

文章编号:1007-5321(2020)04-0095-06

DOI:10.13190/j.jbupt.2019-223

一种基于 Shapelet 算法的指纹定位方法

常紫英¹, 王文涵¹, 李 涛¹, 刘 芬¹, 陈朋朋^{1,2}

(1. 中国矿业大学 计算机科学与技术学院, 徐州 221116; 2. 教育部矿山数字化工程研究中心, 徐州 221116)

摘要: 信道状态信息(CSI)受时空影响较大,导致现有基于CSI的室内定位技术鲁棒性差. 针对这一问题,提出了基于Shapelet算法的指纹定位方法. 在训练阶段将CSI作为原始位置数据,通过 $3\text{-}\sigma$ 异常值处理法和卡尔曼滤波对原始数据进行处理、修正;再使用Shapelet算法提取每个位置的指纹,并建立指纹库;最后使用指纹库构建Shapelet决策树,通过决策树分类实现较为精准的定位. 通过与主成分分析算法以及 k 近邻算法的对比实验,结果表明,该方法在不同时间的定位精度较高,且能保持性能稳定,所需训练集更小.

关键词: 指纹定位; 信道状态信息; Shapelet 算法; 决策树分类

中图分类号: TN92

文献标志码: A

A Fingerprint Localization Method Based on Shapelet Algorithm

CHANG Zi-ying¹, WANG Wen-han¹, LI Tao¹, LIU Fen¹, CHEN Peng-peng^{1,2}

(1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China;

2. Mine Digitization Engineering Research Center of the Ministry of Education, Xuzhou 221116, China)

Abstract: Due to large influence of time and space on channel state information (CSI), the existing CSI-based indoor positioning technology is poor in robustness. Aiming at this problem, a fingerprint positioning method based on Shapelet algorithm is proposed. In the training phase, CSI is taken as the original location data, and the original data is processed and corrected by the $3\text{-}\sigma$ anomaly value processing method and the Kalman filter; then the fingerprint of each location is extracted and the fingerprint database is established by using the Shapelet algorithm; Finally, the fingerprint database is used to construct the Shapelet decision tree, and the decision tree classification is used to achieve more accurate positioning. Compared with the principal components analysis algorithm and k -nearest neighbor algorithm, It is shown that the method has higher positioning accuracy and stable performance at different times, and the training set is smaller.

Key words: fingerprint location; channel state information; Shapelet algorithm; decision tree classification

随着社会生活以及服务业的发展,人们对定位的范围提出了更加精确的要求. 传统的GPS定位技术由于其局限性,无法实现更小范围的室内定位.

常见的室内定位技术包括蓝牙(Bluetooth)^[1],红外线^[2],射频识别技术^[3]等,由于设备的局限性,这些技术并没有被广泛应用. 相较于上述设备,行动热

收稿日期: 2019-10-10

基金项目: 国家自然科学基金项目(51674255, 51874302)

作者简介: 常紫英(1995—),女,硕士生.

通信作者: 陈朋朋(1983—),男,教授, E-mail: chenp@cumt.edu.cn.

点(WiFi)的优势很明显:①随着无线网络的发展,WiFi设备随处可见;②相较于红外线来说,WiFi信号是多路径传播,不会受到光照等环境因素的影响,对环境适应性更强;③与射频识别技术相比,WiFi具有网络服务功能,便于信息的即时解析,且传输效率高。WiFi室内定位技术有很多种实现方法,包括到达时间法(TOA,time of arrival)^[4]、到达时间差法(TDOA,time difference of arrival)^[5]、到达角度法(AOA,angle of arrival)^[6]等。这些算法的实现都需要附加额外的硬件设备。此外,Singh等^[7]还提出了基于接收信号强度指示(RSSI,received signal strength indication)的定位,该方法虽然摆脱了额外硬件设备的限制,但无线信号的传播受环境影响较为严重,导致定位精确度低。为了弥补以上不足,Wang等^[8]提出了一种指纹定位方法,该方法分为2个步骤:第1步需在离线环境下进行,主要目标是构建位置指纹库,需要在各个定位点采集数据样本,经过训练得到各个点的指纹;第2步为实时定位,即根据位置指纹库来判断当前用户的位置。

基于Shapelet算法的指纹定位方法,其主要创新点有以下几个方面:

1) 在数据预处理阶段,使用了 $3\text{-}\sigma$ 异常值处理以及卡尔曼滤波,减少了周围环境及人员走动带来的干扰;

2) 将Shapelet算法做适当改进,用于提取信道状态信息(CSI,channel state information)子序列,作为各位置的指纹,构建指纹库;

3) 通过与现有方法进行比较,证实该方法在使用小样本进行训练的情况下能够实现精确的定位,且保证其鲁棒性。

1 基本原理

1.1 CSI

在无线通信领域,CSI就是通信链路的信道属性,它描述了信号在每条传输路径上的衰落因子,即信道增益矩阵 \mathbf{H} 中每个元素的值,如信号散射(scattering)、环境衰弱(fading multipath fading or shadowing fading)、距离衰减(power decay of distance)等信息。这表明在不同路径下,距离不同,所接收到的WiFi信号的各个子载波信息(CSI)会各不相同。根据这一原理,很多算法将不同位置点的CSI样本作为该位置点的指纹用于实现定位。而数据表明,对于代表各位置点的CSI序列,其主要差别

仅体现在某些子载波或子载波构成的子序列。基于Shapelet算法的指纹定位方法就是根据这一原理,采集每个定位点的CSI数据,提取其中的振幅信息,选取最能区别于其他位置点的子序列Shapelet作为该位置的指纹,从而创建指纹库来实现定位。

1.2 Shapelet 算法

Shapelet^[9]是一种时序数据挖掘方法。该方法指出在对2种时序数据进行分类时,会有一个最能体现其差异的子序列能实现最好的分类效果,这个子序列被称为Shapelet,其基本流程如下:

1) 数据集 D 中包含 M 、 N 两个类别的时间序列,使用滑动窗口 k (取所有可能值)在所有序列上截取子序列,所有子序列构成Shapelet候选项集 S ;

2) 依次取出候选项集的每一个候选项 S_i 求出其与数据集 D 中每一个时间序列 D_i 的距离并依据大小排序;

3) 取所有距离值 d 对数据集 D 进行分类,求出相应的信息增益,取信息增益最大的距离值作为最佳分割点 b ,此时的信息增益为2)中候选项 S_i 的信息增益;

4) 将各个候选项的信息增益值进行比较,最大的信息增益值所对应的候选项 S_i 为选定的Shapelet。

根据应用,对于此方法主要做出的改进有2点:①针对每一个位置,在生成候选项的距离数据集时,只取该位置数据所对应候选项,而非候选项集中的全部;②由于CSI的规律性,在求候选项与每个CSI序列距离的时候,不再需要截取每个等长的子序列来比较,只需选与候选项位置对应的子序列即可,这样在很大程度上降低了时间复杂度。

2 方法流程

基于CSI的指纹定位技术主要有2个难点:①如何从CSI中提取出代表各个位置的指纹?Bassiouny等^[10]使用主成分分析方法对数据集进行降维,在定位点样本空间中找到了最大限度表示该样本的一组正交向量作为该样本的指纹。Shapelet算法在提取每个位置的特征时,将每个位置信道状态信息中与其他位置差别最大的特征子序列作为该位置的指纹,以达到最好的定位效果。②如何对用户所在的位置进行识别?该阶段主要通过分类算法来实现,主要有决策树、 K 近邻算法、神经网络以及深度学习等。基于①中位置指纹所构建

的指纹数据库,使用决策树分类法实现对用户位置的判定,具体流程如图1所示. 其主要包括数据

采集与预处理、Shapelet 特征提取以及决策树定位等流程.

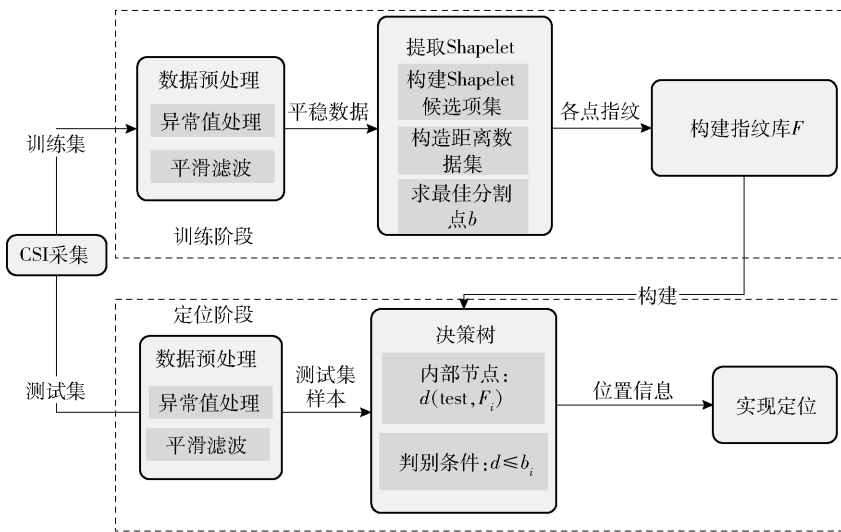


图1 基于 Shapelet 算法的指纹定位方法框架

2.1 数据采集与预处理

在实验区域内布署 WiFi 路由器,并设定好定位点,将 WiFi 接收器分别采集各点的信道状态信息,并为这些样本添加位置标签. 每个位置点采集的样本要足够多,每个样本 h 包含 30 个子载波 $(h_1, h_2, \dots, h_{30})$,以从中提取振幅信息. 由于数据采集环境中,其他的 WiFi 信号、物体反射以及人员站位和走动都会对信号产生影响和干扰,导致出现数据发生偏差或缺失等情况. 为此,采用了 $3\text{-}\sigma$ 异常值处理法以及卡尔曼平滑滤波对该原始数据进行预处理.

在使用 $3\text{-}\sigma$ 法对异常值进行处理时,要将每个位置点所有样本的第 $j(1 \leq j \leq 30)$ 个子载波构建成一个矩阵 $H_{1j}, H_{2j}, \dots, H_{nj}$,求其均值 $\mu = \frac{1}{n} \sum_{i=1}^n H_{ij}$,

标准差 $\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (H_{ij} - \mu)^2}$;对其中的每个数据值进行判断,将所有 $H_{ij}(H_{ij} > \mu + 3\sigma \parallel H_{ij} < \mu - 3\sigma)$ 作为异常值;然后用均值来代替每个异常值. 经过这样的处理,将原始数据中干扰引发的异常值进行修正. 之后采用了卡尔曼滤波对数据进行平滑处理.

2.2 Shapelet 特征提取及构造指纹库

经过异常值处理和卡尔曼滤波等预处理后的 CSI 数据趋于平稳,接下来将对每一个位置点的数据进行特征提取,构建最终的指纹库.

1) 构建 Shapelet 候选项集

选取每个定位点的 10 个数据样本作为训练集 D ,将每个数据样本作为一个序列 D_{ij} ,设置候选项序列的最小长度和最大长度. 从第 i 个定位点的训练集 $D_i(i = 1, 2, \dots, 16)$ 中依次取序列 $D_{ij}(j = 1, 2, \dots, 10)$ 作为原始序列,在 D_{ij} 上截取所有可能长度的子序列 s_k ,长度 l 介于最小长度与最大长度之间. 所有的子序列 s_k 构成第 i 个定位点的 Shapelet 候选项集 S_i .

2) 构造距离数据集

从第 i 个定位点的候选项集 S_i 中取所有子序列 s ,依次构造其距离数据集.

取第 k 个子序列 s_k ,计算 s_k 与 D 中所有序列 D_{ij} 的距离 d_{ij} . 在 D_{ij} 上截取与 s_k 相同位置的子载波序列 X ,则

$$d_{ij} = \sum [s_k(a) - X(a)]^2 \quad (1)$$

将 d_{ij} 按大小顺序存入数据集 P .

3) 求最佳分割点

依据 2) 中的数据集 P ,选取相邻数据的均值作为分割点 $f = \frac{P_i + P_{i+1}}{2} (i = 1, 2, \dots, 160)$,将满足 $d_{ij} \leq f$ 的样本序列 D_{ij} 划分至数据集 D_1 中;反之,划分至 D_2 中,并求该分割方式下的信息增益,如

$$G = E(D) - \frac{l_1}{l} E(D_1) - \frac{l_2}{d} E(D_2) \quad (2)$$

其中: l, l_1, l_2 分别为数据集 D, D_1, D_2 的大小;

$E(D)$ 为 D 的熵,其表达式为

$$E(D) = -\frac{a}{d}\log\left(\frac{a}{d}\right) - \frac{q}{d}\log\left(\frac{q}{d}\right) \quad (3)$$

其中: a 为数据集 D 中第 i 个定位点样本的个数, q 为数据集 D 中不属于第 i 个定位点的样本个数.

根据式(3),依次求得所有分割点的信息增益 G ,取信息增益最大值为第 k 个Shaplet候选选项的最大信息增益 g_k ,取相应的分割点 f 为第 k 个Shapelet候选选项的最佳分割点 b_k .

4) 构造 Shapelet 位置指纹库

依次求得各个候选选项的最大信息增益 g_k 及最佳分割点 b_k 后,取 g_k 最大的候选选项子序列为第 i 个定位点的特征子序列 S_i ,并标记其对应的最佳分割点最佳分割点 b_i .依次求得每个定位点的Shapelet,并存入位置指纹库 F .至此,完成了位置指纹库的构造,训练阶段结束.

5) 构造 Shapelet 决策树

根据决策树的定义,需要为决策树选定内部节点和叶子节点.2.2节所构造的位置指纹库便是决策树构造的依据.叶子节点依次为每个定位点,内部节点依次为其对应的定位点特征子序列 F_i 与目标序列的距离 d ,如果该距离值满足($d \leq b_i$),则有向边指向定位点;否则,有向边指向另一分支.最终构造Shapelet决策树,如图2所示.

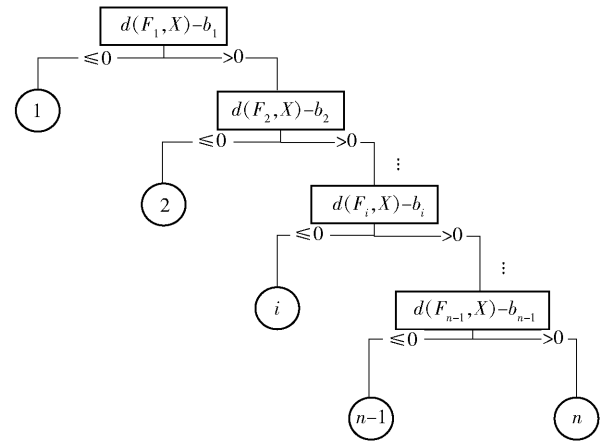


图 2 决策树模型

2.3 决策树定位

方法中使用了决策树分类来实现定位,其主要方法是将定位数据与构建的决策树内部节点,即每个位置点的指纹特征,进行匹配.若与某一特征匹配成功,则可认定其位置在该点.操作步骤如下:

1) 对定位样本进行预处理操作,去除干扰与

噪音;

2) 根据图2所示的决策树,将每一个定位样本依次与16个定位点匹配.具体依据式(1)求出与每个定位点指纹 F_i 的距离 d ;

3) 如果 $d \leq b_i$,则判定该样本是在 i 点采集到的,实现定位.

3 结果分析与比较

3.1 实验设计

本实验分别在2个实验场地进行,如图3所示.

① 4.8 m×4.8 m的空旷环境,选位置点16个,相邻点间距均为1.2 m;② 4 m×6 m的休闲区,其中有4组桌椅以及走廊过道,共选位置点12个,相邻点间距同为1.2 m.在以上2个场地中每个位置点分别采集CSI数据,分3个时间段进行:第1天上午实验

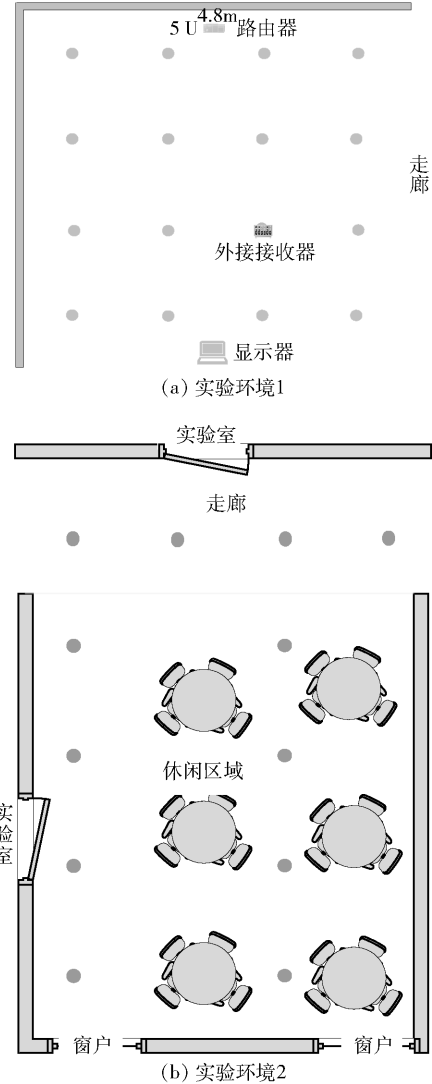


图 3 实验环境及设备布置

环境 1 采集的数据为 A_1 , 实验环境 2 采集的数据为 B_1 ; 第 1 天下午实验环境 1 采集的数据为 A_2 , 实验环境 2 采集的数据为 B_2 ; 第 2 天上午实验环境 1 采集的数据为 A_3 , 实验环境 2 采集的数据为 B_3 . 分别提取 CSI 数据中的振幅数据, 并对这些数据样本进行标记.

离线建立指纹库阶段, 在 A_1 、 B_1 中为每个定位点选取 10 个训练样本, 对这些样本进行预处理, 并根据上述方法提取位置指纹, 分别构造决策树 T_1 、 T_2 . 在实验中, 分别测试了基于 Shapelet 算法的指纹定位与基于主成分分析 (PCA, principal components analysis) 算法及 k 近邻 (KNN, k -nearest neighbor) 算法的指纹定位结果, 并分析比较了定位精确度和平均定位误差, 其中 PCA 与 KNN 算法的所使用的训练集为 A_1 、 B_1 .

3.2 实验结果分析

1) 实验环境 1 下的结果分析

在对数据 A_1 的测试中, 在定位误差小于 1 m 时, 基于 Shapelet 算法定位方法的精确度达到 99.58%, 基于 PCA 算法与 KNN 算法的定位方法高达 100%, 其中误差累计分布如图 4 所示.

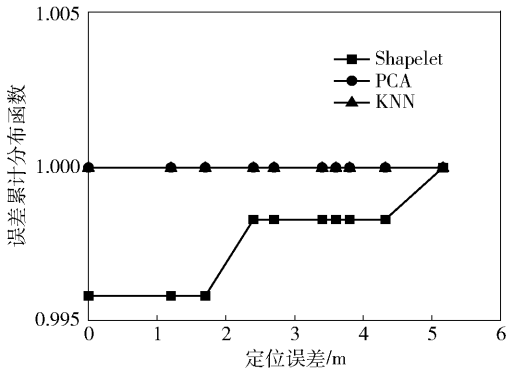


图 4 A_1 定位误差分析

在对测试数据 A_2 的实验中, 定位误差小于 1 m 时, 基于 Shapelet 算法的定位的精确度为 91.21%, 基于 PCA 算法的方法进行定位精确度为 92.24%, 使用 KNN 算法进行定位精确度为 97.74%. 定位误差累计分布如图 5 所示.

将数据 A_3 作为测试数据进行定位, 定位误差小于 1 m 时, 基于 Shapelet 算法的定位精确度依旧很高, 为 89.34%; 而使用 PCA 算法和 KNN 算法的定位方法定位精度达 89% 时, 定位误差分别为 4 m 和 3 m. 定位误差累计分布如图 6 所示.

根据上述各定位结果计算出每次实验的平均

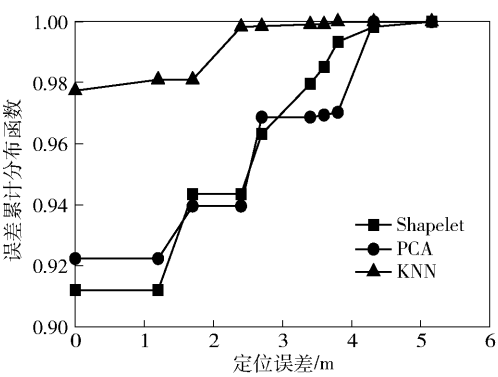


图 5 A_2 定位误差分析

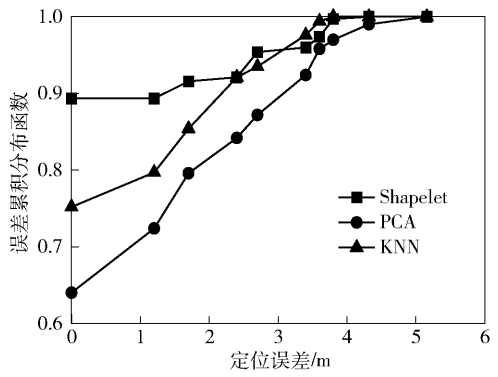


图 6 A_3 定位误差分析

定位误差, 如表 1 所示. 对于同一天采集的数据 A_2 , 3 种定位方法的定位误差都比较小; 对于第 2 天采集的数据 A_3 来说, 基于 Shapelet 算法的定位方法平均误差达到了 25.8 cm, 而使用基于 PCA 和 KNN 的指纹定位方法平均误差分别为 91.20 cm 和 65.48 cm.

表 1 实验环境 1 的平均定位误差				cm
数据	Shapelet	PCA	KNN	
A_1	1.40	0	0	
A_2	17.74	15.43	3.33	
A_3	29.44	91.20	65.48	

2) 实验环境 2 下的结果分析

在环境 2 下, 仅对数据 B_2 、 B_3 进行了定位测试, 结果如图 7、图 8 所示. 在复杂环境下, 不同时间段内基于 Shapelet 算法的指纹定位在 1.2 m 内的定位精度依旧保持稳定, 分别为 90.34% 和 87.34%; 基于 PCA 算法的定位精度分别为 93.20% 和 59.2%; 基于 KNN 算法的定位精度分别为 89.2% 和 73.3%. 在该环境下的平均误差如表 2 所示.

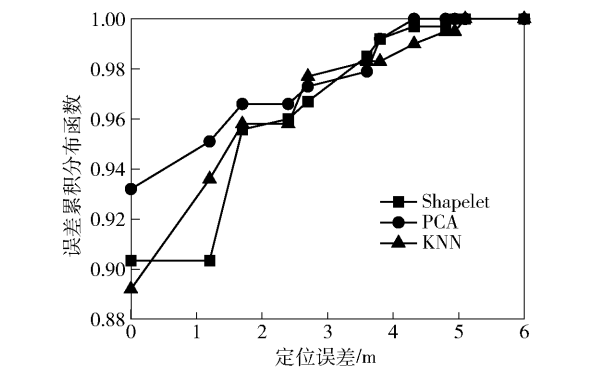


图 7 B_2 定位误差分析

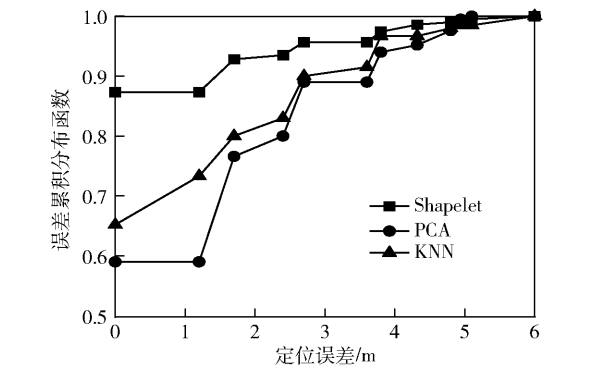


图 8 B_3 定位误差分析

表 2 实验环境 2 的平均定位误差				cm
数据	Shapelet	PCA	KNN	
B_2	24. 58	18. 43	24. 30	
B_3	35. 92	110. 50	92. 27	

上述实验结果表明,基于 Shapelet 算法的指纹定位方法不同时间的定位精度较高,且能保持稳定.

4 结束语

提出了一种基于 Shapelet 算法的指纹定位方法,在相对理想的环境下搭建常用的数据采集平台,在 WiFi 信号中提取 CSI 数据,并对该数据进行处理修正;通过 Shapelet 特征提取方法提取各个定位点的指纹,构建指纹库;使用 Shapelet 决策树方法实现定位,使得不同时间段的平均定位精度达到 90. 23%,明显优于 PCA 算法以及 KNN 算法. 实验结果表明,Shapelet 特征提取算法对于不同时间段的数据,定位准确率较高;Shapelet 算法所需要的训练集数据需求量小. 针对这两点可以看出,使用 Shapelet 算法实现定位在生活中适用性更强.

参考文献:

[1] Kalbandhe A A, Patil S C. Indoor positioning system using bluetooth low energy[C] // 2016 International Conference on Computing, Analytics and Security Trends (CAST). Pune: IEEE, 2016: 451-455.

[2] Naz A, Hassan N U, Pasha M A, et al. Single LED ceiling lamp based indoor positioning system [C] // 2018 IEEE 4th World Forum on Internet of Things (WF-IoT). Singapore: IEEE, 2018: 682-687.

[3] Ni L M, Liu Y, Lau Y C, et al. LANDMARC: indoor location sensing using active RFID[C] // Proceedings of the First IEEE International Conference on Pervasive Computing and Communications. [S. l.]: IEEE, 2003: 407-415.

[4] Kim S, Jang S H, Chong J W. Hybrid RSS/TOA wireless positioning with a mobile anchor in wireless sensor networks[C] // LANC 2011 - Proceedings of the 6th Latin America Networking Conference. Quito: ACM, 2012: 1-8.

[5] Poursheikhali S, Zamiri-Jafarian H. TDOA based target localization in inhomogenous underwater wireless sensor network[C] // 2015 5th International Conference on Computer and Knowledge Engineering (ICCKE). Mashhad: IEEE, 2015: 1-6.

[6] Zhu Xiao, Tan Guanghua, Li Renfa, et al. Joint TOA AOA localization based on UWB for wireless sensor networks[J]. J Comput Res Dev, 2013, 50: 453-460.

[7] Singh A P, Singh D P, Kumar S. NRSSI: new proposed RSSI method for the distance measurement in WSNs[C] // 2015 1st International Conference on Next Generation Computing Technologies (NGCT). Dehradun: IEEE, 2015: 296-300.

[8] Wang Xuyu, Gao Lingjun, Mao Shiwen, et al. CSI-based fingerprinting for indoor localization: a deep learning approach[J]. IEEE Transactions on Vehicular Technology, 2016, 66(1): 763-776.

[9] Ye L, Keogh E. Time series shapelets: a new primitive for data mining [C] // Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris: ACM, 2009: 947-956.

[10] Basiouny Y, Arafa M, Sarhan A M. Enhancing Wi-Fi fingerprinting for indoor positioning system using single multiplicative neuron and PCA algorithm [C] // 2017 12th International Conference on Computer Engineering and Systems (ICCES). Cairo: IEEE, 2017: 295-305.