

多特征融合在半监督流形约束定位方法

钱政, 严亮, 孙顺远

(江南大学 物联网工程学院, 轻工先进控制教育部重点实验室, 江苏 无锡 214122)

摘要: 针对无线保真(WiFi)和低功耗蓝牙(BLE)指纹定位方法需要大量标记训练样本以及单模定位精度和稳定性难以满足大规模定位场景需求的问题, 提出一种融合WiFi和BLE信号的半监督流形约束定位方法. 实验结果表明: 该方法与单一特征相比, 每一维度归一化方差稳定在0.08以下, 定位精度约提高25个百分点; 使用分别构建流形约束的半监督学习方法时, 能使定位过程中所需标记样本数量减少约90%. 因此, 该方法能极大减少需标记的样本数量, 并有效提高定位的稳定性和精度.

关键词: 多特征融合; 半监督学习; 流形正则化; 无线保真(WiFi); 低功耗蓝牙

中图分类号: TP391 **文献标志码:** A **文章编号:** 1671-5489(2024)05-1219-09

Semi-supervised Manifold Constraint Localization Method with Multi-feature Fusion

QIAN Zheng, YAN Liang, SUN Shunyuan

(Key Laboratory of Advanced Process Control for Light Industry of Ministry of Education, School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, Jiangsu Province, China)

Abstract: Aiming at the problems that wireless fidelity (WiFi) and bluetooth low energy (BLE) fingerprint localization methods required a large number of labeled training samples and that the accuracy and stability of single-mode localization were difficult to meet the requirements of large-scale localization scenarios, we proposed a semi-supervised manifold constraint localization method that fused WiFi and BLE signals. The experimental results show that compared with a single feature, the normalized variance of each dimension of the proposed method is stable below 0.08, and the accuracy of localization is improved by about 25 percentage points. When the semi-supervised learning method is used to construct manifold constraints separately, the number of labeled samples required in the localization process can be reduced by about 90%. Therefore, this method can greatly reduce the number of required label samples, and effectively improve the stability and accuracy of localization.

Keywords: multi-feature fusion; semi-supervised learning; manifold regularization; wireless fidelity; bluetooth low energy

近年来, 随着基于智能终端的各种定位应用和集成定位服务的出现以及无线通信技术的普及, 使用智能手机进行室内实时定位备受关注. 作为大多数智能终端和设备集成的无线保真(WiFi)和低功耗蓝牙(BLE)无线模块因其合适的有效距离及完善的通信设施而被视为潜在的理想方法. 目前, 广泛应

收稿日期: 2023-03-03.

第一作者简介: 钱政(1999—), 男, 汉族, 硕士研究生, 从事室内定位的研究, E-mail: 6221913046@stu.jiangnan.edu.cn. 通信

作者简介: 孙顺远(1985—), 男, 汉族, 博士, 副教授, 从事无线传感网络与室内定位的研究, E-mail: hzrobin@stu.jiangnan.edu.cn.

基金项目: 国家自然科学基金(批准号: 61702228)和江苏省自然科学基金(批准号: BK20170198).

用的基于接收信号强度(received signal strength, RSS)的位置指纹定位方法通过采集来自多个无线传感器信标的 RSS 信号推断用户位置,但由于无线信号在复杂的室内环境中传输时易被干扰,导致 WiFi 和 BLE 技术的定位精度降低、稳定性减弱,虽然 BLE 的定位精度略高于 WiFi,设备功耗更低,但由于蓝牙设备普及程度较低,因此高密度地部署蓝牙设备会增加定位成本.在单一定位技术无法满足需要的情况下,多技术融合定位成为新的研究热点.

文献[1]提出了 WiFi 和蓝牙融合的定位算法,解决了单一定位算法精度不足的问题,利用欧氏向量距离作为相似度标准对实验场地划分的网格进行合并,定位时首先计算单模定位技术各自结果,然后在决策层加权融合两种定位结果,在一定程度上提升了定位精度.文献[2]提出了基于蓝牙和 WiFi 通信的混合室内定位方法,先根据蓝牙热点部署位置将定位区域划分为互不相交的子区域,然后在每个子区域各自建立自己的 WiFi 指纹库,利用缩减后的指纹库完成定位,但由于蓝牙信号的波动导致区域划分存在模糊地带,因此该部分区域定位精度不佳.文献[3]提出了一种融合蓝牙、WiFi 和音频的定位方案,首先使用 WiFi 指纹定位得到初步定位结果,然后利用蓝牙和音频定位技术对 WiFi 定位结果进行修正,但该方法需要大量声波定位设备,整体成本较高.文献[4]提出了根据蓝牙信号在空气中的传播模型先将 RSS 转换成待定位点与蓝牙信号发射点的距离,然后利用 RSS 相对于蓝牙设备的邻近性过滤原始 WiFi 指纹,得到新的 WiFi 指纹库再利用其进行定位计算,由于蓝牙近邻法确定的区域过于粗略且最终定位计算仅依靠 WiFi 数据,使定位效果也不太理想.上述使用多传感器信号进行定位的方法仅将传感器数据简单融合,没有有效提取异构设备间的关联特征.文献[5]采用了 WiFi、BLE、行人航迹推算(pedestrian dead reckoning, PDR)融合定位系统,先建立卷积神经网络(convolutional neural networks, CNN)分别训练 WiFi 和 BLE 定位模型,使用训练好的模型进行定位,再利用扩展 Kalman 滤波融合 WiFi、BLE、PDR 的位置信息,但其并未在训练模型前关联不同设备的特征.文献[6]对 WiFi 和蓝牙信号进行互相关特征提取,并采用半监督学习减少标定训练数据,与单一特征相比,融合特征在鲁棒性和定位精度上均有提升.文献[7]考虑到由于 WiFi 和蓝牙具有不同的传播特性和有效距离,对它们在半监督学习模型中分别构建流形正则化约束并融入到统一模型中,通过调整模型参数的方式结合了不同类型数据的优势,并使模型具备学习未标记样本的能力.

基于上述方法的优点,本文提出一种能减少采集样本时的标记数据量,同时提高定位精度的指纹定位方法,为加快模型训练速度,并能对未标记样本进行学习,引入了半监督极限学习机(semi-supervised extreme learning machine, SELM),与基于单信号的半监督流形方法不同,本文在半监督学习模型中同时考虑了 WiFi 和蓝牙信号以及两者的融合特征信号,提出了多特征融合的半监督极限学习机(multi-feature fusion Semi-supervised extreme learning machine, FFSELM)模型应用于实际定位问题.

1 数据处理与特征提取

基于 RSS 指纹定位的基本思想是通过采集信号接入点(access point, AP)的发射信号强度和位置坐标信息,利用指纹匹配或机器学习模型算法计算用户当前位置.其主要分为两个阶段:离线训练阶段和在线定位阶段.在离线阶段,从参考点(reference point, RP)处收集各 AP 的 RSS 数据,从而建立 RSS 指纹与对应位置之间的模型映射关系;在线阶段,将采集到的 RSS 实时信号作为该模型的输入即可计算实时位置坐标.根据指纹定位的系统框图如图 1 所示,其中 rss_m^m 表示 r^{th} RP 处采集到来自 m^{th} AP 的 n^{th} RSS 值, $\hat{r}ss_m^m$ 表示在线阶段从 m^{th} AP 处采集的 RSS 值.

1.1 数据预处理

由于现实环境中 WiFi 和蓝牙的信号易受多径传播、障碍物遮挡以及随距离增大而强度大规模衰减的影响,使手机会接收到易引起定位结果出现波动的异常值信号,因此,采集的指纹库数据需进行预处理.环境中信号近似服从正态分布,假设 RP 的总数为 r ,在每个参考点采集 n 次,记第 i 个 RP 处接收到的来自第 j 个 AP 的 n 次信号数据为 $(rss_{i1}^j, rss_{i2}^j, \dots, rss_{in}^j)$,则其均值 μ 和标准差 σ 的计算公式分别为

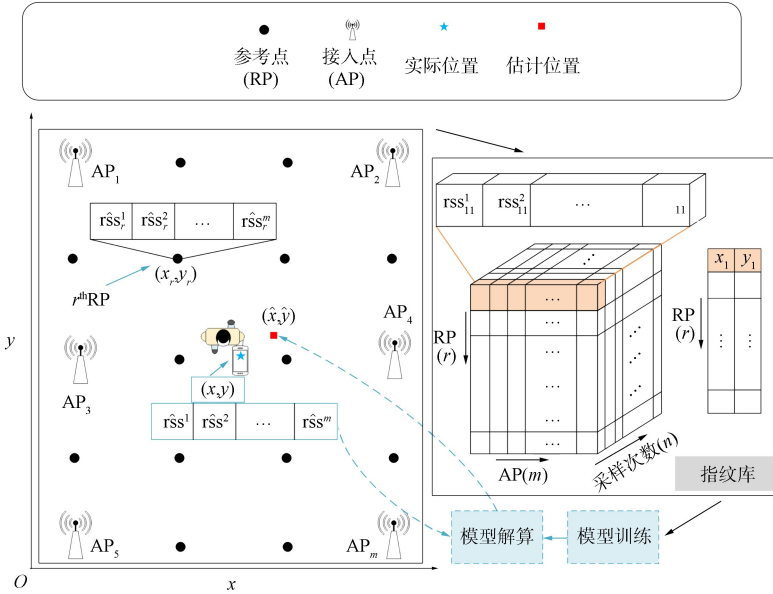


图 1 指纹定位系统框图

Fig. 1 Block diagram of fingerprint positioning system

$$\mu_i = \frac{1}{n} \sum_{k=1}^n rssi_{ik}^j, \quad \sigma_i = \sqrt{\frac{1}{n} \sum_{k=1}^n (rssi_{ik}^j - \mu_i)^2}, \tag{1}$$

概率密度函数为

$$f(rss) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left\{-\frac{(rss - \mu)^2}{2\sigma^2}\right\}. \tag{2}$$

通过计算同一 RP 处采集到来自相同 AP 的信号序列的均值 μ 和标准差 σ , 根据式(2)计算 $f(rss)$ 确定高概率区间, 对不在区间内的信号进行剔除, 完成数据预处理.

1.2 AP 优选

考虑到室内场景中广泛分布的 WiFi 设备可能来自于其他楼层或建筑物, 且同一局部区域内可能存在大量的 AP 节点, 它们提供的不稳定和冗余信号可能会对定位精度产生负面影响, 因此需进行剔除. 选择策略在离线阶段主要分两步进行: 先通过计算 AP 的信号稳定度排除随时间变化波动较大的不稳定节点; 然后计算 AP 之间的相似度剔除冗余的节点.

AP 的信号稳定度可通过计算在所有 RP 处采集到的 RSS 信号标准差的均值衡量. AP_j 在所有 RP 处的标准差均值为

$$\text{Std}(AP_j) = \frac{1}{r} \sum_{i=1}^r \sigma_i = \frac{1}{r} \sum_{i=1}^r \sqrt{\frac{1}{n} \sum_{k=1}^n (rssi_{ik}^j - \mu_i)^2}, \tag{3}$$

则信号稳定度可表示为

$$G(AP_j) = \left(\eta + \frac{1}{r} \sum_{i=1}^r \sqrt{\frac{1}{n} \sum_{k=1}^n (rssi_{ik}^j - \mu_i)^2} \right)^{-1}, \tag{4}$$

其中 η 为防止分母为零的极小正数. $G(AP_j)$ 的值越大, 说明该 AP 的信号稳定度越好, 通过设置阈值, 将信号稳定度小于该阈值的 AP 剔除.

引入 Pearson 相关系数计算 AP 间的相似度, 通过计算 AP 之间的线性相关性找出在各 RP 处信号强度相近的 AP 对, Pearson 相关系数计算公式为

$$\rho(a, b) = \frac{\sum_{i=1}^r (rssi_{a,i} - \overline{rssi_a})(rssi_{b,i} - \overline{rssi_b})}{\sqrt{\sum_{i=1}^r (rssi_{a,i} - \overline{rssi_a})^2} \sqrt{\sum_{i=1}^r (rssi_{b,i} - \overline{rssi_b})^2}}, \tag{5}$$

其中 a 和 b 分别表示两个 AP, $\text{rss}_{a,i}$ 表示 AP_a 在第 i 个 RP 处的信号强度, $\overline{\text{rss}_a}$ 和 $\overline{\text{rss}_b}$ 为参考点处接收信号强度的均值. $\rho(a, b)$ 的取值范围为 $[-1, 1]$, $\rho(a, b)$ 值越大, 表示两个 AP 之间的相似度越高. 利用式(5)计算所有 AP 之间的相似度, 当相似度介于 $0.8 \sim 1$ 间时可认为两个 AP 为强相似, 保留其中信号稳定度较大的 AP.

1.3 融合特征提取

信号的互相关理论揭示了不同传感器获取的信号在不同时刻的线性关系, 其互相关信号取决于两个异构传感器信号序列的相似程度, 可采用互相关函数描述这种相似程度, 两个连续信号 X, Y 的互相关函数定义为

$$R_{XY}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x^*(t)y(t+\tau)dt, \tag{6}$$

其中 $x(t)$ 表示随机信号序列, $x^*(t)$ 表示随机信号的共轭, τ 表示时间延迟. 通过对连续信号进行采样, 再利用累加的形式得到两个离散信号序列的离散化计算公式为

$$R_{XY}(\tau) = \sum_{t=-\infty}^{\infty} x^*(t)y(t+\tau). \tag{7}$$

在实际定位场景中, WiFi 和蓝牙模块实际上是在一个时间周期 T 内对环境中的 AP 进行扫描, 通常 T 的范围为 $1 \sim 2$ s. 在 T 时间内, 蓝牙通过获取处于广播状态的广播信道包并进行解析得到信号强度等信息, 大约只需 3 ms 即可完成该过程. 在 T 时间内, 获取到的若干 RSS 信息具有时间顺序关系, 假设分别经过 T_1 和 T_2 时间后, WLAN 设备采集的 WiFi 信号序列为

$$\mathbf{r}_w = (r_{w_1}, r_{w_2}, \dots, r_{w_p}), \tag{8}$$

BLE 信号序列为

$$\mathbf{r}_b = (r_{b_1}, r_{b_2}, \dots, r_{b_q}), \tag{9}$$

其中 \mathbf{r}_w 表示 T_1 时间后采集到来自 p 个 WiFi 信号的 RSS, \mathbf{r}_b 表示 T_2 时间后采集到来自 q 条 BLE 信号的 RSS.

将互相关理论应用于 WiFi 和蓝牙信号之间, 通过计算互相关序列得到两者之间的融合特征, 它们之间的互相关性可表示为

$$\hat{R}_{r_b r_w}(\tau) = \begin{cases} \sum_{k=1}^{N-\tau} r_{b_{k+\tau}} r_{w_k}, & \tau \geq 0, \\ \hat{R}_{r_w r_b}(-\tau), & \tau < 0, \end{cases} \tag{10}$$

其中 $\tau = \dots, -2, -1, 0, 1, 2, \dots$ 表示时延参数, $N = \max\{p, q\}$. 当信号序列 \mathbf{r}_w 和 \mathbf{r}_b 的有限样本数分别为 p 和 q 时, 它们的互相关样本序列数为 $2\max\{p, q\} - 1$, 记经过互相关计算后的互相关序列为

$$\mathbf{r}_c = (r_{c_1}, r_{c_2}, \dots, r_{c_{2\max\{p, q\}-1}}). \tag{11}$$

本文通过引入互相关理论对 WiFi 和蓝牙信号进行互相关特征提取得到融合序列 \mathbf{r}_c , 将其作为融合特征与 \mathbf{r}_w 和 \mathbf{r}_b 进行组合, 得到 5 种组合特征, 结果列于表 1.

表 1 指纹特征序列

Table 1 Fingerprint feature sequence

编号	指纹特征	编号	指纹特征
S ₁	$\mathbf{r}_w = (r_{w_1}, r_{w_2}, \dots, r_{w_p})$	S ₄	$\mathbf{r}_c = (r_{c_1}, r_{c_2}, \dots, r_{c_{2\max\{p, q\}-1}})$
S ₂	$\mathbf{r}_b = (r_{b_1}, r_{b_2}, \dots, r_{b_q})$	S ₅	$\mathbf{r}_f = (r_b, r_w, r_c)$
S ₃	$\mathbf{r}_{wb} = (r_w, r_b)$		

2 基于融合信息的半监督学习模型

2.1 极限学习机模型

极限学习机模型(ELM)是一种不涉及迭代调整的具有随机权重的神经网络^[8], 通过随机生成隐藏节点的输入权重和偏差使其具有极快的学习速度. ELM 的输出函数可表示为

$$F_{ELM}(x) = \sum_{i=1}^{N_h} \beta_i g(\omega_i x + b_i) = \mathbf{h}(x)\boldsymbol{\beta}, \tag{12}$$

其中 $g(x)$ 为激活函数, β_i 为输出权重, ω_i 为输入权重, b_i 是偏差, $\mathbf{h}(x) = (g(\omega_1 x + b_1), \dots, g(\omega_{N_h} x + b_{N_h}))$ 表示隐藏层网络的输出, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_{N_h})^T$ 表示所有隐藏节点的输出权重. 训练样本的输出可表示为

$$\mathbf{T} = \mathbf{H}\boldsymbol{\beta}, \quad \mathbf{T} = (t_1, t_2, \dots, t_n), \quad \mathbf{H} = (h(x_1), \dots, h(x_n))^T. \tag{13}$$

通过求解 $\arg\min_{\boldsymbol{\beta}} \|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\|^2$ 得到

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^+ \mathbf{T}, \tag{14}$$

其中 \mathbf{H}^+ 为矩阵 \mathbf{H} 的广义逆矩阵.

由于 ELM 在学习速度和泛化性能方面优于梯度的方法, 因此备受关注, 目前已提出了许多版本改进的 ELM. 例如: 域自适应极限学习机(DAELM)可通过利用来自目标区域的有限数量的标记数据进行漂移补偿提高分类器的鲁棒性^[9]; 结合流行正则化的其他领域自适应 ELM 算法已被应用于视觉分类和图像分类问题. 而且改进的 ELM 算法涵盖了监督学习、半监督学习、无监督学习以及不平衡学习问题, 通过扩展 ELM 可获得更高的预测精度、更稳定的性能和训练速度.

2.2 半监督流形约束定位方法

为降低在离线阶段采集大量标记样本的工作量, 同时保证指纹定位算法的预测精度, 可采用半监督极限学习机模型, 并引入图的 Laplace 算子对未标记样本作流形约束. 流形约束是一种用于从标记和未标记样本中学习的几何框架^[10], 这种方法依赖于数据的局部几何特性, 建立在流形假设^[11]基础上, 其认为在一批样本中, 如果两点在局部空间中邻近, 则他们的条件概率也相近. 在定位场景中, 如果从位置 l_1 和 l_2 采集到的两个信号向量 \mathbf{s}_{l_1} 和 \mathbf{s}_{l_2} 相似, 则他们的条件概率 $P(l_1 | \mathbf{s}_{l_1})$ 和 $P(l_2 | \mathbf{s}_{l_2})$ 也相似.

根据文献[11]中流形假设给定定位场景中约束未标记样本的半监督学习模型为

$$f^* = \arg\min_f \sum_{i=1}^l \|f_i - y_i\|^2 + \gamma f^T \mathbf{L} f, \tag{15}$$

其中: 第一项为已标记数据的拟合误差, l 表示样本数; 第二项为流形约束, γ 表示平滑度参数, \mathbf{L} 表示高维信号空间的图 Laplace 矩阵. 其得到的最优解在物理空间上等同于一条尽可能经过更多标记点的平滑轨迹.

2.3 基于融合特征的半监督流形约束定位方法

与基于单信号的半监督流形方法不同, 本文在半监督学习模型中同时考虑了 WiFi 和蓝牙信号, 以及两者的融合特征信号. WiFi 和蓝牙具有不同的传播特性及有效距离, 如图 2 所示, 顶点表示获取 RSS 向量的位置, 顶点间的连线表示 RSS 向量间彼此相似, 图 2(A)和(B)分别表示 WiFi 和 BLE 的近邻关系. 由于 WiFi 比 BLE 的有效传播距离更大, 因此它具有更复杂的近邻关系. 例如, 对于 WiFi 信号的 6 号顶点和 9 号顶点间存在链接, 而 BLE 却不存在. 因此在构建半监督学习模型时, 分别对 WiFi 和 BLE 构建流形正则化约束.

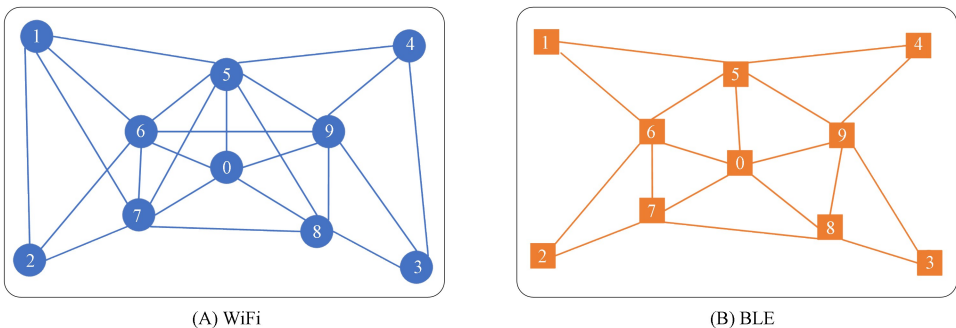


图 2 信号样本的相似关系

Fig. 2 Similarity relations of signal samples

根据结构风险最小化原则, FFSELM 使用图 Laplace 正则化找到高维信号空间中标记和未标记样本的结构关系. 为构建基于 v 个标记样本和 u 个未标记样本的邻接图, 用权重 ω_{ij} 表示样本 i 和 j 之间的相似度, 图 Laplace 矩阵 \mathbf{L} 可表示为 $\mathbf{L} = \mathbf{D} - \mathbf{W}$, 其中 \mathbf{D} 为对角矩阵, $D_{ii} = \sum_{k=1}^{v+u} \omega_{ik}$, $\mathbf{W} = (\omega_{ij})_{(u+v) \times (u+v)}$, $\omega_{ij} = \exp\{-|r_i - r_j|^2 / (2\sigma^2)\}$. 为控制复杂性的同时考虑经验风险, FFSELM 最小化了拟合误差、WiFi 和 BLE 以及两者融合特征的单平滑度惩罚项:

$$\operatorname{argmin}_f \left\{ \frac{1}{2} \|\mathbf{f} - \mathbf{T}\|^2 + \lambda_1 \mathbf{f}^T \mathbf{L}_1 \mathbf{f} + \lambda_2 \mathbf{f}^T \mathbf{L}_2 \mathbf{f} + \lambda_3 \mathbf{f}^T \mathbf{L}_3 \mathbf{f} \right\}, \quad (16)$$

其中第一项表示标记样本的拟合误差, 后三项分别表示对 WiFi 和 BLE 以及两者融合特征的流形约束项, 用系数 λ_k 控制流形结构的复杂性. 将 ELM 训练样本的输出 $\mathbf{f} = \mathbf{H}\boldsymbol{\beta}$ 代入式(16), 可得

$$\operatorname{argmin}_{\boldsymbol{\beta}} l(\boldsymbol{\beta}) = \operatorname{argmin}_{\boldsymbol{\beta}} \left\{ \frac{1}{2} \|\mathbf{J}\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\|^2 + \lambda_1 (\mathbf{H}\boldsymbol{\beta})^T \mathbf{L}_1 \mathbf{H}\boldsymbol{\beta} + \lambda_2 (\mathbf{H}\boldsymbol{\beta})^T \mathbf{L}_2 \mathbf{H}\boldsymbol{\beta} + \lambda_3 (\mathbf{H}\boldsymbol{\beta})^T \mathbf{L}_3 \mathbf{H}\boldsymbol{\beta} \right\}, \quad (17)$$

其中: 输出矩阵 \mathbf{H} 已知且包含标记样本和非标记样本, 维度为 $(u+v) \times N_h$; $\mathbf{J} = \operatorname{diag}(\delta_1, \delta_2, \dots, \delta_{u+v})$ 为指示矩阵, 当第 i 个样本为包含位置坐标的标记样本时, $\delta_i = 1$, 否则 $\delta_i = 0$; \mathbf{T} 为实际位置坐标组成的矩阵, v 个元素对应实际位置坐标, 其余 u 个元素等于 0.

对式(17)求导, 可得

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = 0 \Rightarrow (\mathbf{J}\mathbf{H}\boldsymbol{\beta} - \mathbf{T})^T \mathbf{J}\mathbf{H} + \lambda_1 (\mathbf{H}\boldsymbol{\beta})^T \mathbf{L}_1 \mathbf{H} + \lambda_2 (\mathbf{H}\boldsymbol{\beta})^T \mathbf{L}_2 \mathbf{H} + \lambda_3 (\mathbf{H}\boldsymbol{\beta})^T \mathbf{L}_3 \mathbf{H} = 0, \quad (18)$$

从而可得

$$\boldsymbol{\beta} = ((\mathbf{J} + \lambda_1 \mathbf{L}_1^T + \lambda_2 \mathbf{L}_2^T + \lambda_3 \mathbf{L}_3^T) \mathbf{H})^+ \mathbf{J}\mathbf{T}. \quad (19)$$

当式(19)中 $\lambda_1, \lambda_2, \lambda_3$ 为 0 时, 模型退回式(14)的形式, 相当于忽略训练样本中的无标定样本, 算法变为有监督学习.

综上所述, 对于标记样本 $\{r_{wi} r_{bi} r_{ci}, (x_i, y_i) \mid i=1, 2, \dots, v\}$ 和未标记样本 $\{r_{wi} r_{bi} r_{ci} \mid i=1, 2, \dots, u\}$, FFSELM 算法流程如下:

- 1) 随机选择 ω_i 和 i , 设置隐藏层节点数;
- 2) 使用所有标记样本和未标记样本计算隐藏层输出矩阵 \mathbf{H} ;
- 3) 分别计算 WiFi, BLE 及两者融合特征的 Laplace 矩阵 $\mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3$, 以 \mathbf{L}_1 为例, $\mathbf{L}_1 = \mathbf{D}' - \mathbf{W}'$,

$$D'_{ii} = \sum_{k=1}^{v+u} \omega'_{ik}, \omega'_{ij} = \exp\{-|r_{wi} - r_{wj}|^2 / (2\sigma^2)\};$$

- 4) 根据式(19)计算输出权重矩阵 $\boldsymbol{\beta}$;
- 5) 利用训练得到的模型 $\mathbf{f} = \mathbf{H}\boldsymbol{\beta}$ 进行预测.

3 实验分析

3.1 实验场景

为验证本文算法的性能, 选用江南大学物联网工程学院 C 区一楼走廊作为实景测试场地, 如图 3 所示. 其中低功耗蓝牙信标是以 NRF52832 为主控芯片, 采集终端设备为基于 Android12 的红米 K30 Pro 手机, 采用网格取点的方式, 间隔 1 m, 共取得 156 个采样点. 结合自主开发的 WiFi 和 BLE 采集客户端, 共采集数据 7 864 条. 由于楼层间 WiFi 信号的干扰, 采样过程中共扫描到 14 个 WiFi 信标, 经过预处理和 AP 优选, 确定 8 个稳定的 WiFi 节点和 11 个 BLE 节点, 并剔除 226 条异常值信号, 剩余有效数据 7 638 条, 将其按 1 : 1 的比例分为测试集和训练集.

通过 MATLAB R2019a 仿真软件对算法性能进行仿真实验, 运行环境为 Windows 10 操作系统, 硬件环境为 Intel(R) Core(TM) i5-9400f CPU@2.9 GHz, 16 GB 内存.

3.2 特征稳定性对比分析

将数据集中缺失的信号用最小值补齐(BLE 为 -110 dBm, WiFi 为 -100 dBm), 通过融合特征提取得到 21 维 $((2 \times 11 - 1)$ 维)互相关特征. 其中互相关特征采用 MATLAB 中的 Xcorr 函数作近似计

算. 从采样点中随机选择一个点, 将 3 种数据样本的每一维度先进行归一化处理, 然后计算方差, 实验结果如图 4 所示. 通过对比 3 条特征曲线的归一化方差值可知, WiFi 特征和蓝牙特征的特征起伏不定, 其中 WiFi 的特征曲线波动最大, 而融合的互相关特征曲线最稳定. 可见, 相对于单模的定位信号, 互相关特征更稳定, 更有利于提高定位精度.

3.3 特征有效性对比分析

对模型中流形约束系数的确定, 本文使用文献[12]中提出的改进麻雀搜索算法, 以训练集的平均定位误差作为适应度函数, 分别对参数 $\lambda_1, \lambda_2, \lambda_3$ 进行拟合. 针对表 1 所列的 5 种指纹特征, 以训练集中标记样本占总样本数的 1/4 进行实验, 基于本文 FFSELM 算法进行分析对比, 实验结果如表 2 和图 5 所示. 表 2 中 t_t 和 t_p 分别表示训练时间和预测时间, p_t 和 p_p 分别表示误差在 4 m 内的训练精度和测试精度. 图 5 中 Cross-Corr 表示基于互相关理论融合的特征.

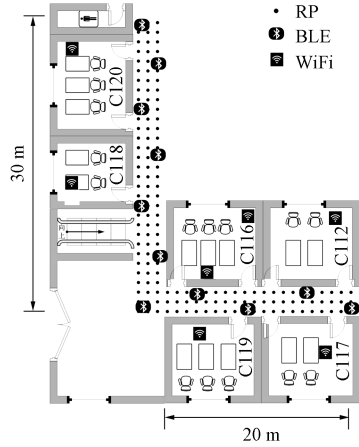


图 3 实验场景示意图

Fig. 3 Schematic diagram of experimental scene

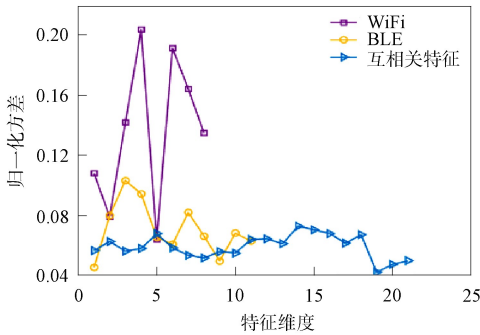


图 4 不同信号的特征稳定性对比

Fig. 4 Comparison of characteristic stability of different signals

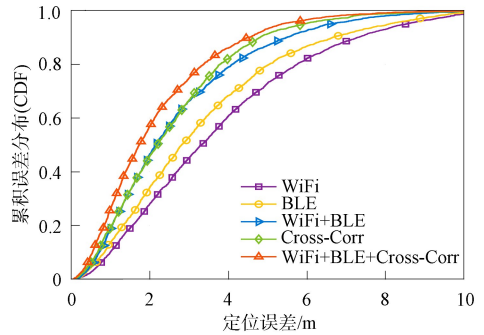


图 5 不同指纹特征下的定位精度对比

Fig. 5 Comparison of positioning accuracy under different fingerprint characteristics

表 2 不同指纹特征下的实验结果对比

Table 2 Comparison of experimental results under different fingerprint characteristics

编号	λ_1	λ_2	λ_3	t_t/s	t_p/s	$p_t/\%$	$p_p/\%$
S ₁	0.08	0	0	6.754	0.105	65.4	60.6
S ₂	0	0.11	0	6.797	0.103	74.3	68.2
S ₃	0.06	0.18	0	11.012	0.125	83.2	78.9
S ₄	0	0	0.24	7.074	0.107	85.3	81.9
S ₅	0.04	0.11	0.16	13.652	0.125	90.2	86.9

在算法训练时间上可见, 单独使用蓝牙或 WiFi 特征在单一流形约束模型上训练时间总体较接近, 分别为两者构建流形正则化约束项后, 训练时间有明显增加, 主要原因是额外的流形约束项在 Laplace 矩阵的计算上需要消耗时间, 在结合 3 种特征的训练时间最长的结果上也可见这一点; 由于 ELM 模型的复杂度对输入特征的维度不敏感, 在互相关特征增加维度的情况下, 训练时间并没有明显增大. 从定位精度的结果可见, 单独使用 WiFi 定位的精度最低, 传播距离更近的 BLE 的定位精度要优于 WiFi; 分别为 BLE 和 WiFi 建立流形约束, 以及进行融合特征提取后, 定位精度都有明显提升, 且精度较接近, 可见融合特征能充分发挥互相关特征稳定性好这一优点; 结合 3 种特征进行流形约束的定位精度最高, 与单一特征相比, 定位精度约提升 25 个百分点.

3.4 标定量对定位精度的影响

为分析 FFSELM 算法在不同标定数据量下对定位精度的影响, 将训练集样本中标记样本 v 和未标记样本 u 比例进行调整, 取 $v/(u+v)$ 分别为 $1/10, 1/8, 1/6, 1/4, 1/2$ 进行训练, 然后测试模型精度, 实验结果如图 6 所示. 随着标记样本在训练集中的比例增大, 定位精度逐步提高. 当训练集中标记样本占总样本的 $1/2$ 时, 定位误差在 2 m 内的概率为 74% , 3 m 的定位精度达 87.2% , 在实际环境中 3 m 的定位精度已经具有很好的实用性; 当占比为 $1/10$ 时, 定位误差在 5 m 内的概率为 81.7% . 可见, 本文 FFSELM 算法能利用较少的标记样本实现较高的定位精度.

3.5 算法对比分析

为验证本文算法的性能, 选取 ELM 算法、BP 神经网络、融合半监督极限学习机 (fusion semi-supervised extreme learning machine, FSELM) 算法^[6] 以及基于 WiFi 和蓝牙多模融合 (Multi-Fusion) 定位算法^[11] 进行对比. 在训练集样本总数不变的条件下, 将 $v/(u+v)$ 分别取 $1/8, 1/6, 1/4, 1/2$ 进行实验, 其中由于 ELM, BP 和 Multi-Fusion 算法无法实现半监督学习, 因此对这 3 种算法只选用标记样本进行实验, 且 ELM 和 BP 算法是以 BLE 和 WiFi 的指纹特征向量合并组成的新向量作为训练集进行训练. 实验结果如图 7 所示.

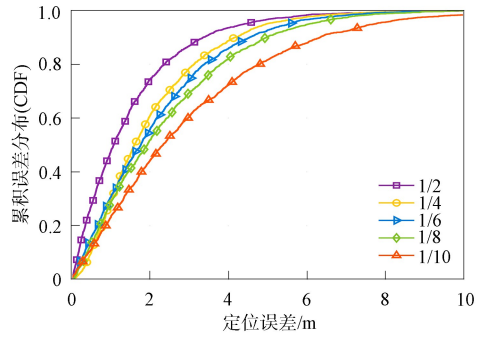
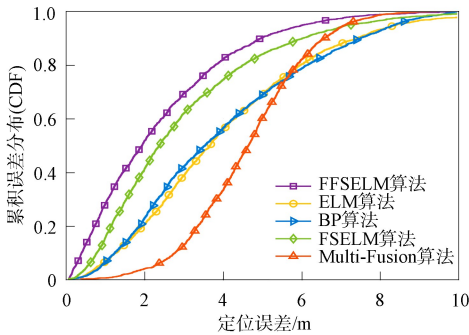
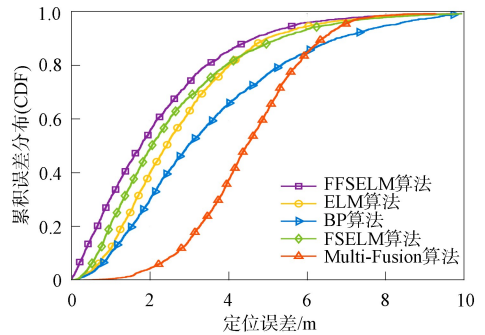


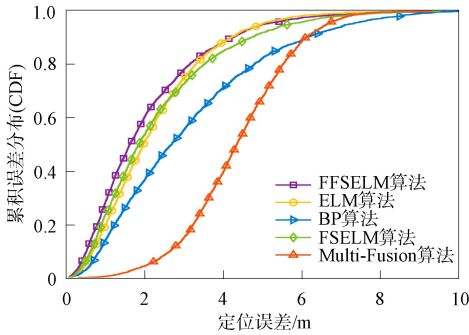
图 6 不同标定数据量下的定位精度对比
Fig. 6 Comparison of positioning accuracy under different calibration data



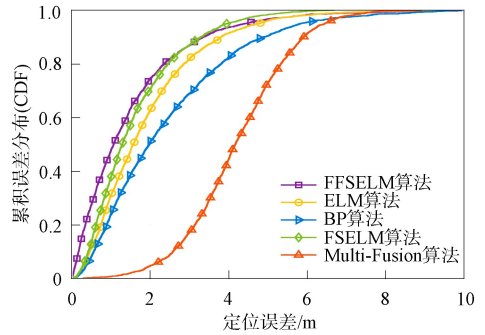
(A) 1/8 标定样本下的对比



(B) 1/6 标定样本下的对比



(C) 1/4 标定样本下的对比



(D) 1/2 标定样本下的对比

图 7 不同标定数据量下各算法的定位精度对比

Fig. 7 Comparison of positioning accuracy of various algorithms under different calibration data

由图 7 可见, 随着标记样本占比的不断增大, 5 种算法的定位精度均有提升. FFSELM 算法在标定样本占比为 $1/8$ 时, 在 5 m 内的定位精度达 90% , 相比于 FSELM 算法, 定位精度约提升 6 个百分点. 由于 FFSELM 算法在建立流形约束时, 不仅考虑了 WiFi 和 BLE 不同程度的约束效果, 并对两者

的互相关信号也建立了约束,使模型从多方面形成了互补关系,从而取得了更理想的效果.相比于不具备半监督学习能力的 ELM 和 BP 算法,由于本文算法保留了 ELM 算法的优势,同时引入半监督流形约束,充分利用了对未标记样本进行学习的能力,因而定位精度有明显提升,在不同标定样本数量下,提升幅度最大可达 22 个百分点.对于 Multi-Fusion 算法,由于其对网格内的数据取均值进行计算,使该算法对标定训练样本的数量不敏感,因而在不同标定样本量下定位精度变化不明显.

综上所述,针对 WiFi 单模定位方式定位精度较低以及 BLE 单模难以进行大规模定位的问题,提出了一种基于特征融合的半监督流形约束定位算法.在离线指纹库建立阶段,首先对采集到的信号进行预处理,剔除对定位精度产生负面影响的离异值,然后采用 AP 优选的方式对实际环境中的 WiFi 信号进行过滤,选择稳定性好、信号强的 AP 节点.将采集到的 WiFi 和 BLE 指纹特征向量进行特征融合,然后在半监督极限学习机的基础上对 3 种特征向量分别建立流形约束项,从多方面对未标定样本进行学习.实验结果表明,互相关特征提取可有效减少环境对信号的影响,提高信号的稳定性,在半监督学习模型结合不同类型数据的优势可提升定位精度.与其他定位算法相比,本文算法定位精度明显提升,且半监督流形约束方法极大减少了标定数据采集的工作量.

参 考 文 献

- [1] 王睿,赵方,彭金华,等.基于 WI-FI 和蓝牙融合的室内定位算法[J].计算机研究与发展,2011,48(增刊2):28-33. (WANG R, ZHAO F, PENG J H, et al. Combintion of WI-FI and Bluetooth for Indoor Localization [J]. Journal of Computer Research and Development, 2011, 48(Suppl 2): 28-33.)
- [2] BANUIKEVIC A, JENSEN C S, LU H. Hybrid Indoor Positioning with Wi-Fi and Bluetooth: Architecture and Performance [C]//2013 IEEE 14Th International Conference on Mobile Data Management. Piscataway, NJ: IEEE, 2013: 207-216.
- [3] 卢志泳.室内融合定位的研究与应用[D].广州:中山大学,2015. (LU Z Y. The Design and Implementation of Multiple Technology Fusion Indoor Localization [D]. Guangzhou: Sun Yat-sen University, 2015.)
- [4] KANARIS L, KOKKINIS A, LIOTTA A, et al. Fusing Bluetooth Beacon Data with Wi-Fi Radiomaps for Improved Indoor Localization [J]. Sensors, 2017, 17(4): 812-1-812-15.
- [5] TIAN Z S, XIAO Z Y, HUANG Y D. Multi-source Fusion Localization Technology Based on Convolutional Neural Networks [C]//2022 14th International Conference on Wireless Communications and Signal Processing (WCSP). Piscataway, NJ: IEEE, 2022: 1081-1085.
- [6] 黄正宇,蒋鑫龙,刘军发,等.基于融合特征的半监督流形约束定位方法[J].浙江大学学报(工学版),2017,51(4):655-662. (HUANG Z Y, JIANG X L, LIU J F, et al. Fusion Feature Based Semi-supervised Manifold Localiation Method [J]. Journal of Zhejiang University (Engineering Science), 2017, 51(4): 655-662.)
- [7] JIANG X L, CHEN Y Q, LIU J F, et al. FSELM: Fusion Semi-supervised Extreme Learning Machine for Indoor Localization with Wi-Fi and Bluetooth Fingerprints [J]. Soft Computing: A Fusion of Foundations, Methodologies and Applications, 2018, 22(11): 3621-3635.
- [8] HUANG G B, ZHOU H M, DING X J, et al. Extreme Learning Machine for Regression and Multiclass Classification [J]. IEEE Transactions on Systems Man Cybernetics Part B (Cybernetics), 2012, 42(2): 513-529.
- [9] ZHANG L, ZHANG D. Domain Adaptation Extreme Learning Machines for Drift Compensation in e-Nose Systems [J]. IEEE Transations on Instrumentation Measurement, 2015, 64(7): 1790-1801.
- [10] CHAI X Y, YANG Q. Reducing the Calibration Effort for Probabilistic Indoor Location Estimation [J]. IEEE Transactions on Mobile Computing, 2007, 6(6): 649-662.
- [11] GU Y, CHEN Y Q, LIU J F, et al. Semi-supervised Deep Extreme Learning Machine for Wi-Fi Based Localization [J]. Neurocomputing, 2015, 166: 282-293.
- [12] 印雷,顾德,刘飞.基于改进麻雀搜索算法优化的 DV-Hop 定位算法[J].传感技术学报,2021,34(5):670-675. (YIN L, GU D, LIU F. Improved Sparrow Search Algorithm Based DV-Hop Localization in WSN [J]. Chinese Journal of Sensors and Actuators, 2021, 34(5): 670-675.)