

设备自适应的无线信号特征提取与定位方法*

谷洋^{1,2,3}, 蒋鑫龙^{1,2,3}, 刘军发^{1,2}, 陈益强^{1,2}

¹(中国科学院 计算技术研究所 普适计算研究中心, 北京 100190)

²(北京市移动计算与新型终端重点实验室, 北京 100190)

³(中国科学院大学 信息科学与工程学院, 北京 100190)

通讯作者: 谷洋, E-mail: guyang@ict.ac.cn

摘要: 近年来,基于 Wi-Fi 的无线定位研究日益受到关注.然而,在实际应用中,由不同终端设备的差异引起的定位偏差是一个重要问题.针对此问题提出了一种免标定、无监督的,SSDR(signal strength difference ratio)解决方法.考虑采集训练数据的设备和测试数据的设备之间信号存在差异,首先将信号指纹特征进行去线性化处理以获取新的特征;然后结合 AP(access point)对定位结果的影响,提出了基于 AP 影响因子计算距离的标准;最后根据新的特征和距离计算准则消除不同设备之间的差异以实现定位.在真实的室内无线环境下的实验结果表明,所提出的 SSSR 方法相比于传统的直接基于信号强度和欧式距离计算准则的定位方法而言,可以提高 10%~20%的定位精度,增强了无线定位系统的实际可用性.

关键词: 设备差异性;特征融合;AP 影响因子;Wi-Fi 室内定位;机器学习

中文引用格式: 谷洋,蒋鑫龙,刘军发,陈益强.设备自适应的无线信号特征提取与定位方法.软件学报,2014,25(Suppl.(2)): 12-20. <http://www.jos.org.cn/1000-9825/14018.htm>

英文引用格式: Gu Y, Jiang XL, Liu JF, Chen YQ. Device adaptive wireless signal feature extraction and localization method. Ruan Jian Xue Bao/Journal of Software, 2014, 25(Suppl.(2)): 12-20 (in Chinese). <http://www.jos.org.cn/1000-9825/14018.htm>

Device Adaptive Wireless Signal Feature Extraction and Localization Method

GU Yang^{1,2,3}, JIANG Xin-Long^{1,2,3}, LIU Jun-Fa^{1,2}, CHEN Yi-Qiang^{1,2}

¹(Department of Pervasive Computing, Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

²(Beijing Key Laboratory of Mobile Computing and Pervasive Device, Beijing 100190, China)

³(Department of Information Science and Engineering, University of Chinese Academy of Sciences, Beijing 100190, China)

Corresponding author: GU Yang, E-mail: guyang@ict.ac.cn

Abstract: In recent years, research on Wi-Fi based indoor localization draws increasing attention. However, in practical applications, the localization error caused by device variance is a severe problem. In this paper, a new calibration-free and unsupervised method, SSSR (Signal Strength Difference Ratio) is proposed to solve this issue. Considering the signal variance between training devices and testing devices, SSSR first removes the linear effect of fingerprint to get new features. It then puts forward a distance calculation criterion with AP impact factor according to the effect of AP. Finally SSSR eliminates the variance of devices and realizes indoor localization based on the new features and distance calculation criterion. The experiment deployed in real indoor wireless environment shows, compared with traditional indoor localization methods, the proposed SSSR can increase the indoor localization accuracy by 10%~20%, which greatly improves the practical usability of indoor localization system.

Key words: device variance; feature fusion; AP impact factor; Wi-Fi indoor localization; machine learning

近年来,随着智能终端的普及和网络通信的日益发展,人们通过移动终端在室内获取所需服务的方式已经

* 基金项目: 国家自然科学基金(61173066); 广东省中国科学院全面战略合作项目(2011A090100001)

收稿时间: 2013-06-15; 定稿时间: 2013-08-21

十分普遍.室内定位就有许多系统方案,如:辅助 GPS(assisted GPS,简称 AGPS)定位、超声波定位、蓝牙 (Bluetooth)定位、Wi-Fi 定位、红外线定位、移动通信基站定位等.其中,基于 Wi-Fi 的室内定位^[1]由于其简便易行的优点成为室内定位的主要方法之一.

但是,基于 Wi-Fi 的室内定位在实际应用中涉及到设备差异性的问题,即不同的智能终端使用的 Wi-Fi 模块、封装材料、封装方式等不尽相同,导致了终端获取的 Wi-Fi 信号强度存在明显差异,这些差异甚至可以高达 25dbm^[2].已有工作表明,在不对设备的差异性做任何处理的情况下,使用最近邻(nearest neighbor,简称 NN)的指纹方法进行定位,可使得房间级的定位精度降到 10%^[3],严重降低了室内定位系统的实际可用性.

因此为了保证室内定位的精度,从实际应用的角度解决不同终端的信号差异性,提出了免标定、无监督的数据处理方法和基于 AP(access point) 影响因子的距离计算准则,使得定位精度相比于传统的基于信号强度和标准欧式距离计算准则的定位得到大幅提升.本文第 1 节介绍无线定位领域设备差异性处理的相关工作.第 2 节详细介绍提出的 SDDR(signal strength difference ratio)方法,包括特征融合和 AP 影响因子(计算距离的准则).第 3 节介绍在真实的室内环境下定位实验,并详细对比分析不同方法的实验结果;第 4 节总结本文的工作.

1 相关工作

对基于指纹模型的 Wi-Fi 室内定位而言,由于训练阶段和测试阶段使用的设备不同,从而造成离线训练数据和在线定位数据的差异,导致定位精度的下降.因此针对无线定位的设备差异性,问题已展开了很多相关工作.

众多实验研究均发现不同的硬件设备获取的信号强度之间存在线性变化的关系^[4-6].Haeberlen 等人^[5]对于这种线性关系提出了 3 种处理设备差异性的方法:第 1 种是手动标定方法,在训练阶段采集足够多的不同设备的数据对,使用最小二乘的方法求解出线性变换的参数(斜率,截距);当测试数据到来时,用求出的斜率和截距将测试信号向训练设备采集的信号空间转化,使用该方法可以较大幅度的提升定位精度;第 2 种是准自动标定(quasi-automatic calibration)方法,该方法通过 Markov 定位技术选择置信度高的数据,并通过这些高置信度的数据求解斜率和截距这两个参数,以此对测试数据进行处理;第 3 种是自动标定方法(automatic calibration),利用历史观测数据估计变化的参数,并使用 EM(expectation maximization)方法选择使得估计位置出现的概率最大的参数,而后实现定位.相比于 Haeberlen 的工作,Kjaergaard^[3]提出了免标定的双曲线指纹定位方法,该方法将获取到的指纹向量的分量进行成对相除的转化,以此消除了线性变换中斜率项的影响,使用该方法定位精度可以得到一定的提升.与 Kjaergaard 的想法类似,Dong^[6]提出了将指纹向量的分量进行成对相减的方法,以此消除线性变换中截距项的影响.使用此方法后,可以将基于 NN 的定位精度提升 20%,将基于 Bayesian Inference(BI)方法的定位精度提升 12%.相比于以往的对数据直接进行调整的工作,Tsui 等人^[7]在传统的离线训练阶段和在线定位阶段之间增加了一个在线的调整阶段,首先通过测试数据选出数据库中最相似的数据,并利用选出来的数据和测试数据求解线性变化的斜率和截距,当另有测试数据到来时,利用已求解出的两个参数对测试数据进行转化,转化后再选出与之最相似的数据,重新求解这两个参数,以此循环该操作,当测试数据越来越多时,求解的斜率和截距这两个参数就越准确,定位精度也得到不断的提升.Kim 等人^[8]通过观察训练阶段和定位阶段信号的差异,提出了使用 AP 峰值的方法,该方法认为信号强度从训练阶段到定位阶段的映射是一种线性映射,即在训练阶段有最强信号强度的位置,在定位阶段也仍然是信号强度最强的位置,以此解决设备差异给信号强度带来的影响.Figuera 等人^[9]提出了 3 种设备标定的方法以解决设备的差异性,包括:非线性调整、敏感阈值校正和 Generalized Regression Neural Network(GRNN)方法.通过这 3 种方法处理,在使用 5 种不同设备的情况下,可以实现平均误差距离(mean distance error,简称 MDE)为 3m 的定位精度.相比通过线性变化的关系解决设备差异性问题的方法,核方法和流形学习方法也被用来解决该问题.Park 等人^[10]通过使用具有较大核宽的核估计方法来减少概率估计的差异,在使用合适的核宽时,即使设备间存在较大差异,也可实现 3m~4m 的定位精度.利用源数据和部分标定目标数据,Sun 等人^[11]通过流形对齐(manifold alignment)的方法解决了设备的差异性,保证了有效的定位精度.

虽然,已有一些解决室内定位领域中设备的差异性问题的研究工作,但是这些方法还存在一些不足.传统的

免标定设备差异性处理方法^[3,6]没有同时解决斜率和截距的影响,因此定位精度还可以得到提升;同时消除斜率和截距影响的处理方法^[7]需要一个在线调整阶段,在实际应用中不够便捷;而通过手工标定消除设备差异性的方法^[5]需要人工采集大量数据,不具有实际可用性.因此为了消除线性变化的影响,解决不同设备的差异性,提高实际应用中室内定位系统的定位精度,文章提出了 SDDR 方法.通过融合消除线性变换中截距和斜率影响的特征,以综合性的消除不同设备带来的信号差异;同时,提出了基于 AP 影响因子的距离计算准则,克服 AP 的差异性带来的定位误差,以提高定位精度.

2 本文工作

2.1 特征融合

众多研究工作^[4-6]表明,由于离线阶段和在线阶段的设备不同而导致的接收信号强度的差异可以用线性变化进行描述.因此,假设理想条件下的标准设备获取到的信号指纹向量为 $V^0 = (RSS_{j1}^0, RSS_{j2}^0, \dots, RSS_{jk}^0, \dots, RSS_{jN}^0)$, 其中上标 0 表示理想条件的标准设备,下标 j 表示第 j 条指纹向量, N 为采集到的 AP 个数.对于不同的设备 D_i , 在相同位置获取到的信号指纹向量为 $V^{D_i} = (RSS_{j1}^{D_i}, RSS_{j2}^{D_i}, \dots, RSS_{jk}^{D_i}, \dots, RSS_{jN}^{D_i})$, 根据文献[4], 不同的设备和理想条件下的标准设备获取的信号指纹满足如下关系:

$$V^{D_i} = a^{D_i} \times V^0 + b^{D_i} + \varepsilon \quad (1)$$

把指纹向量展开成分量的形式,则有:

$$RSS_{jk}^{D_i} = a^{D_i} \times RSS_{jk}^0 + b^{D_i} + \varepsilon \quad (2)$$

$$V^{D_i} = (a^{D_i} \times RSS_{j1}^0 + b^{D_i} + \varepsilon, a^{D_i} \times RSS_{j2}^0 + b^{D_i} + \varepsilon, \dots, a^{D_i} \times RSS_{jN}^0 + b^{D_i} + \varepsilon) \quad (3)$$

令 $mRSS_j^{D_i}$ 表示一组指纹向量的均值,则有:

$$mRSS_j^{D_i} = \text{mean}(RSS_{j1}^{D_i}, RSS_{j2}^{D_i}, \dots, RSS_{jN}^{D_i}) = \frac{RSS_{j1}^{D_i} + RSS_{j2}^{D_i} + \dots + RSS_{jN}^{D_i}}{N} = \frac{a^{D_i} \times (RSS_{j1}^0 + RSS_{j2}^0 + \dots + RSS_{jN}^0)}{N} + b^{D_i} + \varepsilon \quad (4)$$

将指纹向量的分量减去其均值,得到新的指纹向量为

$$V^{D_i,1} = (RSS_{j1}^{D_i} - mRSS_j^{D_i}, RSS_{j2}^{D_i} - mRSS_j^{D_i}, \dots, RSS_{jN}^{D_i} - mRSS_j^{D_i}) \\ = \left(\begin{array}{c} a^{D_i} \times RSS_{j1}^0 - \frac{a^{D_i} \times (RSS_{j1}^0 + RSS_{j2}^0 + \dots + RSS_{jN}^0)}{N}, \dots, \\ a^{D_i} \times RSS_{jN}^0 - \frac{a^{D_i} \times (RSS_{j1}^0 + RSS_{j2}^0 + \dots + RSS_{jN}^0)}{N} \end{array} \right) \quad (5)$$

由此可以看到,在不需要任何标定和监督的情况下,获取到的新的指纹向量消除了截距 b^{D_i} 对不同设备造成的差异.

在消除截距的影响后的工作是消除斜率项对信号强度变化的影响.这一部分借鉴了 Kjaergaard^[3]的工作,用双曲线相除的方法去除斜率项的影响.具体可表示为

$$V^{D_i,2} = \left(\log \left(\frac{RSS_{j1}^{D_i}}{mRSS_j^{D_i}} \right), \log \left(\frac{RSS_{j2}^{D_i}}{mRSS_j^{D_i}} \right), \dots, \log \left(\frac{RSS_{jN}^{D_i}}{mRSS_j^{D_i}} \right) \right) \quad (6)$$

虽然使用该处理方法并不能完全的消除斜率的影响,但是在截距项的影响不明显时,该方法是可以起到很好的作用的.

最终,数据的新特征是融合消除截距和斜率的影响的两部分特征,既新特征为

$$V_{new}^{D_i} = [V^{D_i,1} \quad V^{D_i,2}] = \left(\begin{array}{c} RSS_{j1}^{D_i} - mRSS_j^{D_i}, RSS_{j2}^{D_i} - mRSS_j^{D_i}, \dots, RSS_{jN}^{D_i} - mRSS_j^{D_i}, \\ \log \left(\frac{RSS_{j1}^{D_i}}{mRSS_j^{D_i}} \right), \log \left(\frac{RSS_{j2}^{D_i}}{mRSS_j^{D_i}} \right), \dots, \log \left(\frac{RSS_{jN}^{D_i}}{mRSS_j^{D_i}} \right) \end{array} \right) \quad (7)$$

新的特征融合了消除截距和斜率影响的特征,解决了单独消除截距影响和斜率影响的不足,并避免了手动

标定数据的困难和在线调整数据的不便捷.

2.2 AP影响因子距离计算准则

在获取到融合的特征后,并不直接用来定位,因为选取 AP 的方式以及距离计算的准则都会影响到定位结果.

在传统的室内定位中,即使对于同一设备,训练阶段和测试阶段采集到的 AP 也会存在差异.如若在训练阶段和测试阶段使用不同的移动终端,设备的差异性会因为获取到的 AP 存在的明显区别而变得更大.因此,在实际定位应用中,要考虑到 AP 的影响,克服 AP 的差异性导致的定位误差.现有的室内定位方法较多采用确定 AP 的方式,然后对训练阶段和测试阶段缺失的 AP 进行补齐,常用的补齐方法是使用-100dbm 作为缺省值^[12],然而该方法会导致定位错误,见表 1.

表 1 采集 AP 数据对比表

位置	AP							
	AP1	AP2	AP3	AP4	AP5	AP6	AP7	...
P1			-45	-71	-66	-34		
P2	-79	-63	-54	-61				
P3	-80	-68	-56	-70				
P4	-81	-70	-58	-90				
P5			-63	-82		-59		
...
P?	-80	-66	-57	(?)				
P??	-75	-63		-72			-88	

表 1 中有两个待定位置 P?和 P??.假设待定位置 P?点原本与 P2 点相邻(相邻是指在真实的物理空间中,如果点 A 与点 B 的距离比点 A 到其他点的距离都小,则认为 AB 两个点是相邻的),若直接将其缺失的 AP4 信号值赋值为-100,则其位置会与 P4 点相邻,而 P?与 P4 点在物理位置上可能相差甚远,这就增加了定位误差.

如果不使用补齐 AP 的方法,而使用训练数据和测试数据共同收集到的 AP 进行误差计算(见表 1),仍然会导致定位错误.假设待定位置 P??点原本与 P3 点相邻,但是通过选取共有的 AP,直接按照均方误差(mean square error,简称 MSE)准则计算距离,会发现 P??点与 P3 的距离为

$$\frac{(-75 - (-80))^2 + (-68 - (-63))^2 + (-70 - (-72))^2}{3} = 18,$$

而 P??点与 P1 的距离为 $\frac{(-70 - (-71))^2}{1} = 1$,比较距离可以得出 P??点与 P1 更近的结论,这就也导致了定位的误差.

因此,在实际应用中要充分考虑 AP 的影响.假定训练数据中的某条数据 $\{(x_i, y_i), RSS_{i1}, RSS_{i2}, \dots, RSS_{ip}, \dots, RSS_{iN}\}$,其中 i 表示第 i 条数据, (x_i, y_i) 为训练数据采集位置的坐标, N 为收集到的 AP 的个数, RSS_{ip} 为第 p 个 AP 的信号强度;某条测试数据为 $\{RSS'_{j1}, RSS'_{j2}, \dots, RSS'_{jq}, \dots, RSS'_{jM}\}$,其中 M 为收集到的 AP 个数, RSS'_{jq} 为第 q 个 AP 的信号强度.其中两条数据共同含有的 AP 个数为 N_{common} ,由共同的 AP 重新构成的训练数据则表示为

$$\{(x_i, y_i), RSS_{i1}^{common}, RSS_{i2}^{common}, \dots, RSS_{ip}^{common}, \dots, RSS_{iN_{common}}^{common}\},$$

重新构成的测试数据表示为

$$\{RSS_{j1}^{tcommon}, RSS_{j2}^{tcommon}, \dots, RSS_{jq}^{tcommon}, \dots, RSS_{jN_{common}}^{tcommon}\}.$$

通过获取共同的 AP 消除 AP 种类的影响之后,用第 2.1 节的方法获取融合的特征.在通过融合的特征计算测试点和训练点之间的距离时,需要考虑到匹配上的 AP 个数的影响,因为在实际定位应用中面临这样的困境:匹配上的 AP 个数少,计算得到的距离小;匹配上的 AP 个数多,计算得到的距离大;而实际情况是距离大(匹配上 AP 个数多)的位置与测试位置更为接近.

因此,文章提出 AP 影响因子:

$$AP_{coeff} = \frac{N}{N_{common}} \times \frac{M}{N_{common}} \quad (8)$$

在计算距离时,首先判断 AP 影响因子的倒数 $\frac{1}{AP_{coeff}}$ 是否大于 0.25(0.25 影响因子的取值来源于假设训练数据和测试数据中均有两个 AP,并有一个共同的 AP 时,AP 影响因子的倒数 $\frac{1}{AP_{coeff}} = \frac{1}{2} \times \frac{1}{2} = 0.25$),当 $\frac{1}{AP_{coeff}} \geq 0.25$ 时,证明匹配的数据的置信度较高,将 AP 影响因子作用到由特征得到的距离 D_{ij} (通过特征计算距离 D_{ij} 按照传统的欧式距离计算方法)上计算新的距离 D_{ij}^{new} :

$$D_{ij}^{new} = AP_{coeff} \times D_{ij} \quad (9)$$

当共同的 AP 在训练数据和测试数据所占的比例大时, AP_{coeff} 比较小,当 AP_{coeff} 作用于距离 D_{ij} 时,会进一步增强匹配的置信度;与之类似当共同的 AP 的训练数据和测试数据所占的比例小时(但仍大于 0.25), AP_{coeff} 比较大,当 AP_{coeff} 作用于距离 D_{ij} 时,通过增大距离降低匹配的置信度,以此提高定位结果的可靠性.当 $\frac{1}{AP_{coeff}} < 0.25$,证明匹配上的 AP 在训练数据和测试数据中占的比重很小(该匹配的置信度低),则新的距离表示为

$$D_{ij}^{new} = D_{ij} + 10000 \quad (10)$$

经过该处理后,即使由特征计算得到的距离 D_{ij} 较小,但在新距离计算时引入了一个较大的偏置(该偏置可根据用户的实际需求设置一个较大值即可)来降低该结果候选的概率,因此可以在一定程度上降低定位误差.

综上,SSDR 方法的流程图如图 1 所示.

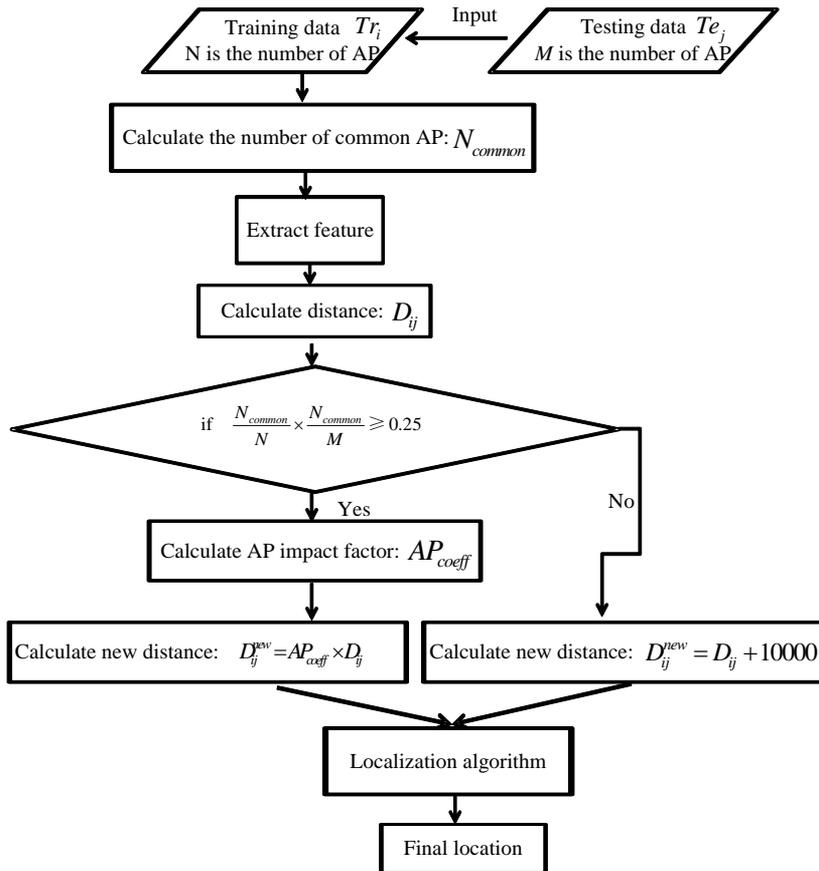


图 1 算法流程图

3 实验及分析

3.1 实验数据采集

为了处理不同的采集设备带来的信号差异,文章使用了 4 种不同的智能终端:HTC Sensation(Sensation), HTC Desire(Desire), 小米 2 代(MI 2),华为 S8600(S8600).在某研究所的一楼大厅 84m×35m 的范围内进行数据采集,数据的采集距离间隔为 2m~4m,该环境的布局如图 2 所示.实验并不需要对 AP 的位置进行刻意的布置,直接使用环境中现存的 AP 即可(也不需要知道这些 AP 的位置),最终统计这个实验环境可以得到 89 个 AP(包括了较为稳定存在的 20 个 AP 和偶然出现的 AP).

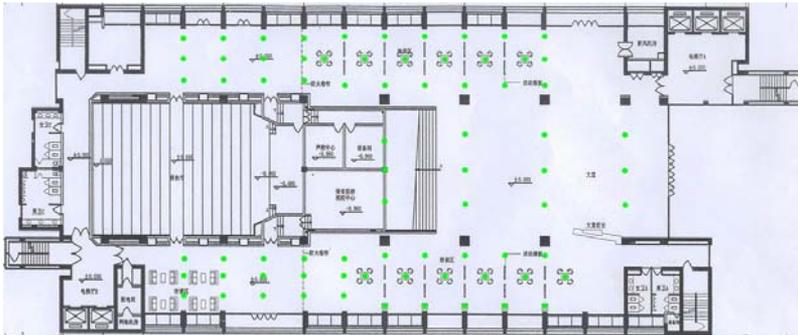


图 2 环境布局图

图中的点表示训练数据采集的位置.在该实验环境下,于每天的不同时段,手持多种终端设备在相应的位置进行数据采集,每个位置进行多次数据采集,最终采集到 6 000 条数据.

3.2 实验分析

实验首先分析了不同的手机在相同的时间(环境下)获取的信号强度的变化,如图 3 所示.从图 3 中可以看到,对于不同的设备即使在相同的环境下,获取的信号指纹也存在着明显的差异,这些差异正是由于引言所述的 Wi-Fi 模块、封装材料、封装方式等因素不尽相同而导致的.比较 MI2 与 Sensation 可发现,其信号强度差异高达 20dbm.因此,如果不进行信号差异性处理,必然会导致定位精度的下降.

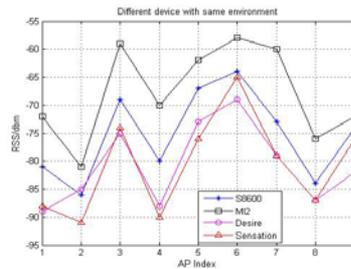


图 3 不同设备相同环境采集的信号指纹对比图

实验中主要使用了已存在的 3 种解决设备差异性的方法 Signal Strength Ratio (SSR)^[3],Signal Strength Difference(SSD)^[6],手动标记方法(Manual);没有经过差异性处理的方法 OS(original signal),可以学习数据分布特性和抽象特征的深度学习方法(deep leaning,简称 DL)^[13],以及具有快速学习能力和随机特征映射的极速学习机(extreme learning machine,简称 ELM)^[14]等 6 种方法与提出的方法进行比较.对于 DL 和 ELM 的参数设置如下:

DL:1) 使用栈自动编码结构(stack auto encoder,简称 SAE)^[15],一个输入层,两个隐层,使用 Softmax 分类器,一个输出层.

2) 两个隐层的隐节点个数均为 100,使用 sigmoid 做隐层的编码函数.

3) 迭代寻找最优值的次数为 400,权重衰减惩罚因子为 3e-3,权重稀疏惩罚因子为 3,隐节点稀疏约束因子为 0.1.

ELM:1) 一个输入层,一个隐层,一个输出层.

2) 隐层的隐节点个数为 1 000,使用 sigmoid 作为隐层的激励函数.

3) 惩罚因子 C 取值为 $2^{[16]}$.

对于不同设备差异性影响,考虑了两种具有代表性的情况:1) 使用小米二代(MI2)采集训练数据,其他 3 种设备采集测试数据.实验中发现 MI2 手机采集到的信号强度明显高于其他设备(由图 3 观测可得),即与其他设备之间的差异性很大;2) 使用 HTC Sensation 采集训练数据,其他 3 种设备采集测试数据.实验中发现 Sensation 的数据与 HTC Desire、S8600 和 MI2 的信号差异性程度依次增强.因此,这两种组合可以检验不同程度的设备差异时算法的定位能力.具体实验结果如下:

(1) MI2 作为训练数据采集设备

方法 SSR,SSD,Manual,OS,SSDR 均采用 NN 作为定位结果的决策,DL 与 ELM 按照上述的参数设置.具体的实验结果显示如图 4 所示.

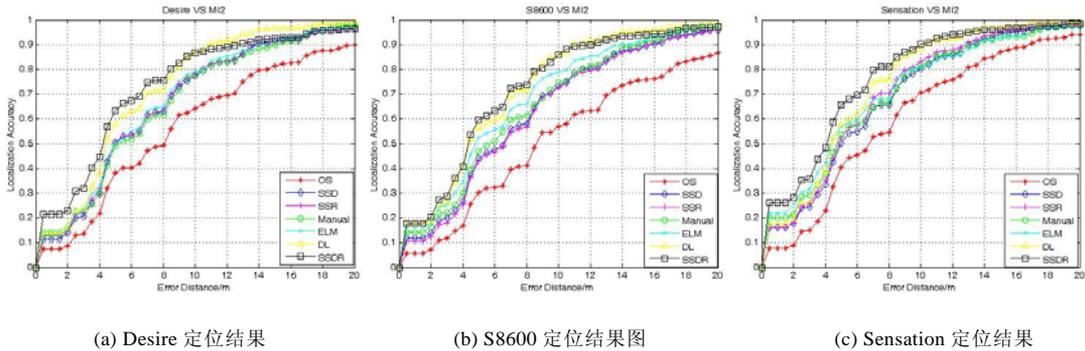


图 4 MI2 为训练数据采集设备,其他 3 种终端作为测试数据采集设备的定位结果

从图 4 的 3 个图可以发现,相比于其他 6 种方法而言,基于 SSSDR 的方法得到的定位精度是最好的,尤其是当定位误差范围较小的时候(8m 以内),相比于没有进行线性变换的 OS 方法来说,可以将定位精度提升 20%~30%,显著地提高了定位结果.

对于 SSR 与 SSD 方法来说,SSR 只解决了设备差异性中斜率项的影响,SSD 只解决了设备差异性中截距项的影响,而由图 3 可以看到,训练设备 MI2 与其他 3 种设备存在明显的线性变化关系,即斜率和截距对变化都有明显影响,因此 SSR 与 SSD 方法不能完全消除线性关系.相比于这两种方法,SSDR 方法结合了消除截距和斜率影响的特征,并引入 AP 影响因子计算距离,在综合消除不同设备影响的同时去除了由于 AP 种类和个数的变化给定位结果带来的影响,因此提高了定位精度.

对于 Manual 方法而言,通过求解的线性参数调整测试数据,减少了设备之间的差异性,因此得到的结果比 OS 方法提升 10%左右.但该方法的结果没有明显的优于 SSD 与 SSR 方法,原因在于标定数据的获取比较费时费力,所以在实验中只采集了少量的标定数据求解线性变化的参数,导致了求得的结果不是最优值;此外该方法在计算距离时没有考虑到 AP 的个数和种类的影响,因此结果不及 SSSDR.

DL 是一种可以充分学习数据分布结构和抽象特征的机器学习方法.DL 的隐层是一个特征提取的过程,数据经过隐层处理后可以得到高层抽象特征,再经过分类器的处理后得到输出结果.对于无线定位的数据来说,虽然不同的设备之间存在一定的差异,但是经过 DL 的特征提取后可以得到反映无线数据内在联系的特征,在一定程度上消除了设备的差异性,因此定位精度得到提升;ELM 是一种单隐层的前馈神经网络,利用随机赋值的输入权重和隐层节点偏置求解全局最优的输出权重,保证了学习能力.对于室内无线定位来说,无线数据经过隐层得到的随机特征可以在一定程度上缓解不同设备带来的信号差异,因而提高了定位精度.但是由于没有考虑到 AP 的影响问题,在多数情况下,这两种方法的结果不及 SSSDR.

(2) Sensation 作为训练数据采集设备

在该实验情况下,方法 SSR,SSD,Manual,OS,SSDR 均采用 NN 作为定位结果的决策,DL 与 ELM 仍然按照上述的参数设置.具体的实验结果显示如图 5 所示.

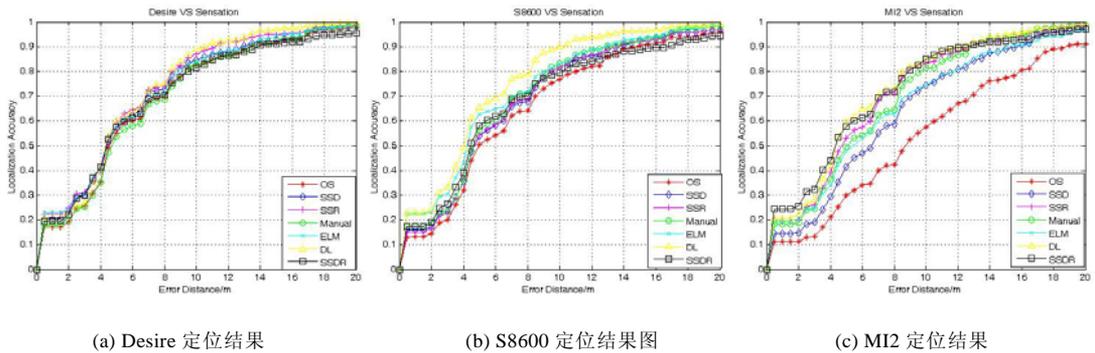


图 5 Sensation 为训练数据采集设备,其他 3 种终端作为测试数据采集设备的定位结果

从上图 5(a)、(b)可以发现,当 Desire 作为测试数据采集设备时,测试数据与训练数据之间的差异不是很明显(可以参见图 3 的不同设备信号对比图),因此各种方法得到的结果比较相近.当 S8600 作为测试数据采集设备时,存在着一定的线性变化,因此经过处理的方法 SSD,SSR,SSDR,Manual 都比 OS 的结果好;而 ELM 和 DL 方法的结果稍微优于 SSSDR,是因为这两种方法的输入数据都标准化到 0 均值和 1 方差,而这种归一化的处理是以矩阵为单元进行的统计操作,可以降低偶然因素的影响,因此定位精度稍优于 SSSDR.然而,对实际的定位系统来说,测试数据是逐条到来的,因此这种统计性的处理方法在实际应用缺少可行性.此外,DL 虽然得到较好的结果,但是其多层学习和反馈调整(fine-tune)是一个耗时的过程,无法保证用户对定位实时性的要求.因此,SSDR 的逐条数据处理方法获得的结果虽稍逊于 DL 和 ELM,但其更具有实际应用价值.当 MI2 作为测试数据采集设备时获取的信号与训练数据存在明显的线性差异,因此相比于其他方法,SSDR 仍然得到了最好的结果.

同时,考虑到 AP 情况对室内定位的影响,实验也分析了 AP 个数对 SSSDR 方法的影响.该实验是以 MI2 为训练数据采集设备,Sensation 为测试数据采集设备,在定位误差距离为 2.5m 时的结果如图 6 所示.

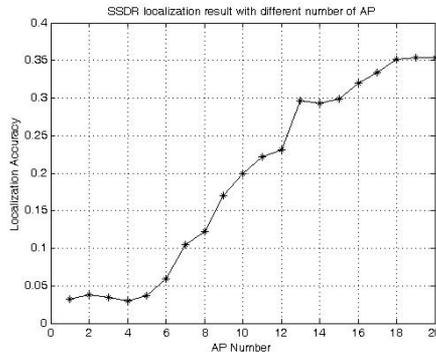


图 6 AP 数目对 SSSDR 方法定位结果的影响

从图 6 中可以看到,随着 AP 数目的增多,定位的精度从 3%提升到 35%.这是因为 AP 数目的增加,表征位置的指纹信息更丰富,使得不同位置间的区分度增加.但是当 AP 个数足够多(18~20)时,定位精度基本保持平稳,因为已有足够充分的 AP 指纹信息表示位置,再增加 AP 只会带来信息的冗余而对定位精度的提升没有明显帮助.

4 总结

训练数据和测试数据采集设备的差异性给实际的室内定位应用系统带来了较大的影响.本文对此提出了 SSSDR 方法,该方法可以在免标定、无监督的情况下消除不同设备的差异性的影响,并通过 AP 影响因子计算距离,避免了由 AP 的差异带来的定位误差.在真实的室内环境下的实验表明,SSDR 方法相比于不做任何处理,直接基于 RSS 进行定位的方法而言,可以提高 10%~20%的精度;相比于已有的 SSD、SSR、手动标定的方

法,DL和ELM这5种方法,也可以得到不同程度的提升.因此,SSDR以定位精度高,简单便捷等优势解决了Wi-Fi室内定位的设备差异性问题,提升了定位系统的实际可用性.

References:

- [1] Bahl P, Padmanabhan VN. RADAR: An in-building RF-based user location and tracking system. In: Proc. of the INFOCOM 2000, 19th Annual Joint Conf. of the IEEE Computer and Communications Societies. Tel Aviv: IEEE Press, 2000. 775–784.
- [2] Kaemarungsi K. Distribution of WLAN received signal strength indication for indoor location determination. In: Proc. of the 1st Int'l Symp. on Wireless Pervasive Computing. Phuket: IEEE Press, 2006. 6–11.
- [3] Kjaergaard MB, Munk CV. Hyperbolic location fingerprinting: A calibration-free solution for handling differences in signal strength. In: Proc. of the IEEE Int'l Conf. on Pervasive Computing. Hong Kong: IEEE Press, 2008. 110–116.
- [4] Tao P, Rudys A, Ladd AM, Wallach DS. Wireless LAN location-sensing for security applications. In: Proc. of the 2nd ACM Workshop on Wireless Security (WiSe). San Diego: ACM Press, 2003. 11–20.
- [5] Haeberlen A, Flannery E, Ladd AM, Rudys A, Wallach DS, Kavraki LE. Practical robust localization over large-scale 802.11 wireless networks. In: Proc. of the ACM MOBICOM 2004, the 10th Annual International Conf. on Mobile Computing and Networking. Philadelphia: ACM Press, 2004. 70–84.
- [6] Dong FF, Chen YQ, Liu JF, NingQ, Piao SM. A calibration-free localization solution for handling signal strength variance. In: Hutchison D, Kanade T, Kittler J, Kleinberg JM, eds. Proc. of the Int'l Workshop on Mobile Entity Localization and Tracking in GPS-Less Environments (MELT). Orlando: Springer-Verlag, 2009. 79–90.
- [7] Tsui AW, Chuang YH, Chu HH. Unsupervised learning for solving RSS hardware variance problem in wifi localization. Mobile Networks and Applications, 2009,14(5):677–691.
- [8] Kim Y, Shin H, ChonY, Cha H. Smartphone-Based Wi-Fi tracking system exploiting the RSS peak to overcome the RSS variance problem. Pervasive and Mobile Computing, 2013. 406–420.
- [9] Figuera C, Alvarez JLR, JimenezI M, Curieses AG, Wilby M, Lopez JR. Time-Space sampling and mobile device calibration for wifi indoor location systems. IEEE Trans. on Mobile Computing, 2011,10(7):913–926.
- [10] Park JG, Curtis D, Teller S, Ledlie J. Implications of device diversity for organic localization. In: Proc. of the INFOCOM 2011. Shanghai: IEEE Press, 2011. 3182–3190.
- [11] Sun Z, Chen Y, Qi J, Liu J. Adaptive localization through transfer learning in indoor wi-fi environment. In: Proc. of the 2008 7th Int'l Conf. on Machine Learning and Applications. San Diego: IEEE Press, 2008. 331–336.
- [12] Pan JJ, Yang Q, Pan SJ. Online co-localization in indoor wireless networks by dimension reduction. AAAI, 2007. 1102–1107.
- [13] Hinton GE, Osindero S, The Y. A fast learning algorithm for deep belief nets. Neural Computation, 2006,18(7):1527–1554.
- [14] Huang GB, Zhu QY, Siew CK. Extreme learning machine: Theory and applications. Neurocomputing, 2006,70(1-3):489–501.
- [15] Le QV, Ranzato MA, Monga R, Devin M, Chen K, Corrado GS, Dean J, Ng AY. Building high-level features using large scale unsupervised learning. In: Langford J, Pineau J, eds. Proc. of the 29th Int'l Conf. on Machine Learning. Edinburgh: ACM Press, 2012. 81–88.
- [16] Huang GB, Zhou H, Ding X, Zhang R. Extreme learning machine for regression and multiclass classification. IEEE Trans. on System, Man, and Cybernetics—Part B: Cybernetics, 2012,42(2):513–529.



谷洋(1988—),女,吉林省吉林市人,博士生,主要研究领域为普适计算,室内定位,机器学习.

E-mail: guyang@ict.ac.cn



刘军发(1973—),男,博士,副研究员,主要研究领域为普适计算,人机交互,室内定位.

E-mail: liujunfa@ict.ac.cn



蒋鑫龙(1988—),男,博士生,主要研究领域为普适计算,室内定位,基于位置的服务.

E-mail: jiangxinlong@ict.ac.cn



陈益强(1973—),男,博士,研究员,主要研究领域为普适计算,人机交互,室内定位.

E-mail: yqchen@ict.ac.cn