for quality assessment and drove the causal-enhanced prediction and reputation-based truth discovery (CE-PRBTD) algorithm to perform deconfounding correction on quality scoring and reputation updating. Experimental results show that the proposed method outperforms the comparison methods in terms of low-quality user identification F1 score, reputation discrimination, and noise reduction, and exhibits good robustness in cross-domain scenarios. The proposed method can enhance the robustness and rationality of quality assessment in no-ground-truth MCS scenarios.

Key words: mobile crowdsensing, truth discovery, spatio-temporal prediction, deconfounding correction, structural causal model

收稿日期：2026–04–15；**修回日期：**2026–05–28
通信作者：石顺究，2233145030@qq.com
基金项目：国家自然科学基金资助项目（No.61772196）；湖南省自然科学基金资助项目（No.2020JJ4249）；湖南省教育厅科学研究重点项目（No.24A0446, No.24A0753）；湖南工商大学研究生科研创新项目（No.CX2025YB028, No.CX2025YB029）
Foundation Items: The National Natural Science Foundation of China (No.61772196), The Natural Science Foundation of Hunan Province (No.2020JJ4249), The Key Scientific Research Projects of the Hunan Provincial Department of Education (No.24A0446, No.24A0753), The Postgraduate Scientific Research Innovation Project of Hunan University of Technology and Business (No.CX2025YB028, No.CX2025YB029)

0　引言

随着无线通信和智能终端的发展，移动群智感知（mobile crowdsensing, MCS）已成为环境监测、交通监控和公共安全等场景的重要数据获取方式[1-4]。在无真值场景下，如何从多源观测中恢复可信结果，是MCS数据融合与质量保障的核心问题[5-6]。

由于MCS具有开放性和弱约束特征，用户上报数据易受恶意行为、设备差异和环境噪声影响而产生偏差。相关研究虽从隐私保护[7]、异常检测、任务分配和激励机制等方面进行了改进，但数据质量偏差仍会沿处理链路传播，进而影响系统整体性能。因此，研究无真值场景下兼顾有效性、公平性与可解释性的质量保障方法具有重要意义。

在无真值场景下，质量保障通常依赖真值发现（truth discovery, TD）思想：更接近真实值的用户应获得更高权重，从而在聚合中发挥更大作用并抑制噪声数据[8]。为此，许多方法采用“真值发现–声誉估计”的迭代框架，通过联合估计任务真值与用户权重实现低质量数据抑制。然而，现有方法仍存在两个关键瓶颈：一是预测误差会在质量评分和声誉更新中累积传播；二是环境扰动引起的系统性偏差易被误归因于用户可靠性，导致声誉估计偏置。

针对上述问题，本文提出了因果增强深度时空真值发现（causal-enhanced deep spatio-temporal truth discovery, CE-DSTTD）框架。该框架设计多尺度时空Transformer（multi-scale spatio-temporal Transformer, MSTT）预测参照生成模块，为后续评估提供更稳定的误差代理；随后构建数据质量结构因果模型（structural causal model, SCM），利用因果结构学习与效应估计识别环境因素对质量的因果贡献，并基于后门调整对质量评分与声誉更新进行去混淆校正，从而在保持识别能力的同时提升评估公平性与结果可解释性。本文主要贡献如下。

（1）提出了CE-DSTTD框架，将预测参照生成、质量评估、声誉更新与真值发现纳入统一闭环，缓解预测误差传播与环境混淆误归因问题。

（2）设计面向因果校正的MSTT预测参照生成模块，提出含趋势一致性约束L_{corr}与空间连续性约束L_{smooth}的多目标增强损失，将误差代理的时序稳定性与空间连续性纳入训练目标，使预测残差在统计特性上满足CE-PRBTD因果建模的输入需求，从而降低预测噪声对下游环境效应估计与声誉更新的干扰。

（3）设计面向质量评估的环境去混淆机制，利用结构因果模型思想估计环境因素对质量代理的系统性影响，并在评分与声誉更新前进行校正，从而减轻环境因素导致的误归因，提高用户识别效果与声誉区分能力。

1　相关工作

本节从真值发现、时空预测和因果质量评估3个方面综述相关工作，分别对应结果聚合、参考值构造和误差归因分析，为本文方法提供研究基础。

1.1　真值发现

MCS作为新兴的泛在数据采集范式，近年来已引起学术界与产业界的持续关注[9]。MCS依赖参与者协同上报数据，其长期质量保障是移动计算的重要问题[10-14]。在无真值或弱监督条件下，TD通过对多源观测加权聚合，同时估计任务真值与参与者权重[15-16]。围绕“真值估计–权重学习”这一核心机制，TruthFinder、CRH、CATD及统一最大似然框架等方法形成了较成熟的TD体系[17-19]。针对MCS的隐私与可审计需求，PPTD、L-PPTD、RDPP-TD、TP-MCS和FedTruthFinder等方法进一步从隐私保护、可验证性和联邦部署等方面扩展了该方向[20-25]。

然而，现有方法在质量建模上存在两方面不足：（1）多数工作仅以地理位置、电量、带宽等客观指标刻画参与者质量，对主观行为与策略性虚假上报的系统研究相对有限[26-30]；（2）现有方法普遍将观测偏差归因于用户质量，缺乏对误差来源的显式区分：当观测偏差同时受用户行为与环境扰动影响时，权重学习与声誉更新将产生系统性偏置。因此，如何从相关性聚合进一步过渡到面向误差来源的归因分析，仍是MCS真值发现的关键问题。

1.2　时空预测

在真值不可得的MCS场景中，常用策略是以时空预测模型生成参考估计，并以观测残差构造质量特征，进而实现异常检测与低质量数据识别。Li等[9]提出了PRBTD方法，通过相关性聚焦时空Transformer网络（correlation-focused spatio-temporal Transformer network, CFSTTN）预测真值并提取稳定的时空蕴含关系，结合用户声誉识别恶意用户和低质量数据。然而，该类方法的质量判别主要依赖统计相关性，未显式建模环境混淆误差，可能在

数据过滤与声誉更新中积累系统性偏差。

城市时空预测模型已历经多代演进：DeepST与ST-ResNet分别以卷积神经网络和多分支残差结构建模时空依赖，奠定了早期深度时空预测的基础[31-32]。DeepSTN+以ConvPlus结构增强跨区域长程依赖[33]。此后，ST-TSNet、ASTTN与SSL-ST-MFormer从局部–全局依赖、自适应图结构与自监督表征等角度进一步提升性能[34-36]。MSSTFormer、MLCAFormer与DAT-STAN引入多尺度与结构学习以应对复杂时空模式[37-39]。UrbanGPT与LLM-TFP则探索大语言模型用于少样本预测，但稳定性与可解释性仍有待进一步验证[40-41]。

然而，MCS质量评估中的残差往往同时混叠用户行为误差与环境扰动，仅提升预测精度并不足以保证评估公平性。因此，仍需引入面向环境混淆的显式建模与校正机制。

1.3 因果质量评估

为避免将环境因素对观测质量的影响误归因于用户能力，质量评估需要在统计相关性分析基础上进一步引入因果归因视角。本节从因果推断基础框架、深度因果学习及MCS时空场景中的因果方法三方面概述研究进展。

在因果推断基础框架方面，Pearl[42]提出的SCM及do演算以有向无环图刻画变量间的因果生成机制，通过阻断混淆路径实现因果调整与去偏。在效应估计层面，倾向得分匹配等经典方法形成了成熟的操作化范式[43]。Hernán等[44]系统地阐明了因果推断的关键识别假设与操作条件。PC算法等支持从观测数据中学习因果图[45-46]。然而，上述方法主要面向低维静态场景，在MCS数据的高维、时空相关与非独立同分布数据中，可识别性保证与计算效率面临挑战。

在方法体系方面，因果表征学习、深度因果发现与深度因果推断三类范式将神经网络引入高维非结构化数据的因果建模[47]，为提升可识别性与可解释性提供了新的工具链。但现有工作多聚焦于医疗、推荐等独立个体数据，对MCS场景中时空自相关与动态环境混杂的系统探索仍不充分。

在MCS应用层面，无真值质量保障研究常通过深度时空模型构建预测参照以支撑真值发现。PRBTD构建了基于预测参照和声誉更新的闭环机制，但质量评估仍依赖预测残差的统计相关性，缺乏对环境混杂的显式因果处理[9]。Andreou等[48]提出了同化因果推断，利用贝叶斯数据同化由观测效应反向追溯潜在原因，支持动态因果交互的在线追踪，为时变混杂的刻画提供了新路径，但其与MCS质量评估及声誉更新机制的结合仍有待进一步探索。

综合来看，现有研究存在3点不足：一是经典方法在高维时空数据中面临识别性与效率挑战；二是深度因果学习缺乏面向混杂环境的效应分解；三是因果分析尚未有效嵌入声誉更新过程。基于此，本文提出了CE-DSTTD框架，将预测参照构造与因果去混淆纳入统一闭环，以提升质量评估与声誉更新的稳健性与可解释性。本文的主要符号及其含义如表1所示。

表1 主要符号及其含义

符号	含义
t	时间索引
T_h	历史输入序列长度
H,W	空间网格尺寸
N	空间位置数
C	输入通道数
d	隐表示/嵌入维度
X_{t+1}	监督目标
$\hat{X}_{t+1}$	预测参照
U	用户能力变量
E	环境因素变量
R	区域特征变量
q	数据质量变量
e_{data}	误差代理
Z_{adj}	后门调整集
λ	因果校正权重
λ_{mix}	Mixup数据增强混合系数
λ_{wd}	权重衰减系数
δ_{es}	早停容忍阈值
B	批大小
U^-、U^+	低质量用户集合、高质量用户集合
B_h	每个用户的历史buffer长度
$\hat{\beta}_E$	环境因素边际效应回归系数
Z_0	时序嵌入输出的中间特征表示
N_{epoch}	训练总轮数
E_w	Warmup轮数
E_{mix}	应用Mixup数据增强的训练轮数上限

2　因果增强深度时空真值发现方法

2.1　问题分析与框架构建

MCS 场景下的数据质量保障通常缺乏可用真值，现有方法多采用“预测参照–偏差构造–质量评分–声誉更新–真值发现”的闭环流程，但仍面临预测误差传播和环境混淆误归因两个问题。为此，本文提出了 CE-DSTTD 框架，以预测参照生成与因果校正的联合优化为核心：MSTT 模块在 CE-PRBTD 的误差代理质量约束下，联合优化时序趋势一致性与空间结构连续性；CE-PRBTD 模块则基于该误差代理构建 SCM，通过后门调整校正环境因素的系统性影响。两模块在损失设计层面深度耦合，共同决定滑动窗口回归的估计稳定性与后门调整集的可靠性，构成以因果校正质量为导向的端到端闭环。整体流程如图 1 所示。

2.2　MSTT 模型

为降低预测误差在后续质量评估中的传播风险，本文设计 MSTT 作为预测参照生成模块。MSTT 结合多尺度卷积、可学习位置编码与 Transformer 编码器，增强局部结构建模、长程依赖捕获和空间异质性适配能力。

2.2.1　模型设计与结构

在 CE-DSTTD 框架中，预测参照生成模块的核心职责并非追求最低预测误差，而是为后续因果校正提供具备特定质量属性的误差代理 e_{data}。现有时空预测模型（如 ST-ResNet、ASTTN 等）以最小化预测误差为优化目标，其残差分布未针对质量评估场景加以约束，存在以下不足。

（1）残差噪声放大效应。局部预测抖动被质量评分机制（式(15)）非线性放大，使误差代理在用户间产生虚假质量差异，影响声誉更新。

（2）空间连续性缺失。相邻区域预测误差的系统性跳变会被误判为用户质量差异，干扰区域代理特征的估计，影响后门调整集 Z_{adj} 的可靠性。

（3）时序趋势不一致。跨时刻残差缺乏趋势约束，导致滑动窗口内的误差漂移，使环境效应系数 $\hat{\beta}_E$（式(18)）产生偏置。

上述 3 项约束由 CE-PRBTD 的因果建模结构决定，而非通用时空预测场景的优化目标。为此，MSTT 模型在损失函数上引入时序趋势一致性约束 L_{corr} 与空间连续性约束 L_{smooth}（式(5)）。在结构上采用“局部表征–多尺度融合–全局依赖–异质性适配”的层次化架构，并通过 Pre-LayerNorm 与随机深度保障训练稳定性。各模块具体设计如下。

为将历史观测序列映射为统一维度的空间特征图，并为后续的多尺度融合与全局依赖建模提供输入，本文采用卷积式时序特征嵌入模块。设历史输入序列为 $X_{t-T+1:t} \in \mathbb{R}^{T\times H\times W\times C}$，其中 T 为历史长度，$H\times W$ 为空间网格尺寸，C 为输入通道数。实现中将时间维 T 视作卷积输入通道，经两层卷积逐级提取时空局部模式，得到中间表示 $Z_0 \in \mathbb{R}^{H\times W\times d}$。

定义 1　时序嵌入。给定输入 $X_{t-T+1:t}$，时序嵌

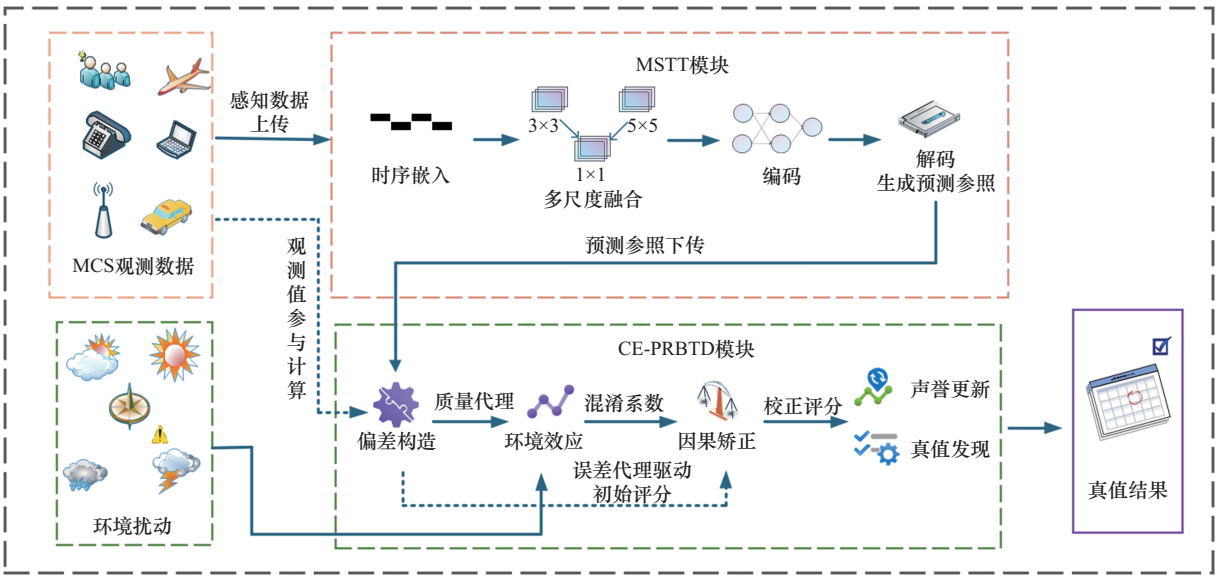


图 1　CE-DSTTD 框架

入定义为：

$$\widetilde{Z_0} = \sigma(\mathrm{BN}(\mathrm{Conv}_1(X_{t-T+1:t}))),$$
$$Z_0 = \sigma(\mathrm{BN}(\mathrm{Conv}_2(\widetilde{Z_0}))) \tag{1}$$

其中，Conv_1和Conv_2为卷积算子，BN为批归一化（batch normalization, BN），$\sigma(\cdot)$为激活函数。

给定特征图，采用双分支卷积与1×1卷积完成多尺度特征融合，并引入残差连接保持表征稳定。

定义 2 多尺度特征融合。给定特征图$Z_0 \in \mathbb{R}^{B\times d\times H\times W}$，多尺度特征融合定义为：

$$Z_0^{(3)} = \mathrm{Conv}_{3\times 3}(Z_0; W_3),$$
$$Z_0^{(5)} = \mathrm{Conv}_{5\times 5}(Z_0; W_5),$$
$$Z_{0\,\mathrm{cat}} = \mathrm{Concat}\left(Z_0^{(3)}, Z_0^{(5)}\right) \in \mathbb{R}^{B\times 2d\times H\times W},$$
$$Z_{0\,\mathrm{fused}} = \mathrm{GELU}(\mathrm{BN}(\mathrm{Conv}_{1\times 1}(Z_{0\,\mathrm{cat}}; W_f))) + Z_0 \tag{2}$$

其中，W_3, W_5, W_f为可学习参数，残差连接用于稳定梯度并保持表征一致性。

MCS场景具有强空间异质性，仅依赖固定正弦位置编码难以适配非均匀空间结构。为此，本文引入二维可学习位置编码（learnable 2D positional encoding, LPE），为各空间网格构建可训练的位置先验，并通过端到端训练学习不同区域的重要性及其差异化模式。

定义 3 LPE。对于空间位置(h,w)，LPE定义为：

$$\mathrm{LPE}(h,w) = \mathrm{Embed}_{\mathrm{row}}[h] + \mathrm{Embed}_{\mathrm{col}}[w] \tag{3}$$

其中，$\mathrm{Embed}_{\mathrm{row}} \in \mathbb{R}^{H\times d}$和$\mathrm{Embed}_{\mathrm{col}} \in \mathbb{R}^{W\times d}$为可学习参数矩阵，通过反向传播优化。

在获得融合特征并叠加位置编码后，MSTT使用前置层归一化（Pre-LayerNorm, Pre-LN）Transformer编码器建模跨时间与区域的全局依赖关系。对输入序列表示Z，单层编码器为：

$$Z' = Z + \mathrm{MHA}(\mathrm{LN}(Z)),$$
$$Z'' = Z' + \mathrm{FFN}(\mathrm{LN}(Z')) \tag{4}$$

其中，MHA（multi-head attention）为多头自注意力，FFN（feed-forward network）为前馈网络，LN（layer normalization）为层归一化。

Pre-LN结构能够缓解深层Transformer训练中的梯度不稳定问题，同时引入随机深度以提升泛化稳定性并降低过拟合风险。通过LPE与Transformer编码器的结合，MSTT能够在保持空间差异感知的同时捕捉跨区域长程依赖，为后续解码预测提供更完整的全局信息支撑。

2.2.2 解码预测与训练策略

为将Transformer编码后的全局表示映射为最终预测结果，MSTT采用渐进式多尺度解码模块输出预测图$\hat{X}_{t+1}$，以增强空间连续性表达并缓解局部噪声对输出的影响。为提升预测质量并使预测参照更适合作为后续质量评估的基准，本文采用多目标增强损失函数，从绝对误差、趋势一致性与空间结构3个方面联合优化模型。训练损失定义为：

$$L = \alpha L_{\mathrm{mse}} + \beta L_{\mathrm{mae}} + \gamma L_{\mathrm{corr}} + \delta L_{\mathrm{smooth}} \tag{5}$$

其中，L_{mse}与L_{mae}分别用于约束预测的均方误差与绝对误差，L_{corr}用于强化预测与真实序列在变化趋势上的一致性，L_{smooth}用于约束空间连续性并提升结构稳定性。

在训练策略上，本文采用AdamW优化器并结合Warmup与Cosine调度以改善收敛过程，同时引入Mixup数据增强、梯度裁剪与早停机制提升训练稳定性与泛化能力。完整训练与推断流程如算法1所示。在推断阶段，MSTT输出预测参照$\hat{X}_{t+1}$供后续因果增强模块使用，并进一步构造误差代理$e_{\mathrm{data},t+1}$（对应式(11)），为CE-PRBTD模块提供对预测误差的可用刻画，从而支撑后续质量评分、声誉更新与真值发现过程的稳定运行。

算法1 MSTT

输入 历史序列$\{X_{t-T+1:t}\}$与监督目标$\{X_{t+1}\}$，测试长度N_{test}，总轮数N_{epoch}，批大小B，学习率η_0，权重衰减λ_{wd}，Warmup轮数E_w，Mixup参数$(\alpha_{\mathrm{mix}}, E_{\mathrm{mix}})$，Mixup混合系数$\lambda_{\mathrm{mix}}$，梯度阈值$g_{\max}$，早停参数$(P, \delta_{\mathrm{es}})$，损失权重$(\alpha,\beta,\gamma,\delta)$

输出 最优参数θ^*，预测参照$\hat{X}_{t+1}$，误差代理e_{data}，RMSE/MAE/R^2

1) 数据划分：训练/验证/测试 ← Split($\{X_{t-T+1:t}, X_{t+1}\}, N_{\mathrm{test}}$)，构建DataLoader
2) 初始化MSTT模型f_θ，设增强损失$L = \alpha L_{\mathrm{mse}} + \beta L_{\mathrm{mae}} + \gamma L_{\mathrm{corr}} + \delta L_{\mathrm{smooth}}$（式(5)）
3) 初始化AdamW$(\eta_0, \lambda_{\mathrm{wd}})$与调度器Warmup$(E_w)$ + Cosine
4) best ← $+\infty$, $\theta^* \leftarrow \theta$，早停器ES ← init$(P, \delta_{\mathrm{es}})$
5) for { epoch = 1,2,⋯,E } do
6) for $\{(X_{t-T+1:t}, X_{t+1}) \in B_{\mathrm{tr}}\}$ do
7) if epoch ⩽ E_{mix} then
8) $(\tilde{X}, Y_a, Y_b, \lambda_{\mathrm{mix}}) \leftarrow$ Mixup$(X_{t-T+1:t}$,

X_{t+1}; α_{mix})

9) $\hat{X} \leftarrow f_\theta(\tilde{X})$, loss $\leftarrow \lambda_{\text{mix}} L(\hat{X}, Y_a) + (1-\lambda_{\text{mix}}) L(\hat{X}, Y_b)$

10) else

$\hat{X} \leftarrow f_\theta(X_{t-T+1:t})$, loss $\leftarrow L(\hat{X}, X_{t+1})$

11) 反向传播更新θ，梯度裁剪 $\| \nabla_\theta \| \leqslant g_{\max}$

12) end if

13) end for

14) 验证：val $\leftarrow \frac{1}{|B_{\text{val}}|} \sum L(f_\theta(X_{t-T+1:t}), X_{t+1})$；调度器 step

15) if val < best then best $\leftarrow$ val, $\theta^* \leftarrow \theta$

16) 更新 ES(val)

17) if ES = stop then break

18) end if

19) end if

20) end for

21) 加载最优模型：$\theta \leftarrow \theta^*$

22) 测试推断：$\hat{X}_{t+1} \leftarrow f_\theta(X_{t-T+1:t})$ //对测试集逐批计算并反归一化

23) 生成误差代理：$e_{\text{data},t+1} \leftarrow \| X_{t+1} - \hat{X}_{t+1} \|$（式(11)）

24) return $\theta^*, \hat{X}_{t+1}, e_{\text{data},t+1}$, RMSE, MAE, R^2

综上所述，MSTT 多目标损失中的L_{corr}确保误差代理在时序窗口内的趋势稳定性，直接提升 CE-PRBTD 滑动窗口回归（式(18)）对环境效应系数$\hat{\beta}_E$的估计精度；L_{smooth}的空间连续性约束则压缩区域间系统性预测偏差，降低区域代理f_{region}的测量噪声，从而增强后门调整集Z_{adj}的可靠性。

2.3　CE-PRBTD 算法

在预测辅助真值发现中，观测偏差同时受用户能力与环境扰动影响。若直接依据观测偏差更新声誉，环境扰动可能被误归因于用户能力。为此，本文提出了 CE-PRBTD 算法，通过 SCM 估计环境因素对质量评分的系统性影响，并在声誉更新前进行去混淆校正。

2.3.1　因果建模与基本假设

本文在可观测代理变量条件下近似估计环境因素对质量相关量的系统性影响，并将其用于修正声誉更新偏差。设U表示用户能力相关因素，E表示环境因素，R表示区域特征，q表示质量相关量。本文考虑式(6)所示结构关系。

$$q = g(U, E, R, \varepsilon) \tag{6}$$

其中，ε为未建模随机扰动。由于U和R难以直接观测，本文以其可观测代理构造调整变量集合Z_{adj}，用于估计环境因素E对q的边际影响。

图 2 给出了数据质量评估阶段的变量关系与环境混淆建模思路。图 2 表明，用户U与R不仅会影响q，还会影响环境因素E，因此若直接依据原始误差代理进行质量评分与声誉更新，容易将环境变化导致的系统性偏差误判为用户能力不足。因此，本文对环境效应进行近似调整，并据此构造去混淆校正项。

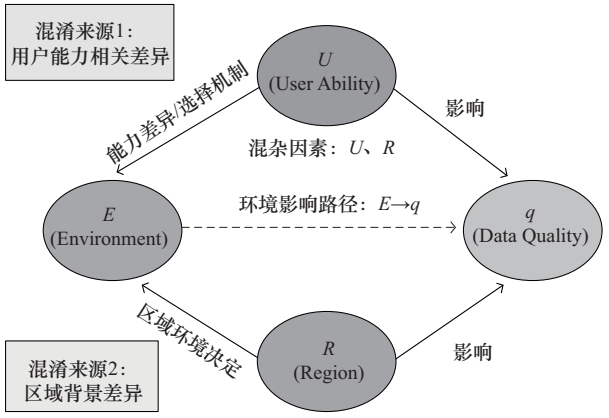


图2　数据质量评估阶段的变量关系与环境混淆建模思路

为保证上述估计具备可操作性，本文作如下近似识别假设。

假设 1　条件可交换性。给定可观测调整变量集合$Z_{\text{adj}} = \{U, R\}$后，不存在同时影响环境因素$E$与质量相关量$q$的重要未观测混杂因素。等价地，在潜在结果意义下，环境暴露与质量结果在给定Z_{adj}后满足条件独立，即：

$$q(e) \perp E \,\big|\, Z_{\text{adj}} \tag{7}$$

该假设在本文中以近似形式成立：通过引入能够覆盖主要混杂路径的可观测代理变量组合，旨在减弱而非完全消除未观测混杂对声誉更新的影响，并在 3.5.5 节通过受控仿真对假设违背程度的影响进行量化分析。

假设 2　一致性。当观测到环境因素$E=e$时，其观测到的质量相关量等于对应环境水平下的潜在结果，即：

$$q = q(e) \tag{8}$$

该假设意味着本文定义的环境变量具有明确且

一致的含义，不同环境水平下的观测结果可视为相应潜在结果在实际暴露下的实现。

假设 3 正值性。对于任意满足支持条件的(u, r)组合，不同环境水平均具有非零出现概率，即：

$$P(E = e \mid U = u, R = r) > 0 \tag{9}$$

该假设保证在给定调整变量后，不同环境水平下的样本具有可比性，从而避免后门调整退化为外推估计。

为增强上述假设在本文场景中的合理性，本文以历史声誉作为用户能力代理，以环境质量指数作为环境代理，以区域标识及其派生统计量作为区域代理。其中，TaxiBJ 场景中的环境质量指数依据时段属性以规则映射方式构造，UCI PM2.5 场景则由 5 类气象变量经 Z-score 标准化后加权合成，具体构造细节在 3.1 节叙述。

代理变量的充分性取决于其是否能近似封锁所有同时影响环境因素 E 与质量相关量 q 的主要混杂路径。在 MCS 场景中，潜在的未观测混杂来源主要有 3 类：用户端因素（如设备精度异质性、移动轨迹规律）长期稳定，历史声誉作为多期上报偏差的加权积累，可有效吸收其长期影响；区域端因素（如基础设施差异、局部信号遮挡）在地理上具有稳定性，区域标识及其派生统计量可覆盖区域固定效应，缓解地理异质性导致的误归因；短期动态因素（如突发事件、极端天气异常峰值）是残存未观测混杂的主要来源，本文采用滑动窗口机制对 $\hat{\beta}_E$ 进行周期性更新，使环境效应估计能够适应动态变化，其鲁棒性将在 3.5.5 节进行定量验证。

在上述因果假设与代理变量框架确立后，本文进一步形式化表达环境混杂导致误归因的观测生成过程，采用的结构化关系为：

$$U \to X, E \to X, R \to X, X \to q \tag{10}$$

其中，X 为用户观测，q 为由观测偏差构造得到的质量相关量。在无真值场景中，本文以观测 X_{t+1} 与预测参照 $\hat{X}_{t+1}$ 的差异构造数据误差代理。

$$e_{\text{data},t+1} = \| X_{t+1} - \hat{X}_{t+1} \| \tag{11}$$

用于在线质量评估、环境校正与声誉更新。在离线评估中，为验证误差分解与归因效果，进一步定义真实误差为：

$$e_{t+1}^{\text{true}} = \| X_{t+1} - X_{t+1}^{\text{true}} \| \tag{12}$$

其中，e_{t+1}^{true} 表示时刻 t+1 的真实误差。式(11)用于算法在线闭环，式(12)用于离线验证，不参与实际部署时的在线质量评估过程。

由于误差代理同时受用户能力、区域背景和环境扰动影响，不能直接作为用户能力差异的度量指标。为此，本文在后续环境效应估计模块中结合用户能力代理、环境因素代理和区域特征代理进行局部窗口建模，估计环境因素的边际影响，并据此修正原始质量评分，以减弱环境扰动对声誉更新的干扰。代理变量的选取策略在上述充分性分析中已作定性论证，其对系统性能的实际影响，以及假设 1 在不同违背程度下的鲁棒性，将分别在 3.5.2 节的消融实验和 3.5.5 节的受控仿真中进行定量验证。

2.3.2 环境效应估计与质量校正

在上述因果建模基础上，本文进一步将抽象变量 U、E、R 和 q 转化为可计算和可学习的代理特征，并在此基础上估计环境因素对质量代理的系统性影响，进而构造去混淆校正项。对于每条数据记录，定义特征集合为：

$$F_t = \left\{ f_{\text{user},t+1}, f_{\text{env},t+1}, f_{\text{region},t+1}, e_{\text{data},t+1}, q_{\text{data},t+1} \right\} \tag{13}$$

其中，$f_{\text{user},t+1}$ 为用户能力代理，$f_{\text{region},t+1}$ 为区域特征代理，$f_{\text{env},t+1}$ 为环境因素代理，$e_{\text{data},t+1}$ 为由式(11)得到的误差代理，$q_{\text{data},t+1}$ 为用于环境效应估计的质量代理。

为增强可复现性，本文将多个可观测环境变量归一化加权，构建环境质量指数。

$$f_{\text{env}} = \sum_{m=1}^{K_e} \omega_m \operatorname{Norm}(z_m), \sum_{m=1}^{K_e} \omega_m = 1, \omega_m \geq 0 \tag{14}$$

其中，z_m 可取时段、天气、网络质量等，$\operatorname{Norm}(\cdot)$ 为归一化函数，K_e 为环境因素的测量变量个数。对于 TaxiBJ 场景，f_{env} 依据时段属性以规则映射方式构造（详见 3.1.1 节）；对于 UCI PM2.5 场景，f_{env} 则由风速、降雨量、温度偏离等可观测气象变量合成。进一步，将误差代理映射为区间[0,1]上的质量代理。

$$q_{\text{data},t+1} = \frac{1}{1 + e_{\text{data},t+1}} \tag{15}$$

其中，$q_{\text{data},t+1}$ 为用于“环境效应估计”的可回归质量代理，$q_{\text{raw},t+1}$ 为真值发现内部迭代机制得到的质量评分，两者均位于区间[0,1]，语义相近但来源不同。前者用于表征观测偏差在多大程度上可能受到环境因素影响，后者用于真值发现与声誉更新。

CE模块将利用估计的环境影响对$q_{\mathrm{raw},t+1}$进行去混淆校正，并输出校正后的质量评分。

在此基础上，本文进一步估计环境因素对质量代理的系统性影响，并分析环境、用户与区域三类因素对质量波动的相对贡献。本节的目标不是在强干预意义下严格恢复环境因素对真实质量的平均因果效应，而是在可观测代理变量条件下，构造可操作的ACE估计形式，以近似刻画环境因素对质量相关量的影响强度，并为后续环境校正提供依据。

定理1　基于调整的环境效应估计形式。在假设1~假设3成立的条件下，考虑图2所示的数据质量生成机制，其中U、E、R与q分别表示用户能力、环境因素、区域特征与质量相关量，其对应的数据质量因果图记为：

$$G=(U,E,R,q,\mathrm{Edges}) \tag{16}$$

若取调整集$Z_{\mathrm{adj}}=\{U,R\}$，则Z_{adj}满足后门准则：Z_{adj}不包含E的任何后代节点；Z_{adj}阻断从E到q的所有后门路径。

在后门调整思想下，环境因素E对质量相关量q的平均因果效应可由式(17)近似表征。

$$\mathrm{ACE}=E_{U,R}\left[E\left(q\middle|E=e',U,R\right)-E\left(q\middle|E=e,U,R\right)\right] \tag{17}$$

其中，当U,R为连续变量时，式(17)中的求和相应替换为积分形式。

证明　由图2可知，用户能力U与区域特征R同时影响环境因素E与质量相关量q，因此从E到q存在后门路径。U与R同时影响E与q，因此存在后门路径$E\leftarrow U\rightarrow q$与$E\leftarrow R\rightarrow q$。调整集$Z_{\mathrm{adj}}=\{U,R\}$包含上述路径上的非碰撞节点，因此能够阻断$E$与$q$之间的所有后门路径。

进一步，根据假设1，给定$\{U,R\}$后不存在未观测混杂。根据建模设定，U与R表示相对稳定的先验属性，不受E影响，即因果图中不存在$E\rightarrow U$或$E\rightarrow R$的边，因此Z_{adj}不包含E的任何后代节点。

综上所述，调整集$Z_{\mathrm{adj}}=\{U,R\}$满足后门准则。根据Pearl的后门定理，在假设1~假设3成立的条件下，环境因素E对质量相关量q的平均因果效应是可识别的，并可由式(17)给出的后门调整式表示。证毕。

在实现中，本文采用线性回归模型估计环境因素的边际效应系数。

$$q_{\mathrm{data},t+1}=\beta_0+\beta_U f_{\mathrm{user},t+1}+\hat{\beta}_E f_{\mathrm{env},t+1}+\beta_R f_{\mathrm{region},t+1}+\varepsilon_{t+1} \tag{18}$$

若将环境二值化$E\in\{0,1\}$（如按分位数将环境划分为“低/高”）时，则平均处理效应定义为：

$$\mathrm{ATE}=E\left[q\middle|\mathrm{do}(E=1)\right]-E\left[q\middle|\mathrm{do}(E=0)\right] \tag{19}$$

在不含交互项的线性设定下，$\hat{\beta}_E$与ATE在模型假设下具有一致的效应解释。为量化估计不确定性，本文采用bootstrap重采样（1 000次）计算置信区间。

为解释分析“用户/环境/区域”对质量波动的相对贡献，并据此自适应确定调整强度，本文基于线性回归模型的方差贡献进行归因。

$$\mathrm{Contribution}_i=\frac{|\beta_i|\mathrm{Var}(f_i)}{\sum_{j\in\{U,E,R\}}|\beta_j|\mathrm{Var}(f_j)},i\in\{U,E,R\} \tag{20}$$

该分解一方面用于解释用户、环境和区域因素对质量波动的相对贡献，另一方面用于确定因果调整强度。

基于$\hat{\beta}_E$的估计结果，CE模块对从真值发现阶段的原始质量评分q_{raw}进行环境校正，得到校正后的质量评分为：

$$q_{\mathrm{adj},t+1}=q_{\mathrm{raw},t+1}-\lambda\hat{\beta}_E(f_{\mathrm{env},t+1}-\overline{f_{\mathrm{env}}}) \tag{21}$$

其中，$\overline{f_{\mathrm{env}}}$为滑动窗口内的环境均值，$\lambda\in[0,1]$为调整权重。为保持数值合理性，输出裁剪为：

$$q_{\mathrm{adj},t+1}\leftarrow\mathrm{clip}(q_{\mathrm{adj},t+1},[0,1]) \tag{22}$$

此外，为解释校正结果，本文进一步给出理想环境条件下的参考质量估计。对于给定数据点，估计其在理想环境条件$f_{\mathrm{env}}^{\mathrm{ideal}}$下的质量。

$$q_{\mathrm{cf}}=q_{\mathrm{obs}}+\mathrm{ATE}_{\mathrm{env}}(f_{\mathrm{env}}^{\mathrm{ideal}}-f_{\mathrm{env}}^{\mathrm{obs}}) \tag{23}$$

在无未观测混杂、调整集有效且线性设定近似合理等条件下，式(21)可降低f_{env}波动对质量评分的系统性影响，使质量评分更接近用户能力相关部分。

2.3.3　因果增强的真值发现与在线更新机制

CE-PRBTD在PRBTD迭代更新框架的基础上引入环境效应校正，并采用滑动窗口机制适应时变环境，流程如算法2所示。

算法 2 CE-PRBTD

输入 数据流 D，初始声誉 $\{R_u^{(0)}\}$，滑动窗口大小 W，更新周期 Δt，最小样本量 $M_{\min}$，调整权重 λ，时间步 T

输出 声誉 r，质量评分 $\{q_{\text{adj}}\}$

1) 初始化 $r \leftarrow r^{(0)}$，滑动窗口缓存 $W \leftarrow \varnothing$，环境效应估计 $\hat{\beta}_E \leftarrow 0$

2) 对每个时间步 t=1,2,…,T

3) 预测参照：用 MSTT 得到预测参照 $\hat{X}_{t+1}$

4) 误差代理：由观测 X_{t+1} 与 $\hat{X}_{t+1}$ 计算误差代理 $e_{\text{data},t+1}$（式(11)）

5) 质量代理：将 $e_{\text{data},t+1}$ 映射为可回归质量代理 $q_{\text{data},t+1} \in [0,1]$（式(15)）

6) 缓存更新：将 $(q_{\text{data},t+1}, f_{\text{env},t+1}, f_{\text{user},t+1}, f_{\text{region},t+1})$ 写入 W，仅保留最近 W 个样本

7) 因果效应更新：若 $t \bmod \Delta t = 0$ 且 $|W| \geqslant M_{\min}$，用式(18)在 W 上更新 $\hat{\beta}_E$

8) 原始质量评分：按 PRBTD 的质量评估机制（参考文献[9]）得到原始质量评分 $q_{\text{raw},t+1}$

9) 因果校正：按式(21)得到 $q_{\text{adj},t+1}$，并按式(22)裁剪到[0,1]

10) 回灌更新：用 $q_{\text{adj},t+1}$ 更新声誉 r，并以更新后的 r 执行真值发现得到本步估计真值，时间步结束

由此，CE-PRBTD 实现了由预测参照生成、误差代理构造、环境效应校正、质量评估、声誉更新和真值发现组成的在线迭代过程。其中，MSTT 模块用于提供稳定的预测参照，CE 模块用于削弱环境扰动导致的误归因，两者共同提升无真值场景下质量评估与声誉更新的稳健性。

为兼顾时变性与稳定性，环境效应采用滑动窗口周期更新。

$$\text{ATE}_{\text{env}}^{(t)} = \begin{cases} \text{Estimate}(F_{t-W:t}) \\ \hat{\beta}_E(t-1) \end{cases} \tag{24}$$

当 t 能被 Δt 整除，并且滑动窗口 $[t-W, t]$ 内的样本数量不少于阈值 $M_{\min}$ 时，触发环境效应更新。该机制能够在保持估计稳定性的同时适应环境关系的动态变化过程。

调整权重采用基于归因贡献的自适应策略，表示为：

$$\lambda = \frac{\text{Contribution}_{\text{env}}}{\text{Contribution}_{\text{env}} + \text{Contribution}_{\text{user}}} \tag{25}$$

为避免历史缓冲区缺失带来的不稳定性，本文采用一致性补全机制。

$$q_{\text{other}} = \begin{cases} Q[u_{\text{other}}], & u_{\text{other}} \in Q \\ R[u_{\text{other}}], & u_{\text{other}} \in R \setminus Q \\ 0.5, & \text{其他} \end{cases} \tag{26}$$

同时，对声誉与质量进行裁剪以避免数值不稳定。

$$R_{\text{safe}} = \text{clip}(R, [\epsilon, 1-\epsilon]), \epsilon = 10^{-6} \tag{27}$$

通过上述滑动窗口更新、自适应调整与一致性补全机制，CE-PRBTD 能够在动态环境下保持较好的在线运行稳定性。值得注意的是，在现实 MCS 场景中，工人–任务矩阵通常呈长尾稀疏特征，可能导致滑动窗口内样本不足，使 $\hat{\beta}_E$ 置信区间扩大。对此，上述机制从 3 个层面形成协同保护：线性回归替代传统分层加权（式(18)）避免稀疏层估计退化；$M_{\min}$ 阈值确保样本不足时沿用历史估计；自适应 λ（式(25)）在估计可靠性下降时自动收缩校正幅度。上述协同效果将在 3.5.6 节通过 $\hat{\beta}_E$ 置信区间宽度的定量分析加以验证。

2.4 理论分析

为评估 CE-DSTTD 框架在无真值 MCS 场景中的可部署性与理论稳定性，本节对 CE-PRBTD 在线闭环过程进行复杂度与性质分析。CE-PRBTD 在每个时间步都需要调用预测模块 MSTT 生成预测参照并构造误差代理，因此下述时间复杂度分析同时计入 MSTT 前向预测和 CE-PRBTD 的总体开销。

2.4.1 时间复杂度分析

定理 2 时间复杂度上界。设数据集包含 T 个时间步，每个时间步平均有 K 个用户，历史 buffer 大小为 B_h，空间位置数为 N，特征维度为 d，滑动窗口内用于因果效应估计的样本量为 M，回归特征维度为 p，环境效应每 Δt 个时间步更新一次（Δt 为常数），则 CE-PRBTD 在 CE-DSTTD 框架内的端到端时间复杂度上界为：

$$O\left(TL(N^2d + Nd^2) + TKB_h c_{\text{imp}} + \frac{T}{\Delta} c_{\text{reg}}(M,p)\right) \tag{28}$$

其中，$TL(N^2d + Nd^2)$ 对应 MSTT 的多层 Transformer 编码器前向传播，$TKB_h c_{\text{imp}}$ 对应 PRBTD 中 implication/buffer 的逐用户历史匹配计算，

$\frac{T}{\Delta}c_{\mathrm{reg}}(M,p)$对应滑动窗口内环境效应的回归估计。

证明　对 CE-PRBTD 的三部分计算开销逐项分析如下。

（1）预测参照生成开销。在单层 Transformer 编码器中，多头自注意力计算注意力权重的时间复杂度为$O(N^2d)$，前馈网络对各空间位置执行线性变换的时间复杂度为$O(Nd^2)$。因此，L层编码器在单个时间步的时间复杂度为$O\left(L(N^2d+Nd^2)\right)$，累计$T$步的时间复杂度为$O\left(TL(N^2d+Nd^2)\right)$。

（2）真值发现相关迭代开销。每个时间步需对当前K条用户数据与各用户的B_h条历史记录进行匹配或相似度计算，单步代价为$O(KB_hc_{\mathrm{imp}})$，累计时间复杂度为$O\left(TKB_hc_{\mathrm{imp}}\right)$。

（3）因果效应估计开销。每Δt个时间步触发一次窗口回归估计，单步代价为$c_{\mathrm{reg}}(M,p)$，累计时间复杂度为$O\left(\frac{T}{\Delta}c_{\mathrm{reg}}(M,p)\right)$。

上述三部分计算在每个时间步中串行执行，合并各项即可得到定理所述的总体时间复杂度上界。证毕。

在典型参数L=4、d=128、N=1 024设置下，$O(TL\cdot(N^2d+Nd^2))$为主导开销。CE 附加项$c_{\mathrm{reg}}(M,p)=O\left(Mp^2+p^3\right)$在实验参数$\Delta t$=5、$p$=4 条件下（每 5 步触发一次，回归特征维度为4），单步均摊代价远小于 Transformer 主导项，不改变总体时间复杂度量级。为进一步说明 CE-DSTTD 相对现有方法的理论效率，表2给出了各代表性方法的时间复杂度对比。其中，TruthFinder 时间复杂度根据文献[49]中的表 3 转换得出（$|S||V|$在 MCS 场景下对应K^2N）；CATD 原文未给出显式表达式，根据文献[19]中的算法 1 推导得出；经典方法内部迭代轮数已合并嵌入O符号。

表2　CE-DSTTD 与代表性方法的时间复杂度对比

方法	时间复杂度
TruthFinder	$O\left(TK^2N\right)$
CATD	$O\left(TK^2N\right)$
PRBTD	$O\left(TL(N^2d+Nd^2)+TKB_hc_{\mathrm{imp}}\right)$
CE-DSTTD	$O\left(TL(N^2d+Nd^2)+TKB_hc_{\mathrm{imp}}+\frac{T}{\Delta}c_{\mathrm{reg}}(M,p)\right)$

2.4.2　收敛性分析

在完成时间复杂度分析后，进一步考察 CE-PRBTD 中声誉更新过程的稳定性。由于校正后的质量评分最终通过声誉更新作用于真值发现结果，因此声誉序列是否稳定收敛，直接关系到整个闭环机制的长期可用性。

定理3　声誉更新序列的收敛性。设用户u在第t轮的质量评分$q_u(t)\in[\epsilon,1-\epsilon]$（其中$\epsilon>0$），其声誉更新式可写为：

$$R_u(t)=0.5+\phi(q_u(t),R_u(t-1)) \tag{29}$$

其中，映射$\phi(q,r)$关于变量r满足一致 Lipschitz 条件，即对任意$q\in[\epsilon,1-\epsilon]$及$r_1,r_2$，有$|\phi(q,r_1)-\phi(q,r_2)|\leqslant\rho|r_1-r_2|$，$0<\rho<1$。

在本文参数设置下，可取$\rho=\max\{\alpha(1-k_1),\beta(1-k_2)\}<1$，则声誉更新序列$\{R_u(t)\}$满足：

$$\left|R_u(t+1)-R_u(t)\right|\leqslant\rho\left|R_u(t)-R_u(t-1)\right| \tag{30}$$

因此，声誉更新序列$\{R_u(t)\}$收敛，当$q_u(t)$收敛到稳态值时，$\{R_u(t)\}$收敛到相应更新映射的唯一不动点。

证明　由式(29)可得：

$$R_u(t+1)=0.5+\phi(q_u(t+1),R_u(t)) \tag{31}$$

当$q_u(t)$稳定后，由$\phi(q,\cdot)$的 Lipschitz 连续性可得：

$$\begin{aligned}&|R_u(t+1)-R_u(t)|=\\&\left|\phi(q_u(t),R_u(t))-\phi(q_u(t),R_u(t-1))\right|\leqslant\\&\rho|R_u(t)-R_u(t-1)|\end{aligned} \tag{32}$$

递推可得：

$$\left|R_u(t+1)-R_u(t)\right|\leqslant\rho^t\left|R_u(1)-R_u(0)\right| \tag{33}$$

由于$0<\rho<1$，故$\{|R_u(t+1)-R_u(t)|\}$几何收敛于0。进一步，对任意$m>n$，有：

$$\begin{aligned}&|R_u(m)-R_u(n)|\leqslant\sum_{k=n}^{m-1}|R_u(k+1)-R_u(k)|\leqslant\\&|R_u(1)-R_u(0)|\sum_{k=n}^{m-1}\rho^k\end{aligned} \tag{34}$$

由几何级数收敛性可知，$\{R_u(t)\}$为 Cauchy 序列，从而收敛。

对任意固定$q_u\in[\varepsilon,1-\varepsilon]$，映射$F(R)=0.5+\phi(q_u,R)$为压缩映射。根据 Banach 不动点定理，$F$在定义域内存在唯一不动点。因此，当$q_u(t)$收敛

于稳态值时，$\{R_u(t)\}$收敛于该唯一不动点。证毕。

2.4.3 鲁棒性与公平性分析

为避免对强假设的依赖，本文给出与上述建模一致的保守性质说明。若在局部窗口内q_{raw}可写为：

$$q_{\text{raw}} = \alpha + \gamma f_{\text{env}} + s\left(f_{\text{user}}, f_{\text{region}}\right) + \eta \tag{35}$$

且$\hat{\beta}_E \approx \gamma$，则校正后评分$q_{\text{adj}}$中与$f_{\text{env}}$线性相关的系统性项被削弱，其随环境变化的敏感度由$\gamma$降为$(1-\lambda)\gamma$。因此在固定$(f_{\text{user}}, f_{\text{region}})$下，环境波动导致的评分方差满足近似上界：

$$\operatorname{Var}\left(q_{\text{adj}} \mid f_{\text{user}}, f_{\text{region}}\right) \lesssim (1-\lambda)^2 \operatorname{Var}\left(q_{\text{raw}} \mid f_{\text{user}}, f_{\text{region}}\right) \tag{36}$$

进一步，若两名用户在能力与区域特征上相同$(f_{\text{user}}^A = f_{\text{user}}^B, f_{\text{region}}^A = f_{\text{region}}^B)$，仅环境不同，则校正后期望差异满足近似关系：

$$\left|\mathrm{E}\left[q_{\text{adj}}^A\right] - \mathrm{E}\left[q_{\text{adj}}^B\right]\right| \approx (1-\lambda)\left|\hat{\beta}_E\right|\left|f_{\text{env}}^A - f_{\text{env}}^B\right| \tag{37}$$

这表明其可通过λ调节环境差异引起的不公平性上界。

综上所述，本节从3个方面说明了CE-PRBTD的理论性质：在计入MSTT预测参照开销的情况下，端到端时间复杂度仍处于可控范围；声誉更新序列在Lipschitz压缩条件下具有良好的收敛性；通过环境校正项可对评分方差和用户间由环境差异引起的不公平性进行显式约束。上述分析为后续实验中关于效率、稳定性、鲁棒性和公平性的结果提供了理论依据。

3 实验结果与分析

本文设计了多项实验，以验证CE-DSTTD在预测精度、用户区分能力、机制可解释性和系统公平性等方面的有效性。实验主要从预测参照有效性、真值发现效果、机制验证、鲁棒性和参数敏感性等方面评估本文方法。具体的实验环境和超参数设置如表3所示。

表3 实验环境

软/硬件配置	版本/超参数
CPU	Intel Xeon Platinum 8481C
GPU	NVIDIA RTX 4090D
内存	80 GB
操作系统	Ubuntu 22.04
Python	3.10
PyTorch	2.1.0
CUDA	12.1

3.1 数据集说明

为保证真值发现设定一致，本文统一采用“用户–任务–观测”定义：用户为具有固定质量类型的虚拟参与者，任务为某一时间片上的待测对象，观测为多个用户对同一任务的独立上报。基于此，本文在TaxiBJ与UCI PM2.5数据集上构建MCS仿真场景。

3.1.1 TaxiBJ数据集的MCS仿真场景构建

TaxiBJ数据集包含城市网格流量序列，不具备原生的多用户上报结构。本文将“网格–时间片”对定义为感知任务，以真实流量值为任务真值，MSTT预测输出为参考值，每个时间片抽取100名虚拟参与者生成多源独立上报。用户类型全程固定，正常、低质量、恶意用户比例为70%、20%、10%，分别叠加小幅高斯噪声、大幅高斯噪声和随机固定偏移，并统一叠加时段相关环境扰动以构造混淆。用户能力以历史声誉为代理，环境质量指数依据时段属性以规则映射方式构造：早晚高峰时段（7:00—9:00及17:00—19:00）对应较差的感知环境，取值为0，其余时段取值为1，并叠加基于区域标识的小幅扰动项以反映局部空间差异。该方案基于北京交通流量的已知周期性规律，以确定性规则替代参数估计，避免了在有限样本下进行加权合成的不稳定性。真实流量仅用于离线评估，不参与在线推断过程。

3.1.2 UCI PM2.5数据集的MCS仿真场景构建

UCI PM2.5数据集包含单站点逐小时PM2.5浓度及同期气象记录，不具备二维网格结构。本文将单时间片PM2.5浓度定义为感知任务，为100名虚拟参与者生成独立观测。由于不具备二维网格结构，以Persistence预测和滑动平均替代MSTT作为参考预测，从而隔离预测模型差异对CE模块评估的影响。用户噪声设置与TaxiBJ数据集一致，并额外引入偏移型与随机型攻击及环境放大系数以增强混淆。环境质量指数由风速（I_{ws}）、降雨量（I_r）、温度偏离、气压偏离及露点偏离5类气象变量构造：各变量首先进行Z-score标准化以消除量纲差异，随后依据领域先验知识赋予固定权重进行加权求和，构造综合不利气象强度指标Z_{adv}，5类变量权重依次为0.35、0.35、0.15、0.10和0.05，其中风速与降雨量因对PM2.5传感精度影响最为显著而被赋予较高权重。最终通过Sigmoid变换映射为环境质量指数并裁剪至区间(0.1, 0.9)，即$E = \text{clip}(\sigma(-Z_{\text{adv}}), 0.1, 0.9)$。为评估

变量间共线性的潜在影响，本文计算了5类气象变量的Pearson相关系数矩阵，结果表明各变量间相关性均处于较低水平，共线性对合成指数的干扰在可接受范围内。

综上所述，两个数据集均在任务层构造了多用户独立上报结构：TaxiBJ侧重于验证空间异质条件下的预测辅助真值发现能力；UCI PM2.5侧重于验证真实环境混淆条件下去混淆校正的有效性，二者共同支撑本文方法的实验评估。

3.2　评价指标

3.2.1　时空预测指标

本文实验指标包括均方根误差（root mean square error, RMSE）、平均绝对误差（mean absolute error, MAE）和决定系数R^2，用于评估模型的预测性能和效果。设测试集共有N个样本，真实值为y_i，预测值为$\hat{y}_i$，均值为：

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{38}$$

均方根误差为：

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{39}$$

平均绝对误差为：

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{40}$$

决定系数为：

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \tag{41}$$

3.2.2　恶意用户识别指标(F1)

设参与用户集合为U。在合成/可标注实验中，定义低质量（或恶意注入）用户集合为$U^- \subset U$（正类），其余为$U^+ = U \setminus U^-$（负类）。算法输出每个用户的声誉$r_u \in [0,1]$。采用阈值τ判定低质量用户集合，即：

$$\hat{U}^-(\tau) = \{u \in U \mid r_u < \tau\} \tag{42}$$

定义：

$$\mathrm{TP} = \left|\hat{U}^- \cap U^-\right|, \mathrm{FP} = \left|\hat{U}^- \setminus U^-\right|, \mathrm{FN} = \left|U^- \setminus \hat{U}^-\right| \tag{43}$$

则有：

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}},\quad \mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}},\quad \mathrm{F1} = \frac{2\mathrm{Precision}\,\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \tag{44}$$

阈值τ在验证集上通过网格搜索选取，使F1最大。

3.2.3　声誉区分度指标

为衡量声誉对好/坏用户的区分程度，定义声誉区分度（reputation distance, RepDist）为两类用户声誉均值之差。

$$\mathrm{RepDist} = \frac{1}{|U^+|}\sum_{u \in U^+} r_u - \frac{1}{|U^-|}\sum_{u \in U^-} r_u \tag{45}$$

该值越大表示声誉对低质量用户的压制与对高质量用户的提升越明显，区分效果越好。

3.2.4　降噪效果

在无真值场景中，算法依赖误差代理e_{data}（式(11)）进行质量评估与声誉更新；在离线评估场景中，可进一步利用真实误差$e_{\mathrm{data}}^{\mathrm{true}}$（式(12)）对质量增强带来的去噪收益进行量化。降噪（noise reduction, NRed）仅在可获得真值的离线数据上计算。

设评估集中共有M条观测，其去噪前的平均真实误差为：

$$\bar{e}_{\mathrm{before}}^{\mathrm{true}} = \frac{1}{M}\sum_{m=1}^{M} e_{\mathrm{data},m}^{\mathrm{true}} \tag{46}$$

算法输出校正后的质量评分$q_{\mathrm{adj}} \in [0,1]$，并据此筛选出高质量数据观测集合$S$。去噪后的平均真实误差定义为：

$$\bar{e}_{\mathrm{after}}^{\mathrm{true}} = \frac{1}{|S|}\sum_{m \in S} e_{\mathrm{data},m}^{\mathrm{true}} \tag{47}$$

据此，降噪效果定义为去噪后平均真实误差相对去噪前的下降率。

$$\mathrm{NRed} = 1 - \frac{\bar{e}_{\mathrm{after}}^{\mathrm{true}}}{\bar{e}_{\mathrm{before}}^{\mathrm{true}}} \tag{48}$$

该指标越大，表示质量增强带来的真实误差下降越明显，即降噪效果越好。

3.3　MSTT模型的性能评估

本文从收敛性、预测精度、跨域泛化和稳定性等方面评估MSTT模型的性能。预测模块的改进能够降低参考值误差，从而为后续因果校正与声誉更新提供更可靠的输入。

3.3.1　模型收敛性分析

图3给出了MSTT与基线PRBTD-CFSTTN的训练与验证损失曲线。PRBTD-CFSTTN在epoch 15~20出现明显振荡，表明其在复杂时空依赖学习中存在优化不稳定性。MSTT引入Pre-LayerNorm

与随机深度机制后，训练损失由初始约 0.15 单调下降至 0.02 以下，验证损失稳定收敛于 0.015 附近，训练与验证损失全程同步下降且差距极小，表明模型在收敛过程中未出现过拟合，泛化性能良好。结果表明，多尺度融合与位置编码机制有助于提升模型训练的稳定性和收敛性。

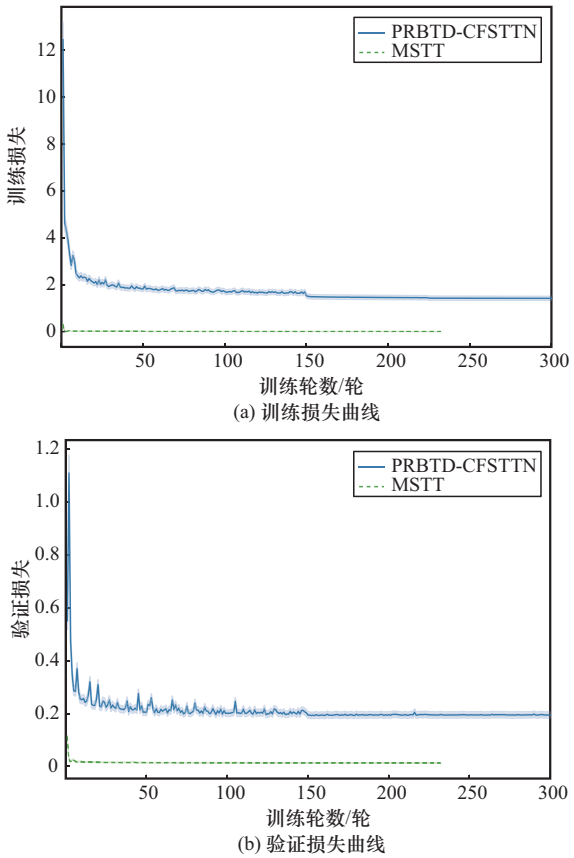


图3　训练与验证损失曲线

为进一步评估 MSTT 在测试集上的最终预测性能，本文选取验证损失最优检查点于测试集上评估。MSTT 取得 RMSE=14.74、MAE=8.99，预测偏差较小；R^2=0.981 8、方差解释度（Variance Score=0.981 9），表明模型对目标变量方差具有较强解释能力。综合收敛曲线与测试指标可知，MSTT 在训练稳定性与预测精度两方面均表现出色，为后续因果校正模块的集成提供了可靠的预测基准。

3.3.2　各模型对比

为对比不同预测模型在 TaxiBJ 数据集上的性能差异，本文以 RMSE 为评价指标，对多种代表性方法进行测试，结果如图 4 所示。本文对比的预测方法包括：以历史同时段时空信息加权均值作为预测值的统计基线方法 HA；通过自回归与差分滑动平均对时序动态进行线性建模的经典统计方法 ARIMA[50]；捕捉多元时间序列间线性相互依赖关系的向量自回归模型 VAR；采用卷积神经网络建模空间依赖、循环神经网络捕捉时序依赖的早期深度时空模型 DeepST[31]；基于残差卷积网络分别对时间接近性、周期性与趋势性进行建模，并融合外部信息动态聚合多分支输出的 ST-ResNet 方法[32]；PRBTD 采用的基于时空 Transformer 的预测主干网络 CFSTTN[9]。

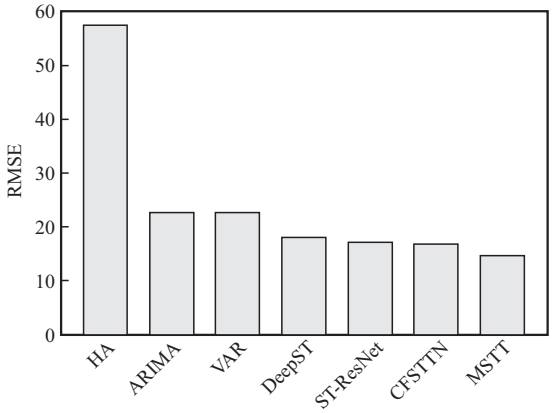


图4　各模型在 TaxiBJ 数据集上的 RMSE 指标对比

由图 4 可知，HA 的 RMSE 最高，仅依赖历史均值难以捕捉复杂时空动态；ARIMA 与 VAR 的 RMSE 约为 23，较 HA 明显改善，但受限于线性假设，仍无法有效建模非线性时空依赖。深度学习方法表现更优，DeepST 与 ST-ResNet 的 RMSE 降至 17~19，对时空相关性的捕捉能力更强。PRBTD 采用的 CFSTTN 在此基础上进一步提升，RMSE 为 16.89，而本文提出的 MSTT 取得最优结果，RMSE 为 14.74，较 CFSTTN 降低 12.7%，较 ST-ResNet 等次优模型降低约 15%，验证了多尺度特征融合与注意力机制的有效性。

3.3.3　跨域迁移学习效果

图 5 展示了各模型在不同训练–测试域组合下的性能表现，用于评估跨域迁移设置下模型性能的变化。具体而言，图 5 中通过 R^2、RMSE 和 MAE 这 3 个评价指标，分析模型在不同训练–测试域组合下的性能表现。

从图 5(a)可以看出，当训练和测试数据来自同一域时（即图 5(a)中的对角线元素），模型的表现最优，R^2值普遍超过 0.98，尤其在 BJ15 测试集上，

R^2达到了最高值0.987 7，表明模型在同域数据上的预测精度较高。然而，在跨域测试场景下，尽管模型性能有所下降，R^2值仍维持在较高水平，分布在0.970 0~0.985 1，表明模型在域间迁移场景下仍具有较好的泛化能力。

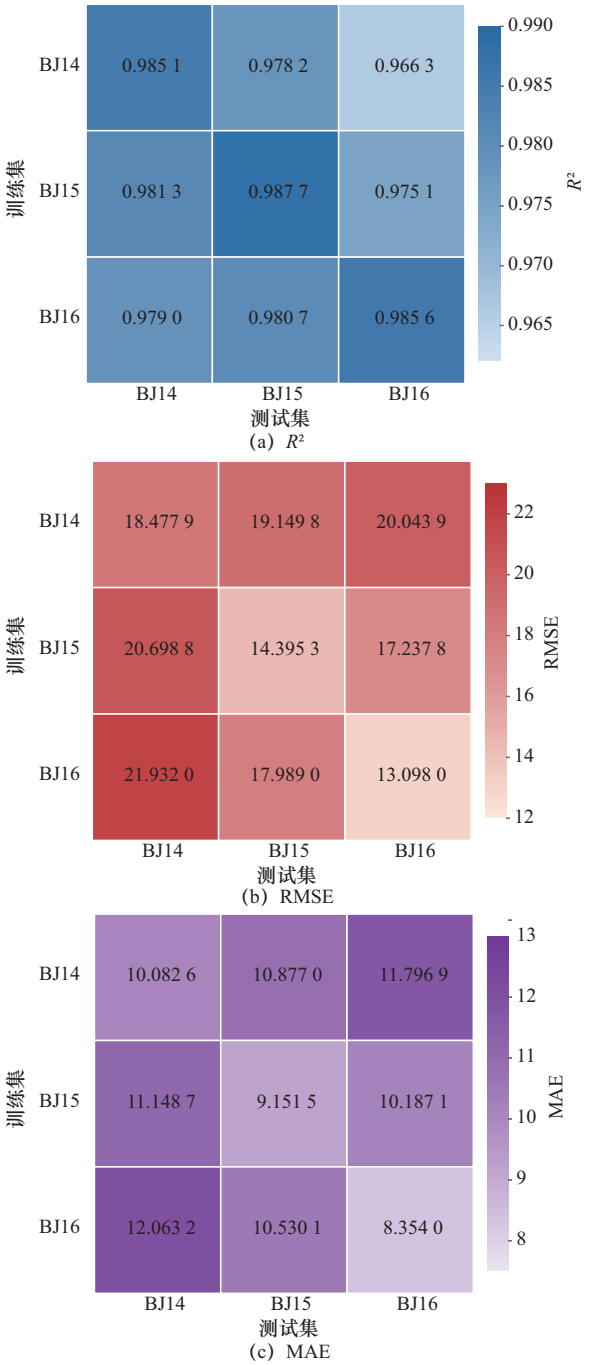


图5　不同训练–测试域组合下的性能表现

图5(b)和图5(c)进一步印证了这一趋势，表明域内测试的误差显著低于跨域测试。例如，BJ15域内测试的RMSE为14.40，而在跨域测试中，BJ14和BJ16的RMSE分别上升至20.70和17.24。结果表明，尽管存在域偏移，本文模型仍具备较强的泛化能力，能够有效应对跨域迁移问题。总体来看，模型在不同测试域下仍能保持较好的预测精度与稳定性。

表4进一步汇总了模型在不同评测设置及训练年份/数据集条件下的R^2表现。通过比较各类评测设置和不同训练年份下的结果，可以更全面地考察模型在多种环境中的稳定性与鲁棒性。这些结果也为后续分析模型的适应性提供了重要依据，有助于从整体上评估模型在不同条件下的性能表现。

表4　不同实验设置下的R^2结果汇总

实验维度	具体分组	R^2值
评测设置	前向泛化	0.973 2
	反向泛化	0.980 3
	域内测试	0.986 1
训练年份/数据集	BJ14	0.972 3
	BJ15	0.978 2
	BJ16	0.979 8

从表4中可以看出，模型在不同分组中的R^2值均维持在较高水平，表明本文模型具有较强的拟合能力和稳定的解释性能。具体而言，域内测试的R^2值为0.986 1，前向与反向泛化的R^2值分别为0.973 2和0.980 3，说明在分布发生变化时，模型性能有所回落，但整体保持良好。此外，反向泛化优于前向泛化，显示了模型对不同时间方向的分布漂移的适应性差异。

图6进一步量化分析了不同跨年泛化场景相对于域内测试的R^2性能下降幅度。横轴表示“训练年份→测试年份”的迁移方向，纵轴为R^2性能下降百分比。通过图6可以观察到跨年泛化导致的R^2性能下降呈现明显的不对称性和时间跨度依赖性。

具体而言，相邻年份间的跨年泛化性能下降幅度相对较小：BJ16→BJ15的下降幅度约为0.7%，BJ15→BJ14约为0.6%，表明模型在短时间跨度下具有良好的迁移能力。然而，当训练和测试数据的时间跨度增大时，性能下降幅度显著增加：BJ15→BJ16的下降幅度达到约1.0%，而BJ14→BJ16的跨年泛化导致了最大的性能衰减，下降幅度高达约1.9%。

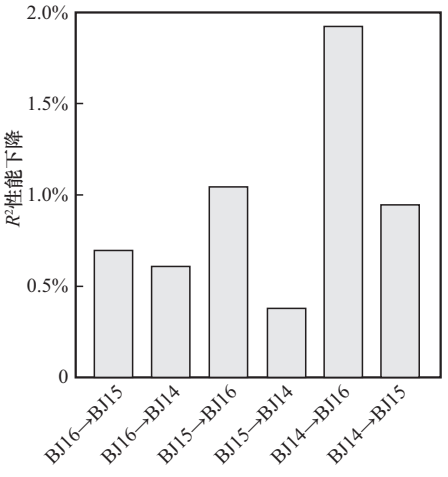


图6　跨年泛化相对域内性能的下降幅度

值得注意的是，反向泛化（从较新年份向较旧年份迁移）整体上表现出比前向泛化更小的性能下降。这一现象可能源于交通基础设施和出行模式随时间演变，导致未来数据包含的模式更加复杂多样，因而向过去泛化时，模型已学习到的特征仍然适用。

总体而言，即使在最困难的BJ14→BJ16跨年泛化场景下，性能下降幅度也仅为1.9%，这表明MSTT模型具有较好的时间域泛化能力和鲁棒性，能够有效应对交通流量预测中常见的时间分布偏移问题。

3.4　因果推断实验分析与结果

3.4.1　*与主流方法对比*

本文将CE-PRBTD与7种主流方法进行对比，分别是预测辅助类（PRBTD和CNB）、经典真值发现类（TD、CATD和CRH），以及近期真值发现类（DTD和HGEM），分别用于评估因果增强模块的独立增益、预测辅助机制的整体贡献及与最新方法的竞争力。所有主流方法在验证集上独立调优，结果如图7所示。其中，DTD通过联合建模数据可信度与激励竞价机制提升MCS场景下的聚合精度[51]；HGEM利用超图结构对MCS稀疏数据中的参与者协作关系进行建模，在EM框架下改进可靠性评估[52]。

由图7可知，CE-PRBTD在F1-Score和RepDist上均取得最优结果，其中F1-Score仅略优于PRBTD和HGEM，而RepDist提升更为明显。HGEM基于超图拓扑的EM迭代在F1-Score上表现尚可，但其可靠性估计以局部邻居平均为主，缺乏对环境混淆误差的显式建模，导致RepDist偏低。DTD通过数据可信度迭代更新在RepDist上优于HGEM，但在复杂时空场景下的识别边界仍不及预测参照类方法。经典真值发现类方法（TD、CATD和CRH）在两项指标上均表现较弱，表明仅依赖统计一致性难以有效区分复杂时空场景中的用户质量。

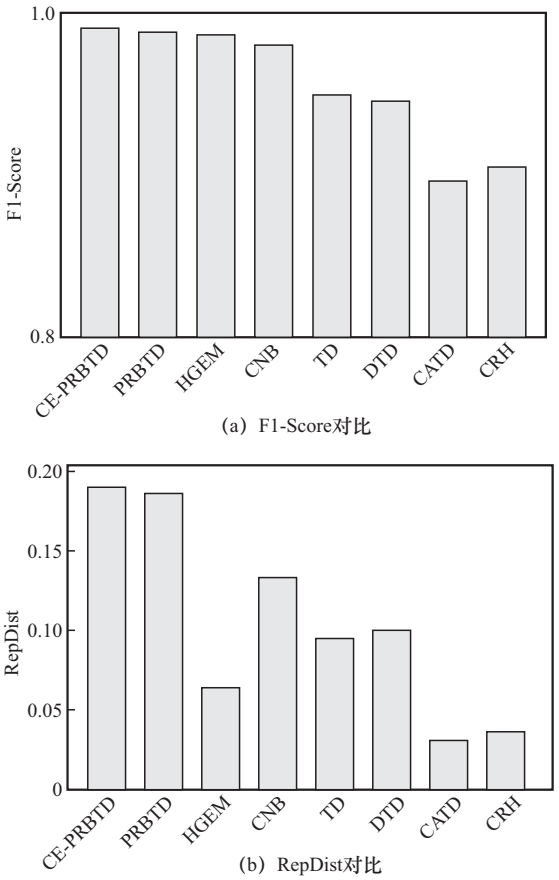


图7　各主流方法对比

两项指标上的非对称提升与CE模块的设计定位一致：F1-Score作为二分类准确率，其上限受限于用户噪声分布的重叠程度；RepDist作为正常与恶意用户的声誉均值差距，对环境校正更为敏感。CE模块通过后门调整剥离环境扰动对质量评分的系统性影响，使正常用户在恶劣环境下的声誉不被错误压低，从而拉大声誉分布间距。但当恶意用户的上报偏移量与低质量用户在数值上存在重叠时，仅凭环境校正难以在二分类边界上产生显著突破，因此F1-Score的提升相对有限。CE模块在不同混淆强度下的具体收益边界将在3.5.5节的受控仿真中进一步验证。

为进一步验证上述结论在真实环境混淆条件下

的有效性，本文在UCI PM2.5数据集上开展跨数据集实验，结果如表5所示。

表5　UCI PM2.5数据集上CE-PRBTD与PRBTD的对比结果

预测模式	方法	F1-Score	RepDist	NRed
Persistence	PRBTD	0.988 9	0.863 2	0.549 2
Persistence	CE-PRBTD	0.995 4	0.869 5	0.539 4
Moving-Avg	PRBTD	0.977 8	0.779 0	0.534 7
Moving-Avg	CE-PRBTD	0.992 3	0.785 3	0.524 2

在两种预测模式下，CE-PRBTD在F1-Score与RepDist指标上均优于PRBTD，说明去混淆校正对用户识别与声誉区分具有一致的正向作用。与此同时，NRed指标略有下降，反映出该方法在降噪效果与归因合理性之间存在一定权衡：部分由环境因素引起但不可归责于用户的高误差样本被保留，从而提高了用户层评估的公平性，但相应降低了纯粹的误差压缩幅度。

综合TaxiBJ与UCI PM2.5实验结果可以看出，CE模块通过环境校正减轻误差归因，提升声誉区分能力与评估稳健性，该机制在不同数据集与预测参照下保持一致。其收益在以下两类场景中最为显著：环境扰动与用户能力差异存在系统性混淆且不同质量用户的误差分布因环境产生偏移，以及评估目标侧重声誉排序的公平性与稳定性而非单纯的二分类精度。反之，当环境扰动较弱或用户噪声差异足够显著时，PRBTD基线能提供较好的声誉区分能力，CE模块的边际收益相应有限。因此，CE-PRBTD尤其适用于存在显著环境异质性的真实MCS部署场景，如气象条件剧烈变化的室外传感网络或交通高峰期的城市感知系统。

3.4.2　泛化性评估与分布偏移分析

为评估所提方法在分布偏移条件下的泛化能力，本文在不同训练–测试域组合下构建9组迁移实验，并从用户识别、声誉区分能力及降噪效果3个维度进行分析。实验结果如表6与图8所示。

表6　迁移场景下的平均性能对比

指标	相同年份平均值	跨年份平均值
F1-Score	0.966 7	0.961 2
RepDist	0.238 6	0.229 0
NRed	0.087 6	0.105 2

由表6可知，3项指标在跨域条件下呈现出不同的变化规律。F1-Score由0.966 7小幅降至0.961 2，降幅有限，说明模型在用户识别层面具有较强鲁棒性；RepDist由0.238 6降至0.229 0，降幅相对更明显，反映出数据分布偏移会削弱声誉学习对用户质量差异的刻画能力，这与跨域场景下环境扰动对误差归因干扰加剧相一致；NRed则呈现反向趋势，由0.087 6升至0.105 2，表明跨域场景中环境异质性增强，CE模块的去混淆校正筛除了更多受环境干扰的低质量样本，降噪收益反而有所扩大。

综合来看，本文方法在分布偏移条件下整体保持良好的泛化能力：用户识别性能基本稳定，声誉区分能力略有下降但下降幅度有限，而去混淆校正机制在跨域环境异质性更强的场景中体现出更大的校正价值。上述结果表明，CE模块的核心收益在于改善误差归因和提升声誉结构质量，其作用在环境扰动更显著的跨域场景中尤为突出。

3.4.3　效率分析

为保证不同方法之间比较的一致性，本文仅统计质量评估与真值发现阶段的运行开销，不含

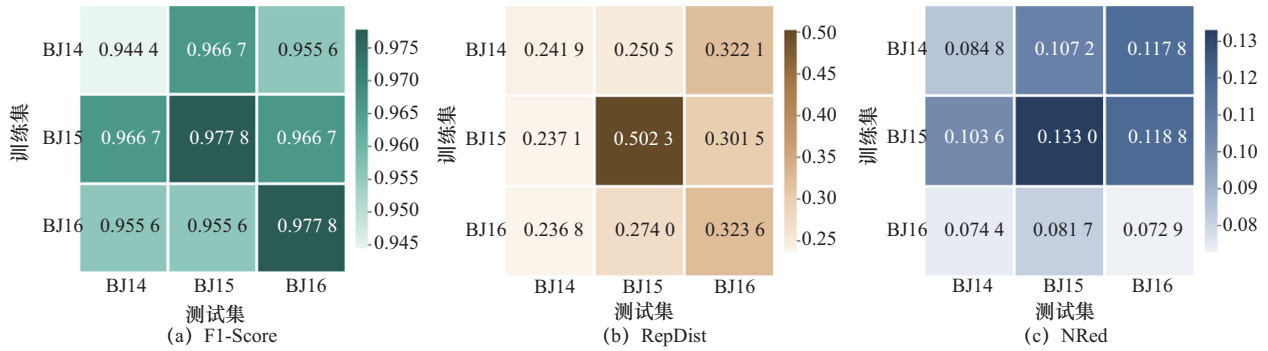


图8　模型在跨域场景下的多维度性能

MSTT的GPU推断开销（0.125 s/步），实验结果如表7所示。

表7　各方法质量评估与真值发现阶段效率对比

方法	TD阶段耗时/s
CE-PRBTD	247.4
PRBTD	178.3
CNB	0.06
WEI	0.07
TD	0.16

由表7可知，CNB、WEI、TD等统计类基线方法仅涉及轻量矩阵运算，耗时均低于0.2 s；CE-PRBTD与PRBTD因引入时空预测参照与逐用户历史匹配机制，算法复杂度本质上高于统计类基线方法，两类方法功能层次不同，耗时差距主要源于算法设计。CE-PRBTD相较PRBTD额外开销增幅约38.8%，涵盖因果效应估计、滑动窗口回归及质量去混淆校正，与CE模块带来的性能增益相比处于合理水平。当前实现未引入区域级并行加速，效率优化将作为后续工作。

为进一步验证CE-DSTTD在大规模任务场景中的实际可行性，本文固定T=119、N=32，将用户数K从100增至200，结果如表8所示。随着K从100增至200，CE-DSTTD耗时从247.4 s增至879.2 s，增幅约3.6倍，与用户规模扩大带来的计算开销增长趋势一致。各规模重复实验标准差均低于2 s，结果稳定可信。在K=200的大规模设定下，系统单次评估CE-DSTTD耗时约879.2 s，满足离线MCS质量评估的时效要求。

表8　CE-DSTTD在不同用户规模下的运行时间

任务规模	用户数K	时间步T	CE-DSTTD耗时/s
小规模	100	119	247.4
中规模	150	119	496.8
大规模	200	119	879.2

3.5　消融实验

3.5.1　预测模块消融实验

为评估MSTT各关键设计对预测性能的贡献，本文开展消融实验。图9给出了不同变体在RMSE与MAE两项指标上的对比结果。消融实验设置包括：完整模型MSTT、无多尺度融合（移除多尺度特征融合模块，以单尺度卷积替代）、无位置编码（移除可学习位置编码）、MSE+MAE损失（以MSE与MAE的组合作为训练目标），以及“仅MSE损失。

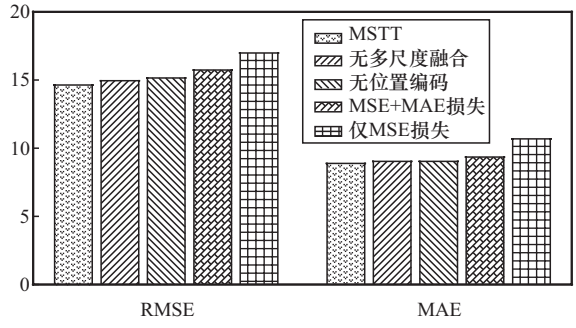


图9　各模块消融实验性能对比

图9表明，完整模型MSTT在RMSE和MAE上均取得最优结果。移除多尺度特征融合或可学习位置编码后性能小幅下降，说明两者均有助于时空依赖与空间异质性建模；采用仅MSE损失或MSE+MAE损失时性能退化更明显，说明多目标损失对误差控制更为关键。总体上，结构与损失设计在模型中发挥互补作用。

3.5.2　因果增强模块消融实验

为验证因果增强模块的独立贡献，本文从模块有效性、估计器选择与环境扰动鲁棒性3个维度开展消融实验。表9汇总了CE-PRBTD在3种估计器配置下与PRBTD基线的核心指标对比结果。

表9　去混淆机制中不同环境效应估计器的敏感性分析

方法配置	F1-Score	RepDist	NRed
PRBTD	0.988 9	0.186 5	0.457
CE_Linear	0.990	0.189 3	0.631
CE_IPW	0.990	0.189 7	0.648
CE_DR	0.991	0.190 1	0.667

由表9可知，引入CE模块后，各配置在F1-Score、RepDist和NRed这3项指标上均优于PRBTD基线：F1-Score由0.988 9升至0.990~0.991，NRed由0.457升至0.631~0.667，说明性能提升主要来自去混淆校正。3种估计器差异较小，综合复杂度、稳定性与理论一致性，后续默认采用CE_Linear。

此外，本文以CE-DSTTD为完整模型，并构

建无因果增强、无Transformer和基础预测模型3种变体，进一步分析各组成部分的贡献，结果如图10所示。

在预测性能方面，完整模型CE-DSTTD的RMSE与MAE最优；去除因果增强模块后误差略有上升；无Transformer时性能退化更为显著；基础预测模型性能退化幅度最大。各配置下R^2均维持较高水平，表明消融操作主要影响误差控制能力，对整体拟合质量影响有限。在鲁棒性方面，完整模型CE-DSTTD在F1-Score和NRed指标上表现最优，因果增强模块对分类判别与抗噪能力贡献显著；各配置的类别间分数方差差异较小，说明不同方法在类别一致性上总体相当。

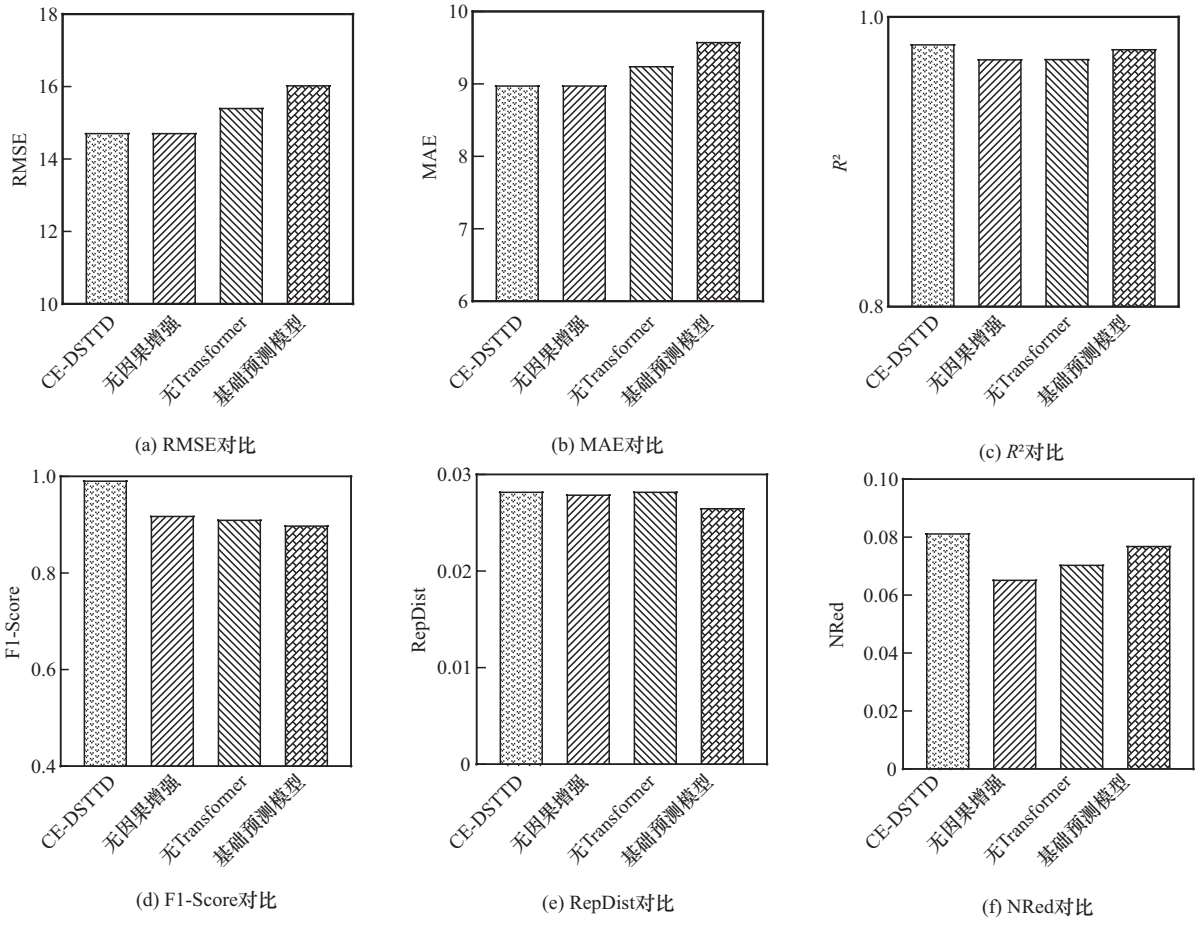
3.5.3　因果调整权重λ敏感性分析

为评估式(21)和式(25)中因果调整权重λ对系统性能的影响，本文在6组独立实验上对$\lambda\in\{0, 0.1,\cdots,1.0\}$进行敏感性实验，结果如图11所示。

图11表明，λ在识别精度、声誉区分与降噪效果间存在权衡：F1-Score在$\lambda\in[0,0.2]$时均高于0.99，至$\lambda\approx0.7$累计下降约2.2%，且置信区间明显拓宽；RepDist呈倒U形，在$\lambda\in[0.3,0.5]$时达峰；NRed先降后升，于$\lambda\approx0.3$最低且高λ区波动增大。综合来看，$\lambda\in[0.1,0.2]$时折中最优，与式(32)一致，即适度校正可在将评分方差压缩至原始的$(1-\lambda)^2$倍的同时保留区分信号。

为检验上述结论的跨数据集一致性，本文在UCI PM2.5数据集上复现相同实验设置，仍取$\lambda\in\{0,0.1,\cdots,1.0\}$，评价指标与TaxiBJ数据集保持一致。考虑到该实验主要用于验证关键区间而非刻画完整趋势，结果如表10所示，并据此为后续统一参数设置提供定量依据。

由表10可知，F1-Score由λ=0时的0.988 9提升至λ=0.2时的0.989 9，随后随着λ继续增大而明显下降；RepDist整体波动较小，但在λ=0.2时达到0.871 1；NRed则由0.549 2降至λ=0.3时的0.539 1，之后小幅回升。该结果进一步表明，λ取0.1~0.2时能够在识别精度、声誉区分与降噪效果之间取得更稳定的平衡。考虑到参数设置一致性，本文后续实验统一取λ=0.1。



(a) RMSE对比　(b) MAE对比　(c) R^2对比

(d) F1-Score对比　(e) RepDist对比　(f) NRed对比

图10　完整消融实验

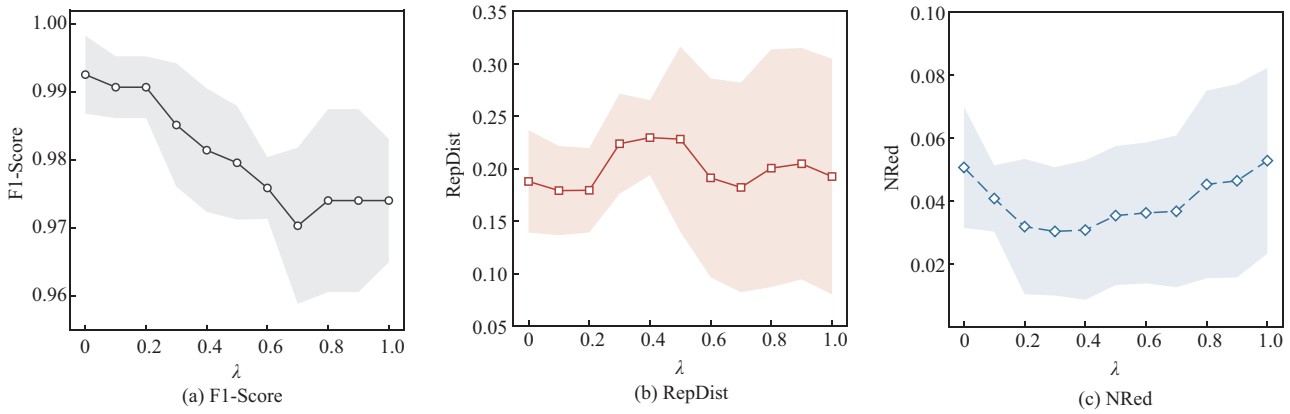


图 11　权重 λ 敏感性分析

表 10　UCI PM2.5 数据集 λ 敏感性结果

λ	F1-Score	RepDist	NRed
0	0.988 9	0.863 3	0.549 2
0.1	0.989 2	0.867 7	0.543 4
0.2	0.989 9	0.871 1	0.540 4
0.3	0.961 2	0.869 2	0.539 1
0.5	0.964 5	0.865 3	0.536 4
0.7	0.959 0	0.866 7	0.540 7
1.0	0.952 3	0.872 9	0.548 6

3.5.4　恶意用户注入下的鲁棒性实验

为进一步评估其在强对抗场景下的鲁棒性，本文在 BJ16 数据集上模拟 6 类恶意攻击：随机噪声攻击（random_noise）、系统性偏差攻击（systematic_bias）、策略性攻击（strategic）、合谋攻击（collusive）、时间依赖攻击（time_dependent）和区域定向攻击（region_targeted）。恶意用户比例设置为 5%、10%、15%、20% 和 25%，每种配置独立重复 3 次，共 90 组实验。表 11 给出了不同攻击类型下的平均检测结果。

表 11　不同攻击类型下 CE-PRBTD 的检测性能

攻击类型	F1-Score 均值	F1-Score 标准差	RepDist
systematic_bias	0.929	0.030	0.059
collusive	0.914	0.027	0.052
random_noise	0.907	0.026	0.043
strategic	0.882	0.053	0.013
time_dependent	0.864	0.060	0
region_targeted	0.850	0.071	−0.011

由表 11 可知，CE-PRBTD 在 6 类攻击下的 F1-Score 均值为 0.891±0.054，且不同攻击类型间差异显著。总体上，systematic_bias 最易检测（F1-Score =0.929），region_targeted 最难检测（F1-Score =0.850）。进一步分析表明，strategic 场景下 F1-Score 随着恶意用户比例下降最快，而 random_noise 和 systematic_bias 下降较缓。即使在 25% 的高恶意用户比例下，systematic_bias 和 collusive 场景中的 F1-Score 仍高于 0.88。RepDist 结果与此一致：systematic_bias 和 collusive 具有较高区分度，而 time_dependent 和 region_targeted 接近 0 甚至为负，说明全局因果校正对局部时空攻击的刻画仍有限。

为在真实气象混淆场景下进一步验证上述鲁棒性结论，本文在 UCI PM2.5 数据集上开展不同恶意用户比例下的鲁棒性实验，并与 PRBTD、CNB、TD 和 WEI 基线方法进行对比，结果如图 12 所示。

图 12 表明，随着恶意用户比例由 5% 增至 25%，各方法的 F1-Score 和 RepDist 整体下降，NRed 与 F1 下降幅度整体上升，并在 15% 附近出现明显拐点。CE-PRBTD 在所有恶意用户比例下均保持最高 F1-Score 和 RepDist：F1-Score 由 0.990 4 降至 0.947，RepDist 由 0.887 降至 0.451；对应的 F1 下降幅度仅为 4.38%，显著低于 PRBTD 的 8.81% 和 WEI 的 12.63%。相比之下，PRBTD 在 NRed 上略优于 CE-PRBTD（25% 时分别为 0.638 和 0.617），说明因果增强的主要收益不在于进一步放大降噪效果，而在于缓解误归因并保持更稳健的声誉可分性。

3.5.5　因果假设违背下的鲁棒性

假设 1 是 CE 模块后门调整的理论前提。为验证其在不同程度违背时的稳健性，本节引入隐变量 $Z_t \sim N(0,1)$，并通过混淆强度 ξ 对 CE 模块的可观

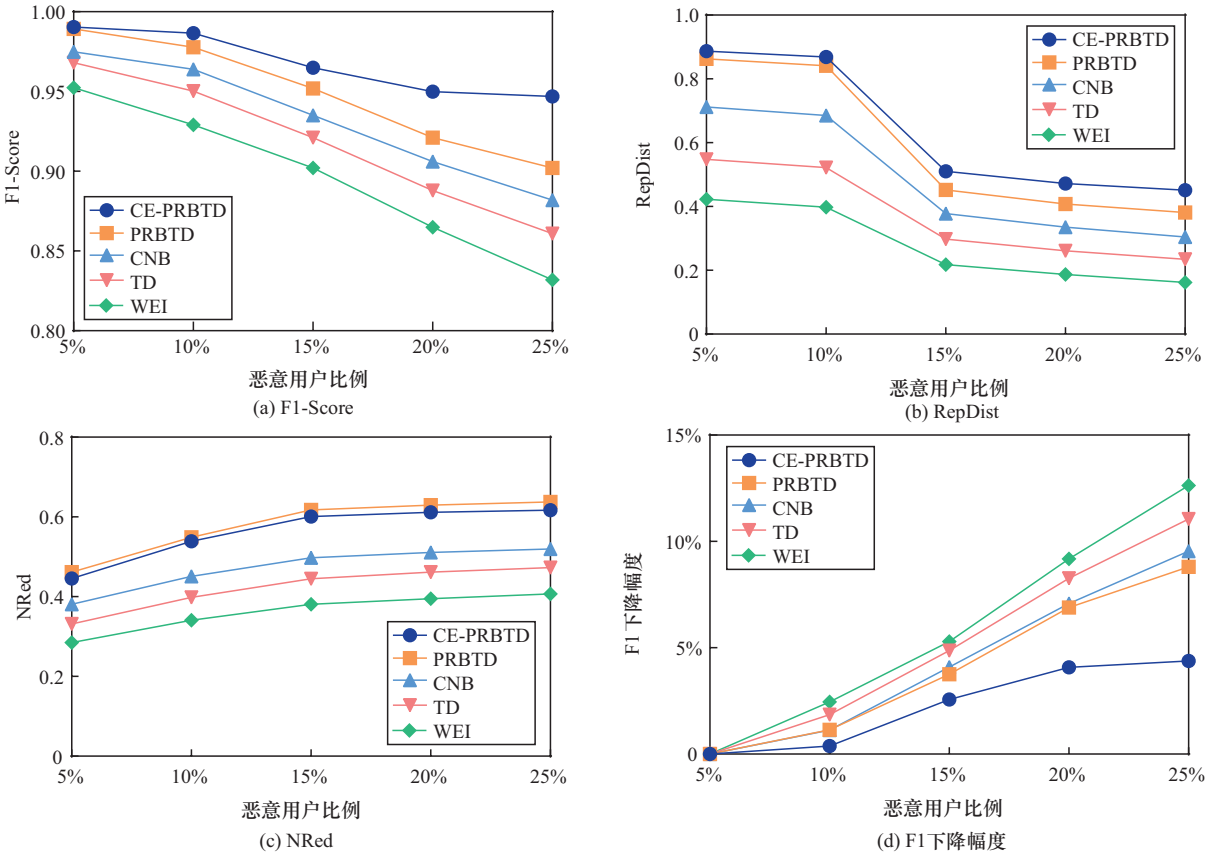


图 12　在 UCI PM2.5 数据集上的鲁棒性验证

测环境变量施加系统性扰动，同时按比例放大用户感知噪声。当$\xi=0$时，假设 1 完全成立；随着ξ增大，未观测混杂程度逐步加剧。由于真实数据集无法精确控制未观测混杂强度，本节采用基于 SCM 构造的合成数据集进行受控实验，并以 NRed 作为评价指标，结果如表 12 所示。

表 12　　因果假设违背敏感性分析

混淆强度ξ	CE-PRBTD	PRBTD
0	0.667	0.457
0.1	0.645	0.419
0.2	0.665	0.390
0.3	0.613	0.374
0.5	0.586	0.366
0.8	0.543	0.351

随着ξ增大，两种方法的噪声降低率均下降，PRBTD 由 0.457 降至 0.351，CE-PRBTD 由 0.667 降至 0.543。CE-PRBTD 在各混淆强度下始终优于 PRBTD，且两者之间的性能差距在中等混淆强度（$\xi\leqslant0.5$）下保持相对稳定。结果表明，本文引入的因果校正机制对中等程度的因果假设违背具有一定的鲁棒性，间接支持了代理变量选取策略的合理性：以历史声誉与区域特征作为后门调整集，能够有效覆盖 MCS 场景中的主要混杂路径，使校正效果在中等残存混杂条件下仍保持显著优势。当未观测混杂极强时，两种方法性能均出现明显下滑，提示在强混杂场景下引入设备型号标识、用户移动模式特征等更多可观测辅助变量，或采用代理变量因果学习方法，是未来值得探索的方向。

3.5.6　数据稀疏度敏感性分析

为验证 CE-PRBTD 在数据稀疏场景下的鲁棒性，本文设计了观测保留率ρ变化实验。具体地，对每个时间片的用户上报数据以概率$\rho\in\{1.0,0.7,0.5,0.3,0.2,0.1\}$进行随机保留，模拟从稠密到极端稀疏的连续退化过程。每组配置在 6 个实验场景中各独立重复 5 次，共计 30 次实验取均值。实验结果如表 13 所示，其中$\hat{\beta}_E$ CI 宽度为基于 Bootstrap 重采样（1 000 次）计算的 95% 置信区间宽度均值，反

映因果效应系数估计的不确定性程度。

表 13 不同观测保留率下的 F1 性能对比

观测保留率ρ	样本数	CE-PRBTD	PRBTD	$\hat{\beta}_E$ CI 宽度
1.0	3 000	0.977 8	0.966 7	0.011 6
0.7	2 048	0.954 1	0.951 1	0.018 1
0.5	1 471	0.939 0	0.928 9	0.023 0
0.3	844	0.908 9	0.913 3	0.050 6
0.2	555	0.899 7	0.902 0	0.129 4
0.1	249	0.891 8	0.884 5	0.102 1

随着ρ由 1.0 降至 0.1，CE-PRBTD 的 F1 整体仅下降 8.8%。$\hat{\beta}_E$ CI 宽度在$\rho\geq0.5$时不超过 0.023，ρ进一步降低时阶段性扩大至 0.129 4，ρ=0.1 时收窄至 0.102 1，与$M_{\min}$降级触发频率上升一致。在中等稀疏场景（ρ=0.2~0.3）下，受$\hat{\beta}_E$估计可靠性下降影响，CE-PRBTD 与 PRBTD 差异收窄至 0.005 以内。在稠密端（$\rho\geq0.5$）与极端稀疏端（ρ=0.1），CE-PRBTD 均保持明显优势，F1 降幅与$\hat{\beta}_E$CI 宽度增幅的显著不对称表明$M_{\min}$阈值与自适应λ有效隔离了高方差$\hat{\beta}_E$估计向质量评分的传导。上述结果表明，CE-PRBTD 在长尾稀疏的真实 MCS 场景下能够有效保障因果去混淆机制的估计可靠性。

4 结束语

本文针对移动群智感知无真值场景中的预测误差传播与环境混淆问题，提出了 CE-DSTTD 框架：以 MSTT 提升预测参照质量，以结构因果模型与后门调整实现去混淆校正，从而提升质量评估的稳健性、公平性与可解释性。实验结果表明，CE-DSTTD 在低质量用户识别、声誉区分与降噪效果上均优于对比方法，CE 模块在环境异质性显著的场景中收益尤为突出，稀疏场景下的因果估计稳定性与跨域泛化能力也得到了验证。未来将针对强未观测混杂与跨域场景，进一步探索更鲁棒的因果识别与泛化方法。

参考文献：

[1] 蒋伟进, 吕斯健, 陈晓红. 群智感知中面向移动群体的参与者选择优化模型[J]. 控制理论与应用, 2022, 39(2): 343-351.
Jiang W J, Lyu S J, Chen X H. Participant selection optimization model for mobile groups in crowdsensing[J]. Control Theory & Applications, 2022, 39(2): 343-351.

[2] Dong X W, Zhang W, Zhang Y S, et al. Optimizing task location privacy in mobile crowdsensing systems[J]. IEEE Transactions on Industrial Informatics, 2022, 18(4): 2762-2772.

[3] 蒋伟进, 杨璇, 李碧霞. 基于可解释贡献异常检测与动态裁剪的联邦学习投毒攻击防御方法[J]. 计算机学报, 2025, 48(12): 2855-2874.
Jiang W J, Yang X, Li B X. Federated learning poisoning attack defense method based on interpretable contribution anomaly detection and dynamic pruning[J]. Chinese Journal of Computers, 2025, 48(12): 2855-2874.

[4] Capponi A, Fiandrino C, Kantarci B, et al. A survey on mobile crowdsensing systems: challenges, solutions, and opportunities[J]. IEEE Communications Surveys & Tutorials, 2019, 21(3): 2419-2465.

[5] Restuccia F, Ghosh N, Bhattacharjee S, et al. Quality of information in mobile crowdsensing: survey and research challenges[J]. ACM Transactions on Sensor Networks, 2017, 13(4): 1-43.

[6] Liu Y T, Kong L H, Chen G H. Data-oriented mobile crowdsensing: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2019, 21(3): 2849-2885.

[7] 蒋伟进, 陈艺琳, 韩裕清, 等. K-Modes 聚类数据收集和发布过程中的混洗差分隐私保护方法[J]. 通信学报, 2024, 45(1): 201-213.
Jiang W J, Chen Y L, Han Y Q, et al. Shuffled differential privacy protection method for K-Modes clustering data collection and publication[J]. Journal on Communications, 2024, 45(1): 201-213.

[8] 朱伊波, 方贤进, 张朋飞, 等. 本地差分隐私下面向离群点的真值发现算法研究[J]. 电子学报, 2025, 53(5): 1541-1558.
Zhu Y B, Fang X J, Zhang P F, et al. A study of truth discovery algorithms for forward outliers under local differential privacy[J]. Acta Electronica Sinica, 2025, 53(5): 1541-1558.

[9] Li J J, Gu B, Gong S M, et al. Can we enhance the quality of mobile crowdsensing data without ground truth?[J]. IEEE Transactions on Mobile Computing, 2025, 24(5): 4451-4465.

[10] 蒋伟进, 吴玉庭, 陈艺琳, 等. 基于 Stackelberg 博弈的社交移动群体任务分配控制策略[J]. 控制理论与应用, 2025.
Jiang W J, Wu Y T, Chen Y L, et al. Task allocation control strategy of social mobile groups based on Stackelberg game[J]. Control Theory & Applications, 2025.

[11] Sahraoui Y, Kerrache C A, Amadeo M, et al. A cooperative crowdsensing system based on flying and ground vehicles to control respiratory viral disease outbreaks[J]. Ad Hoc Networks, 2022, 124: 102699.

[12] Wang Y, Gao H, Ngai E C H, et al. A coverage-aware high-quality sensing data collection method in mobile crowd sensing[J]. IEEE Transactions on Mobile Computing, 2025, 24(4): 3025-3040.

[13] Xie X, Bai T, Guo W W, et al. Cooperative computing for mobile crowdsensing: design and optimization[J]. IEEE Transactions on Mobile Computing, 2024, 23(5): 6437-6454.

[14] 田有亮, 田茂清, 高鸿峰, 等. 面向群智感知应用的基于协作的位置认证方案[J]. 通信学报, 2022, 43(9): 121-133.
Tian Y L, Tian M Q, Gao H F, et al. Cooperation-based location authentication scheme for crowdsensing applications[J]. Journal on Communications, 2022, 43(9): 121-133.

[15] Li Y L, Li Q, Gao J, et al. Conflicts to harmony: a framework for resolving conflicts in heterogeneous data by truth discovery[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(8): 1986-1999.

[16] Meng C S, Jiang W J, Li Y L, et al. Truth discovery on crowd sensing of correlated entities[C]//Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems. New York: ACM Press, 2015: 169-182.

[17] Yin X X, Han J W, Yu P S. Truth discovery with multiple conflicting information providers on the web[J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(6): 796-808.

[18] Xiao H P, Wang S Y. A joint maximum likelihood estimation framework for truth discovery: a unified perspective[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(6): 5521-5533.

[19] Li Q, Li Y L, Gao J, et al. A confidence-aware approach for truth discovery on long-tail data[J]. Proceedings of the VLDB Endowment, 2014, 8(4): 425-436.

[20] Miao C L, Jiang W J, Su L, et al. Cloud-enabled privacy-preserving truth discovery in crowd sensing systems[C]//Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems. New York: ACM Press, 2015: 183-196.

[21] Miao C L, Su L, Jiang W J, et al. A lightweight privacy-preserving truth discovery framework for mobile crowd sensing systems[C]//Proceedings of the IEEE INFOCOM 2017-IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2017: 1-9.

[22] Wang T C, Xu N, Zhang Q, et al. A lightweight privacy-preserving truth discovery in mobile crowdsensing systems[J]. Journal of Information Security and Applications, 2024, 83: 103792.

[23] Wu L J, Xie W K, Tan W, et al. RDPP-TD: reputation and data privacy-preserving based truth discovery scheme in mobile crowdsensing[PP]. V1. (2025-05-07)[2026-04-15]. arXiv: arXiv.2505.04361.

[24] Jia R J, Ma J, You Z Y, et al. Transparent and privacy-preserving mobile crowd-sensing system with truth discovery[J]. Sensors, 2025, 25(7): 2294.

[25] Wang L Y, Fan G H, Han X. Federated truth discovery for mobile crowdsensing with privacy-preserving trustworthiness assessment[J]. IEEE Data Engineering Bulletin, 2023, 46(1): 124-144.

[26] Liu S Z, Zheng Z Z, Wu F, et al. Context-aware data quality estimation in mobile crowdsensing[C]//Proceedings of the IEEE INFOCOM 2017-IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2017: 1-9.

[27] Xiao Z, Lim H B, Ponnambalam L. Participatory sensing for smart cities: a case study on transport trip quality measurement[J]. IEEE Transactions on Industrial Informatics, 2017, 13(2): 759-770.

[28] Han K, Huang H, Luo J. Quality-aware pricing for mobile crowdsensing[J]. IEEE/ACM Transactions on Networking, 2018, 26(4): 1728-1741.

[29] Jin H M, Su L, Chen D Y, et al. Thanos: incentive mechanism with quality awareness for mobile crowd sensing[J]. IEEE Transactions on Mobile Computing, 2019, 18(8): 1951-1964.

[30] Gao G J, Xiao M J, Wu J, et al. Truthful incentive mechanism for nondeterministic crowdsensing with vehicles[J]. IEEE Transactions on Mobile Computing, 2018, 17(12): 2982-2997.

[31] Zhang J B, Zheng Y, Qi D K, et al. DNN-based prediction model for spatio-temporal data[C]//Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM Press, 2016: 1-4.

[32] Zhang J B, Zheng Y, Qi D K. Deep spatio-temporal residual networks for citywide crowd flows prediction[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 31(1): 1655-1661.

[33] Lin Z Q, Feng J, Lu Z Y, et al. DeepSTN+: context-aware spatial-temporal neural network for crowd flow prediction in metropolis[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 1020-1027.

[34] Peng Z Z, Huang X H. Spatial-temporal transformer network with self-supervised learning for traffic flow prediction[C]//Proceedings of the Spatio-Temporal Representation Learning Workshop at NeurIPS 2022. New York: ACM Press, 2022: 1-8.

[35] Feng A S, Tassiulas L. Adaptive graph spatial-temporal transformer network for traffic forecasting[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. New York: ACM Press, 2022: 3933-3937.

[36] Li Z T, Hu Z, Han P, et al. SSL-STMFormer self-supervised learning spatio-temporal entanglement transformer for traffic flow prediction[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(11): 12130-12138.

[37] Qiu Z C, Wu H, Teng G Q, et al. Multi-scale spatial-temporal transformer for traffic flow prediction[J]. Scientific Reports, 2026, 16: 3531.

[38] He H Y, Long Z T, Zhang Y C, et al. Spatio-temporal transformer traffic prediction network based on multi-level causal attention[J]. PLoS One, 2025, 20(9): e0331139.

[39] Yin K Y, Niu X Z, Zhu J H, et al. DAT-STAN: dual-module adaptive transformer and spatio-temporal attention network for large-scale traffic flow prediction[J]. Tsinghua Science and Technology, 2025, doi: 10.26599/TST.2025.9010062.

[40] Li Z H, Xia L H, Tang J B, et al. UrbanGPT: spatio-temporal large language models[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2024: 5351-5362.

[41] Cheng H T, Gong Z B, Wang C. LLM-TFP: integrating large language models with spatio-temporal features for urban traffic flow prediction[J]. Applied Soft Computing, 2025, 177: 113174.

[42] Pearl J. Causality[M]. Cambridge University Press, 2009.

[43] Rosenbaum P R, Rubin D B. The central role of the propensity score in observational studies for causal effects[J]. Biometrika, 1983, 70(1): 41-55.

[44] Hernán M A, Robins J M. Causal inference[M]. Boca Raton, FL: CRC, 2010.

[45] Spirtes P, Glymour C, Scheines R. Causation, prediction, and search[M]. The MIT Press, 2001.

[46] Peters. Elements of causal inference: foundations and learning algorithms[M]. The MIT Press, 2017.

[47] Deng Z Z, Zheng X L, Tian H, et al. Deep causal learning: representation, discovery and inference[J]. ACM Computing Surveys, 2026, 58(2): 1-36.

[48] Andreou M, Chen N, Bollt E. Assimilative causal inference[J]. Nature Communications, 2026, 17: 1854.

[49] Waguih D A, Berti-Equille L. Truth discovery algorithms: an experimental evaluation[PP]. V1. (2014-09-23) [2026-04-15]. arXiv: arXiv.1409.6428.

[50] Williams B M, Hoel L A. Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results[J]. Journal of Transportation Engineering, 2003, 129(6): 664-672.

[51] Kang Y C, Liu A F, Xiong N N, et al. DTD: an intelligent data and bid dual truth discovery scheme for MCS in IIoT[J]. IEEE Internet of Things Journal, 2024, 11(2): 2507-2519.
[52] Wang P F, Jiao D, Yang L Y, et al. Hypergraph-based truth discovery for sparse data in mobile crowdsensing[J]. ACM Transactions on Sensor Networks, 2024, 20(3): 1-23.

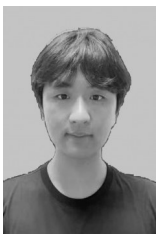
[作者简介]



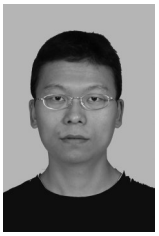
蒋伟进（1964– ），男，博士，湖南工商大学教授、博士生导师，主要研究方向为信息安全、网络安全、群智感知。



石顺究（2004– ），女，湖南工商大学硕士生，主要研究方向为移动群智感知、数据质量保障。



欧阳宇明（2003– ），男，湖南工商大学硕士生，主要研究方向为移动群智感知、异质智能体。



许宇胜（1976– ），男，北方导航控制技术股份有限公司高级工程师，主要研究方向为结构设计及仿真分析、群智感知、系统控制。



刘溪溪（2002– ），女，湖南工商大学硕士生，主要研究方向为移动群智感知、联邦学习。



崔新雨（2002– ），女，湖南工商大学硕士生，主要研究方向为移动群智感知、联邦学习。



刘志华（1999– ），男，湖南工商大学硕士生，主要研究方向为群智感知、任务分配。