

基于接收信号强度非齐性分布特征的半监督学习室内定位指纹库构建

李世宝* 王升志 刘建航 黄庭培 张鑫

(中国石油大学(华东)计算机与通信工程学院 青岛 266580)

摘要: 室内定位中半监督学习的指纹库构建方法能够降低人力开销,但忽略了高维接收信号强度(RSS)数据不均匀的非齐分布特点,影响定位精度,针对此问题该文提出一种基于RSS非齐性分布特征的半监督流形对齐指纹库构建方法。该算法运用局部RSS尺度参数以及共享近邻相似性构造权重矩阵,得到精确反映RSS数据流形结构的权重图,利用该权重图通过求解流形对齐的目标函数最优解,实现运用少量标记数据对大量未标记数据的位置标定。实验结果表明,该算法可以显著降低离线阶段数据采集的工作量,同时可以取得较高的定位精度。

关键词: 无线局域网;室内指纹定位;半监督流形对齐;非齐性分布;指纹库构建

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2019)10-2302-08

DOI: [10.11999/JEIT180599](https://doi.org/10.11999/JEIT180599)

Semi-supervised Indoor Fingerprint Database Construction Method Based on the Nonhomogeneous Distribution Characteristic of Received Signal Strength

LI Shibao WANG Shengzhi LIU Jianhang HUANG Tingpei ZHANG Xin

(College of Computer and Communication Engineering, China University of
Petroleum (East China), Qingdao 266580, China)

Abstract: The radio map construction is time consuming and labor intensive, and the conventional semi-supervised based methods usually ignore the influence of the uneven distribution of high-dimensional Received Signal Strength (RSS). In order to solve that problem, a semi-supervised radio map construction approach which is based on the nonhomogeneous distribution characteristic of RSS is proposed. The approach utilizes the RSS local scale and common neighbors similarities to calculate the weighted matrix. Thus, the weighted graph that reflects accurately the structure of RSS data manifold is presented. In addition, the weighted graph is used to find the optimal solution of the objective function to calibrate the locations of plenty of unlabeled data by a small number of labeled RSS. The extensive experiments demonstrate that the proposed method is capable of not only constructing an accurate radio map at a low manual cost, but also achieving a high localization accuracy.

Key words: Wireless Local Area Network (WLAN); Indoor fingerprinting localization; Semi-supervised manifold alignment; Nonhomogeneous distribution; Radio map construction

1 引言

随着移动互联网的发展以及智能手机的普及,基于位置的服务(Location Based Service, LBS)业务需求不断增长,室内无线定位技术近年来得到极

大的关注^[1]。其中,基于无线局域网(Wireless Local Area Network, WLAN)的指纹定位法^[2]与基于到达时间(或时间差)定位、到达角度(或角度差)定位、基于传播模型的定位方法相比,该方法对传输带宽要求低而且不需要额外的硬件部署,同时具有较高的定位精度,因而得到广泛的应用。

WLAN位置指纹定位技术主要分为离线数据采集以及在线位置解算两个阶段^[3]。离线阶段,在待定位区域内按一定规则选择若干参考点(Reference Point, RP),遍历每个参考点进行信号采集,将接收信号强度(Received Signal Strength, RSS)向量以及参考点的地理位置信息存储到指纹数据库中,完成指纹库(Radio Map, RM)的构建。由于无线信

收稿日期: 2018-06-20; 改回日期: 2019-02-28; 网络出版: 2019-03-30

*通信作者: 李世宝 lishibao@upc.edu.cn

基金项目: 国家自然科学基金(61972417, 61601519, 61872385), 中央高校基本科研业务费专项资金(18CX02134A, 18CX02137A, 18CX02133A, 19CX05003A-4)

Foundation Items: The National Natural Science Foundation of China (61972417, 61601519, 61872385), The Fundamental Research Funds for the Central Universities (18CX02134A, 18CX02137A, 18CX02133A, 19CX05003A-4)

号传播复杂,易受到多径效应、阴影效应等的影响,在某个参考点上接收到的同一个接入点(Access Point, AP)的RSS信号通常表现出复杂的时变特性,因此需要在参考点重复采集数据以提高指纹库的精度。在线阶段,将采集到的RSS信号通过模式匹配算法与指纹库中的信号进行匹配,推算用户的估计位置。由此可见,离线阶段需要采集大量数据进行训练,增加了人力成本和时间消耗,限制了指纹定位法的普及。因此,如何有效减少指纹构建的工作量,保证定位精度,得到了越来越多研究者的关注。

为了降低指纹库构建的消耗,研究人员提出了基于模型的构建方法^[4]、基于半监督流形学习的指纹库构建方法^[5-7]。其中,基于传播模型的方法通常需要已知AP的具体位置,由于多径效应和室内环境变化的影响,此类方法一般无法准确构建指纹库。由于指纹库构建过程中,大量无标记RSS数据可以通过随机路径轻易采集,基于这种优势,文献^[5]运用半监督学习的思想,通过少量已标记的指纹数据的位置信息对大量无标记数据的坐标位置进行标记,进而实现融合指纹库的构建。

半监督学习中的流形学习方法不仅利用RSS值代表的距离度量进行定位,而且考虑RSS数据的流形几何结构特征,同时,可以充分运用标记RSS数据的先验信息,提取非线性定位特征用于定位,因而具有较好的精度,文献^[6]运用流形学习的方法,引入坐标数据集和RSS数据集,通过两个数据集中部分数据的对应关系,通过低维特征向量的相似性得到未标记数据的坐标,但是这种方法对坐标数据集要求较高,需要准确体现室内布局结构。文献^[7]提出基于图的混合半监督流形对齐的指纹库构建方法,通过RSS数据进行拉普拉斯图的构建,结合未标记数据的时间戳信息增强高维RSS的相似性,通过流形对齐标定未标记数据完成指纹库的构建,显著降低了指纹库构建开销,一定程度上保证了定位精度。

然而,当今定位应用场景一般具有复杂的室内

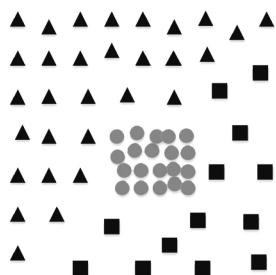
布局结构,而且AP布置位置多样,这种特点导致RSS数据的分布疏密程度不同,即不同参考点数据呈现非齐性的分布特征。基于图的流形对齐方法中权重图的精度将影响算法的性能,此类方法通常采用k近邻图(k-Nearest Neighbor graph, k-NN graph)中的欧式距离构造权重矩阵^[7,8],而通常高维空间中位于同一流形上且疏密程度类似的数据具有较高相似性,仅利用欧式距离大小没有考虑RSS非齐性分布体现的相关性特征,无法准确反映高维RSS数据复杂的近邻关系^[9,10],不利于流形对齐半监督学习过程中对未标记数据的位置标定,进而影响室内定位的性能。针对上述问题,本文提出基于RSS非齐性分布特征的流形对齐指纹库构建算法,通过标记RSS数据和大量用户轨迹数据的RSS非齐性分布特征构造权重图,并结合共同近邻相似性调整高维相似RSS数据在权重图中的连接关系,建立拉普拉斯图并以此通过半监督学习过程实现对用户轨迹数据的位置标定,从而保证在较低开销的条件下,实现精确指纹库的构建。

2 RSS数据非齐性分布特征及相关性分析

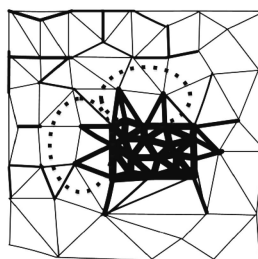
2.1 RSS数据非齐性特征描述

室内定位的应用场景趋于复杂化,此类场景AP数量多、无线传播环境多变。在定位过程中RSS向量的维数即AP个数,这表明最终指纹库中存储的都是高维RSS数据,而高维数据通常具有复杂的分布特征。RSS非齐性特征用于表征RSS数据向量在高维数据空间分布的不均匀性,具体表现为在相同的距离范围内,处于密集区域内的RSS向量具有更多的近邻点,相类似地,数据稀疏区域,近邻点之间距离有较大的跨度。

RSS数据分布非齐性示意图如图1,图中不同的点表示不同类的RSS向量,点之间连线的粗细表示在权重图中点的相似性大小,图1(a)表示高维空间中分布不均匀的3类RSS数据点,同类RSS向量之间相对非同类RSS应该具有更强的连接。图1(b)表示传统方法下的权重图建立结果示意图,如图中



(a) RSS数据分布示意图



(b) RSS数据不均匀对权重图构建的影响

图1 RSS非齐性分布示意图

虚线所示,仅考虑距离度量而忽略RSS非齐性特征会导致不同类别的RSS数据有较紧密地连接,进而会影响构建的权重图精度。

2.2 RSS非齐性分布特征数据分析

本小节将分别用数据疏密程度分析方法和数据可视化算法线性判别分析^[11](Linear Discriminant Analysis, LDA)对复杂场景中的RSS数据进行分析,得到高维RSS数据不均匀分布的非齐性特征,进而挖掘其中隐含的数据相关性。

根据文献^[12]提出的数据疏密程度参数进行分析, ρ_i 计算如式(1)

$$\rho_i = \sum_j \chi(d_{ij} - d_c) \quad (1)$$

式中,当 $x < 0$ 时, $\chi(x) = 1$, 在其它情况下, $\chi(x) = 0$ 。 d_c 表示截断半径,与数据点 i 距离小于 d_c 的数据个数表示数据疏密程度 ρ_i 。

分析TUT数据集^[13]、UJIIndoorLoc数据集^[14]以及本文办公楼定位场景中RSS指纹数据,计算得到数据疏密的分布如图2所示,3个数据集的4分位距均超过10,表明RSS数据的疏密分布均有较大跨度。图3表示在TUT数据集的2楼采集129个参考点,不同参考点处的数据疏密程度。可见,不同参考点数据的疏密程度参数差异较大,整体RSS数据呈现非齐性的分布特征,但是在一定参考点区域内,多个RSS数据具有相似的疏密程度参数,RSS数据非齐性分布的疏密特征隐含着高维RSS数据的相关性。

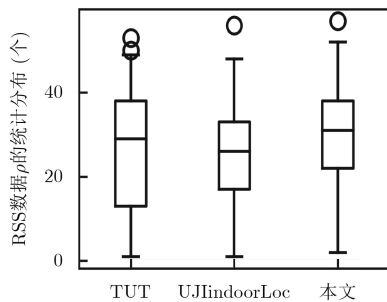


图2 3个数据集RSS数据疏密分布

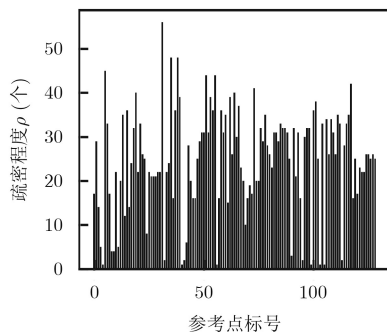


图3 TUT数据集各参考点RSS疏密程度

进一步地,分析高维RSS数据的可视化2维投影数据的分布特点。通过LDA算法对UJIIndoorLoc数据集2楼3个房间的定位数据进行处理,LDA算法是有监督的常规Fisher线性判别分析算法,对不同类别的数据具有保持能力,该算法在降维过程中通过保持类内方差最小、类间方差最大得到高维数据在低维空间的投影。其中,类内散度 S_W 与类间散度 S_B 定义为

$$S_B = \frac{1}{N} \sum_{i=1}^N (\mathbf{r}_i - \bar{\mathbf{r}}) (\mathbf{r}_i - \bar{\mathbf{r}})^T \quad (2)$$

$$S_W = \frac{1}{P} \sum_{i=1}^N \sum_{j=1}^n (\mathbf{r}_i(j) - \mathbf{r}_i) (\mathbf{r}_i(j) - \mathbf{r}_i)^T \quad (3)$$

式(2),式(3)中 N 表示参考点个数, P 表示所有的RSS向量数量, $\mathbf{r}_i$ 和 $\bar{\mathbf{r}}$ 分别代表第 i 个指纹点的RSS均值和所有RSS向量的均值。 $\mathbf{r}_i(j)$ 表示参考点 i 的第 j 个RSS值。

图4是高维RSS数据的低维投影结果。RSS数据分为明显的3簇,而且不同簇的数据在低维投影中呈现明显的分布不均匀现象。房间1和房间3的RSS数据紧密,房间2数据稀疏。数据非齐性分布特征体现了簇内相似RSS数据的相关性,同一簇内相似性高的数据具有更相近的疏密程度。

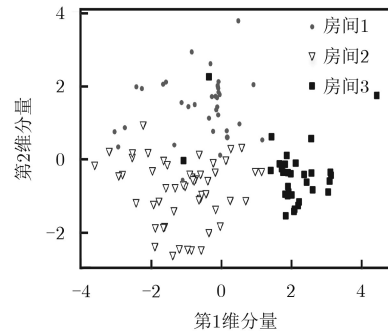


图4 3个房间的RSS数据在LDA下2维空间分布

综上所述,定位场景中的RSS数据在高维空间与低维投影空间中均呈现出明显的非齐性分布特征。RSS数据不均匀分布现象在多个定位场景中普遍存在,非齐相关性体现了相似RSS数据的固有特征,本文旨在设计基于RSS数据非齐性分布相关性的权重图构造算法,提高流形对齐过程中未标记RSS数据位置标定的准确性。

3 基于RSS非齐性分布特征的流形对齐指纹库构建

本文提出一种基于RSS非齐性分布特征的半监

督流形对齐室内定位指纹库构建算法。在离线指纹采集阶段,选定定位区域设置参考点数目为 N ,多条用户轨迹采集RSS数量为 M ,AP个数为 n ,采集的所有RSS表示为 $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{N+M})^T$,其中, $\mathbf{r}_i = (r_{i1}, r_{i2}, \dots, r_{in})$, $i \in [1, N+M]$,表示在 i 点的所有AP的RSS数据。 $\mathbf{R}^f = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)^T$ 表示固定参考点采集的RSS数据, $\mathbf{R}^u = (\mathbf{r}_{N+1}, \mathbf{r}_{N+2}, \dots, \mathbf{r}_M)^T$ 表示随机采集的用户轨迹的RSS数据。位置坐标表示为 $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{N+M})^T$,其中, $\mathbf{P}^f = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N)^T$ 表示与 $\mathbf{R}^f$ 相对应指纹点的位置坐标, $\mathbf{P}^u$ 表示未标记数据的初始坐标,初值设为任意值。

首先,运用标记数据 $\mathbf{R}^f$ 和未标记数据 $\mathbf{R}^u$ 组成混合数据矩阵 $\mathbf{R}$,根据RSS数据非齐性分布特征,利用尺度参数构造自适应数据分布疏密程度的近邻权重图矩阵 $\mathbf{W}_d$,并通过共同近邻相似性进行调整,得到RSS流形的几何结构模型,即权重图 $\mathbf{G} = \{\mathbf{R}, \mathbf{W}_d\}$ 。其次,根据 $\mathbf{W}_d$ 构建拉普拉斯图,通过流形对齐得到未标记数据的位置标定,结合指纹数据完成流形指纹库构建。最后,通过传统定位匹配方法KNN实现用户位置估计。

3.1 基于RSS非齐性分布特征的权重图构造

对在定位场景中采集的混合离线RSS数据矩阵 $\mathbf{R}$ 进行如下的权重图构造:

(1) 局部数据尺度相似矩阵构造

根据第2节所述,定位场景中的RSS数据分布具有非齐性的特征,通过 $\mathbf{rss}_i$ 邻居点的距离统计特征计算反映RSS数据非齐性特征的局部尺度^[15]因子,令局部尺度参数 δ_i 为

$$\begin{aligned} \delta_i &= \frac{1}{k} \sum_{m=1}^k d_s(\mathbf{rss}_i, \mathbf{rss}_m) \\ &= \frac{1}{k} \sum_{m=1}^k \left(\frac{\sum_{p=1}^n |\mathbf{rss}_{ip} - \mathbf{rss}_{mp}|}{\sum_{p=1}^n (\mathbf{rss}_{ip} + \mathbf{rss}_{mp})} \right), \\ &\quad \mathbf{rss}_m \in \text{NN}_k(\mathbf{rss}_i) \end{aligned} \quad (4)$$

其中, δ_i 表示 $\mathbf{rss}_i$ 与其 k 个近邻(Nearest-Neighbor)点的平均sorensen距离^[16], $\text{NN}_k(\mathbf{rss}_i)$ 代表 $\mathbf{rss}_i$ 的 k 个近邻点集合, $k \in \mathbb{N}^+$ 。考虑局部数据尺度的影响, $\mathbf{rss}_i$ 到 $\mathbf{rss}_j$ 的距离调整为 $d_{ij} = d_s(\mathbf{rss}_i, \mathbf{rss}_j) / \delta_i$,相反, $\mathbf{rss}_j$ 到 $\mathbf{rss}_i$ 的距离表示为 $d_{ji} = d_s(\mathbf{rss}_j, \mathbf{rss}_i) / \delta_j$ 。一般地,均方距离计算为 $d_s^2(\mathbf{rss}_i, \mathbf{rss}_j) / \delta_i \delta_j$,则得到 $\mathbf{rss}_i$ 与 $\mathbf{rss}_j$ 间径向基函数表示的相似性大小为

$$A_{ij} = \begin{cases} \exp\left(-\frac{d_s^2(\mathbf{rss}_i, \mathbf{rss}_j)}{\alpha \delta_i \delta_j}\right), \\ \mathbf{rss}_i \in \text{NN}_k(\mathbf{rss}_j) \text{ 或 } \mathbf{rss}_j \in \text{NN}_k(\mathbf{rss}_i) \\ 0, \text{ 其它} \end{cases} \quad (5)$$

其中, α 是径向基函数的调整因子。当调整之后的距离 d_{ij} 与 d_{ji} 均较小的情况下,式(5)取得较大的值,即只有 d_s 相对于 δ_i 和 δ_j 较小时, $\mathbf{rss}_i$ 与 $\mathbf{rss}_j$ 才更相似,保证了高维空间中相近的 $\mathbf{rss}_i$ 与 $\mathbf{rss}_j$ 具有相同量级的尺度与疏密程度。

(2) 共同近邻相似性

高维空间中相近的RSS向量有更多的共同邻居点,为了进一步防止高维空间中不相似RSS向量间的过度连接,计算共同邻居相似性为

$$S(\mathbf{rss}_i, \mathbf{rss}_j) = |\text{NN}_k(\mathbf{rss}_i) \cap \text{NN}_k(\mathbf{rss}_j)| \quad (6)$$

式(6)表示 $\mathbf{rss}_i$ 与 $\mathbf{rss}_j$ 的 k 个近邻点集合的交集。设 $\text{NN}_k(\mathbf{rss}_i)$ 中的近邻点是按与 $\mathbf{rss}_i$ 距离大小升序排列,则存在 $\mathbf{rss}_i$ 与 $\mathbf{rss}_j$ 两点之间的共同近邻点的个数与 $\mathbf{rss}_i$, $\mathbf{rss}_m$ 之间的共同近邻点个数相同,但是共同近邻点的排序不同的情况。考虑RSS数据的排序体现的相似性等级,得到基于等级的共同近邻相似性为

$$\begin{aligned} \hat{S}(\mathbf{rss}_i, \mathbf{rss}_j) &= \sum_{\mathbf{rss}_x} (k - m + 1)(k - n + 1), \\ \mathbf{rss}_x &\in \{\text{NN}_k(\mathbf{rss}_i) \cap \text{NN}_k(\mathbf{rss}_j)\} \end{aligned} \quad (7)$$

其中, m 和 n 分别为共同近邻点在集合 $\text{NN}_k(\mathbf{rss}_i)$, $\text{NN}_k(\mathbf{rss}_j)$ 中的位置。将式(7)归一化之后代入式(5),得到 $\mathbf{rss}_i$ 与 $\mathbf{rss}_j$ 权重大小为

$$W_{ij} = \begin{cases} \exp\left(-\frac{d_s^2(\mathbf{rss}_i, \mathbf{rss}_j)}{\hat{S}(\mathbf{rss}_i, \mathbf{rss}_j) \delta_i \delta_j}\right), \\ \mathbf{rss}_i \in \text{NN}_k(\mathbf{rss}_j) \text{ 或 } \mathbf{rss}_j \in \text{NN}_k(\mathbf{rss}_i) \\ 0, \text{ 其它} \end{cases} \quad (8)$$

最终,得到自适应局部疏密程度的权重图 $\mathbf{G} = \{\mathbf{R}, \mathbf{W}_d\}$,其中,权重矩阵定义为 $\mathbf{W}_d = (\mathbf{W}_{ij})_{(N+M) \times (N+M)}$ 。

上述权重矩阵优化算法的示意图如图5, x , y , z 表示3个RSS向量点,均有 $d_s(x, y) = d_s(x, z)$ 。图5(a)中,两簇RSS数据具有相近的疏密程度,由于 x 与 y 有更多的共同邻居,根据式(8)求得 x 与 y 连接权重更大。图5(b)中两簇疏密分布差异较大的RSS数据 $\delta_z < \delta_x$ 且 $\delta_x \approx \delta_y$, x 与 z 相对尺度参数差异较大,根据式(8)求得 x , z 的连接权重更小,得到 x 与 z 之间的连接较弱。经过式(8)优化调整,得到相似性结果更符合数据分布特点。

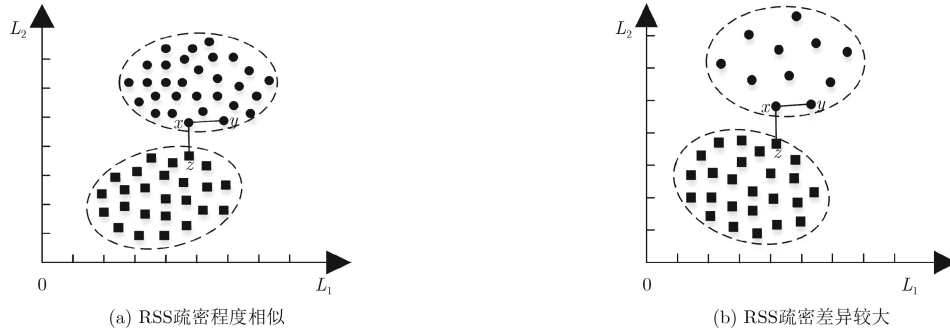


图5 算法示意图

3.2 基于半监督学习流形对齐的指纹库构建算法

离线阶段采集的高维RSS数据 $\mathbf{R}$ 的低维映射位置坐标设为 $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_{N+M})$, 为了在保持上述流形结构 $\mathbf{G} = \{\mathbf{R}, \mathbf{W}_d\}$ 的前提下, 求解未标记RSS数据 $\mathbf{R}^u$ 的低维映射坐标, 本文引入拉普拉斯特征映射^[17]作为流形对齐^[18]过程中的降维算法, 通过优化 $\sum_{i,j}^{N+M} \|\mathbf{c}_i - \mathbf{c}_j\|^2 \mathbf{W}_d$ 得到最优映射, 存在

$$\begin{aligned} \sum_{i,j}^{N+M} \|\mathbf{c}_i - \mathbf{c}_j\|^2 \mathbf{W}_d &= \sum_{i,j}^{N+M} (\mathbf{c}_i^T \mathbf{c}_i - 2\mathbf{c}_i^T \mathbf{c}_j + \mathbf{c}_j^T \mathbf{c}_j) \mathbf{W}_d \\ &= 2 \sum_i^{N+M} D_{ii} \mathbf{c}_i^T \mathbf{c}_i - 2 \sum_{i,j}^{N+M} \mathbf{c}_j^T \mathbf{c}_j \mathbf{W}_d \\ &= 2\text{tr}(\mathbf{C} \mathbf{D}_d \mathbf{C}^T) - 2\text{tr}(\mathbf{C} \mathbf{W}_d \mathbf{C}^T) \\ &= 2\text{tr}[\mathbf{C} (\mathbf{D}_d - \mathbf{W}_d) \mathbf{C}^T] \\ &= 2\text{tr}(\mathbf{C} \mathbf{L}_d \mathbf{C}^T) \end{aligned} \quad (9)$$

其中, $\text{tr}(\cdot)$ 表示矩阵的迹, $\mathbf{L}_d = \mathbf{D}_d - \mathbf{W}_d$, 对角阵 $\mathbf{D}_d = \sum_{j=1}^{N+M} \mathbf{W}_d(i, j)$, $\mathbf{L}_d$ 是对称的半正定矩阵。

在RSS数据流形中, 已知子集 $\mathbf{R}^f$ 先验的低维流形坐标, 即已知 $\mathbf{R}^f$ 与 $\mathbf{P}^f$ 的数据对应信息。进一步, 可以构造式(10)的凸可微的目标函数

$$F(\mathbf{C}) = \arg \min_{\mathbf{C}} \sum_{i=1}^N \mu \|\mathbf{c}_i - \mathbf{p}_i\|^2 + \mathbf{C}^T \mathbf{L}_d \mathbf{C} \quad (10)$$

其中, 第1项表示指纹数据与其低维流形坐标的映

射误差, 是均方损失函数。第2项是正则化流形约束项, μ 是权衡因子。

由于 $\frac{\partial \mathbf{C}^T \mathbf{L}_d \mathbf{C}}{\partial \mathbf{C}} = (\mathbf{L}_d + \mathbf{L}_d^T) \mathbf{C}$, 上述2次问题运用标准线性代数的解法可以直接求解, 对式(10)右侧求导可以得到式(11)的最优解

$$\mathbf{C} = \begin{pmatrix} \mu \mathbf{I}_k + \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{pmatrix}^{-1} \begin{pmatrix} \mu \mathbf{I}_k \\ 0 \end{pmatrix} \mathbf{P} \quad (11)$$

其中, $\mathbf{L} = \begin{pmatrix} \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{pmatrix}$ 表示联合拉普拉斯矩阵,

$\mathbf{W}_d = \begin{pmatrix} \mathbf{W}_{ll} & \mathbf{W}_{lu} \\ \mathbf{W}_{ul} & \mathbf{W}_{uu} \end{pmatrix}$ 表示近邻权重矩阵, 前 l 行 l 列表示标记RSS数据之间的近邻权重, 最后 u 行 u 列表示随机路径上采集RSS数据间的邻居连接权重, $\mathbf{W}_{lu}$ 表示标记RSS数据与随机路径上未标记RSS数据的邻居权重。 $\mathbf{I}_k = \text{diag}(\varphi_1, \varphi_2, \dots, \varphi_{N+M})$ 是一个指示对角矩阵, 当 rss_i 坐标已知时, $\varphi_i = 1$, 其他情况下, $\varphi_i = 0$ 。最终, 求得映射矩阵 $\mathbf{C}$ 的后 M 行为未标记RSS数据的低维映射坐标。

4 实验测试及性能分析

4.1 实验环境和数据采集

为验证本文提出的指纹库构建算法的性能, 实验场景设于某科研楼17楼, 平面图如图6, 该场景为典型的室内定位场景, 有多个办公室、实验室、走廊等, 区域大小为70.2 m × 18 m, 阴影区域为

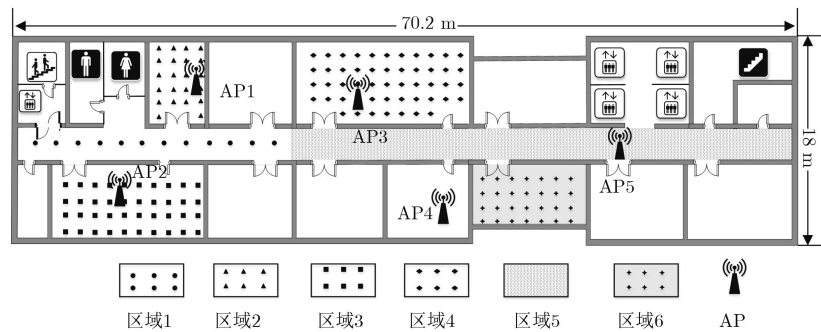


图6 实验环境平面图

选取的6个定位区域。实验选用5个型号为TP-LINK-WR886N的AP, 使用搭载Android 7.1操作系统的小米5X进行指纹采集, 考虑到智能手机wifi模块的性能, 设定RSS数据采集频率为1 Hz。实验选取412个指纹点, 指纹密度约为每平方米1.4个指纹, 同时采集随机路径上1145个零散RSS数据作为未标记数据。

4.2 定位性能分析

图7表示本文所提算法构建的指纹库, 与基于局部线性嵌入的流形对齐算法^[6](Locally Linear Embedding Manifold Alignment, LLE-MA)、混合半监督流形对齐定位算法LE-MA^[7]以及传统方法RADAR^[19]构建的指纹库定位误差的累积概率分布图。当定位误差为4.0 m时, 本文方法定位误差累积概率为77.5%, LLE-MA, LE-MA, RADAR算法的累积概率分别为64.0%, 71.7%, 62.5%, 相对上述3种算法, 精度分别提升13.5%, 5.8%, 15.0%, 由此可知, 本文所提算法具有更优的定位性能。主要原因是在复杂的定位场景中, 本文方法利用RSS数据非齐性分布的数据疏密相关性特征构建拉普拉斯图, 充分挖掘高维RSS数据的相似性近邻关系。传统流形算法主要运用欧式距离构建权重矩阵, 受数据疏密程度影响较大。

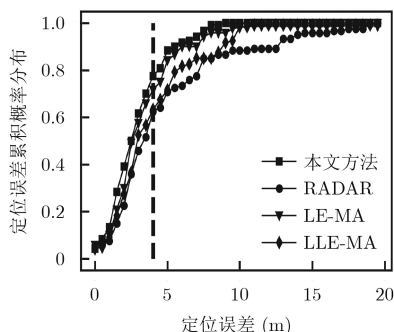


图7 不同算法定位性能

图8表示采集指纹减少一定百分比情况下, 定位精度相对于使用全部指纹定位的精度下降比率。当指纹减少80%, KNN精度下降103%, LLE-MA精度下降50%, LE-MA和本文所提算法分别下降33%, 31%。相对于传统建指纹库的方法, 本文算法可以在保持精度的情况下大大降低指纹点的采集数量, 相较于LE-MA本文算法对指纹比例下降有较好的鲁棒性。

不同比例指纹数下的构建时间消耗如图9。按照传统的指纹库构建方式, 为了消除信号波动的影响需要在每个参考点4个方向采集指纹, 通常按1 Hz每个方向采集90次, 所以采集412个指纹点共

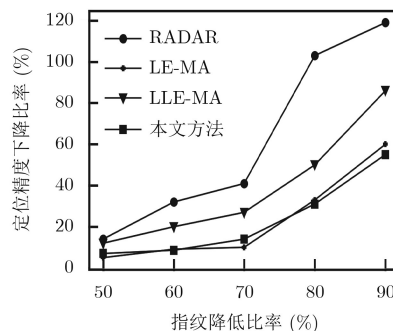


图8 不同比例指纹数下的定位精度下降比率

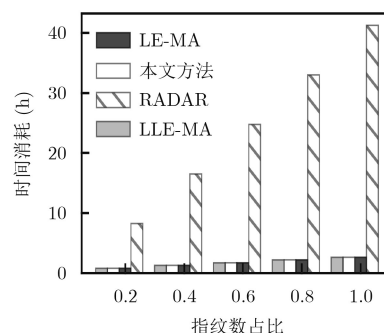


图9 不同算法构建指纹库时间消耗

需要41.2 h。本文所提的指纹库构建算法, 在使用20%指纹点, 每个指纹点仅采集20个RSS样本情况下精度已经较理想, 此时, 构建的时间消耗约为0.8 h, 大大减少了构建时间消耗。同时, 本文算法进行RSS数据标记的运算时间不足1 min, 相对于指纹库构建的时间消耗可以忽略不计。

AP数量是影响指纹定位算法性能的一个重要因素, 图10是定位误差为4 m时, 不同算法定位精度与AP数量的关系。由图可以得出, 当AP数量增加, 上述4种算法的定位精度均升高, 且本文算法保持相对较优的定位性能。当AP数量持续增加的情况下, 引入更多定位信息的同时带来了更多的干扰, 定位精度的变化趋于平缓。同时, 更多的AP带来了更多的运算消耗, 增加了计算复杂度, 因此, 本文选取5个AP进行上述定位算法的性能分析。

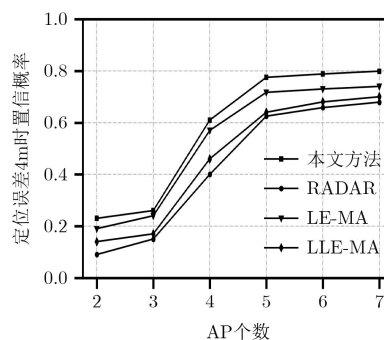


图10 AP个数对定位精度的影响

4.3 参数对定位精度影响分析

图11表示共同近邻相似性中邻居数 k 对定位精度的影响。如图所示,当选择邻居个数为13时,定位的平均误差为2.83 m取得较好的定位性能,当近邻数较小时,导致实际相似数据计算得到的相似性偏小,无法准确体现RSS数据的相关性特征,误差较大。当近邻数增大到19,定位误差增大至3.22 m。原因是近邻数增大, RSS向量间会有更多的共同邻居,共同近邻相似性偏大,权重矩阵对RSS数据疏密程度的体现下降,导致RSS数据的近邻点距离跨度较大,近邻点连接的准确率降低,影响最终指纹库构建精度。因此,本文中選擇 $k=13$ 。

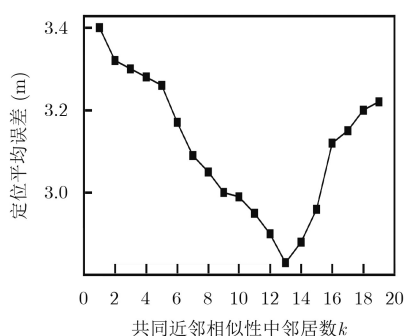


图 11 共同近邻相似性中邻居数对定位性能影响

5 结束语

本文所提基于RSS非齐性分布特征的半监督学习室内定位指纹库构建算法,首先分析了复杂定位场景中RSS数据疏密不均匀的非齐性分布特征,然后利用混合RSS数据中反映非齐性特征的局部尺度参数以及共同近邻相似性优化权重图,最后通过流形对齐挖掘RSS权重图中的近邻相似性,实现指纹库的精确构建。实验表明,本文所提算法所构建的指纹库定位性能优于传统方法以及其他流形算法,同时保证了较低的指纹库构建开销。下一步工作将深入探索指纹库构建的流形对齐过程中,其他流形学习算法(比如Hessian映射)对输入RSS数据流形的局部几何特征的提取性能,进一步提升流形指纹库的精度和稳定性。

参考文献

- [1] MELAMED R. Indoor localization: Challenges and opportunities[C]. 2016 IEEE/ACM International Conference on Mobile Software Engineering and Systems, Austin, USA, 2016: 1–2.
- [2] 徐玉滨, 邓志安, 马琳. 基于核直接判别分析和支持向量回归的WLAN室内定位算法[J]. 电子与信息学报, 2011, 33(4): 896–901. doi: [10.3724/SP.J.1146.2010.00813](https://doi.org/10.3724/SP.J.1146.2010.00813).
- [3] KHALAJMEHRABADI A, GATSIS N, and AKOPIAN D. Modern WLAN fingerprinting indoor positioning methods and deployment challenges[J]. *IEEE Communications Surveys & Tutorials*, 2017, 19(3): 1974–2002. doi: [10.1109/COMST.2017.2671454](https://doi.org/10.1109/COMST.2017.2671454).
- [4] JI Yiming, BIAZ S, PANDEY S, et al. ARIADNE: A dynamic indoor signal map construction and localization system[C]. The 4th International Conference on Mobile Systems, Applications and Services, Uppsala, Sweden, 2006: 151–164.
- [5] OUYANG R W, WONG K K S, LEA C T, et al. Indoor location estimation with reduced calibration exploiting unlabeled data via hybrid generative/discriminative learning[J]. *IEEE Transactions on Mobile Computing*, 2012, 11(11): 1613–1626. doi: [10.1109/TMC.2011.193](https://doi.org/10.1109/TMC.2011.193).
- [6] SOROUR S, LOSTANLEN Y, VALAEE S, et al. Joint indoor localization and radio map construction with limited deployment load[J]. *IEEE Transactions on Mobile Computing*, 2015, 14(5): 1031–1043. doi: [10.1109/TMC.2014.2343636](https://doi.org/10.1109/TMC.2014.2343636).
- [7] ZHOU Mu, TANG Yunxia, TIAN Zengshan, et al. Semi-supervised learning for indoor hybrid fingerprint database calibration with low effort[J]. *IEEE Access*, 2017, 5: 4388–4400. doi: [10.1109/ACCESS.2017.2678603](https://doi.org/10.1109/ACCESS.2017.2678603).
- [8] WANG Jin, TAN N, LUO Jun, et al. WOLoc: WiFi-only outdoor localization using crowdsensed hotspot labels[C]. IEEE Conference on Computer Communications, Atlanta, GA, USA, 2017: 1–9.
- [9] CHAPELLE O and ZIEN A. Semi-supervised classification by low density separation[C]. The Tenth International Workshop on Artificial Intelligence and Statistics, Bridgetown, Barbados, 2005: 57–64.
- [10] STEVENS J R, RESMINI R G, and MESSINGER D W. Spectral-density-based graph construction techniques for hyperspectral image analysis[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(10): 5966–5983. doi: [10.1109/TGRS.2017.2718547](https://doi.org/10.1109/TGRS.2017.2718547).
- [11] ZHU Manli and MARTINEZ A M. Pruning noisy bases in discriminant analysis[J]. *IEEE Transactions on Neural Networks*, 2008, 19(1): 148–157. doi: [10.1109/TNN.2007.904040](https://doi.org/10.1109/TNN.2007.904040).
- [12] RODRIGUEZ A and LAIO A. Clustering by fast search and find of density peaks[J]. *Science*, 2014, 344(6191): 1492–1496. doi: [10.1126/science.1242072](https://doi.org/10.1126/science.1242072).
- [13] LOHAN E S, TORRES-SOSPEDRA J, LEPPÄKOSKI H, et al. Wi-Fi crowdsourced fingerprinting dataset for indoor positioning[J]. *Data*, 2017, 2(4): 32. doi: [10.3390/](https://doi.org/10.3390/)

- [data2040032](#).
- [14] TORRES-SOSPEDRA J, MONTOLIU R, MARTÍNEZ-USÓ A, *et al*. UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems[C]. 2014 International Conference on Indoor Positioning and Indoor Navigation, Busan, South Korea, 2014: 261–270.
- [15] ZELNIK-MANOR L and PERONA P. Self-tuning spectral clustering[C]. Advances in Neural Information Processing Systems, Vancouver, British Columbia, Canada, 2005: 1601–1608.
- [16] TORRES-SOSPEDRA J, MONTOLIU R, TRILLES S, *et al*. Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems[J]. *Expert Systems with Applications*, 2015, 42(23): 9263–9278. doi: [10.1016/j.eswa.2015.08.013](#).
- [17] BELKIN M and NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. *Neural Computation*, 2003, 15(6): 1373–1396. doi: [10.1162/089976603321780317](#).
- [18] HAM J, LEE D D and SAUL L K. Semisupervised alignment of manifolds[C]. The Tenth International Workshop on Artificial Intelligence and Statistics, Bridgetown, Barbados, 2005: 120–127.
- [19] BAHL P and PADMANABHAN V N. RADAR: An in-building RF-based user location and tracking system[C]. The Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Tel Aviv, Israel, 2000, 2: 775–784.
- 李世宝：男，1978年生，副教授，研究方向为移动计算、无线传感器网络、干扰对齐等。
- 王升志：男，1994年生，硕士生，研究方向为无线定位技术。
- 刘建航：男，1978年生，副教授、博士，研究方向为无线局域网、车联网。
- 黄庭培：女，1980年生，讲师、博士，研究方向为无线传感器网络。
- 张 鑫：男，1993年生，硕士生，研究方向为无线定位技术。