

文章编号:2095-6134(2021)06-0817-08

跨时间迁移的多源无线信号指纹融合定位方法^{*}

史达亨^{1,2†}, 刘立刚¹, 周斌¹, 卜智勇¹

(1 中国科学院上海微系统与信息技术研究所 中国科学院无线传感网与通信重点实验室, 上海 200050;

2 中国科学院大学, 北京 100049)

(2021 年 1 月 25 日收稿; 2021 年 5 月 11 日收修改稿)

Shi D H, Liu L G, Zhou B, et al. Fingerprinting localization of cross-temporal transferred and multi-source wireless signal fusion[J]. Journal of University of Chinese Academy of Sciences, 2021, 38(6): 817-824.

摘 要 针对传统指纹定位方法中无线信号覆盖范围有限以及信号时变特性难以处理的问题, 提出利用多种无线信号进行指纹定位的方法, 并采用测地线流核 (geodesic flow kernel, GFK) 的迁移学习方法减轻信号时变性对定位精度的不利影响。首先, 通过对多源无线信号特征进行多轮随机采样融合, 从而构建多个数据集, 获得更丰富多样的指纹特征; 其次, 对不同时间的指纹特征进行迁移的 GFK 进行融合, 使得迁移从 2 个领域扩展到多个领域; 最后, 使用多个数据集训练基分类器, 并从多个基分类器得到最终的预测结果, 从而提高位置分类器的泛化性能。实验结果表明, 本方法的定位精度比传统方法更高。

关键词 指纹定位; 多源融合; 迁移学习; 集成学习; 测地线流核

中图分类号: TN911.7; TP391.41 **文献标志码:** A **doi:** 10.7523/j.issn.2095-6134.2021.06.012

Fingerprinting localization of cross-temporal transferred and multi-source wireless signal fusion

SHI Daheng^{1,2}, LIU Ligang¹, ZHOU Bin¹, BU Zhiyong¹

(1 Key Laboratory of Wireless Sensor Network and Communications of CAS, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai 200050, China;

2 University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract To address the problem of limited coverage area of wireless signals and difficulty in dealing with time-varying characteristics of wireless signals in traditional fingerprinting localization, we propose a method of using multi-source wireless signals for fingerprinting localization, and the accuracy of positioning, which is affected by the time varying of signals, is mitigated by geodesic flow kernel. Firstly, we construct our datasets by a multi-round random sampling of multiple wireless signal sources, which provides a richer and more diverse fingerprint features. Secondly, we fuse geodesic flow kernels from fingerprint features of different times, so that we extend transferring methods from two domains to multiple domains. Finally, base classifiers are trained on multiple

^{*} 上海市自然科学基金 (19ZR1467200) 和国家重点研发计划 (2020YFB0905900) 资助

[†] 通信作者, E-mail: sdaheng@mail.sim.ac.cn

datasets, and the predicted position are obtained from all base classifiers, so as to elevate the generalization of the model. Simulation results show that the proposed method outperforms the traditional approaches in terms of positioning accuracy.

Keywords fingerprinting localization; multi-source fusion; transfer learning; ensemble learning; geodesic flow kernel

通常室外环境下利用卫星导航定位系统可以提供精确的地理位置,但卫星信号微弱,易被遮蔽,例如在城市峡谷、隧道及室内等场合定位误差过大而无法使用。指纹定位方法基于无线信号传播对环境的依赖性,在不同位置的无线电信号具有独特的可分辨特征,因此可利用这些信号特征将位置对应起来。指纹定位的这些特点使其适合应用在室内等复杂环境。其中文献[1]首次提出使用信号强度信息进行定位,它将定位分为2个阶段:离线阶段,在选定参考点采集无线信号并提取信号特征来构建指纹库;在线阶段,实时分析采集的信号特征并与指纹库中的数据进行匹配来实现定位。文献[2-3]将位置估计作为概率问题来研究,通过概率方法估计位置。文献[4-5]采用机器学习中的支持向量机和集成学习方法对离线阶段的指纹数据进行训练,在在线阶段进行分类预测。文献[6-8]利用深度学习以及卷积神经网络训练无线信号指纹模型后进行位置估计。

传统的指纹定位方法都假设采集的无线信号数据的分布基本保持不变,但由于人员的流动以及环境变化等因素的影响,经过一段时间后采集的信号分布并不相同,这使得经过一段时间后由于指纹库不能及时更新导致定位误差变大。针对这种问题,文献[9-10]提出采用众包的方式更新指纹库,这种方法利用公众采集数据,持续更新指纹库,但公众的位置信息可能存在误差,且这种误差会积累在指纹库中。文献[11-13]提出结合惯性导航或地磁传感器进行多源融合定位以补偿指纹定位由于信号时变性带来的误差影响,但惯性导航方法会产生积累误差,需要惯性测量单元具有较高的精度。在室内定位中常用的行人航位推算(pedestrian dead reckoning, PDR)方法受到惯导的积累误差使得定位误差逐步加大,文献[14-19]通过融合无线信号指纹定位与地磁传感器等进行多源融合定位,以补偿惯导航位推算的积累误差。

由于上述指纹定位方法的不足,文献[20]研究了将迁移学习应用在指纹定位中,利用半监督隐马尔可夫模型将已知模型迁移到新模型,从而

减少无线信号指纹在不同时间校准的工作量。文献[21]则应用迁移学习,将完整构建的指纹库迁移到其他的室内环境,使得其无需重新构建完整的无线指纹库即可定位。迁移学习的核心问题是找到新问题与原问题的相似性,以便实现知识的迁移。领域自适应问题是迁移学习的研究内容之一,它侧重于解决特征空间一致、类别空间一致,仅特征分布不一致的问题。指纹变化具有与其相同的问题,故可应用领域自适应的方法减少指纹变化带来的影响。文献[22]提出一种数据边缘分布自适应的方法(transfer component analysis, TCA)。TCA方法的目标是通过最大均值差异学得一个可再生希尔伯特空间上的特征映射,使得映射后的数据分布接近。但TCA方法只考虑了数据的边缘分布自适应。文献[23]提出同时适配数据的边缘分布与条件分布的联合分布自适应方法(joint distribution adaptation, JDA)。JDA方法的边缘分布适配与TCA相同,在适配条件分布时通过生成伪标签来近似边缘分布。但JDA方法未考虑边缘与条件分布自适应在领域自适应中的贡献并不一定是相等的。文献[24]提出了平衡分布自适应方法(balanced distribution adaptation, BDA)。BDA方法在JDA方法的基础上添加了平衡因子来动态地调整边缘分布与条件分布的重要性。虽然BDA方法首次给出了边缘分布与条件分布在领域自适应中贡献度的定量估计,但并未解决平衡因子的精确计算问题,在某些情况下效果与JDA类似,而且正确性无法保证。

相对于数据分布的适配,流形学习也是领域自适应的一个重要的方向。它的基本假设是,现有的数据是从高维空间中采样出来的。所以,它具有高维空间中的低维流形结构。而且由于在流形空间中的特征通常都有着良好的几何性质,可以避免扭曲,因此可以将原始空间中的特征变换到流形空间中。文献[25]提出采样测地线流(sampled geodesic flow, SGF)方法,它利用两点间测地线在流形空间中距离最短,将源域与目标域看作流形空间中的2个点,在两点间求出测地线

就能将源域变换到目标域。文献[26]在SGF方法的基础上提出测地线流核(geodesic flow kernel, GFK)方法。GFK方法提出一种核学习方法, GFK方法学得2个领域子空间中的差异以及共同特征的增量变化,并给出2个领域之间不变性的低维表示。通过这种低维表示对数据进行变换,可实现不同领域的迁移。

针对上述问题,本文提出一种结合多种无线信号源的跨时间测地线流核融合定位方法。首先,对多源无线信号进行多轮次的样本及特征的随机采样融合,该方法不仅为指纹库的构建提供了比单一信号源更丰富的信号特征,同时也增加了样本及其特征的多样性,减少定位精度在单一信号受到多径衰落和快衰落时的影响。其次,每轮中将不同时间采集的数据作为源域和目标域计算测地线流核并加权融合,使得最终的核具有多个时间所采集数据的不变性的低维表示,并且根据权值的不同,每个测地线流核对最终的核具有不同的贡献程度,融合不同时间的指纹特征不变性减少各种变化因素对指纹定位的影响。最后使用融合后的测地线流核对每一轮的源域数据进行变换,再对变换后的数据训练一个基分类器,则最终的集成分类器由所有的基分类器组成。定位结果则是所有基分类器预测结果中投票数量最多的位置。多个基分类器之间的多样性给集成分类器带来更好的泛化能力。

1 系统模型与问题描述

假设位置空间 $L = \{l_1, \dots, l_n\}$ 。 L 定义为坐标 x, y 的集合 $L = \{l_1 = (x_1, y_1), \dots, l_n = (x_n, y_n)\}$, 其中每一个元组 (x_i, y_i) 代表一个参考点。

参考点收集到的一种无线信号强度向量定义为 $\mathbf{RSS} = (rss_1, \dots, rss_m)$, m 为不同信道采集到的信号强度数量。对 \mathbf{RSS} 的样本、特征进行 K 轮随机采样融合, 则一轮多源融合最终的 \mathbf{RSS} 样本为

$$\mathbf{RSS}_k = \begin{pmatrix} rss_{11}^1 & \dots & rss_{1m}^1 \\ \dots & \dots & \dots \\ rss_{n1}^k & \dots & rss_{nm}^k \end{pmatrix}, k \in [1, K], n \text{ 为在参}$$

考点采集的无线信号样本数量。

假设不同时间采集的 \mathbf{RSS}_k^i 值作为不同的源域 S 及目标域 T , 其中上标 $i \in [1, N_t]$ 表示信号的采集时间, N_t 为采集信号时间点个数。如图1所示将源域 S 和目标域 T 的数据嵌入到 Grassmann

流形中, 记 \mathbf{R}^D 上全部 d 维子空间的 Grassmann 流形为 $G_{D,d}$, 则 $G_{D,d}$ 可看作商空间 $SO(D)/(SO(d) \times SO(D-d))$, 其中 $SO(D)$ 是 $D \times D$ 的李群。所以源域与目标域可以看作流形空间 $G_{D,d}$ 上的2个点, 那么两点的测地线可看作是一个单参数的指数流, $t \mapsto \exp(t\mathbf{B})$, 其中 t 为经过的时间, $\mathbf{B} = \begin{pmatrix} \mathbf{0} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix}$, \mathbf{A} 为测地线流从起点在单位时间内到达终点的方向和速度。计算2个点间的测地线流核即可学得源域与目标域的领域不变性表示。

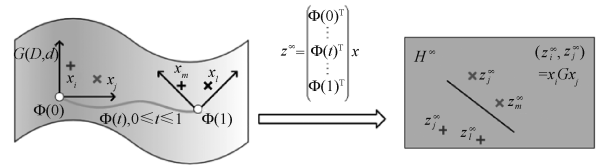


图1 GFK方法

Fig. 1 Principle of GFK

定义融合测地线流核权值 $W_{S_i T_j}, i, j \in [1, N_t]$ 分别代表多个源域和目标域中不同的领域。则跨时间融合测地线流核可定义为 $\mathbf{G} = \sum W_{S_i T_j} \text{GFK}(S_i, T_j)$, 每一轮的训练集可定义为 $\mathbf{RSS}_k = (\mathbf{RSS}_k^T \mathbf{G})^T$, T 代表矩阵的转置。对每轮的训练集训练一个基分类器 f , 将所有基分类器组成一个强分类器 $F = (f_1, f_2, \dots, f_k)$ 。位置估计问题则为给定一组在线阶段收集到的 \mathbf{RSS} 向量, 经融合后测地线流核变换后再用强分类器内所有的基分类器估计其在位置空间 L 中所属的参考点位置, 最终的位置为强分类器 F 中所有基分类器结果中投票数量最多的结果。

2 跨时间多源无线信号指纹融合

2.1 多源信号数据

室内定位通常使用 Wi-Fi、蓝牙等信号的 RSS 值作为指纹特征, 但这些方法需要部署大量的节点且覆盖范围有限。在室内以及室外均有大量的无线信号。这些无线信号的主要类型包括电视台的广播电视信号、广播信号、移动基站的通信信号、雷达信号等。对于室内可接收的信号而言, 相比于常用的 Wi-Fi 信号, 数字广播电视信号 (digital video broadcast, DVB)、调频广播信号 (frequency modulation, FM)、移动通信信号 (cellular mobile communication, CMC) 可以提供比 Wi-Fi 信号更广的覆盖范围, 且这些无线信号处

于许可频段,可以免受周围设备发射的同频干扰。

如图 2 所示,本文通过将多个广域无线信号的 RSS 数据融合在一起以提供更丰富的指纹特征。

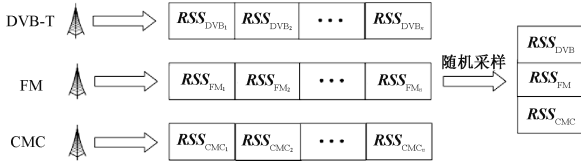


图 2 同时接收 DVB、FM、CMC 的过程

Fig. 2 Receiving DVB, FM and CMC simultaneously

以 DVB、FM 与 CMC 信号为例,在离线阶段信道化扫描这 3 种信号,设信道数分别为 α, β 和 γ 。当确定信道数后,在每个参考点采集 3 种信号源的接收信号强度。在构建指纹特征向量时,分别对 3 种信号源的 N 个数据样本进行 K 次随机抽样。首先在原始数据集中的 N 个样本随机抽取 n 个样本,其中 $n \leq N$ 。再对每个样本在 M 个信道采集的数据随机抽取 m 个数据,其中 $m \leq M, M \in \{\alpha, \beta, \gamma\}$ 。最后将 3 种信号源数据随机抽样后的 $n \times m$ 维数据拼接起来作为一个新数据集,记为

$$RSS_k = [RSS_{DVB}^{n \times m} \quad RSS_{FM}^{n \times m} \quad RSS_{GSM}^{n \times m}]. \quad (1)$$

其中 $k \leq K$ 。经过 K 轮随机抽样后最终的数据集为 $RSS = [RSS_1 \quad RSS_2 \quad \cdots \quad RSS_K]$ 。

多源无线信号融合的时间复杂度为 $O(n \times k)$, n 为从 N 个样本随机抽取的样本数, K 为重复随机抽取的轮数。

2.2 跨时间测地线流核

1) 计算测地线流核

如图 3 所示,当在不同时间采集到无线信号的 RSS 数据后,由于环境及信号的时变性导致不同时间点的 RSS 数据分布发生了变化。为降低数据分布变化对定位精度的影响,需要计算不同时间的无线信号指纹数据共同的不变性特征。由于需要在多个时段间迁移,故需计算多个测地线流核。

记 $P_{S_i}, P_{T_i} \in \mathbf{R}^{D \times d}$ 为第 i 个源域 S_i 与目标域 T_i 的子空间基,其中 $i = 1, 2, \dots, N_i, D, d$ 为指纹的特征维度以及它的子空间维度。在 $\Phi_i(0) = P_{S_i}, \Phi_i(1) = P_{T_i}$ 约束下的多个测地线流可以表示为 $\Phi_i: t \in [0, 1] \rightarrow \Phi_i(t) \in G(d_i, D_i)$ 。对于其他的 t , 测地线流可表示为

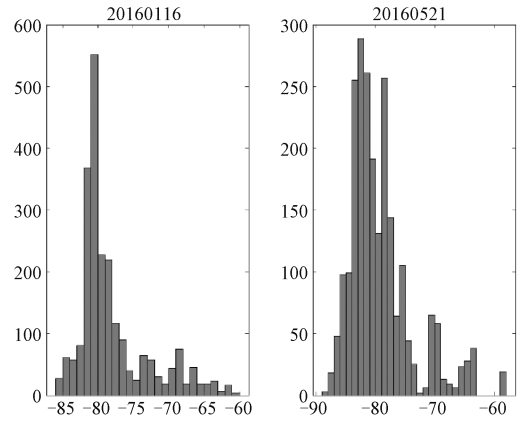


图 3 不同时间 RSS 数据分布对比

Fig. 3 Distribution of RSS between different times

$$\Phi_i(t) = P_{S_i} U_{1,i} \Gamma_i(t) - R_{S_i} U_{2,i} \Sigma_i(t). \quad (2)$$

其中: $R_{S_i} \in \mathbf{R}^{D \times (D-d)}$ 为 P_{S_i} 的正交补空间, $U_{1,i} \in \mathbf{R}^{d \times d}, U_{2,i} \in \mathbf{R}^{(D-d) \times d}$ 是 $P_{S_i}^T P_{T_i}, R_{S_i}^T P_{T_i}$ 经奇异值分解得到的正交矩阵。 Γ_i, Σ_i 为 $d \times d$ 的对角矩阵, 对角元素为 $\cos \theta_{a,i}, \sin \theta_{a,i}$, 其中 $a = 1, 2, \dots, d$ 。 $\theta_{a,i}$ 为 P_{S_i}, P_{T_i} 的主角度, $0 \leq \theta_{1,i} \leq \theta_{2,i} \leq \dots \leq \theta_{d,i} \leq \frac{\pi}{2}$ 。它们测量了子空间之间的覆盖角度。

将第 i 个领域的 2 个原始特征向量 $x_{i,p}, x_{i,l}$ 投影到 $\Phi_i(t)$ 得到它们的无穷维子空间 $z_{i,p}^\infty, z_{i,l}^\infty$, 计算 $z_{i,p}^\infty, z_{i,l}^\infty$ 的内积得到测地线流核

$$\begin{aligned} \langle z_{i,p}^\infty, z_{i,l}^\infty \rangle &= \int_0^1 (\Phi_i(t)^T x_{i,p})^T (\Phi_i(t)^T x_{i,l}) dt \\ &= x_{i,p}^T G_i x_{i,l}. \end{aligned} \quad (3)$$

由式(5)可得半正定矩阵 G_i

$$G_i = [P_{S_i} U_{1,i} \quad R_{S_i} U_{2,i}] \begin{bmatrix} \Lambda_{i,1} & \Lambda_{i,2} \\ \Lambda_{i,2} & \Lambda_{i,3} \end{bmatrix} \begin{bmatrix} U_{1,i}^T & P_{S_i}^T \\ U_{2,i}^T & R_{S_i}^T \end{bmatrix}. \quad (4)$$

其中 $\Lambda_{i,1}$ 到 $\Lambda_{i,3}$ 为对角矩阵, 对角元素为

$$\begin{aligned} \lambda_{1a}^i &= 1 + \frac{\sin(2\theta_{a,i})}{2\theta_{a,i}}, \quad \lambda_{2a}^i = \frac{\cos(2\theta_{a,i}) - 1}{2\theta_{a,i}}, \\ \lambda_{3a}^i &= 1 - \frac{\sin(2\theta_{a,i})}{2\theta_{a,i}}. \end{aligned} \quad (5)$$

将所有得到的核组合到一起得到在不同时间中迁移的测地线流核集合

$$G = [G_1, \dots, G_n, \dots, G_N, G_1, \dots, G_m, \dots, G_M]. \quad (6)$$

2) 计算分布差异度量

不同时间的 RSS 数据分布差异不同,在对测地线流核进行融合的时候需要考虑 2 个域的测地

线流核对最终核的重要程度。分布差异度量是度量 2 个数据分布之间的差异,它可以判断 2 个数据分布之间是否相似。

Wasserstein 距离^[27]是一种衡量 2 个概率分布距离的度量方法。

$$D_p(S,T)=\left(\inf_{\mu\in\Gamma(P,Q)}\int\rho(x,y)^p\mathrm{d}\mu(x,y)\right)^{\frac{1}{p}}. \quad (7)$$

其中: $\Gamma(S,T)$ 是以 S,T 为边缘分布的所有可能的联合分布的集合, $(x,y)\sim\mu$ 代表从中采样出的一对样本, p 为 p 范数。针对这对样本,将 x 搬到 y 处所需的距离是 $\rho(x,y)$, 然后计算在该联合分布 μ 下样本对距离的期望。在所有可能的联合分布期望的下确界即 2 个分布之间的距离。即使 2 个分布的支撑集没有重叠或者重叠很少, Wasserstein 距离仍然能反应 2 个分布的远近。

3) 多核融合

相比于单核模型,多核模型具有更高的灵活性。经过多个核函数映射后的高维空间是由多个特征空间组合而成的组合空间。组合空间可以组合各个子空间不同的特征映射能力,能够将异构数据中的不同特征分量分别通过合适的单个核函数进行映射,最终使得数据在新的组合空间中能够得到更加准确、合理的表达,进而提高样本数据的分类正确率。

本文对核矩阵进行线性加权平均,权值为归一化后的分布差异度量值的倒数。记权值为 $W_{S_iT_j}, i=1,\cdots,N_i, j=1,\cdots,N_t$, 最后的核矩阵为

$$\mathbf{G}=\sum_{i=1}^N\sum_{j=1}^N W_{S_iT_j}\mathbf{G}_{S_iT_j}. \quad (8)$$

2.3 跨时间无线信号指纹定位

在 2.1 节中通过对多源信号数据随机抽样 K 次得到 K 个新数据集 $\mathbf{RSS}_k, k\in[1,K]$ 后,首先分别对 K 个新数据集计算其跨时间融合测地线流核 (multi-kernel fusion GFK, MKFGFK), 即 $\mathbf{G}_k=\text{MKFGFK}(\mathbf{RSS}_k)$ 。其次通过融合后的测地线流核对数据进行变换,即 $\mathbf{X}_k^{\text{new}}=\mathbf{X}_k\mathbf{G}_k$, 得到变换后的训练数据。再使用决策树 (decision tree, DT) $F_k=(\mathbf{X}_k^{\text{new}}, \mathbf{Y}_k)$ 进行训练,其中 \mathbf{Y}_k 为第 k 轮的样本标签。决策树训练时,每个节点从其所有特征集合中随机选择 c 个特征构成一个子集,再从这个子集中选择最优划分。最后在预测阶段,对所有基决策树分类器的预测结果进行投票,结果出现最多的类别作为最终的预测类别

定位结果。

2.4 可行性分析

GFK 方法是一种子空间学习方法,子空间学习方法假设源域与目标域变换到子空间后具有相似的分布。当将源域和目标域变换到 Grassmann 流形中,特征变换和分布适配通常都有等效的数值形式,因此可以很好地被求解。

对于高维的广域多源无线信号指纹特征数据,通过子空间变换得到相似的分布而不需要对不同领域的数据分布进行对齐。计算源域与目标域的不变性特征的低维表示即可学习不同时间无线信号指纹内在的共同特征,且计算简便。GFK 方法是一种无监督迁移学习方法,不需要目标域具有标签,也不需要像 JDA 方法生成伪标签,通过 GFK 方法将特征变换到子空间还具有降维的特性。

领域自适应问题通常是在 2 个域之间进行迁移, GFK 方法也仅是对 2 个领域进行迁移,但 GFK 方法最终计算出的半正定矩阵核适用多核学习方法。当涉及多个领域之间的迁移可以通过多核学习的方式将不同领域之间的不变性表示进行融合从而达到多领域自适应的目的。

算法的平均时间复杂度为 $O(n\log n)$, n 为待迁移的领域数量。

3 实验结果与分析

3.1 数据集介绍

实验中使用的是 AmbiLoc^[28] 数据集。数据集包含一年中 23 个时段收集到的 DVB-T、FM、CMC 信号 RSS 值。DVB 信号的频率范围为 498~602 MHz,信道宽度为 8 MHz,共分为 6 个信道。FM 信号的频率范围为 87.5~108.5 MHz,以 100 kHz 为一个频段,共分为 210 个频段。CMC 为频率范围在 925~960 MHz 之间的 GSM900 信号,以 200 kHz 为一个频段,共分为 175 个频段,详见表 1、表 2。参考点设置详见表 3 与图 4。

表 1 DVB-T, FM, CMC 信号采样设置
Table 1 Sampling settings of DVB, FM, and CMC signals

信号源	中心频率/MHz	采样率/MHz
FM	98	21
CMC	935,950	20
DVB	498,522,554,578,594,602	9.14

表 2 RSS 指纹数据参数

Table 2 RSS fingerprint parameters			
信号源	频段/MHz	信道宽度/kHz	指纹个数
FM	87.5 ~ 108.5	100	210
CMC	925 ~ 960	200	175
DVB	498 ~ 602	8 000	6

表 3 数据集参考点设置

Table 3 Settings of reference points				
测试建筑	楼层	参考点数	平均间隔/m	时段数
办公楼	1,0,-2	33,36,16	4.6	23

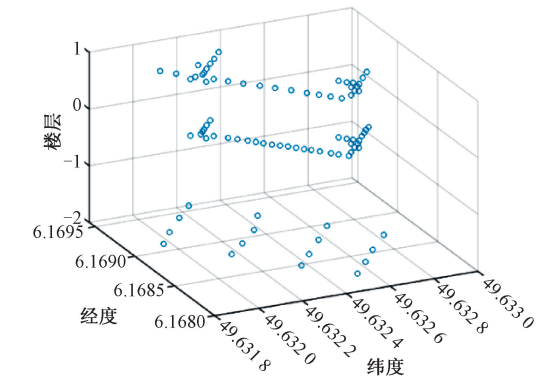


图 4 办公楼参考点分布位置

Fig. 4 Location of reference points in an office building

3.2 实验设置

实验选择样本数量最多的办公楼数据集,以源域作为训练集,目标域作为测试集。实验将办公楼中每个楼层的参考点进行分组,办公楼的-2层参考点分为4组,0层分为4组,1层分为3组。首先对分组进行粗定位,再对组内的参考点进行精确定位。图5展示了-2层参考点的分组划分。其中蓝色点为参考点,4个参考点为一组。决策树的最大划分数量为54,最小叶节点数量为1,共进行15轮训练。实验分别对无迁移、TCA、JDA与BDA方法的距离误差结果进行对比,其中距离误差使用Haversine方法进行计算。

3.3 结果与分析

数据集中办公楼-2层、0层、1层的定位距离误差累积分布函数(cumulative distribution function,CDF)如图6所示。从实验结果中看,本文提出的跨时间多测地线流核融合方法比直接使用决策树方法及其他迁移学习方法都有更好的表现。

从图6中可以看出MKFGFK方法与决策树的定位距离误差的比较中积累分布为90%时分

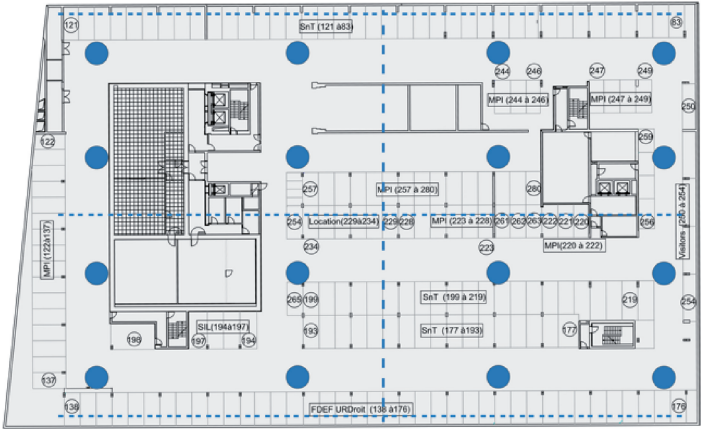


图 5 办公楼-2层参考点分组

Fig. 5 Groups of reference points on -2nd floor

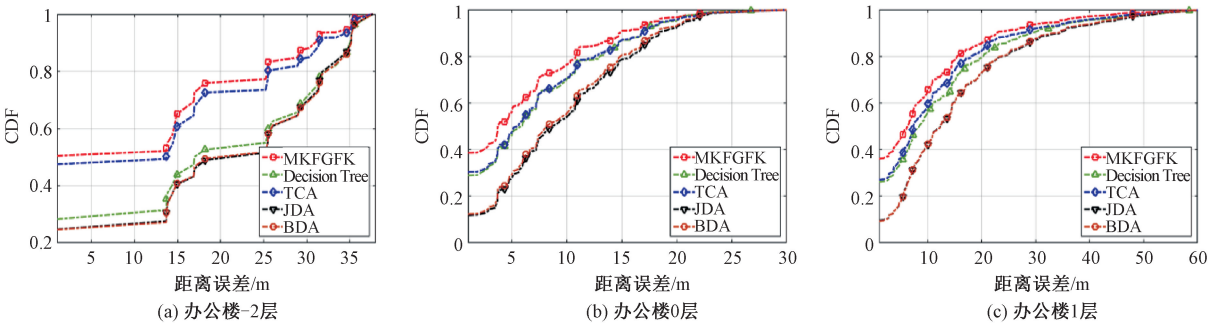


图 6 办公楼定位距离误差对比

Fig. 6 Comparison of positioning distance errors in office buildings

别提高 10.86%、12.81% 和 22.67%。在与 TCA 之间的定位距离误差的比较中积累分布为 90% 时分别提高 6.19%、12.81% 和 15.97%。与 JDA 之间的距离误差比较中在积累分布为 90% 的时候分别提高 10.86%、19.62% 和 30.39%。在与 BDA 之间的距离误差比较中在积累分布为 90% 的时候分别提高 10.86%、20.06% 和 34.05%。

表 4 为使用 MKFGFK、决策树、TCA、BDA 与 JDA 方法进行定位的平均距离误差表。从表中可以看出,由于 0 层的参考点及样本数量比其他楼层多,使得其平均距离误差相对较小。在-2 层参考点及样本数量较少的情况下,MKFGFK 方法的平均距离误差仍可比直接使用决策树方法减小 7.38 m。BDA 方法通过计算平衡因子来选取适配边缘分布与条件分布的重要性。

表 4 平均定位距离误差

Table 4 Average distance error of positioning m

	-2 层	0 层	1 层
MKFGFK	11.29	5.65	9.25
DT	18.67	6.92	11.99
TCA	12.39	6.82	11.02
BDA	19.64	9.32	15.33
JDA	19.65	9.65	15.44

实验结果中 BDA 与 JDA 性能基本相同,表明 BDA 退化为了 JDA,又由于 JDA 的伪标签准确率较低,使得结果均差于其他方法。TCA 方法适配了 2 个领域数据的边缘分布,但由于不需要预测伪标签,结果好于 BDA 与 JDA。MKFGFK 方法无需考虑数据的边缘分布或条件分布的相似性,通过融合多个领域的不变性特征,利用不变性特征直接对数据进行变换,从而得到更稳定的定位结果。

4 总结

本文提出一种多源无线信号跨时间测地线流核融合的指纹定位方法,以降低指纹定位在无线信号受到环境变化及信号时变等因素影响下定位误差。本方法通过融合多种来源的无线信号,相比传统方法使用单一信号源增加了信号指纹特征的数量。跨时间的测地线流核融合将不同时段之间无线信号的不变性特征都融合在一起,增强了指纹定位在时变环境下的稳定性。分类器集成通过对多个随机采样得到的样本及特征训练基分类器,学习不同样本间的差异性从而提高整体分类

器的泛化性能,获得更高的定位精度。

参考文献

[1] Bahl P, Padmanabhan V N. RADAR: an in-building RF-based user location and tracking system [C] // Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No. 00CH37064). March 26-30, 2000, Tel Aviv, Israel. IEEE, 2000: 775-784.

[2] Roos T, Myllymäki P, Tirri H, et al. A probabilistic approach to WLAN user location estimation [J]. International Journal of Wireless Information Networks, 2002, 9(3): 155-164.

[3] 杨如民,陈敏,余成波. 基于贝叶斯概率优化的 Wi-Fi 室内定位算法[J]. 计算机应用与软件, 2021, 38(2): 97-102,144.

[4] Brunato M, Battiti R. Statistical learning theory for location fingerprinting in wireless LANs [J]. Computer Networks, 2005, 47(6): 825-845.

[5] 李梦梦. 基于机器学习的 WIFI 室内定位技术研究[D]. 西安: 西安科技大学, 2020.

[6] Wang X Y, Gao L J, Mao S W, et al. DeepFi: deep learning for indoor fingerprinting using channel state information[C]// 2015 IEEE Wireless Communications and Networking Conference (WCNC). March 9-12, 2015, New Orleans, LA, USA. IEEE, 2015: 1666-1671.

[7] 王玉环. 基于深度学习的 RSSI 室内指纹定位研究[D]. 北京: 北京交通大学, 2020.

[8] 俞敬. 基于卷积神经网络的新型 CSI 室内定位方法研究 [D]. 天津: 天津工业大学, 2019.

[9] Wang B, Chen Q Y, Yang L T, et al. Indoor smartphone localization via fingerprint crowdsourcing: challenges and approaches[J]. IEEE Wireless Communications, 2016, 23(3): 82-89.

[10] 季玉凤. 基于众包的室内 Wi-Fi 指纹地图构建方法 [D]. 江苏镇江: 江苏大学, 2020.

[11] 章裕润,吴飞,毛万葵. 基于移动众包的地磁传感器阵列室内指纹定位技术[J]. 传感器与微系统, 2019, 38(12): 36-39.

[12] Kim Y, Chon Y, Cha H. Smartphone-based collaborative and autonomous radio fingerprinting [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2012, 42(1): 112-122.

[13] 邓英. 基于多传感器信息融合的室内定位方法研究[D]. 重庆: 重庆理工大学, 2020.

[14] 汪振,陆奎. 基于粒子滤波融合 Wi-Fi 和 PDR 的室内定位研究[J]. 电脑知识与技术, 2020, 16(9): 39,42.

[15] 宋世铭,王继,韩李涛. 一种基于改进粒子滤波算法的室内融合定位方法[J]. 导航定位学报, 2020, 8(1): 99-106.

[16] 刘勋. 基于多维信息融合的位置指纹室内定位算法研究 [D]. 河北秦皇岛: 燕山大学, 2020.

- [17] 韩冷,戴鹏,阳媛,等. 多源信息融合的室内定位方法[J]. 传感器与微系统, 2020, 39(7): 21-24.
- [18] 段林甫. 多源异构融合定位方法研究[D]. 成都: 电子科技大学, 2018.
- [19] 赵万龙. 多源融合定位理论与方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2018.
- [20] Zheng V W, Xiang E W, Yang Q, et al. Transferring localization models over time[C]//Proceedings of the Twenty-Third National Conference on Artificial Intelligence, Chicago, Illinois, AAAI, 2008, 3: 1421-1426.
- [21] Sorour S, Lostonlen Y, Valae 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.
- [22] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. IEEE Transactions on Neural Networks, 2011, 22(2): 199-210.
- [23] Long M S, Wang J M, Ding G G, et al. Transfer feature learning with joint distribution adaptation[C]//2013 IEEE International Conference on Computer Vision. December 1-8, 2013, Sydney, NSW, Australia. IEEE, 2013: 2200-2207.
- [24] Wang J D, Chen Y Q, Hao S J, et al. Balanced distribution adaptation for transfer learning[C]//2017 IEEE International Conference on Data Mining (ICDM). November 18-21, 2017, New Orleans, LA, USA. IEEE, 2017: 1129-1134.
- [25] Gopalan R, Li R N, Chellappa R. Domain adaptation for object recognition: an unsupervised approach[C]//2011 International Conference on Computer Vision. November 6-13, 2011, Barcelona, Spain. IEEE, 2011: 999-1006.
- [26] Gong B Q, Shi Y, Sha F, et al. Geodesic flow kernel for unsupervised domain adaptation[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. June 16-21, 2012, Providence, RI, USA. IEEE, 2012: 2066-2073.
- [27] Ramdas A, Trillos N G, Cuturi M. On Wasserstein two-sample testing and related families of nonparametric tests[J]. Entropy, 2017, 19(2): 47.
- [28] Popleteev A. Indoor positioning using ambient radio signals: data acquisition platform for a long-term study[C]//2016 13th Workshop on Positioning, Navigation and Communications (WPNC). October 19-20, 2016, Bremen, Germany. IEEE, 2016: 1-5.