

基于移动众包感知的群体时空行为分析与社群网络结构挖掘

杨磊¹ 赵翔¹
YANG Lei ZHAO Xiang

摘要

针对校园环境下 GPS 与 Wi-Fi 双重稀疏、室内多径漂移及高频采集能耗高等挑战, 文章提出了一种高鲁棒性的群体感知与分析方法。构建移动众包感知体系, 设计隐式触发采集系统, 并提出基于信号强度增强的反距离加权 (SE-IDW) 算法。该算法基于对数距离路径损耗模型将 RSSI 映射为能量权重, 结合 DBSCAN 空间聚类与时间分割, 突破传统方法对数据密度的依赖。对 9 595 万条有效轨迹数据的分析表明: 在模拟 90% 数据丢失及连续缺失下, SE-IDW 定位准确率超 71.5%; 能量集中度提升 18.4%, 边界距离误差减少 525.37 m; 识别五类校园群体, 时空共现网络揭示行为与社群强耦合, 解决了稀疏轨迹语义提取难题, 证实物理共现与社群同质性映射, 支撑智慧校园精细化管理。

关键词

移动众包感知; SE-IDW 算法; 稀疏鲁棒性; 时空行为分析; 社群网络挖掘

doi: 10.3969/j.issn.1672-9528.2026.05.030

0 引言

随着物联网终端的普及与移动通信技术的深度融合, 高校校园内人、机、物的连接密度呈指数级增长。这种数据环境的根本性变化正推动智慧校园建设从“基础设施的数字化”向“管理服务的智能化”转型^[1]。在这一转型进程中, 核心挑战在于如何从海量、异构的底层传感信号中提取高层语义价值。校园用户的移动轨迹作为物理空间与数字空间的连接纽带, 其时空分布特征不仅反映了群体的宏观流动规律, 更蕴含了某种特定的生活范式与潜在的社群网络结构。因此, 基于大规模移动感知数据进行精细化的行为分析与社群网络结构挖掘, 对于优化校园资源配置、提升应急响应效率具有重要的工程与社会学价值。

近年来, 移动众包感知 (mobile crowdsensing, MCS) 作为一种新兴的泛在感知范式^[2], 为城市与校园级的数据采集提供了全新的解决方案。与传统依赖固定传感器网络或专有设备的采集方式不同, 移动众包感知利用普通用户携带的智能移动终端作为感知节点, 通过任务分配与群智协同, 实现室内外多模态信号 (如 GPS 与 Wi-Fi 指纹) 的互补与融合。然而, 在实际的大规模校园场景部署中, 移动众包感知面临着严峻的数据质量挑战。一方面, 出于对智能手机电池续航、计算资源的保护, 用户终端往往无法长时间维持 1Hz 或更高频率的理想采样状态, 自适应降频与后台休眠机制导致收集到的轨迹数据呈现出极高的时空稀疏性; 另一方面, 在建筑密集的大学校园内, GPS 信号在室内易发生严重的衰减与多

径漂移, 而 Wi-Fi 接入日志同样面临着采样粒度粗糙、漫游延迟以及连接跳变等问题。这种由于硬件能耗限制与复杂物理环境共同导致的数据稀疏与高噪现象, 成为制约轨迹语义挖掘的瓶颈。

在高度稀疏与噪声干扰的条件下, 现有的轨迹挖掘与停留点检测算法极易失效。传统的基于时空规则或几何中心 (Geometric Center) 的聚类算法^[3], 其核心假设在于轨迹点在停留区域内呈高密度、均匀分布。一旦数据出现随机稀疏或突发性连续缺失, 且伴随 GPS 向室外的系统性多径漂移, 几何中心将严重偏离用户的真实物理驻留位置, 进而引发严重的语义位置映射误差。此外, 目前绝大多数关于轨迹数据挖掘的研究仍停留在对个体移动路径的统计描述层面^[4], 缺乏将物理空间中的高频时空共现升华为社群网络结构的系统性感知与语义映射机制。

为了解决上述校园复杂环境中的多模态数据稀疏、定位漂移以及社群深层关联挖掘不足等问题, 本文提出了一种具有高稀疏鲁棒性的移动众包感知框架与用户行为语义挖掘方法。该方法采用了基于 Android 平台开发的移动端采集系统, 并创新性地引入了基于信号强度增强的反距离加权停留点检测算法 (signal-energy augmented inverse distance weighting, SE-IDW)。该算法深度融合了无线信号空间传播的对数距离路径损耗先验模型, 将底层的接收信号强度指示 (RSSI) 映射为线性能量物理权重, 从而克服了传统几何中心法对数据绝对密度的依赖, 在模拟随机稀疏与真实突发性缺失的极端条件下, 依然能够精准锚定室内外语义位置。基于该框架, 本文利用海量真实轨迹数据建立了多维度的轨迹质量评价指

1. 四川大学电子信息学院 四川成都 610065

标体系（如能量集中度与边界距离），验证了算法在修正系统性空间漂移方面的有效性。进一步研究从高保真停留点中提取多维特征，自适应聚类识别出了五类典型校园用户群体，并构建了全域时空共现网络，验证了行为范式与物理社交社群之间的强耦合机制，将传统的轨迹挖掘提升至社群网络感知的维度。

1 数据获取与预处理

1.1 基于移动众包感知的数据采集

本研究采用移动众包感知模式，即通过普通用户日常携带的智能手机作为分布式感知节点，在真实校园环境自然、持续地采集时空数据。该模式能够以较低成本获取大规模、多源异构的轨迹数据，反映用户群体的真实活动规律。

具体实现上，使用了课题组基于 Android 平台开发的数据采集 APP^[5]。APP 以后台服务方式运行，利用 Wi-Fi Manager 和 LocationManager API 进行异步数据采集。为平衡数据粒度与设备能耗，系统采用事件驱动与定时轮询结合的策略，目标采样频率设为 1 Hz（即每秒触发一次采集请求）采集 Wi-Fi 指纹（含 SSID、BSSID、RSSI）并同步记录 GPS 经纬度与时间戳。数据经本地轻量级压缩后，通过 HTTP POST 协议加密上传至云端服务器。

为验证所提方法的有效性，本研究基于移动众包感知模式开展了实地数据采集。实验招募了 21 名学生志愿者，在四川大学江安校区进行了为期 30 天的持续采集，累计获取原始轨迹数据 1.45 亿条。志愿者在日常学习、生活场景中自然携带手机，数据采集过程不干扰用户正常行为，充分体现了众包感知的自然性、规模性和低成本优势。

在实际的数据采集过程中，Android 系统的后台资源管理、设备自动休眠机制以及传感器本身的响应延迟，使得数据采集很难稳定保持在理论上的 1 Hz 采样频率。数据采集往往表现为不均匀的间歇状态：有时密集，有时因系统限制而中断。此外，校园内复杂的建筑环境造成的信号衰减和多径效应，也给数据带来了噪声和漂移。这种数据的稀疏性、分布不均及噪声干扰，构成了本文 SE-IDW 算法主要解决的技术难点。

1.2 数据清洗与坐标转换

采集到的原始数据会有噪声干扰且坐标系不统一的问题，需经过以下预处理流程：首先是地理围栏过滤，划定校园边界多边形，去掉漂移到校外的异常 GPS 点；其次是坐标系转换，把 WGS-84 坐标转成国内常用的 GCJ-02 坐标系（火星坐标系），可以和高德地图底图的建筑轮廓对齐，方便后续语义映射；最后是信号质量控制，筛掉 RSSI 低于 -90 dBm 的弱信号，降低噪声影响。

2 基于信号强度的反距离加权停留点检测算法

停留点指用户在某地持续停留超过设定时间的空间聚集簇，将杂乱物理轨迹转化为高层语义事件（像上课、自习）的关键抽象单元。面对传统算法在稀疏数据下容易失真的情况，本研究设计了 SE-IDW 算法。

2.1 基于时空约束的轨迹分段与聚类

挖掘轨迹数据时，仅靠空间密度聚类没法区分用户在同一地点的多次独立访问（比如上午上课和下午复习）。所以，本研究采用了引入时间连续性分割的分层聚类思路。

第一步进行空间维度的密度聚类。采用 DBSCAN（density-based spatial clustering of applications with noise）算法对单用户的全天轨迹进行处理^[6]。该算法通过设定的邻域半径 ϵ 与最小包含点数 MinPts，评估每个轨迹点的局部空间密度。能够被连通的高密度核心点集被归并为初始空间簇，这种非参数化特征很适合校园内部不规则的建筑轮廓。

第二步进行时间连续性分割。针对生成的每一个空间簇，将其内部点集按时间戳严格升序排列。算法遍历相邻点对，计算其时间间隔 $\Delta t^{[7]}$ 。如果两个连续采集点之间的时间差超过所设定的时间间隔 Δt_{break} （本研究设为 15 min），则判定用户在此期间发生了物理空间的脱离或是一次行为的终结。据此，单一空间簇被切分为多个时间连续的独立访问子序列，确保每一次语义事件的时序独立性。

第三步加上停留时长语义约束。对进行分割后的独立子序列，计算其首尾时间差。仅当该时间差大于设定的停留时间 $\Delta t_{\text{stay}} = 10 \text{ min}$ 时，该序列才被保留为候选停留点集。

2.2 SE-IDW 加权质心计算模型

在得到候选停留点集后，传统的几何中心法通常采用简单的算术平均计算停留点的质心坐标。但是当轨迹流遇到终端低频降采样或信号盲区导致突然缺失时，轨迹点的空间分布会剧烈倾斜，几何中心容易被边缘群点带偏，向室外发散。为解决这一问题，SE-IDW 算法引入了无线电物理学先验知识重构权重体系。

首先是还原射频信号强度的线性能量。根据无线通信中经典的对数距离路径损耗模型^[8]，接收信号强度 RSSI（dBm）的衰减与物理距离的对数成正比。由于对数尺度会极大压缩强弱信号之间的权重差异，必须将其还原为线性物理功率值，以真实反映用户深入室内核心 AP 辐射区的物理概率。模型按式（1）进行映射转换：

$$P_{\text{linear}}^{(i)} = 10^{\frac{\text{RSSI}_i}{10}} \text{ (mW)} \quad (1)$$

式中： RSSI_i 为第 i 个轨迹点所捕获的最强有效无线接入点（AP）的信号强度。

接着定义结合引力的反距离权重。借鉴反距离加权（IDW）空间插值思想^[9]，距离初始空间簇几何中心越近且信号物理能量越强的点，其未受多径漂移污染的置信度越高。因此，定义第*i*个轨迹点的综合权重 W_i 为：

$$W_i = \frac{P_{\text{linear}}^{(i)}}{(d_i + \varepsilon)^\alpha} \tag{2}$$

式中： d_i 代表该点至初始簇几何中心的欧氏距离； ε 代表取值为0.1的极小平滑因子，旨在消除点位重合（ $d_i=0$ ）时触发的数值除零奇点； α 代表距离衰减指数（遵循空间辐射的反平方定律，取值为2）。

最后，执行语义坐标的加权合成。利用权重序列，对候选子序列内的所有二维坐标点进行加权平均计算，得出修正后的最终语义停留点坐标 $(\widehat{\text{Lat}}, \widehat{\text{Lon}})$ ：

$$(\widehat{\text{Lat}}, \widehat{\text{Lon}}) = \frac{\sum_{i=1}^m W_i \cdot (\text{Lat}_i, \text{Lon}_i)}{\sum_{i=1}^m W_i} \tag{3}$$

该机制在物理空间上会形成“向心引力”效应：即便外围存在大量因多径效应发生漂移的GPS点，只要序列内有少数处于室内深处（高线性能量 P_{linear} ）的关键点，这些高能量点凭借权重优势，就能将偏移的质心拉回到建筑物内部高置信度核心区域，从而修正位置偏差。

3 实验结果与分析

为充分验证所提算法框架在应对稀疏众包数据及发掘群体社群关联方面的效能，本节从数据分布、鲁棒性量化、几何约束评估及高阶语义聚类四个维度展开详尽的实验分析。

3.1 数据分布与轨迹质量评价指标定义

云端服务器累计接收原始轨迹记录约1.45亿条，经前述严密清洗后，提取出9595万条有效轨迹数据。数据的空间与时间分布呈现出显著的校园节律特征。图1展示了用户空间位置热力图，可以看出用户位置与校园路网及建筑布局高度吻合，位置高密度区域集中在教学楼群及宿舍区。

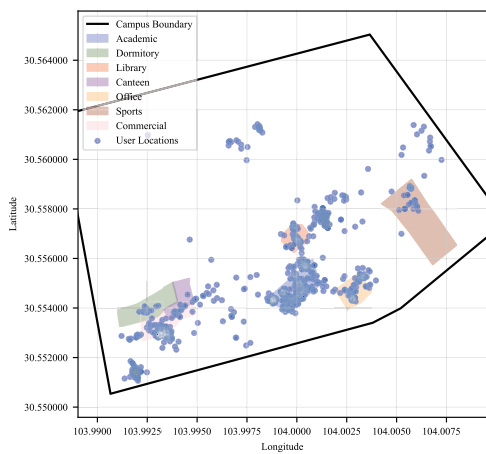


图1 校园用户空间位置热力图

图2为校园活动时间分布热力图，展示了学生群体的“潮汐效应”。图中颜色深浅及右侧色条数值代表该时段内所有用户累计产生的有效轨迹数据点数量（单位：条）。高数值区域表明该时段校园内用户活跃度高，数据采集密集。可以看出，工作日活跃高峰集中在8:00—12:00及14:00—18:00；而周末的活动分布则相对平缓，晚间（19:00后）的活跃度明显高于工作日。

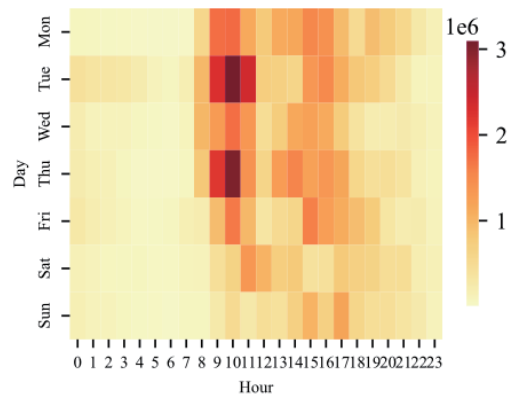


图2 校园活动时间分布热力图

除了基础的语义识别准确率（Accuracy，即预测停留点与人工校验真值一致的比例），本研究还新增了两项物理指标来量化轨迹漂移和空间收敛度：

（1）能量集中度（energy concentration, EC）：该指标用于评估合成停留点及其附件邻域内的射频信号平均功率密度。EC值越高，说明算法合成的质心越逼近真实的室内无线信号源覆盖核心盲区。

（2）边界距离（distance to boundary, DTB）：指算法生成的停留点坐标到其所属物理建筑物多边形真实边界的最短欧氏距离。多径漂移常导致定位点漂移到建筑物外面，所以较小的DTB累积误差能反映算法对向外漂移的约束效果。

3.2 鲁棒性验证

实际部署移动众包网络时，受终端电池续航能力和系统后台调度限制，轨迹数据难免会出现稀疏。本节通过对原始全量轨迹数据进行高强度降采样处理（随机保留50%、20%、10%的数据量），来验证算法在数据严重不足时的性能表现（如表1所示）。

表1 不同数据稀疏度下的算法性能对比

单位：%			
数据保留率	传统几何中心法 准确率	SE-IDW 算法 准确率	相对提升
100	70.1	85.2	+21.5
50	55.1	77.5	+40.9
20	46.0	73.0	+58.7
10	43.0	71.5	+66.3

从实验结果可以看出，数据保留率越低，传统几何中心法因空间均匀密度假设不成立，准确率下降明显。只保留10%数据量（模拟极端低频采样）时，几何中心法准确率降到43.0%。相比之下，SE-IDW算法即使去掉大部分数据，仍能保持71.5%的准确率。

3.3 空间几何约束与漂移修正验证

使用空间几何指标评估后发现，采用SE-IDW算法后，停留点的能量集中度（EC）较几何中心法跃升了18.4%（由12.48增至14.78）。同时，停留点集合与建筑物真实轮廓的累积边界距离（DTB）误差减少了525.37 m。说明新算法能把大量因多径效应漂到室外的轨迹点“拉回”并聚集到建筑功能区内部，从物理层面修正了空间漂移问题。

3.4 语义效用提升与社群网络挖掘

底层物理收敛的改善带动了下游任务的效果提升，使基于停留点的语义行为识别准确率提升了12%^[10]。在此基础上，采用K-means算法对用户进行了自适应聚类分析^[11]。如图3所示，轮廓系数在K=5时达到峰值，并进一步通过t-SNE^[12]可视化验证了聚类效果。

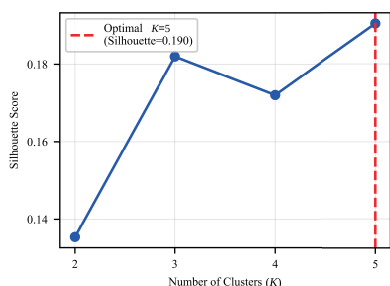


图3 不同聚类数量K的轮廓系数变化图

如图4所示，该图将高维行为特征降至二维展示，横纵坐标为降维后的抽象特征维度，点与点之间越接近说明用户行为模式越相似；图中采用五种基本平面几何形状（圆形○、方形□、上三角△、菱形◇、下三角▽）区分K-means算法生成的五类聚类簇，不同形状代表不同行为群体。五类用户在特征空间中呈现出清晰可分的簇状结构，说明聚类结果在特征空间中具有较好的可分离性。

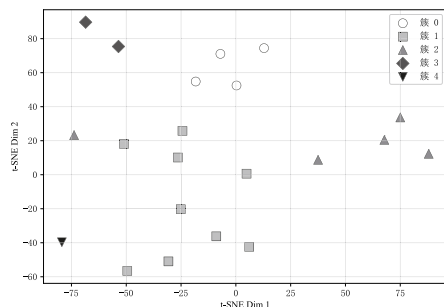


图4 用户行为聚类可视化结果图

同时，通过分析用户行为特征雷达图（图5）中各维度的行为特征，将识别出的五类典型群体定义为学业主导型、生活宅居型等，具体特征定义如表2所示。

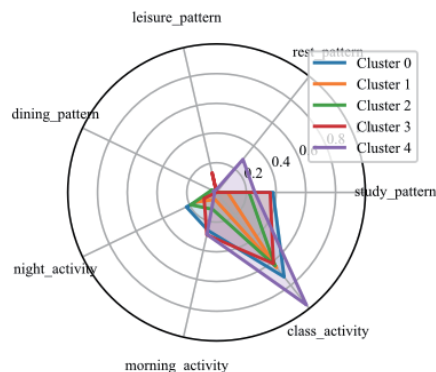


图5 用户行为特征雷达图

表2 校园用户行为群体特征定义

聚类编号	典型群体定义	核心行为特征	空间分布偏好
Cluster 0	学业主导型	课堂时间活动率高，生活规律，两点一线	教学区、图书馆
Cluster 1	生活宅居型	宿舍停留占比>60%，探索指数最低	宿舍区
Cluster 2	科研实践型	驻留时长最长，夜间活动率高，高强度投入	办公区、实验室
Cluster 3	均衡探索型	探索指数最高，活动范围广，积极探索周边	商业区、校园边缘
Cluster 4	社交活跃型	社交场所访问频次高，轨迹呈三角循环模式	食堂区、运动区

为探究群体行为模式在时空共现网络中的分布特征，将识别出的行为聚类结果映射到如图6所示的时空共现网络中。

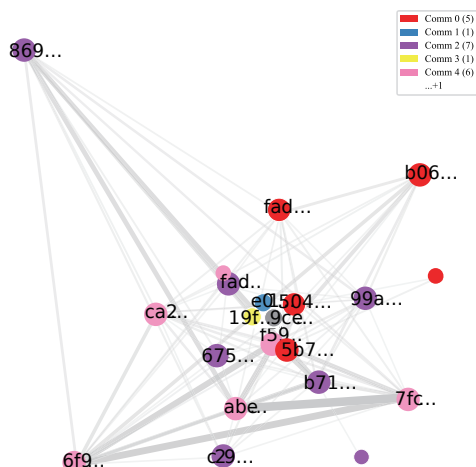


图6 校园时空共现网络拓扑图

该网络基于时空共现原理构建^[13-14]：节点代表校园内的独立用户；连线表示两用户在同一天、同一地点（空间距离

小于 50 m) 且时间重叠的共现行为; 连线的权重则记录二者在统计周期内所有共现事件的累计总时长 (单位: min)。权重越高 (图中表现为连线越粗), 意味着两人在物理空间中共同停留的时间越长, 反映了潜在的社交联系。分析发现, 时空共现网络中联系最为紧密的局部强社交集团, 其内部节点几乎毫无例外地归属于同一 K-means 行为画像簇类。这一发现以大规模数据实证了物理空间的高频“同现”与微观行为模式的高度“同质性”存在极强的正相关耦合, 揭示了维系校园同类社群纽带的核心机制。

4 结论

本文针对智慧校园场景, 提出从端侧高密度采集到语义挖掘的完整框架。核心贡献是提出信号强度增强的反距离加权停留点检测算法 (SE-IDW), 基于 9 595 万条轨迹数据验证其性能。结果表明: 在数据丢失 90% 的极端情况下, SE-IDW 仍保持 71.5% 的定位准确率 (传统方法仅 43.0%), 支撑低功耗“绿色感知”; 算法使信号能量集中度提升 18.4%, 有效修正室内漂移, 语义识别准确率提高 12%。基于高保真停留点识别出五类校园群体, 并揭示时空共现网络与社群结构的强耦合特征。本研究受限于单一校区 21 名志愿者及 30 天周期, 未来将扩大样本、引入联邦学习保护隐私^[15], 并融合能耗数据优化节能策略。

参考文献:

- [1]MIN-ALLAH N, ALRASHED S. Smart campuses: a sketch [J]. Sustainable cities and society, 2020, 59:102231.
- [2]GUO B, WANG Z, YU Z W, et al. Mobile crowd sensing and computing: the review of an emerging human-powered sensing paradigm [J]. ACM computing surveys (CSUR), 2015, 48(1): 1-31.
- [3]ASHBROOK D, STARNER T. Using GPS to learn significant locations and predict movement across multiple users [J]. Personal and ubiquitous computing, 2003, 7: 275-286.
- [4]BARBOSA H, BARTHELEMY M, GHOSHAL G, et al. Human mobility: models and applications [J]. Physics reports, 2018, 734: 1-74.
- [5]ZENG S, CHEN W W, JI Y H, et al. Measurement and calibration of EMF: a study using phone and GBDT for mobile communication signals[J]. Radio science, 2024, 59(2): 1-13.
- [6]ESTER M, KRIEGEL H P, SANDER J, et al. A density-based algorithm for discovering clusters in large spatial databases with noise [C]// Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). New York:ACM,1996: 226-231.
- [7]LI Q N, ZHENG Y, XIE X, et al. Mining user similarity based on location history [C]// Proceedings of the 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. New York: ACM, 2008: 1-10.
- [8]RAPAPORT T S. Wireless communications: principles and practice [M]. 北京: 电子工业出版社, 2018.
- [9]SHEPARD D. A two-dimensional interpolation function for irregularly-spaced data [C]// Proceedings of the 1968 24th ACM National Conference. New York:ACM, 1968: 517-524.
- [10]CHEN F, JING C F, ZHANG H, et al. Wi-Fi log-based student behavior analysis and visualization system [J]. The international archives of the photogrammetry, remote sensing and spatial information sciences, 2022: 493-499.
- [11]MACQUEEN J. Some methods for classification and analysis of multivariate observations [EB/OL].1967[2026-04-17]. <https://www.semanticscholar.org/paper/Some-methods-for-classification-and-analysis-of-MacQueen/ac8ab51a86f1a9ac74dd0e4576d1a019f5e654ed>.
- [12]VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE [J]. Journal of machine learning research, 2008, 9(86): 2579-2605.
- [13]EAGLE N, PENTLAND A S. Reality mining: sensing complex social systems [J]. Personal and ubiquitous computing, 2006, 10: 255-268.
- [14]CRANDALL D J, BACKSTROM L, COSLEY D, et al. Inferring social ties from geographic coincidences[J]. Proceedings of the national academy of sciences, 2010, 107(52): 22436-22441.
- [15]LI J C, MENG Y, MA L C, et al. A federated learning based privacy-preserving smart healthcare system[J].IEEE transactions on industrial informatics, 2021,18(3): 2021-2031.

【作者简介】

杨磊(2001—), 男, 安徽六安人, 硕士研究生, 研究方向: 电磁环境测量。

赵翔(1973—), 女, 四川西昌人, 博士, 教授、博士生导师, 研究方向: 电磁兼容分析与电磁效应评估。

(收稿日期: 2026-03-11 修回日期: 2026-04-28)