

基于Mann-Whitney秩和检验的无线局域网室内映射与定位方法

周 牧 王烟濛* 袁 慧 田增山

(重庆邮电大学通信与信息工程学院 重庆 400065)

摘 要: 该文提出一种基于Mann-Whitney秩和检验的无线局域网(WLAN)室内映射与定位方法。该方法首先根据实际定位精度需求对目标区域中的运动路径进行分段, 同时基于Mann-Whitney秩和检验方法合并相似运动路径片段; 然后, 利用一种基于相似接收信号强度(RSS)序列片段的信号聚类算法, 保证同一聚类中RSS样本的物理邻接关系; 最后, 通过骨干节点的扩散映射, 建立物理与信号空间的映射关系, 实现对运动用户的定位。实验结果表明, 相比于已有WLAN室内映射与定位方法, 该文方法在无需运动传感器辅助和构建位置指纹数据库的条件下, 能够实现更高的映射与定位精度。

关键词: 无线局域网; 室内定位; 行为分析; Mann-Whitney秩和检验; 空间映射

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2019)07-1555-10

DOI: [10.11999/JEIT180392](https://doi.org/10.11999/JEIT180392)

Mann-Whitney Rank Sum Test Based Wireless Local Area Network Indoor Mapping and Localization Approach

ZHOU Mu WANG Yanmeng YUAN Hui TIAN Zengshan

(School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: The Mann-Whitney rank sum test based Wireless Local Area Network (WLAN) indoor mapping and localization approach is proposed. Firstly, according to the localization accuracy requirement, this approach performs the motion paths segmentation in target area, and meanwhile merges the similar motion path segments based on the Mann-Whitney rank sum test. Then, a signal clustering algorithm based on the similar Received Signal Strength (RSS) sequence segments is adopted to guarantee the physical adjacency of the RSS samples in the same cluster. Finally, the backbone nodes based diffusion mapping is used to construct the mapping relations between the physical and signal spaces, and the motion user localization is consequently achieved. The experimental results indicate that compared with the existing WLAN indoor mapping and localization approaches, the proposed one is able to achieve higher mapping and localization accuracy without motion sensor assistance or location fingerprint database construction.

Key words: Wireless Local Area Network (WLAN); Indoor localization; Behavior analysis; Mann-Whitney rank sum test; Space mapping

收稿日期: 2018-04-26; 改回日期: 2019-03-08; 网络出版: 2019-03-28

*通信作者: 王烟濛 hiwangym@gmail.com

基金项目: 国家自然科学基金(61771083, 61704015), 重庆市研究生科研创新项目(CYS17221, CYS18240), 长江学者和创新团队发展计划(IRT1299), 重庆市科委重点实验室专项经费, 重庆市基础与前沿研究计划基金资助项目(cstc2017jcyjAX0380, cstc2015jcyjBX0065), 重庆市高校优秀成果转化(KJZH17117)

Foundation Items: The National Natural Science Foundation of China (61771083, 61704015), The Postgraduate Scientific Research and Innovation Project of Chongqing (CYS17221, CYS18240), The Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), The Fundamental and Frontier Research Project of Chongqing (cstc2017jcyjAX0380, cstc2015jcyjBX0065), The University Outstanding Achievement Transformation Project of Chongqing (KJZH17117)

1 引言

随着无线通信技术的发展^[1-3], 人们对位置服务的需求与日俱增。现有的成熟定位系统, 如全球定位系统(Global Positioning System, GPS)等, 已在室外环境下得到广泛应用。但由于室内环境的复杂性以及多径信号干扰等原因, GPS等室外定位系统在室内环境中的定位性能差强人意。与此同时, 随着无线局域网(Wireless Local Area Network, WLAN)的快速普及, 利用广泛部署的WLAN基础设施来实现室内定位已受到人们的密切关注。由于来自WLAN无线接入点(Access Point, AP)的接收信号强度(Received Signal Strength, RSS)易于测量, 如用于楼内用户定位与跟踪的射频系统(RADAR)^[4]和Horus系统^[5]等采用RSS位置指纹的WLAN室内定位方案已得到普遍应用。这类系统通常包括两个阶段, 在离线阶段, 用户需在目标区域内每个参考点处采集来自不同AP的RSS, 以构建位置指纹数据库, 在在线阶段, 系统根据新采集的RSS与位置指纹数据库的匹配结果, 实现对用户的位置估计^[6]。然而, 位置指纹数据库的构建通常需要消耗大量的人力和时间开销, 从而导致WLAN室内定位系统的普适性较差。

于是, 为了降低位置指纹数据库的构建开销, 人们提出独立于位置指纹的室内映射与定位方法, 如即时定位与映射(Simultaneous Localization And Mapping, SLAM)方法^[7], 其将用户随机采集的RSS样本与运动传感器信息相融合, 以避免繁复的逐点指纹数据采集过程。独立于位置勘测的无线室内定位(Wireless Indoor Localization without site survey, WILL)^[8,9]利用加速度信息对RSS样本进行筛选, 以构造信号空间中的用户运动地图, 并根据用户运动地图与物理环境的映射关系来实现定位。Li等人^[10]提出了一种基于模糊理论的WLAN室内映射方法, 其利用加速计和陀螺仪数据对5种典型用户运动模式进行识别, 并通过挖掘不同用户的姿态特征来优化定位结果。Naik等人^[11]利用专用的探测器监视和识别通信范围内的基站, 并在此基础上构建动态地图以实现室内定位。SmartSLAM系统^[12]基于智能手机进行人员位置跟踪的同时, 实现对未知室内环境平面图的绘制。ActionSLAM系统^[13]通过对具有位置依赖性的人体动作进行标定, 修正运动传感器的漂移误差, 实现准确定位。WiSLAM系统^[14]则是在行人足部安装惯性测量单元(Inertial Measurement Unit, IMU), 通过融合IMU中的Wi-Fi和里程计数据来实现室内跟踪。此外, SignalSLAM系统^[15]通过融合Wi-Fi、蓝牙、长期演进(Long

Term Evolution, LTE)网络和磁传感器数据, 构建用于室内定位和映射的多模式信号图。然而, 基于SLAM的映射与定位方法通常对终端设备有特殊的硬件要求, 且需要利用行之有效的数据融合算法进行特征提取和参数优化^[16], 从而对系统复杂度提出了更高的要求。

与此同时, Zhou等人^[17]利用热点排序方法建立物理和信号空间的映射关系, 其无需运动传感器辅助和构建位置指纹数据库, 并在文献^[18]中利用半监督流形学习方法, 实现了在少量标记点条件下的高精度定位。Wang等人^[19]利用RSS时空特性来实现室内定位和路径映射, 其无需任何特殊硬件设备或平面地图的辅助。文献^[20]则利用RSS在时域和频域上的变化, 通过对少量标定的RSS数据进行学习, 得到行人在运动和静止状态下对应的RSS阈值, 进而实现对目标位置和相关运动行为的估计。Shin等人^[21]通过对连续采集的RSS数据进行语义标定, 构建位置转移拓扑图, 来进行室内定位。然而, 该方法需人为划分目标区域且采集一定数量的标记信号, 使得系统可靠性受到限制。

针对上述问题, 本文提出一种基于Mann-Whitney秩和检验的WLAN室内映射与定位方法。本方法包括客户端、商户端和服务端3个部分, 其中, 客户端用户按照其日常行走习惯在目标区域中随机采集RSS序列; 商户端提供目标区域结构图并根据实际需求设定期望定位精度; 服务端则利用客户端和商户端提供的信息, 建立物理与信号空间的映射关系并完成对客户端的定位。具体而言, 首先, 通过对目标区域物理结构的分析, 预测用户运动趋势并利用A*搜索算法^[22]对运动路径进行建模。同时, 基于Mann-Whitney秩和检验合并相似运动路径片段, 利用得到的物理聚类实现对目标区域的划分。相比于传统基于建筑结构的人为物理空间划分方法^[8,9,18], 本文所提方法考虑了用户的行走习惯, 从而具有更好的物理空间转移逻辑性。然后, 根据连续采集RSS信号在物理空间中的邻接关系, 同样利用Mann-Whitney秩和检验方法对RSS序列片段进行聚类, 由此保证所属相同聚类的RSS样本具有较好的物理邻接性。最后, 为了建立物理与信号空间的映射关系, 将信号空间中出现的频率较高的信号聚类(即信号空间中的骨干节点)映射到物理空间中用户活动频繁的物理聚类(即物理空间中的骨干节点), 同时利用逐层扩散方法对骨干节点的邻接点依次进行映射, 直到所有信号聚类均被映射到物理空间中。此外, 本文所提方法无需运动传感器辅助和构建位置指纹数据库, 且具有较高的映射与定位精度和较强的环境适应性。

本文结构安排如下：第2节详细描述了本文所提基于Mann-Whitney秩和检验的WLAN室内映射与定位方法；第3节通过实验数据验证了所提方法的有效性；第4节总结全文并给出下一步工作。

2 系统描述

本系统包括用户端、商户端和服务端3个部分，其系统架构如图1所示。

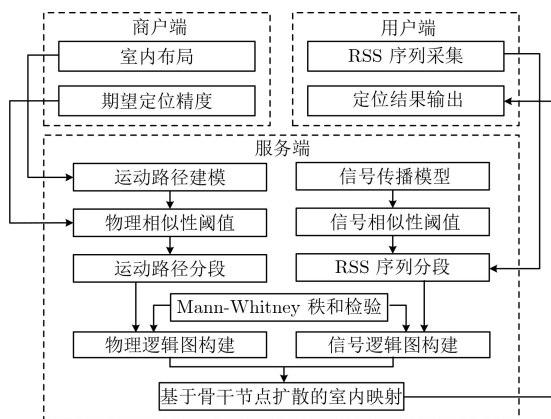


图1 系统架构

2.1 物理空间

2.1.1 运动路径建模

为了保护用户的位置隐私，选择图2所示走廊区域为目标区域并在该区域中布置 $N_{AP}=4$ 个WLAN接入点。同时，为了刻画运动路径的物理属性，将目标区域中的障碍物分为兴趣障碍物(即用户经常使用的物体，如门和椅子)和非兴趣障碍物(即用户较少使用的物体，如空白墙面和窗户)，同时将可行走区域分为功能区域(即兴趣障碍物的邻近区域，如房间和电梯的出入口区域)和非功能区域(即非兴趣障碍物的邻近区域，如不同空白墙面之间的走廊区域)。从而，每条运动路径可表示为两个功能区域之间的用户行走轨迹。为了模拟用户在目标区域中的行走轨迹，本系统采用启发式A*搜索算法[22]对用户运动路径进行建模。具体而言，首先将图2中目标区域结构图转变为图3所示目标区域图

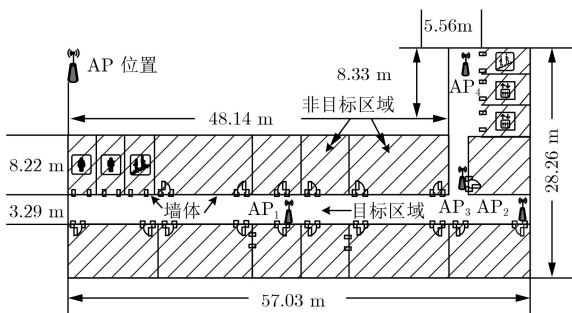


图2 目标区域结构图

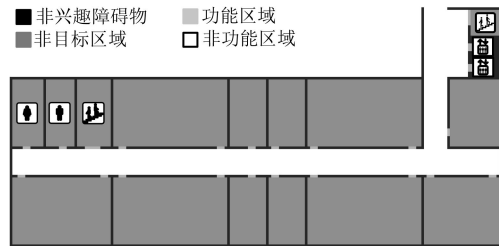


图3 目标区域图像

像，其中，像素宽度为 $w_p=0.3$ m，黑色为非兴趣障碍物，深灰色为非目标区域，浅灰色和白色分别为功能和非功能区域。然后，从起始像素点(即起点位置)出发，依次将其周围8邻域像素点中位于可行走区域的像素点作为候选像素点，同时计算每个候选像素点与起始和终止像素点(即终点位置)之间距离之和，选择距离之和最小的候选像素点为下一像素点。迭代该过程，直到终止像素点被遍历，由此得到的起点与终点间一系列连续的像素点集合即为构建的运动路径。

本系统中，当前与前一像素点的距离用欧氏距离进行度量，并根据每个像素点与其前一像素点的欧式距离累加，得到当前与起始像素点的距离，该距离的定义符合用户对其所行走路径距离的评判习惯。与此同时，由于在实际室内环境中运动路径的终点位置通常位于用户的视距范围之外，当前与终止像素点之间的行走距离往往远大于其欧式距离，因此，本系统利用曼哈顿距离对当前与终止像素点的距离进行估计[23]。此外，考虑用户行走的随机性，在每次运动路径建模时，本系统将随机选择目标区域图像中一定比例的像素点为不可行走的障碍点，于是对于具有相同起始和终止功能区域的多次运动路径建模结果均存在差异。

2.1.2 运动路径建模

由于用户的行走习惯通常为选择两点间的最短路径行走，在图4(a)室内无障碍物环境中，相比于曼哈顿距离，欧式和路径距离更能有效刻画真实与估计位置之间的用户行走距离；而在图4(b)室内有障碍物环境中，相比于欧式和曼哈顿距离，路径距离更能有效刻画真实与估计位置之间的用户行走距离。基于此，本系统利用路径距离来度量用户位置

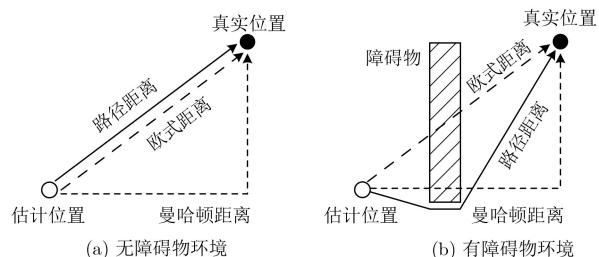


图4 不同环境中的距离度量比较

的估计误差。基于上述距离度量方法,若令与用户真实位置的距离小于期望定位精度 A_r 的区域为可接受定位区域,则可接受定位区域内任意两点之间的距离小于 $2A_r$ 。此时,令物理相似度阈值 $\varepsilon_P = 2A_r$,则同一物理区域内的点彼此之间的距离小于 ε_P 。于是,对于每条运动路径,从其起始像素点开始遍历,将路径距离小于 ε_P 的像素点划分为一个运动路径片段 $T_i = [(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), \dots, (x_{in_i}, y_{in_i})]$,其中, n_i 为 T_i 包含的像素点数(即 T_i 的片段长度)。此外,过短的运动路径片段不能有效地刻画用户运动行为,于是通过设定像素点阈值 $\text{Num}_P = \lceil l_{\min}/w_P \rceil$ 剔除像素点个数小于 Num_P 的运动路径片段,其中,“ $\lceil \cdot \rceil$ ”表示取整运算, $l_{\min}$ 为允许的最短运动路径片段长度。

2.1.3 物理逻辑图构建

为了构建能充分反映用户在目标环境中各子区域间转移关系的物理逻辑图,不同于传统的基于建筑结构的人为划分方法,本系统引入用户的行走习惯,通过合并相似运动路径片段得到物理聚类,实现对目标区域的划分。同时,由于用户运动具有随机性,难以选择某一分布函数对运动路径上位置进行刻画,因此,本系统采用Mann-Whitney秩和检验方法^[24]对两两运动路径片段的相似性进行度量。具体而言,对于任意两条运动路径片段 $T_p = [(x_{p1}, y_{p1}), (x_{p2}, y_{p2}), \dots, (x_{pn}, y_{pn})]$ 和 $T_q = [(x_{q1}, y_{q1}), (x_{q2}, y_{q2}), \dots, (x_{qm}, y_{qm})]$,可得其第1和第2维坐标集合分别为 $[x_{p1}, x_{p2}, \dots, x_{pn}]$ 和 $[x_{q1}, x_{q2}, \dots, x_{qm}]$,以及 $[y_{p1}, y_{p2}, \dots, y_{pn}]$ 和 $[y_{q1}, y_{q2}, \dots, y_{qm}]$,其中, n 和 m 分别为 T_p 和 T_q 的片段长度。若将 T_p 和 T_q 的第 i 维坐标集合分别看作来自总体 C_i^p 和 C_i^q ,则检验 T_p 和 T_q 的第 i 维坐标集合是否相似的问题等价于检验 C_i^p 和 C_i^q 是否相同。令Mann-Whitney秩和检验显著性水平为 α ,构造假设检验问题

$$\left. \begin{aligned} H_0: C_i^p \text{和} C_i^q \text{相似} \\ H_1: C_i^p \text{和} C_i^q \text{不相似} \end{aligned} \right\} \quad (1)$$

构造包含 $n+m$ 个数据的集合 $C_i^t = C_i^p \cup C_i^q$,令 C_i^t 中最小数据的秩为1,次小数据的秩为2,以此类推,得到最大数据的秩为 $n+m$,同时将相同数据的秩定义为其对应的所有秩的平均值。通过对属于 C_i^p 和 C_i^q 的数据的秩分别求和,得到秩和 W_i^p 和 W_i^q ,进而得到 C_i^p 和 C_i^q 的检验统计量 $U_i^p = W_i^p - n(n+1)/2$ 和 $U_i^q = W_i^q - m(m+1)/2$ 。令 $U = \min(U_i^p, U_i^q)$,由于 U 在小样本(即 $n, m \leq 20$)情况下近似服从Mann-Whitney分布^[25],因此根据Mann-Whitney秩和检验临界值表^[26],当 U 大于临

界值表中数据个数分别为 n 和 m 条件下的临界值 U_α 时,接受 H_0 ,即认为 T_p 和 T_q 的第 i 维坐标集合相似;反之,则认为不相似。与此同时,在大样本情况下, U 近似服从正态分布^[17],若 C_i^t 中不存在相同数据,则由正态逼近过程可得检验统计量

$$z = (U - nm/2) / \sqrt{mn(m+n+1)/2} \quad (2)$$

若 C_i^t 中存在相同数据,则式(2)可修正为

$$z = \left(U - \frac{nm}{2} \right) / \sqrt{ \frac{mn(m+n+1)}{12} - \frac{mn \left(\sum_{i=1}^g \tau_i^3 - \sum_{i=1}^g \tau_i \right)}{12(m+n)(m+n+1)} } \quad (3)$$

其中, τ_i 为第 i 种相同数据的个数, g 为相同数据的类数。计算标准正态分布条件下从 $-\infty$ 到 z 的累积分布概率 P' ,然后比较精度双尾 $P = 2(1 - P')$ 与显著性水平 α 的大小,若 $P < \alpha$,则接受 H_1 ;反之则接受 H_0 。

基于上述Mann-Whitney秩和检验方法,可对 T_p 和 T_q 的每一维坐标集合进行检验,若两者的每一维坐标集合均来自相同总体,则认为 T_p 和 T_q 为相似运动路径片段。最后,通过合并相似的运动路径片段,得到 N_P 个物理聚类,进而根据不同物理聚类间的转移关系,将每个物理聚类和对应的转移关系抽象为物理逻辑图的顶点 $v_i^P (i = 1, 2, \dots, N_P)$ 和边 $e_{ij}^P = (v_i^P, v_j^P) (i, j = 1, 2, \dots, N_P; i \neq j)$ 。特别地,为了保证目标区域的连通性,当构建的物理逻辑图为非连通图时,仅保留其最大连通子图。

2.2 信号空间

2.2.1 RSS序列采集与分段

根据用户行走习惯,用户端在目标区域内随机采集 N_{tr} 条RSS序列 $\{\mathbf{RSS}_1, \mathbf{RSS}_2, \dots, \mathbf{RSS}_{N_{tr}}\}$,其中, $\mathbf{RSS}_i = \{\mathbf{rss}_{i1}, \mathbf{rss}_{i2}, \dots, \mathbf{rss}_{iM_i}\} (i = 1, 2, \dots, N_{tr})$, $\mathbf{rss}_{ij} = (\mathbf{rss}_{ij1}, \mathbf{rss}_{ij2}, \dots, \mathbf{rss}_{ijk}) (j = 1, 2, \dots, M_i)$, M_i 为第 i 条RSS序列长度, k 为AP个数, $\mathbf{rss}_{ijl} (1 \leq l \leq k)$ 为 $\mathbf{rss}_{ij}$ 中来自第 l 个AP的RSS值。由于具有较短路径距离的物理位置处通常采集得到相似的RSS样本^[12],于是为了提高用户连续行走过程中采集RSS序列的稳定性,对原始采集RSS序列的每一维RSS样本进行均值滤波,如图5所示。

与运动路径分段过程类似,通过设置信号相似度阈值 ε_S 对RSS序列进行分段。为此,根据信号传播模型中RSS均值与物理距离之间关系^[27],由物理相似度阈值 ε_P 计算得到信号相似度阈值 ε_S 。具体而言,首先,计算物理距离 d_1 和 d_2 处的RSS均值差

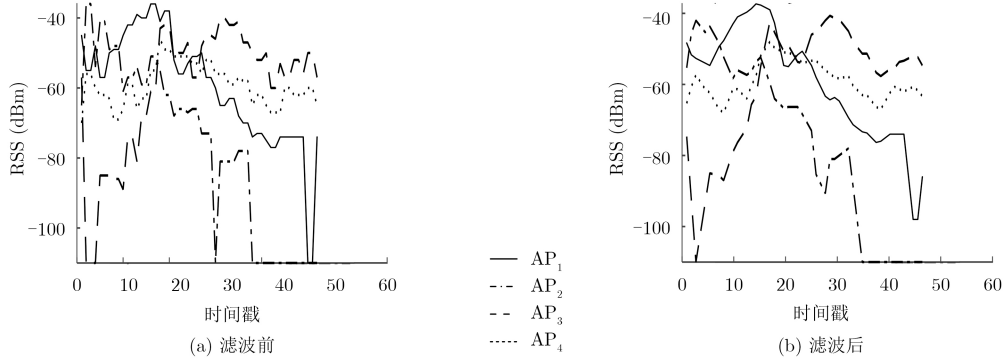


图5 某一RSS序列均值滤波结果

$$\begin{aligned}\Delta P &= \bar{P}_1 - \bar{P}_2 = 10\beta \lg(d_2/d_1) \\ &= 10\beta \lg(1 + \Delta d/d_1)\end{aligned}\quad (4)$$

其中, β 为路径损耗指数^[28], $\Delta d = |d_1 - d_2|$, 当 Δd 给定时, d_1 越小, ΔP 越大。于是, 令 $d_1 = 1$ m, $\Delta d = \varepsilon_P$, 则可计算得到 $\Delta P = 10\beta \lg(1 + \varepsilon_P/d_1) = \varepsilon_S$ 。对于每条RSS序列, 从其第1个RSS矢量开始遍历, 将每一维RSS差值均小于 ε_S 的RSS矢量分为一段 $S_i = \{\text{rss}_{i1}, \text{rss}_{i2}, \dots, \text{rss}_{im_i}\}$, 其中, m_i 为第 i 条RSS序列片段 S_i 包含的RSS矢量数(即 S_i 的片段长度)。

2.2.2 信号逻辑图构建

类似地, 采用Mann-Whitney秩和检验方法对两两RSS序列片段的相似性进行检验。具体而言, 若两条RSS序列片段的每一维RSS矢量集合均来自相同总体, 则认为它们为相似RSS序列片段。然后, 通过合并相似RSS序列片段, 得到 N_S 个信号聚类, 进而根据不同信号聚类之间的转移关系, 将每个信号聚类和对应的转移关系抽象为信号逻辑图中的顶点 v_i^S 和边 $e_{ij}^S = (v_i^S, v_j^S)$ ($i, j = 1, 2, \dots, N_S; i \neq j$)。特别地, 为保证RSS序列采集的连续性, 当构建的信号逻辑图为非连通图时, 仅保留其最大连通子图。

2.3 基于骨干节点扩散的室内映射

由于物理聚类之间的转移关系刻画了用户的运动趋势, 而信号聚类之间的转移关系反映了其对应RSS序列的采集路径, 于是, 当运动趋势能够有效反映目标区域中用户的行走习惯时, 不同物理聚类之间和信号聚类之间的转移关系将存在相似性。基于此, 本系统基于龙骨映射^[8]的思想, 提出一种关于物理与信号空间映射的骨干节点扩散映射方法。本方法首先对物理和信号逻辑图中的骨干节点进行映射, 然后利用逐层扩散方法对骨干节点的邻接点依次进行映射, 直到所有信号聚类均被映射到具有相似转移关系的物理聚类。

首先, 计算物理和信号逻辑图中所有顶点的邻接度^[29], 显然, 某个顶点的邻接度越大, 其在图中的中心性就越明显。令顶点 v 的度为 $\deg(v)$, 其邻

接点 v_i^A ($i = 1, 2, \dots, N_d$) 的度为 $\deg(v_i^A)$, 其中, N_d 为 v 的邻接点个数, 则 v 的邻接度可计算为 $\text{Adj}(v) = \deg(v) + \sum_{i=1}^{N_d} \deg(v_i^A)$ 。

然后, 分别将物理和信号逻辑图中的顶点按邻接度降序排列, 得到集合 $\mathbf{V}_{\text{Ra}}^P = [v_{\text{Ra}-1}^P, v_{\text{Ra}-2}^P, \dots, v_{\text{Ra}-N_P}^P]$ 和 $\mathbf{V}_{\text{Ra}}^S = [v_{\text{Ra}-1}^S, v_{\text{Ra}-2}^S, \dots, v_{\text{Ra}-N_S}^S]$, 其中, $\text{Adj}(v_{\text{Ra}-1}^P) \geq \dots \geq \text{Adj}(v_{\text{Ra}-N_P}^P)$ 且 $\text{Adj}(v_{\text{Ra}-1}^S) \geq \dots \geq \text{Adj}(v_{\text{Ra}-N_S}^S)$ 。令物理和信号逻辑图中邻接度最大的 N_A 个节点所组成的骨干节点集合为 $\mathbf{V}_{\text{Bo}}^P = [v_{\text{Ra}-1}^P, v_{\text{Ra}-2}^P, \dots, v_{\text{Ra}-N_A}^P]$ 和 $\mathbf{V}_{\text{Bo}}^S = [v_{\text{Ra}-1}^S, v_{\text{Ra}-2}^S, \dots, v_{\text{Ra}-N_A}^S]$, 则通过将信号逻辑图中的骨干节点映射到物理逻辑图中与其具有相同邻接度排序的骨干节点, 实现物理和信号逻辑图的骨干节点映射。当完成所有骨干节点映射之后, 继续对骨干节点的邻接点进行映射。此时, 令 $v_{\text{Ra}-1}^P$ 的邻接点集合为 $\mathbf{A}_1^P$ 且 $v_{\text{Ra}-1}^S$ 的邻接点集合中未映射的顶点集合为 $\mathbf{A}_1^S$, 分别对 $\mathbf{A}_1^P$ 和 $\mathbf{A}_1^S$ 中的顶点按邻接度降序排列得到 $\mathbf{A}_{\text{Ra}-1}^P$ 和 $\mathbf{A}_{\text{Ra}-1}^S$, 将 $\mathbf{A}_{\text{Ra}-1}^P$ 和 $\mathbf{A}_{\text{Ra}-1}^S$ 中邻接度排序相同的顶点进行映射, 从而完成物理和信号逻辑图中具有最大邻接度的骨干节点的邻接点映射。特别地, 若 $\mathbf{A}_{\text{Ra}-1}^S$ 中顶点数 $N_{\text{Ra}-1}^S$ 大于 $\mathbf{A}_{\text{Ra}-1}^P$ 中顶点数 $N_{\text{Ra}-1}^P$, 则不对 $\mathbf{A}_{\text{Ra}-1}^S$ 中邻接度序号大于 $N_{\text{Ra}-1}^P$ 的顶点进行映射。以此类推, 直到完成所有骨干节点的邻接点映射。由于某些物理区域中可能没有采集RSS序列, 所以可能存在物理逻辑图中某些顶点未被映射的情况, 此时, 未被映射的顶点将从物理逻辑图中剔除。

2.4 运动用户定位

本系统利用物理与信号聚类之间的映射关系, 对目标区域内用户进行定位。具体而言, 令第 i 个物理聚类所包含位置坐标的均值 $(\bar{x}_i, \bar{y}_i)$ 为其物理聚类中心, 第 j 个信号聚类所包含RSS矢量的均值 $\bar{\text{rss}}_j$ 为其信号聚类中心, 当用户新采集RSS序列为 $\mathbf{RSS}^{\text{new}} = \{\text{rss}_1^{\text{new}}, \text{rss}_2^{\text{new}}, \dots, \text{rss}_M^{\text{new}}\}$ 时, 对于每个

新采集的RSS矢量 $\mathbf{rss}_i^{\text{new}}$ ($i=1, 2, \dots, M$), 计算其与每个信号聚类中心的欧式距离, 选择距离最近的中心所对应信号聚类为 $\mathbf{rss}_i^{\text{new}}$ 的所属聚类, 该信号聚类所映射的物理聚类中心即为用户该时刻的估计位置。

3 实验结果

实验环境选择某实验楼5楼走廊区域, 其平面结构如图2所示。针对图3所示目标区域图像, 在Matlab仿真平台上利用A*搜索算法对同一楼梯出

入口到同一实验室门口之间两次运动路径进行模拟, 同时假定用户在运动过程中不会存在长时间逗留、中途折返等情况。根据用户在目标区域中的行走习惯, 分别针对上班、工作和下班3种假设模式各仿真95条运动路径, 如图6所示。在上班、工作和下班3种假设模式下, 用户的行走习惯分别倾向于从楼梯或电梯出入口走向各个实验室门口、从各个实验室门口走向其它实验室门口和从各个实验室门口走向楼梯或电梯出入口。像素点色度越深, 代表运动路径经过该像素点次数越多。

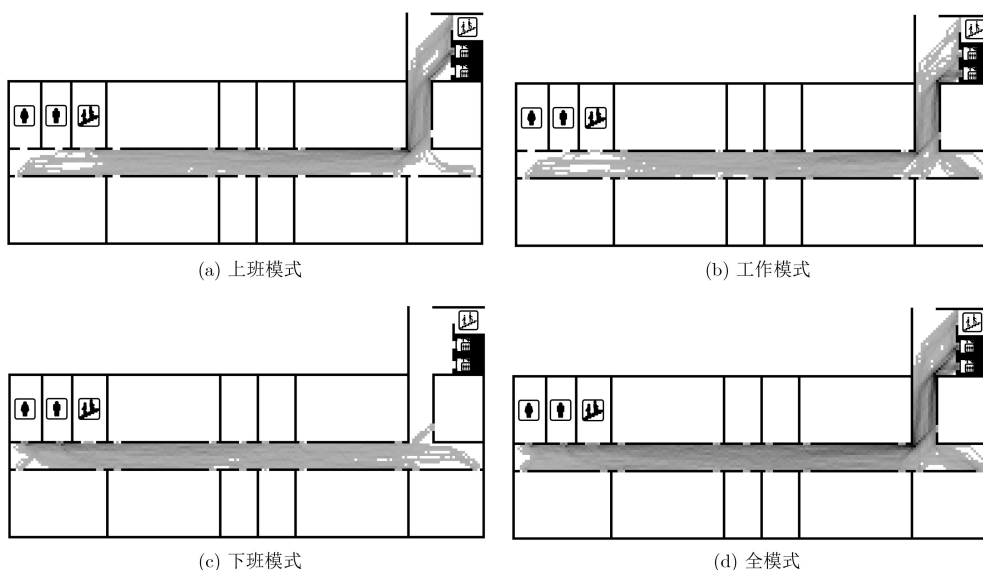


图6 不同假设模式下的运动路径仿真结果

不失一般性, 令 $\varepsilon_p=7$ m并对每条运动路径进行分段。考虑路径长度小于1 m的运动路径片段不能有效刻画用户的运动行为, 故剔除包含像素点个数小于4的运动路径片段(因为 $4 \times w_p = 1.2$ m $>$ 1 m)。在统计学中, 通常把在现实世界中发生几率小于5%的事件称为“不可能事件”, 即小概率原理^[30]。因此, 令显著性水平 $\alpha = 0.05$, 当精度双尾 P 值小于该显著性水平时, 拒绝原假设 H_0 , 即比较的两条运动路径片段不相似。基于此, 利用Mann-Whitney秩和检验方法对运动路径片段进行聚类, 得到图7

所示聚类结果。最后, 根据不同物理聚类之间的转移关系, 构造物理逻辑图。

与此同时, 在图2所示实验环境中, 使用D-Link DAP-2310作为AP发射Wi-Fi信号, 并选择三星S7568手机及自主开发的Wi-Fi信号采集软件对每条运动路径上的RSS数据进行采集。由图8可知, 同一物理聚类中的各个位置点大多位于彼此的视距(Line Of Sight, LOS)范围内。因此, 令式(4)中的参数 β 等于室内LOS环境下典型路径损耗指数值, 即 $\beta=1.6$ ^[31]。从而, 由式(4)可得 $\varepsilon_s=14.4$ dBm并对

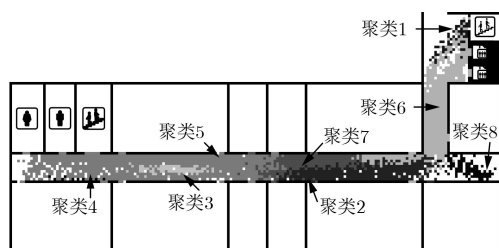


图7 运动路径片段聚类结果

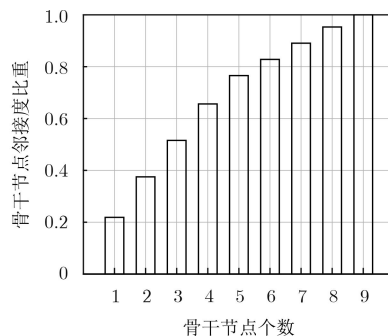


图8 骨干节点集合重要性对比

RSS序列进行分段。同样，利用Mann-Whitney秩和检验方法对RSS序列片段进行聚类，根据信号聚类间的转移关系，构造信号逻辑图。根据骨干节点扩散映射方法，分别选取物理和信号逻辑图中邻接度最大的 N_A 个顶点作为两幅图的骨干节点，其中，骨干节点邻接度之和占所有顶点邻接度总和的比重反映了该骨干节点集合的重要程度，如图8所示。综合考虑骨干节点集合的重要性以及节点扩

散映射的精度，本系统选择 $N_A=5$ ，此时骨干节点邻接度的比重大于75%且骨干节点扩散映射方法具有较高精度。在物理和信号逻辑图中，邻接度排序最高的5个骨干节点的编号分别依次为2, 5, 7, 6, 3和7, 6, 1, 3, 4，如图9(a)所示。当完成所有骨干节点映射之后，继续对骨干节点的邻接点进行映射，最终建立图9(b)中物理与信号空间的映射关系。

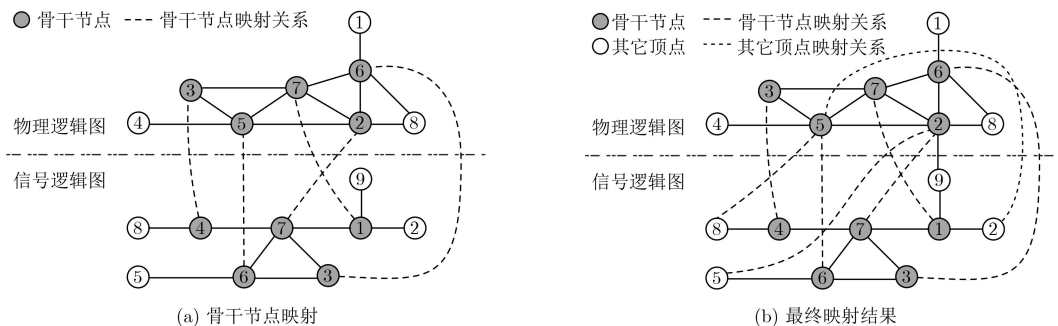


图9 物理与信号逻辑图映射关系

为了验证本文方法的映射与定位精度，分别针对上班、工作和下班3种假设模式下的用户行走习惯，各采集5条具有位置标定的RSS序列用于测试。不失一般性，在图10中根据不同物理区域的功能性对目标区域进行划分，在此基础上，图11比较了本文系统与已有WILL^[8]和文献^[17]所述系统的映射精度，其中，映射精度定义为新采集RSS矢量匹配到用户正确所属子区域的概率。由图可知，本文

系统的平均正确匹配概率为80.4%，优于WILL和文献^[17]所述系统的69.7%和68.5%。

将定位精度定义为用户估计位置与真实位置之间的欧式距离误差，可得本文系统在上班、工作和下班3种假设模式下的误差累积分布函数，如图12所示。由图可计算得到3种假设模式及全模式下的平均误差分别为5.4 m, 3.1 m, 5.2 m和4.5 m。

与此同时，图13比较了本文系统与WILL和文献^[17]所述系统在全模式下的误差累积分布函数。由图可计算得到本文系统的平均误差为4.5 m，小于WILL和文献^[17]所述系统的4.9 m和10 m。此外，由于本文系统同时具有较高的映射精度，所以其拖尾误差远小于另外两个系统。针对上述结果究其原因，WILL系统根据子区域划分个数对信号进行K均值聚类，所得信号聚类难以匹配其所属子区域且K均值聚类性能关于初始聚类中心的选择具有较大随机性^[32]，故该系统的聚类结果不稳定，定位

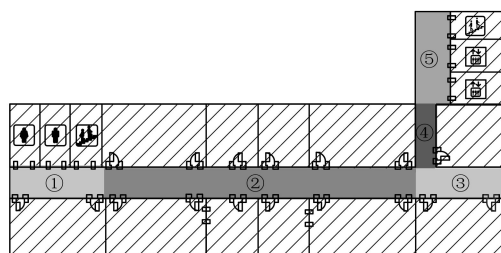


图10 目标区域划分结果

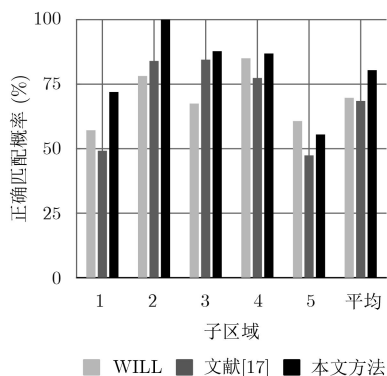


图11 映射精度对比

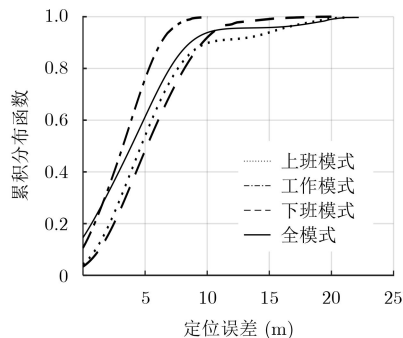


图12 不同假设模式下的定位误差

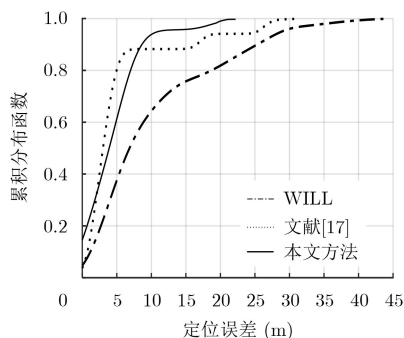


图13 不同系统的定位误差

精度难以得到保证；而文献[17]所述系统假设在信号采集过程中用户行走速率不变，显然其不符合实际室内环境中的用户行走习惯，故该系统在用户行走速率变化较大的情况下，信号聚类结果不可靠，即属于相同聚类的RSS样本可能在物理上不邻近，进而导致定位性能下降。

最后，图14比较了本文系统在不同物理相似度阈值条件下的定位精度。由图可知，当物理相似度阈值过小(如 $\varepsilon_p=3$ m)时，利用Mann-Whitney秩和检验方法得到的物理和信号聚类较多，从而难以保证较高的映射精度，并最终导致定位性能下降；而当物理相似度阈值过大(如 $\varepsilon_p=9$ m)时，每个物理聚类可能对应多个信号聚类，从而导致物理和信号逻辑图中的聚类转移关系不一致，使得映射与定位精度下降。

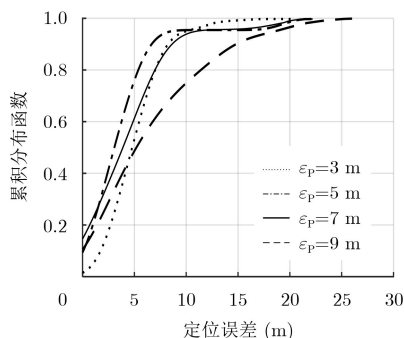


图14 不同物理相似度阈值下的定位误差

4 结束语

本文所提基于Mann-Whitney秩和检验的WLAN室内映射与定位方法不依赖于运动传感器或指纹数据库构建，并对现有WLAN硬件设备不作特殊要求，从而具有较好的普适性和经济性。此外，利用一种基于相似RSS序列片段的信号聚类算法来保证同一聚类中RSS样本的物理邻接关系，并基于骨干节点扩散映射方法建立物理与信号空间的映射关系，进而实现对运动用户的定位。相比于传

统WLAN室内映射与定位方法，本文方法具有更高的映射与定位精度。然而，本文讨论的用户行为较为简单，于是，针对复杂室内环境下的用户运动路径建模和行为特征分析是本文的下一步研究工作。

参考文献

- [1] WANG Xuyu, GAO Lingjun, and MAO Shiwen. Phasefi: Phase fingerprinting for indoor localization with a deep learning approach[C]. 2015 IEEE Global Communications Conference, San Diego, USA, 2015: 1–6. doi: [10.1109/GLOCOM.2015.7417517](https://doi.org/10.1109/GLOCOM.2015.7417517).
- [2] RADU V, LI Jiwei, KRIARA L, *et al.* Poster: A hybrid approach for indoor mobile phone localization[C]. Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, Low Wood Bay, UK, 2012: 527–528. doi: [10.1145/2307636.2307717](https://doi.org/10.1145/2307636.2307717).
- [3] Yaeli A, BAK P, FEIGENBLAT G, *et al.* Understanding customer behavior using indoor location analysis and visualization[J]. *IBM Journal of Research and Development*, 2014, 58(5/6): 3:1–3:12. doi: [10.1147/JRD.2014.2337552](https://doi.org/10.1147/JRD.2014.2337552).
- [4] BAHL P and PADMANABHAN V N. RADAR: An in-building RF-based user location and tracking system[C]. Proceedings of IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Tel Aviv, Israel, 2000: 775–784. doi: [10.1109/INFCOM.2000.832252](https://doi.org/10.1109/INFCOM.2000.832252).
- [5] YOUSSEF M and AGRAWALA A. The Horus WLAN location determination system[C]. The 3rd International Conference on Mobile Systems, Applications, and Services, Seattle, USA, 2015: 205–218. doi: [10.1145/1067170.1067193](https://doi.org/10.1145/1067170.1067193).
- [6] KOWEERAWONG C, WIPUSITWARAKUN K, and KAEMARUNGS K. Indoor localization improvement via adaptive RSS fingerprinting database[C]. 2013 International Conference on Information Networking, Bangkok, Thailand, 2013: 412–416. doi: [10.1109/ICOIN.2013.6496414](https://doi.org/10.1109/ICOIN.2013.6496414).
- [7] DURRANT-WHYTE H and BAILEY T. Simultaneous localization and mapping: Part I[J]. *IEEE Robotics & Automation Magazine*, 2006, 13(2): 99–110. doi: [10.1109/MRA.2006.1638022](https://doi.org/10.1109/MRA.2006.1638022).
- [8] WU Chenshu, YANG Zheng, LIU Yunhao, *et al.* WILL: Wireless indoor localization without site survey[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2013, 24(4): 839–848. doi: [10.1109/TPDS.2012.179](https://doi.org/10.1109/TPDS.2012.179).
- [9] WU Chenshu, YANG Zheng, LIU Yunhao, *et al.* WILL: Wireless indoor localization without site survey[C]. 2012 Proceedings IEEE INFOCOM, Orlando, USA, 2012: 64–72. doi: [10.1109/INFCOM.2012.6195809](https://doi.org/10.1109/INFCOM.2012.6195809).
- [10] LI Chao, JIANG Zhuqing, HUANG Chengkai, *et al.* A

- smartphone-based indoor positioning system using fuzzy theory and WLAN mapping algorithm[C]. 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, Hong Kong, China, 2015: 2177–2181. doi: [10.1109/PIMRC.2015.7343658](https://doi.org/10.1109/PIMRC.2015.7343658).
- [11] NAIK K K and PRASAD M N G. A system for locating users of WLAN using dynamic mapping in indoor and outdoor environment-LOIDS[C]. 2008 11th International Conference on Computer and Information Technology, Khulna, Bangladesh, 2008: 156–160. doi: [10.1109/ICCITECHN.2008.4802971](https://doi.org/10.1109/ICCITECHN.2008.4802971).
- [12] SHIN H, CHON Y, and CHA H. Unsupervised construction of an indoor floor plan using a smartphone[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 2012, 42(6): 889–898. doi: [10.1109/TSMCC.2011.2169403](https://doi.org/10.1109/TSMCC.2011.2169403).
- [13] HARDEGGER M, ROGGEN D, MAZILU S, *et al.* ActionSLAM: Using location-related actions as landmarks in pedestrian SLAM[C]. 2012 International Conference on Indoor Positioning and Indoor Navigation, Sydney, Australia, 2012: 1–10. doi: [10.1109/IPIN.2012.6418932](https://doi.org/10.1109/IPIN.2012.6418932).
- [14] BRUNO L and ROBERTSON P. WiSLAM: Improving footSLAM with Wi-Fi[C]. 2011 International Conference on Indoor Positioning and Indoor Navigation, Guimaraes, Portugal, 2011: 1–10. doi: [10.1109/IPIN.2011.6071916](https://doi.org/10.1109/IPIN.2011.6071916).
- [15] MIROWSKI P, HO T K, YI S, *et al.* SignalSLAM: Simultaneous localization and mapping with mixed Wi-Fi, Bluetooth, LTE and magnetic signals[C]. International Conference on Indoor Positioning and Indoor Navigation, Montbeliard-Belfort, France, 2013: 1–10. doi: [10.1109/IPIN.2013.6817853](https://doi.org/10.1109/IPIN.2013.6817853).
- [16] ZHANG Xiuming, JIN Yunye, TAN H X, *et al.* CIMLoc: A crowdsourcing indoor digital map construction system for localization[C]. IEEE 9th International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Singapore, 2014: 1–6. doi: [10.1109/ISSNIP.2014.6827640](https://doi.org/10.1109/ISSNIP.2014.6827640).
- [17] ZHOU Mu, ZHANG Qiao, WANG Yu, *et al.* Hotspot ranking based indoor mapping and mobility analysis using crowdsourced Wi-Fi signal[J]. *IEEE Access*, 2017, 5: 3594–3602. doi: [10.1109/ACCESS.2017.2674798](https://doi.org/10.1109/ACCESS.2017.2674798).
- [18] ZHOU Mu, ZHANG Qiao, TIAN Zengshan, *et al.* Indoor WLAN localization using high-dimensional manifold alignment with limited calibration load[C]. 2017 IEEE International Conference on Communications, Paris, France, 2017: 1–6. doi: [10.1109/ICC.2017.7997042](https://doi.org/10.1109/ICC.2017.7997042).
- [19] WANG Lin, LIU Wenyan, JING Nan, *et al.* Simultaneous navigation and pathway mapping with participating sensing[J]. *Wireless Networks*, 2015, 21(8): 2727–2745. doi: [10.1007/s11276-015-0944-x](https://doi.org/10.1007/s11276-015-0944-x).
- [20] MUTHUKRISHNAN K, VAN DER ZWAAG B J, and HAVINGA P. Inferring motion and location using WLAN RSSI[C]. Proceedings of the 2nd International Workshop on Mobile Entity Localization and Tracking in GPS-less Environments, Orlando, USA, 2009: 163–182. doi: [10.1007/978-3-642-04385-7_12](https://doi.org/10.1007/978-3-642-04385-7_12).
- [21] SHIN H and CHA H. Wi-Fi fingerprint-based topological map building for indoor user tracking[C]. 2010 IEEE 16th International Conference on Embedded and Real-Time Computing Systems and Applications, Macau, China, 2010: 105–113. doi: [10.1109/RTCSA.2010.23](https://doi.org/10.1109/RTCSA.2010.23).
- [22] HART P E, NILSSON N J, and RAPHAEL B. A formal basis for the heuristic determination of minimum cost paths[J]. *IEEE Transactions on Systems Science and Cybernetics*, 1968, 4(2): 100–107. doi: [10.1109/TSSC.1968.300136](https://doi.org/10.1109/TSSC.1968.300136).
- [23] YU Jie, AMORES J, SEBE N, *et al.* Distance learning for similarity estimation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(3): 451–462. doi: [10.1109/TPAMI.2007.70714](https://doi.org/10.1109/TPAMI.2007.70714).
- [24] MANN H B and WHITNEY D R. On a test of whether one of two random variables is stochastically larger than the other[J]. *The Annals of Mathematical Statistics*, 1947, 18(1): 50–60. doi: [10.1214/aoms/1177730491](https://doi.org/10.1214/aoms/1177730491).
- [25] 吴喜之, 赵博娟. 非参数统计[M]. 第4版, 北京: 中国统计出版社, 2013: 52–56.
- WU Xizhi and ZHAO Bojuan. Nonparametric Test[M]. 4th ed., Beijing: China Statistics Press, 2013: 52–56.
- [26] 王星, 褚挺进. 非参数统计[M]. 第2版, 北京: 清华大学出版社, 2014: 332–334.
- WANG Xing and CHU Tingjin. Non-Parametric Statistics[M]. 2nd ed., Beijing: Tsinghua University Press, 2014: 332–334.
- [27] HATA M. Empirical formula for propagation loss in land mobile radio services[J]. *IEEE Transactions on Vehicular Technology*, 1980, 29(3): 317–325. doi: [10.1109/TVT.1980.23859](https://doi.org/10.1109/TVT.1980.23859).
- [28] RIRI P C, KRISTALINA P, and SUDARSONO A. Cluster-based pathloss exponential modeling for indoor positioning in wireless sensor network[C]. 2016 International Conference on Knowledge Creation and Intelligent Computing, Manado, Indonesia, 2016: 53–59. doi: [10.1109/KCIC.2016.7883625](https://doi.org/10.1109/KCIC.2016.7883625).
- [29] ZHOU Mu, ZHANG Qiao, TIAN Zengshan, *et al.* IMLours: Indoor mapping and localization using time-stamped WLAN received signal strength[C]. 2015 IEEE Wireless Communications and Networking Conference, New Orleans, USA, 2015: 1817–1822. doi: [10.1109/WCNC.2015.712](https://doi.org/10.1109/WCNC.2015.712).

- 7744.
- [30] 刘东海, 陆丽宇. 小概率事件与实际问题的例谈[J]. 科技经济导刊, 2018, 26(12): 114.
- LIU Donghai and LU Liyu. Small probability events and examples of practical problems[J]. *Technology and Economic Guide*, 2018, 26(12): 114.
- [31] 肖莉. 基于路径损耗的WiFi室内定位系统[D]. [硕士论文], 西安电子科技大学, 2017: 19–31.
- XIAO Li. The WiFi indoor positioning system based on path loss model[D]. [Master dissertation], Xidian University, 2017: 19–31.
- [32] 张文君, 顾行发, 陈良富, 等. 基于均值-标准差的 K 均值初始聚类中心选取算法[J]. 遥感学报, 2006, 10(5): 715–721. doi: [10.3321/j.issn:1007-4619.2006.05.017](https://doi.org/10.3321/j.issn:1007-4619.2006.05.017).
- ZHANG Wenjun, GU Xingfa, CHEN Liangfu, *et al.* An algorithm for initializing of K -means clustering based on mean-standard deviation[J]. *Journal of Remote Sensing*, 2006, 10(5): 715–721. doi: [10.3321/j.issn:1007-4619.2006.05.017](https://doi.org/10.3321/j.issn:1007-4619.2006.05.017).
- 周 牧: 男, 1984年生, 教授, 研究方向为无线定位技术、机器学习与人工智能、凸优化理论.
- 王烟濛: 女, 1994年生, 硕士生, 研究方向为室内WLAN定位技术、定位性能评估技术.
- 袁 慧: 女, 1994年生, 硕士, 研究方向为室内Wi-Fi定位技术、定位网络优化.
- 田增山: 男, 1968年生, 教授, 博士生导师, 研究方向为移动通信、个人通信、GPS及蜂窝网定位技术.