

基于 ISODATA 改进 K 均值聚类算法的 NLOS 识别技术*

韦子辉 廖戈 李明轩 周敬仪 董鹏

(河北大学质量技术监督学院 保定 071002)

摘要: 针对超宽带信号中非视距误差造成定位系统定位偏差的问题,提出了一种改进无监督算法的 NLOS 识别技术。本文提取信道脉冲冲激响应波形的 8 种特征参量,选择主成分分析算法对多维特征进行降维处理;采用基于迭代自组织数据分析法改进的 K 均值聚类算法,自适应地选择 K 值来区分视距和非视距信号;最后,结合特征参量的冗余性、相关性对分类结果进行判别。实验结果表明,该方法能有效地识别出 NLOS 信号,且具有较好的环境适应性,识别准确度达到 95%。

关键词: 超宽带定位;非视距识别;无监督算法;信道冲激响应

中图分类号: TP393.1 **文献标识码:** A **国家标准学科分类代码:** 510.99

NLOS identification technique based on K-means clustering algorithm improved by ISODATA

Wei Zihui Liao Ge Li Mingxuan Zhou Jingyi Dong Peng

(School of Quality and Technical Supervision, Hebei University, Baoding 071002, China)

Abstract: To mitigate the issue of positioning deviations in positioning systems caused by non-line of sight (NLOS) errors in Ultra-Wide Ban signals, this study presents an unsupervised clustering method that utilizes the characteristic parameters of the channel impulse response for identifying NLOS signals. The method involves the extraction of eight characteristic parameters from the channel impulse response waveform, followed by the use of the principal component analysis algorithm to reduce the dimension of the multi-dimensional features. An improved K-means clustering algorithm, based on iterative self-organizing data analysis, is then used to select K-values adaptively for distinguishing between LOS and NLOS signals. Finally, the redundancy and correlation of feature parameters are combined to distinguish the classification results. The experimental results demonstrate that this approach effectively identifies NLOS signals with better environmental adaptability and has a recognition accuracy of 95%.

Keywords: indoor localization; NLOS; unsupervised machine learning; CIR

0 引言

近年来,随着社会经济快速发展,手机导航、路径规划、目标追踪等各种基于定位信息提供服务的应用越来越多,人们对定位精度的需求不断增加^[1]。在复杂的室内环境,例如工厂、办公场所、大型商场和医院等场所,传统的全球定位系统(global positioning system, GPS)和北斗导航系统,因卫星信号受地面建筑遮挡、信号强度衰减的限制,导致室内定位精度急剧下降,甚至无法定位目标位置,而室内定位技术弥补了这一空缺^[2]。

超宽带(ultra-wide band, UWB)室内定位技术是一种新型无线定位技术,因其抗多径误差、信号穿透力强、极具空间分辨率、抗干扰能力强等性能优势,能够提供室内定位系统所需的精度和鲁棒性,被广泛应用于室内高精度定位中^[3]。然而,所有的无线定位技术都会存在多径效应,尤其在非视距(non-line of sight, NLOS)环境下,发射信号能够通过穿透、反射、散射或绕射路径到达接收端,使接收端接收的信号存在不同程度的时延和衰减,影响室内定位精度^[4]。通过识别 NLOS 信号可以帮助定位系统剔除受干扰严重的基站,因此,识别 NLOS 信号并减小 NLOS 传播

对定位结果的影响是提高定位精度的重要手段。

针对室内定位 UWB 信号存在 NLOS 传播现象,近年来已有很多学者提出多种方法对信号进行 NLOS 识别,主要有基于距离估计的方法^[5-6],基于位置估计的方法^[7-9]和基于信道冲激响应(channel impulse response, CIR)的方法。其中第 3 种方法,由于 CIR 波形中隐含信号在传输过程中的信道信息^[10],通过提取 CIR 波形中的特征参量,如峰度、峰前延迟、平均超时延迟和均方根时延等,可以帮助我们区分视距(line of sight, LOS)和非视距信号,掌握障碍物情况^[11]。早期的研究中,NLOS 识别依赖于各种信道参数的概率分布函数,随着机器学习的发展,学者们开始将从 UWB 信道脉冲冲激响应中提取的特征作为机器学习的输入对信号进行分类。文献[12]采用支持向量机(support vector machine, SVM)、基于集成学习方法的随机森林(random forest, RF)和基于深度神经网络的多层感知器(multilayer perceptron, MLP)网络进行 NLOS 识别,并比较评估了他们在 UWB 定位中的影响。文献[13]通过相关向量机(relevance vector machine, RVM)技术识别 NLOS 信号,减少了计算复杂度。文献[14]提出了一种基于输入向量机(import vector machine, IVM)的具有特征选择策略的 NLOS 识别方法。上述文献均采用有监督的识别方式,对定位技术的各类应用场景进行了介绍,由于有监督的方式需要提前对学习样本进行标记,使得定位系统或定位算法很难自适应调节或采用迁移学习的方式,以适应环境的不断变化。而无监督学习是一种对不含标记的数据建立模型的机器学习方法,它不需要对信道冲激响应特征参量进行任何严格和明确的标注,环境适应度较高。文献[15]使用一种称为“期望最大化高斯混合模型”的无监督机器学习方法对 LOS 和 NLOS 进行分类,不需要对信道条件进行显式和先验标,实验准确率为 86.5%。文献[16]使用 K-means 算法对特征集的满意度进行评估,仅对有用的特征集进行分类,需要提前对特征参量进行筛选。

因此,本文在现有的研究基础上,针对传统 NLOS 识别算法需要先验数据、使用场景单一、计算复杂度较高等问题,提出了一种改进型无监督学习算法——基于迭代自组织数据分析法(iterative self-organizing data analysis techniques algorithm, ISODATA)改进的 K 均值聚类算法,使用主成分分析(principal component analysis, PCA)算法对特征参量降维作为算法输入,自适应选择当前环境下的 K 值进行分类;并依据特征参量与信道状态识别的相关性、特征参量之间的冗余性对分类后的信道进行 LOS 和 NLOS 判定。实验分析了 6 种典型遮挡环境下,单一特征参量在独立场景对信道识别的准确度与多种特征参量在复杂遮挡环境中的有效性。UWB 定位系统在 NLOS 环境中容易受到多径传播和信号衰减的影响,从而导致定位失效,通过自适应 NLOS 识别,可以帮助系统选择可用的基站,更可靠地适应复杂的环境,提高在各种条件下的性能稳定

性及精确定位。若本文算法在复杂遮挡环境下具有一定准确度,能够较为正确判定 LOS 信道,则说明本算法具有一定有效性和环境适应性。

1 NLOS 识别技术方案设计

在 UWB 定位系统中,电流辐射天线(large current radiator, LCR)在收到信号以后,激励电流导数的平方与天线辐射的功率成正比,天线通过一个电流激励时,系统输出端可以产生一个符合高斯分布的脉冲信号,被称为信道冲激响应(CIR),用式(1)表示:

$$h(t) = \sum_{k=1}^K a_k \cdot \sigma(t)(t - \tau_k) \quad (1)$$

式中:K 为多路径组件的数量; a_k 和 τ_k 分别为第 k 分量的振幅和时延; $\sigma(t)$ 为高斯无线电波形^[17]。

视距传播是准确得到 CIR 波形特征参量的一个必要条件,当基站与标签之间不存在视距路径时,只有反射信号和衍射信号能够到达基站,导致非视距效应。在到达时间差和到达时间定位算法中,对于基站收到的信号,通常使用匹配滤波器获取最大峰值时刻,或者前几位峰值时刻的平均值来表示信号到达时间,但由于多径效应的影响,峰值的位置会产生漂移,导致信号到达估计时间与实际值存在较大误差,NLOS 环境下误差产生示意图如图 1。

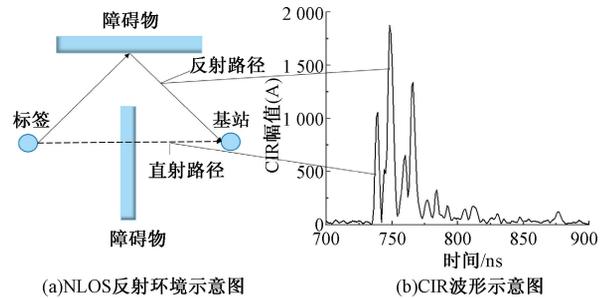


图 1 NLOS 环境误差产生示意图

为能够自适应且准确识别出 NLOS 信号,本文设计了一种基于无监督学习算法的 NLOS 识别技术,方案流程图如图 2。首先采集不同信道下的 CIR 波形,提取信号的多种特征参量,分析不同信道状态下的多种特征参量数据分布情况。其次使用 PCA 算法对特征参量降维,引入基于 ISODATA 改进的自适应 K 均值聚类算法,在不同场景下合理确定 K 值,保证较高的信道状态识别率。最后依据基

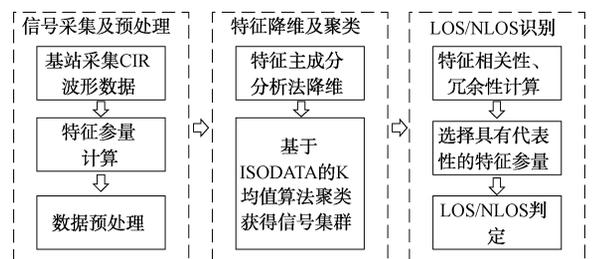


图 2 NLOS 识别技术方案

于特征参量与信道状态识别的相关性、多特征参量之间的冗余性选出的 3 个特征参量,对分类后的 LOS/NLOS 信号进行判断。

2 特征参量提取及处理

2.1 信道脉冲冲激响应特征提取

图 3 和 4 分别是采集的 LOS 和 NLOS 信号 CIR 波形,采集的 CIR 波形由 1 000 个冲激响应连接而成,对低峰值噪声阶段进行滤除留下相对峰值阶段。通过对脉冲冲激响应信号的波形进行分析,直射信号与其他信号都会在 CIR 波形中得到清晰直观的反映。

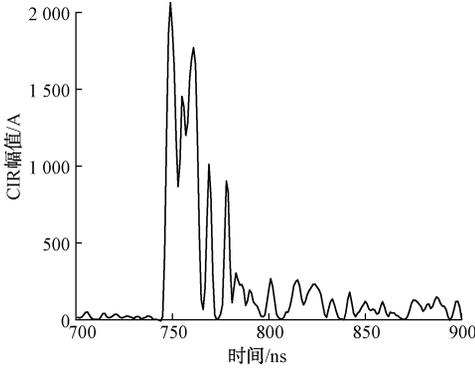


图 3 LOS 环境下 CIR 波形

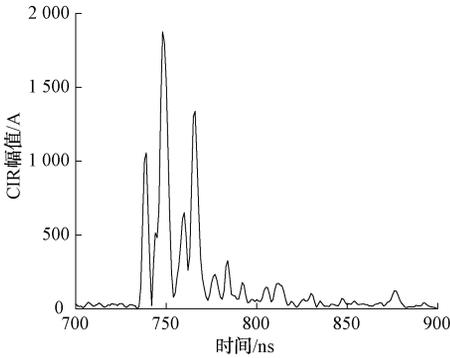


图 4 NLOS 环境下 CIR 波形

为了量化 LOS/NLOS 环境下的 CIR 信号差异,我们采用一些特征参量来描述它。本文提取的特征参量主要有偏度、峭度、最大振幅、总能量、上升时间、平均附加时延、均方根时延、饱和度。

1) 偏度(Skewness):表示脉冲冲激响应信号波形数据分布的参考方向和偏差程度,记为 ske ,其函数模型如式(2)所示。

$$ske = \frac{E[(|h(t)| - \mu_{|h|})^3]}{\sigma_{|h|}^3} \quad (2)$$

式中: $h(t)$ 为信道冲激响应, $\mu_{|h|}$ 为信号冲激响应各采集点振幅均值, $\sigma_{|h|}$ 为信道冲激响应各采集点振幅标准差。

2) 峭度(Kurtosis):表示脉冲冲激响应信号波形数据分布的陡峭程度,记为 k ,其函数模型如式(3)所示。

$$k = \frac{E[(|h(t)| - \mu_{|h|})^4]}{\sigma_{|h|}^4} \quad (3)$$

3) 最大振幅(maximum amplitude):表示脉冲冲激响应信号波形的振幅最大值,记为 r_{max} ,其函数模型如式(4)所示。

$$r_{max} = \max\{|r(t_i)|\} \quad (4)$$

式中: $r(t_i)$ 为 t_i 时刻的振幅。

4) 总能量(total energy):表示脉冲冲激响应信号波形的能量总和,记为 ϵ ,其函数模型如式(5)所示。

$$\epsilon = \sum_{i=1}^N |r(t_i)|^2 \quad (5)$$

5) 上升时间(rise time):表示脉冲冲激响应信号波形从 0.1 倍最大振幅到 0.9 倍最大振幅的时间,记为 t_{rise} ,其函数模型如式(6)所示。

$$t_{rise} = t_{stop} - t_{start} \quad (6)$$

$$\text{其中,} \begin{cases} t_{start} = \min\{t_i : |r(t_i)| \geq 0.1r_{max}\} \\ t_{stop} = \min\{t_i : |r(t_i)| \geq 0.9r_{max}\} \end{cases}$$

6) 平均附加时延(mean excess delay):表示脉冲冲激响应信号波形中多径信道时间色散的一个特性参数,记为 τ_{med} ,其函数模型如式(7)所示。

$$\tau_{med} = \frac{1}{\epsilon} \sum_{i=1}^N (t_i |r(t_i)|^2) \quad (7)$$

7) 均方根时延(root-mean-square delay spread):表示脉冲冲激响应信号波形中多径信道时间色散的另一个特性参数,记为 τ_{rms} ,其函数模型如式(8)所示。

$$\tau_{rms} = \frac{1}{\epsilon} [(t_i - \tau_{med})^2 |r(t_i)|^2] \quad (8)$$

8) 饱和度(saturation):表示脉冲冲激响应信号第一路径附近三个幅值中最大值 d_{1max} 占第一路径能量 d_{max} 的百分比,记为 S ,其函数模型如式(9)所示。

$$S = \frac{d_{1max}}{d_{max}} \quad (9)$$

在实验室条件下,手持标签分别进行 1 000 组 LOS 信道和 1 000 组 NLOS 信道条件下的实验,从 CIR 波形中获取上述特征参量。为统一各特征值量纲,将数据归一化至 $[-1, 1]$,得到信道特征参量分布箱线图如图 5,从箱线图中可以看出各特征参量在 LOS 环境和 NLOS 环境中的均值都存在差异。

2.2 判别参量选择及权重确定

为提高模型的预测精确度和结果的可理解性,采用特征参量选择的方法确定参量,对无监督分类后的不同信号进行判别。特征参量选择可从大规模的特征参量中去掉不相关、冗余或噪声数据,尽可能留下最能代表分类信息的最少特征参量^[18]。特征参量的选择主要需要满足以下几点:

- 1) 高相关性,尽可能代表原大规模数据的信息;
- 2) 弱相关且低冗余性,减少特征参量维度,降低运算量;
- 3) 特征参量标签化,增强数据的可读性,加深对数据理解。

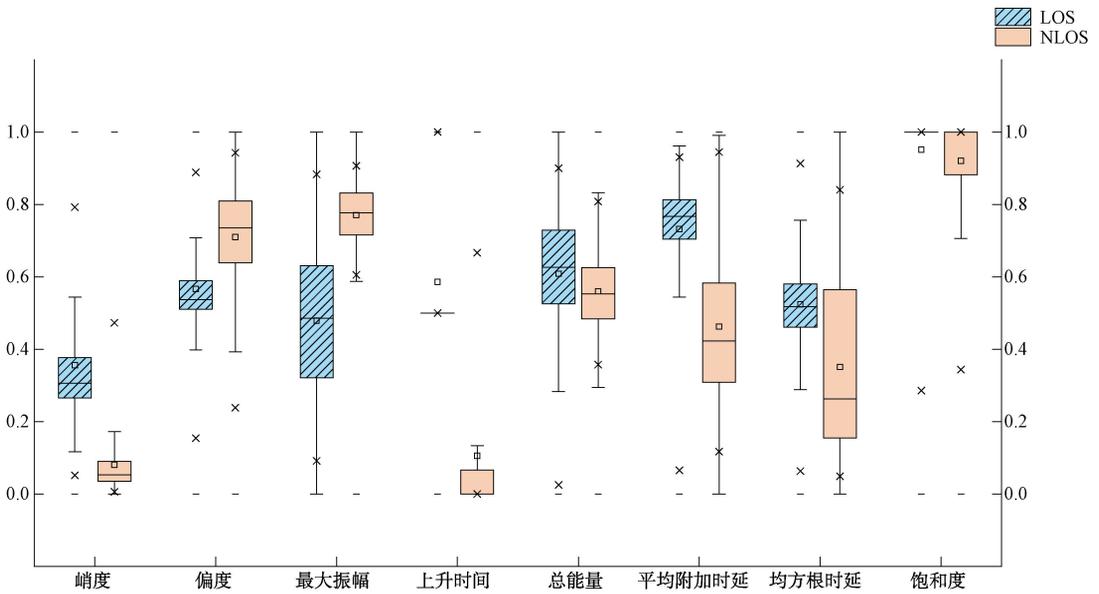


图 5 信道特征参量分布箱线图

首先确定特征参量和分类标签之间的相关性,相关性越高说明特征参量与分类标签之间的联系越紧密,对传播信道状态的区分能力越强,分类效果越好。Relief 算法^[19]是一种通过对特征和类别的相关性给予不同权重的降维算法,其在解决非线性数据分类领域具有优秀的分类效果,并且算法不需要对全局数据进行搜索分类,可以高效快速的完成数据分类,本文使用 Relief 算法对上述特征参量的相关性进行计算。

实验对随机的 6 000 组 LOS 信道和 6 000 组 NLOS 信道数据进行计算,得到 8 种特征参量的权重如表 1 所示。由表 1 可知,最大振幅 r_{max} 、总能量 ϵ 、上升时间 t_{rise} 、平均附加时延 τ_{med} 、饱和度 S 与分类标签的相关性较高,对传播信道状态的区分效果明显。

表 1 特征参量的相关性

特征参量	ske	k	r_{max}	ϵ
相关性	0.157	0.253	17.608	7.003
特征参量	t_{rise}	τ_{med}	τ_{rms}	S
相关性	27.285	12.809	1.837	33.264

为了得到较好的分类特征子集,需要在保证较高相关性的同时保证特征参量之间还要有较高的冗余性,较高的冗余性可以保证特征参量之间的相关性较小,相互间的关联性也越小。互信息特征选择算法 (mutual information feature selection, MIFS) 算法^[20]可以较好的评估各变量之间的依赖性,本文使用 MIFS 算法对上述特征参量之间的冗余性进行计算,结果如表 2 所示。

通过对上述相关性和冗余性分析,得到的最优判别子集特征参量为最大振幅 r_{max} , 上升时间 t_{rise} , 饱和度 S 。

表 2 特征参量的冗余性

	r_{max}	ϵ	t_{rise}	τ_{med}	S
r_{max}	1				
ϵ	0.486	1			
t_{rise}	0.721	0.539	1		
τ_{med}	0.818	0.658	0.501	1	
S	0.702	0.664	0.611	0.726	1

在使用了 K 均值聚类算法实现对信号分类以后,为每一个特征参量选择一个合适的权重会使 LOS/NLOS 定性更加精准,本文对不同特征参量赋予不同权重的依据是每个特征参量与分类标签之间的相关性,即特征参量与分类标签的相关性越高,特征参量的权重就越大,权重计算公式如式(10)。

$$WX_i = \frac{RX_i}{RX_1 + RX_2 + RX_3 + \dots + RX_n} \quad (10)$$

式中: WX_i 为特征 X_i 的权重, RX_i 为特征 X_i 与分类标签的相关性。

2.3 主成分分析算法

本文采用 8 个特征参量对 CIR 信号进行描述,然而过多的特征参量会降低模型识别速度和存储成本。为保存尽可能多的信息同时,减小数据集规模提高计算效率,本文采用 PCA 算法对数据集进行降维。

PCA 算法在找到主成分的过程中应当满足如下两个原则:1) 在新映射的方向上,方差越大的特征参量包含的信息量越大,所以选择出的每个映射点的方差应当尽可能的大。2) 新映射的方向应当彼此正交,这样映射出的坐标点不会存在互相关系,选取的坐标点更有意义。

算法的实现方法^[21]是将 n 维特征映射到 m 维空间 ($n >$

m),形成的 m 维特征被称为主成分,主成分是基于 n 维特征建立的新维度特征,一个二维的 PCA 降维示意图如图 6 所示。设定存在一空间样本集 F ,由 n 个样本组成,每个样本有 p 个变量,定义输出数据为 y (y 通常取值为 1 和 -1),训练样本集 F 的计算模型如式(11)所示。

$$F = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} & y_1 \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} & y_2 \\ x_{31} & x_{32} & x_{33} & \dots & x_{3p} & y_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} & y_n \end{pmatrix} \quad (11)$$

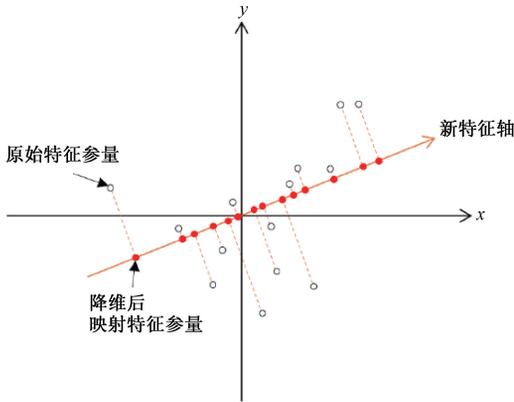


图 6 二维 PCA 算法模型图

且 x, y 经过 PCA 算法 y 由式(12)表示:

$$y = Ax \quad (12)$$

首先对 x, y 求平均值,上式矩阵分别减去对应的平均值实现零均值化,计算得到协方差矩阵,求出协方差矩阵的特征值、特征向量,将特征向量按照特征值的大小从上到下依次排序得到新矩阵,选择新矩阵的前 m 行组成的矩阵即为算法求得特征向量矩阵,最后根据表达式(12)求得矩阵即为求出的主成分矩阵。

图 7 为实验得到的 PCA 主成分图,其主成分由原有 8 个特征参量数据线性组合而来,是原特征参量数据投影到新维度上的全新正交特征,图中箭头表示原始数据投影到新特征轴方向上的相关性,数值为正表示正相关,数值为负表示负相关。由图可知,前 3 个主成分—PC1、PC2 与 PC3 的累计贡献率达到 95.3%,故认为前 3 个主成分对数据的贡献较高,这些主成分包含了数据中最重要的特征,可以作为识别模型的输入,其他成分贡献较低可以舍弃以降低数量维度。

3 基于 ISODATA 的 K 均值聚类算法

K 均值聚类分类算法的性能受 K 值的影响很大,一个合适的 K 值对分类效果有着极大地提升。在非视距信号识别领域,存在数据量较大,聚类效果不集中,数据处理实时性要求高等复杂情况,如果 K 值选择过小,可能发生欠拟合现象,造成非视距信号识别成功率低;而 K 值选择过大时,

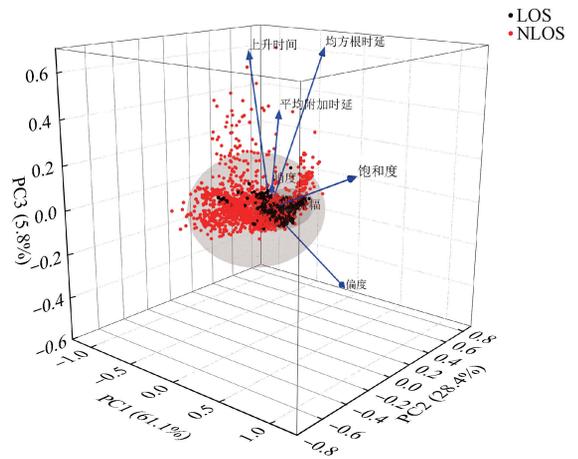


图 7 PCA 降维主成分图

可能发生拟合现象,也可能导致大量视距信号被误判为非视距信号。所以选择一个合适的 K 值是非常重要的。

对 K 均值聚类分类算法的改进主要从 K 值、离群值、初始聚类中心进行考虑:

1) 在聚类的过程中,常规 K 均值聚类分类算法的 K 值是预先选择的,在算法执行的过程中固定无法更改,引入 ISODATA 算法^[22]机制针对 K 值确定的问题进行改进。自适应 K 均值聚类分类算法原理是当某一簇的数据对象过少时,就把这个簇删除,当某一簇数据对象过多且分散程度较大时,将这个簇分为两个簇进行聚类。

2) 在聚类的过程中,K 均值聚类分类算法的分类依据是欧氏距离判断样本的相似度,但是高维度数据对象不适合直接使用此判断依据,通过 PCA 算法对数据对象进行降维处理,会使得聚类效果得到提升。

为此引入 ISODATA 算法对 K 均值聚类算法进行改进,帮助确定初始 K 值,图 8 为改进算法流程图,具体操作步骤如下:

1) 设定初始值:输入初始聚类数目 k_0 , 设定每一个聚类的最多样本数 θ_{max} 和最少样本数 θ_{min} , 两个聚类的中心最小距离 d_{min} , 最大迭代次数 l 。

2) 近邻聚类:将所有样本 x_i 分配到其到距离最小聚类中心所对应的类中。

3) 合并与分裂:

(1)当某个聚类 K_j 的样本数 $n_j < \theta_{min}$, 则丢弃该类,聚类数目 $K = k - 1$, 回到步骤 2);

(2)当某个聚类 K_j 的样本数 $n_j > \theta_{max}$, 则分裂该类,聚类数目 $K = k + 1$, 回到步骤 2);

(3)当某个类别的中心距离 $D_{i,j} < d_{min}$, 则合并两个分类,聚类数目 $K = k - 1$, 在一次迭代中,某一类最多参与合并一次;

4) 重复步骤 2)、3)直到迭代停止,确定合适的 K 值。

5) 利用步骤 4)确定的 K 值作为 K-means 算法的初始 K 值进行样本聚类。

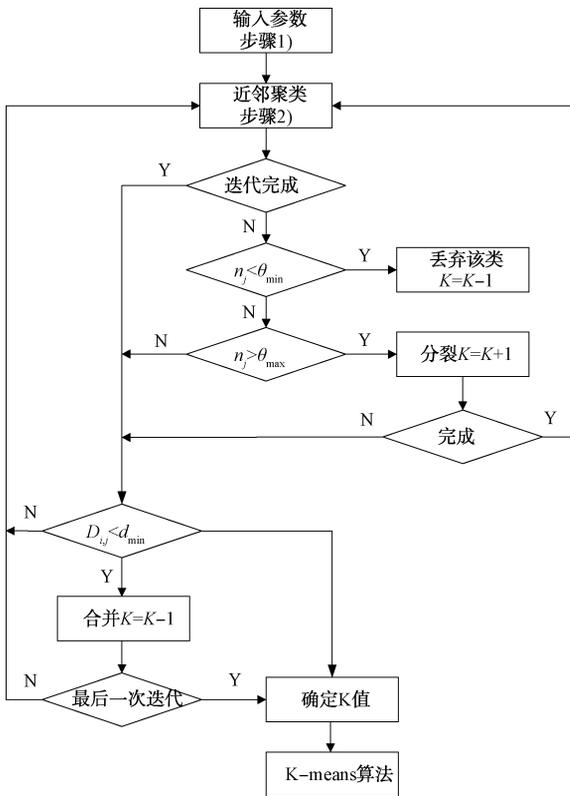


图 8 算法流程图

体环境与参数说明。其次采用准确度(accuracy, A)评估本文所提出的基于 ISODATA 改进 K 均值聚类算法的分类性能,与单个特征参量在独立遮挡环境下的准确度进行对比,并与多层卷积神经网络(convolutional neural network, CNN),SVM,RF 算法进行比较。

如图 9 所示,采集系统由基站、标签、上位机与交换机组成。基站及定位标签均使用 DW1000 射频芯片,基站通过 TCP 协议将数据传至上位机,实现对 CIR 波形的采集、处理、特征提取和保存功能。

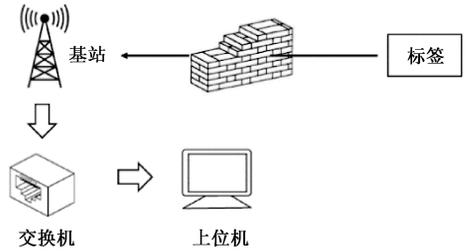


图 9 NLOS 识别系统结构图

在基站到标签的传播路径中,不同遮挡物因具有不同的介电常数与性质,基站接收到的 UWB 信号衰减程度也不同,该算法的输入需要最大程度反映这些衰减特性。为了验证本文算法的性能,具体的实验环境如下:实验室环境(CM1)、墙体遮挡环境(CM2)、工厂遮挡环境(CM3)、玻璃遮挡环境(CM4)、人体遮挡环境(CM5)、大厅 10 m 空旷环境(CM6)。设置基站高度距地面 1.5 m,标签距离地面 1 m,在前 5 种环境(CM1~CM5)中基站均距离标签 2 m,信号采集现场情况如图 10 所示。

4 实验及结果分析

4.1 实验环境

为了验证本文算法的效果,本章首先给出实验的具



(a)实验室环境(CM1)



(b)墙体遮挡环境(CM2)



(c)工厂遮挡环境(CM3)



(d)玻璃遮挡环境(CM4)



(e)人体遮挡环境(CM5)



(f)大厅10 m空旷环境(CM6)

图 10 信号采集实验环境 (CM1~CM6)

4.2 实验性能分析

在机器学习算法领域,混淆矩阵可以衡量一个分类器分类的准确程度^[23],非视距信号识别技术是较为简单的二分类问题,会有两种可能的分类:“是”或者“不是”,分类器做出正确和错误的判混淆矩阵的分类情况显示为 4 种不同的判断结果:

- 1)真阳性(true positive, TP):样本的实际类别为正例,模型预测的结果也是正例。
- 2)真阴性(true negative, TN):样本的实际类别为负例,模型预测的结果也是负例。
- 3)假阳性(false positive, FP):样本的实际类别为负例,模型预测的结果却是正例。
- 4)假阴性(false negative, FN):样本的实际类别为正例,模型预测的结果却是负例。

在信号识别领域,真假阳性和真假阴性如表 3 所示。本文选择使用准确度(A)来评估分类器的性能,计算公式如式(13)所示。

$$A = \frac{TP + TN}{TP + TN + FN + FP} \quad (13)$$

表 4 特征参量的不同信道识别准确度(A)

	<i>ske</i>	<i>k</i>	<i>r_{max}</i>	ϵ	<i>t_{rise}</i>	τ_{med}	τ_{rms}	S
CM1	0.544	0.576	0.807 0	0.909	0.918	0.887	0.559	0.915
CM2	0.695	0.858	0.973	0.956	0.905	0.964	0.577	0.873
CM3	0.681	0.879	0.814	0.827	0.699	0.799	0.786	0.604
CM4	0.590	0.888	0.877	0.814	0.930	0.916	0.924	0.924
CM5	0.547	0.584	0.964	0.957	0.660	0.910	0.710	0.670
CM6	0.532	0.640	0.937	0.893	0.737	0.902	0.834	0.708

采用文中自适应式 K-means 算法,在多种遮挡环境下采集信号 1 000 组,其中 LOS 信号 500 组,NLOS 信号 500 组,最大振幅 *r_{max}*, 上升时间 *t_{rise}*, 饱和度 S, 选择欧氏距离作为算法的距离度量标准,设置一个类别中最少样本数为 200 组,最多样本数为 800 组。如果一个类别中的样本数量少于 200 组,则合并该分组;如果一个类别中的样本数量多于 800 组,则分裂该分组。分类完成后,根据选出的主从基站之间的信号,根据最大振幅 *r_{max}*, 上升时间 *t_{rise}*, 饱

表 3 信号识别混淆矩阵表

混淆矩阵		预测值	
		LOS	NLOS
真实值	LOS	TP	FN
	NLOS	FP	TN

首先,对不同信号特征参量进行单一遮挡环境下的信号分类效果进行实验,采集实验数据对比分析,得到的多种信号特征参量分类效果识别精度如表 4 所示。从表可以看出,以单一特征参量对信号进行分类的效果具有显著的差异性,有的特征参量的信道状态识别率可以保证一个较好的识别准确度,有的特征参量的信道状态识别准确度在部分场景最高可以达到 95% 以上,但在其他场景识别率又会急剧下降,准确度仅能保证在 80% 上下,部分仅有 60%。可以发现采用单个特征参量作为算法输入进行识别,无法保证在每种遮挡环境下均具有良好的识别效果,环境适应性较差,这与定位场景要求的普适性是相违背的,为了实现位技术的多场景使用,得到稳定优秀的信道状态识别准确度,就需要对信道状态识别方法做进一步研究。

和度 S 对分类后的信号集群进行判定,选出最接近主从基站信号特征的信号集群判定为 LOS 信号。

得到的分类效果如图 11 所示,用 ○ 表示识别分类正确的信道,用 △ 表示分类错误的信道,其中空心 △ 表示将 NLOS 错误分类为 LOS,实心 △ 表示将 LOS 错误分类为 NLOS,使用二元混淆矩阵作为分类器识别效果的判定依据,信道状态识别成功率达到 95%。从图中可以看出,LOS 与 NLOS 类别之间具有较为明显的分界,且在 LOS

与 NLOS 类别中均有被错误分类的信号。

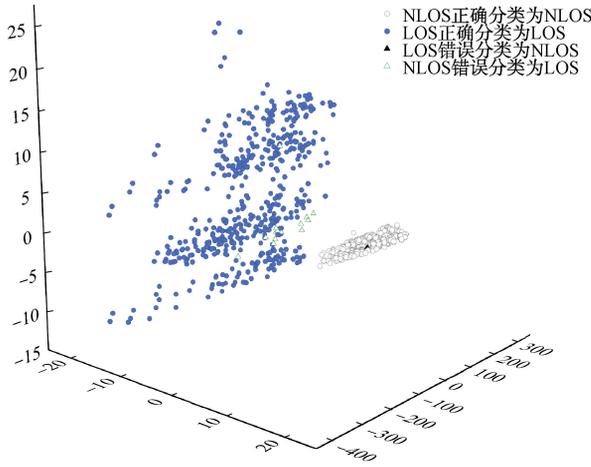


图 11 分类效果图

为了体现算法的先进性,与其他文献提出算法 CNN, SVM, RF 与本文算法识别准确度进行比较如表 5 所示。其中 CNN 算法模型和 RF 算法模型将最大振幅、上升时间、噪声标准差、接收信号功率、峰度、偏度、总能量、均方根时延扩展、峰值时间、平均附加时延、峰均比、能量 12 种特征参量作为算法输入,分类准确度分别为 90% 和 87.43%^[24]。SVM 算法测试了 18 种特征参量在不同组合下的 NLOS 分类效果,最高的分类准确度为 90.16%^[25],各种算法均提高了 NLOS 识别准确率。可以发现大部分文献为提高识别准确度增加了特征参量种类,或采用不同特征参量组合来表示环境特征,但少部分考虑到了数量维度的增加会影响算法运算速率,并引起维数灾难降低分类效果。本文提出的算法在复杂遮挡环境下准确度在 95%,说明采用 CIR 特征参量在 PCA 降维后的 3 个主成分可以敏锐表示当前环境情况,在不减少信息量的情况下具有良好的环境适应性,且对分类性能有一定提升。

表 5 不同识别算法的对比

算法	特征参量种类数	准确度 A/%
CNN	12	90
SVM	18	90.16
RF	12	87.43
本文算法	8(降维至 3)	95

5 结 论

室内定位受限于环境的复杂,信号受到遮挡物的影响,必然会在时间延迟、强度衰减等现象,产生 NLOS 误差,严重影响定位精度,且环境的复杂多变降低了定位系统的环境适应性。本文旨在设计一个自适应的、分类准确度高的 NLOS 识别技术帮助定位系统筛选出适用的基站。在信道脉冲冲激响的基础上,提出了一种基于 ISODATA

改进的 K 均值聚类算法,自适应地确定 K 值进行无监督聚类,使用 PCA 算法降低多特征参量维度后作为该算法的输入参量,最后通过特征参量的相关性、冗余性,选择出最大振幅、上升时间、饱和度三个特征参量,在后续对信号进行无监督分类后实现对不同信号集群进行 LOS/NLOS 定性。本文在多种环境下采集信号,对单一特征量实现 NLOS 识别精度展开分析,在此基础上验证了自适应 K 均值聚类算法的性能。实验结果表明,自适应 K 均值聚类算法识别准确可以实现对信道状态的较高精度识别。然而,仍然存在一些问题需要深入研究,下一步的研究工作可以从以下几方面入手:

本文对多种情况下的脉冲冲激响应信号进行了采集,但尚未解决所有潜在的环境影响因素,这些因素可能会影响信号的质量和稳定性。因此,未来的研究可以考虑扩大实验的场景和范围,或者研究探索其他 CIR 波形特征参量,以更全面地了解信号的行为和性能。

本文实验在二维环境中进行,随着对三维定位的需求不断增长,我们可以进一步探索将实验方案应用于三维定位系统中,或者考虑结合其他定位技术提高三维场景的精度和可靠性。

NLOS 识别的主要目的是提高定位精度,下一步可以结合 NLOS 抑制算法或定位算法,测试采用本文算法是否对 UWB 高精度定位系统可用基站选择提供帮助,并结合动态实验测试定位效果。

参考文献

- [1] 王鹏,侯忠生. UWB 与 IMU 融合的室内动态定位算法[J]. 电子测量技术,2023,46(10):76-83.
- [2] 蓝发籍,刘冉,郭林,等. 融合超宽带方位和距离的移动机器人定位[J]. 电子测量与仪器学报,2023,37(8):155-163.
- [3] YANG X F, ZHAO F., CHEN T J. NLOS identification for UWB localization based on import vector machine[J]. AEU: Archiv fur Elektronik und Ubertragungstechnik; Electronic and Communication, 2018,87:128-133.
- [4] 从飞云,洪纵横,汤子安,等. 基于脉冲响应消歧重构的 UWB 测距优化方法[J]. 仪器仪表学报,2022,43(10):13-21.
- [5] YAN L, LU Y, ZHANG Y. An improved NLOS identification and mitigation approach for target tracking in wireless sensor networks [J]. IEEE Access,2017,5:2798-2807.
- [6] LOPEZ,ONEL L. Polarization diversity-enabled LOS/NLOS identification via carrier phase measurements [J]. IEEE Transactions on Communications,2022,71:1678-1690.
- [7] ZHANG H, WANG Q, YAN C, et al. Research on

- UWB indoor positioning algorithm under the influence of human occlusion and spatial NLOS[J]. *Remote Sensing*,2022,14(24):6338.
- [8] KLMEN B, MIHAEL M. Improving indoor localization using convolutional neural networks on computationally restricted devices[J]. *IEEE Access*, 2018,6:17429-17441.
- [9] CHENG L, WANG Y, XUE M, et al. An indoor robust localization algorithm based on data association technique [J]. *Sensors (Basel, Switzerland)*, 2020, 20(22):6598.
- [10] 畅田田,王威,高婧洁,等. 基于信道测量实验的 NLOS 误差消除方法对比研究[J]. *西北工业大学学报*,2022,40(4):865-874.
- [11] WANG F, TANG H, CHEN J. Survey on NLOS identification and error mitigation for UWB indoor positioning [J]. *Electronics*, 2023, 12 (7), DOI: 10.3390/electronics12071678.
- [12] SANG L C, STENINHAGEN B, HOMBURG D J, et al. Identification of NLOS and multi-path conditions in UWB localization using machine learning methods[J]. *Applied Sciences*,2020,10(11):3980.
- [13] NGUYEN V T, JEONG Y, SHIN H, et al. Machine learning for wideband localization [J]. *IEEE Journal on Selected Areas in Communications*,2015,33(7):1357-1380.
- [14] YANG X F, ZHAO F, CHEN T J. NLOS identification for UWB localization based on import vector machine [J]. *AEUE-International Journal of Electronics and Communications*,2018,87:128-133.
- [15] FAN J, AWAN S A. Non-line-of-sight identification based on unsupervised machine learning in ultra wide band systems[J]. *IEEE Access*,2019,7:32464-32471.
- [16] KIRMAZ, ANIL. LOS/NLOS classification using scenario-dependent unsupervised machine learning[C]. 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC),2021:1134-1140.
- [17] GUI L Q, YUAN W Y, XIAO F. CSI-based passive intrusion detection bound estimation in indoor NLOS scenario [J]. *Fundamental Research*, 2023, 3 (6): 988-996.
- [18] JIANG J, SHANG J F. Feature screening for high-dimensional variable selection in generalized linear models [J]. *Entropy (Basel, Switzerland)*, 2023, 25(6):851.
- [19] WANG H, WANG P, DENG S. Improved relief weight feature selection algorithm based on relief and mutual information[J]. *Information*,2021,12(6):228.
- [20] GAO L Y, WU W G. Relevance assignation feature selection method based on mutual information for machine learning-ScienceDirect [J]. *Knowledge-Based Systems*,2020,209:106439.
- [21] JOLLIFFE, LAN T, CADIMA J. Principal component analysis: A review and recent developments [J]. *Philosophical Transactions of the Royal Society Mathematical Physical & Engineering Sciences*,2016, 374(2065):20150202.
- [22] LU Y, QIU L, LYU X F, et al. Human stampede causative factors and cluster risk; A multi-dimensional analysis based on ISODATA and fuzzy theory [J]. *International Journal of Disaster Risk Reduction*,2021, 66, DOI:10.1016/J.IJDRR.2021.102581.
- [23] DEBAPRASAD D, UPRETI S. A survey and comparative evaluation of actor-critic methods in process control[J]. *The Canadian Journal of Chemical Engineering*,2022,100(9):2028-2056.
- [24] 苏佳,张晶晶,易卿武,等. 基于 CNN-CAM 的 NLOS/LOS 识别方法研究[J/OL]. *无线电工程*,1-11[2024-03-20]. <http://kns.cnki.net/kcms/detail/13.1097.TN.20230921.1857.014.html>.
- [25] ZENG Z, LIU et al. UWB NLOS identification with feature combination selection based on genetic algorithm[C]. 2019 IEEE International Conference on Consumer Electronics(ICCE),Las Vegas,2019:1-5.

作者简介

韦子辉,副教授,博士,主要研究方向为检测技术,室内定位技术。

E-mail: zihui-wei@163.com

廖戈,硕士研究生,主要研究方向为超宽带室内定位技术。

E-mail: gorica@qq.com

李明轩,硕士研究生,主要研究方向为检测技术,室内定位技术。

E-mail: lmx130429@163.com

周敬仪,硕士研究生,主要研究方向为超宽带室内定位技术。

E-mail: jingyizhou0702@163.com

董鹏(通信作者),实验师,主要研究方向为电气工程技术。

E-mail: dongpenghbu@126.com