

基于机器学习的室内UWB三维定位

田心童, 孙静, 张雨晴

南京信息工程大学, 江苏 南京

收稿日期: 2022年3月26日; 录用日期: 2022年4月21日; 发布日期: 2022年4月29日

摘要

随着无线网络的高速发展, 对无线定位技术的要求越来越高。UWB是一种较为优秀的定位技术, 但其容易受信号干扰, 导致其在商场、酒店等室内场合的定位效果不够理想, 故解决信号干扰下的UWB精确定位问题获得极大关注。本文使用的实验数据包含UWB定位的“无干扰数据”和“有干扰数据”。首先进行数据清洗, 第一步根据箱线图找出数据中的异常值并剔除, 第二步将数据进行K-means聚类, 完成相同或相似数据的删除。然后对清洗后保留下的“无干扰数据”和“有干扰数据”使用遗传算法优化的BP神经网络算法(GA-BP)估计Tag的三维坐标, 发现在GA-BP神经网络算法下, 根据“无干扰数据”和“有干扰数据”得到的三维坐标的均方误差都非常接近于0, 说明该模型预测精度较高。同时, 根据GA-BP神经网络算法对10组测试数据进行精确定位, 并建立XGBoost提升算法、支持向量机(SVM)和随机森林、KNN算法四种方法对数据是否在信号干扰下采集进行分类。通过比较四个模型分类准确率和F-score, 确定最终分类模型为经过参数优化的XGBoost算法。

关键词

UWB三维定位, GA-BP神经网络, XGBoost算法

Indoor Ultra-Wideband 3D Positioning Based on Machine Learning

Xintong Tian, Jing Sun, Yuqing Zhang

Nanjing University of Information Science and Technology, Nanjing Jiangsu

Received: Mar. 26th, 2022; accepted: Apr. 21st, 2022; published: Apr. 29th, 2022

Abstract

With the rapid development of wireless communication networks, the requirements for wireless positioning technology are getting higher and higher. UWB is an excellent positioning technology, but it is prone to signal interference, resulting in an unsatisfactory positioning effect in indoor

places such as shopping malls and hotels. Therefore, solving the problem of UWB precise positioning under signal interference has received great attention. The experimental data used in this paper includes “no interference data” and “interference data” for UWB positioning. The first step is to clean the data. The first step is to find outliers in the data according to the boxplot and eliminate them. The second step is to perform K-means clustering on the data to delete the same or similar data. Then use the BP neural network algorithm (GA-BP) optimized by the genetic algorithm to estimate the three-dimensional coordinates of the Tag for the “undisturbed data” and “interference data” retained after cleaning. The MSE of the three-dimensional coordinates obtained from “interference data” and “interference data” is very close to 0, indicating that the model has high prediction accuracy. At the same time, according to the GA-BP neural network algorithm, 10 sets of test data are accurately positioned, and XGBoost is established. Boosting algorithm, SVM, random forest, and KNN algorithm are used to classify whether the data is collected under signal interference. By comparing the classification accuracy and F-score of the four models, it is determined that the final classification model is the parameter-optimized XGBoost algorithm.

Keywords

UWB-3D Positioning, GA-BP Neural Network, XGBoost Algorithm

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着人类社会的不断发展，超宽带(Ultra-wideband, UWB)技术[1]逐渐走入大众的视野，成为了无线通信领域的研究热点。UWB 技术也被称之为脉冲无线电技术。它是通过发送和接收纳秒级脉冲信号来完成数据传输的，并且在传输过程中不需要传统通信技术中的载波，因此其数据传输速率较高，可以实现快速通信[2]。并且，对于通常使用的短距离范围内 UWB 技术来说，传输过程中的功耗非常低，仅仅只有几十 μW 。

除此以外，UWB 技术的定位精度高，目前基于 UWB 的定位技术具备实时的室内外精确跟踪能力，可达到厘米级甚至毫米级定位，这对卫星导航起到一个极好的补充作用。因此，UWB 技术无论是在军事领域还是民用领域都有着广泛的应用[3] [4] [5]。在军事方面，由于 UWB 定位精度高且具有很强的穿透能力，可探测地雷等地下物体、墙壁后方空间的物体，所以其在透视成像雷达、穿地探测雷达等方面得到了越来越多的应用。在民用领域，由于 UWB 技术可以对信号源进行探测，实时跟踪目标，其通常被应用于工厂、工地、隧道等的监管中，不仅可以提高工作效率，而且为工作人员的安全提供了一定的保障；除此以外，该技术还被应用于紧急救援、物流跟踪、医疗设备管控、区域管理、机器人等方面[6]。

无线通信系统的传播条件通常被分成视距传播(LOS)和非视距传播(NLOS)两种环境，LOS 环境下信号的传播是一个无遮挡的直达路径，而在 NLOS 环境下，信号传播路径较为复杂，传播时间更长。因此，在 LOS 环境下，UWB 技术可以达到厘米级的定位精度，并具有良好的抗多径干扰和衰弱的性能以及具有较强的穿透能力[7]。但是，在 NLOS 环境下，干扰信号和遮挡物较多，UWB 通信信号极易受到遮挡，从而产生误差；在较强信号干扰时，飞行时间受到影响，可能导致数据发生异常波动，无法完成室内定位，甚至造成严重事故。这一技术漏洞也是导致 UWB 在许多场合无法使用的原因，其发展和推广也因此受到限制。由此看来，如何解决信号干扰下的 UWB 精确定位问题亟待研究。

2. 方法介绍

2.1. TOF 测距技术

UWB 的定位技术有 TOA 法、TDOA 法和 PDOA 法等, 本文采用 TOF 方法, 它是 UWB 定位法中最常见的定位方法之一。TOF 测距技术属于双向测距技术, 其通过计算信号在标签与基站之间的飞行时间, 再乘以光速从而求出二者之间的距离。TOF 技术的测距过程如图 1 所示:

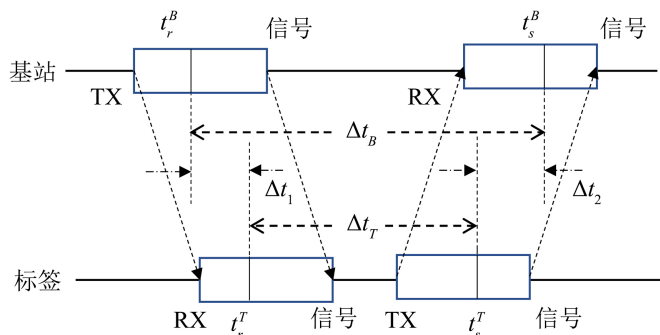


Figure 1. TOF ranging process

图 1. TOF 测距过程

假设基站在 t_s^b 时刻发出一个信号, 在 t_r^t 时刻标签收到该信号, 并于 t_s^t 时刻发出一个应答信号, 基站在 t_r^b 时刻收到该应答信号, 则 Δt_B 表示基站发出信号到收到信号的时间差, Δt_T 表示标签收到信号与发出信号的时间差。则信号传播时间 $\Delta t = \Delta t_B - \Delta t_T$, 从而得到标签和基站之间的距离 $l = \frac{1}{2}c\Delta t$, 其中 c 为光速。通常, 根据 TOF 技术测出的距离和各基站的坐标, 可以实现对标签的定位。

2.2. GA-BP 神经网络

2.2.1. BP 神经网络

BP 网络的产生归功于 BP 算法。BP 算法本质上是一种学习算法, 它将一组样本输入输出问题转化为非线性优化问题, 并通过梯度算法迭代求解权值问题。该网络中存在大量反向误差传播路径, 可以自适应地调整网络参数, 以满足不同对象对不同特性的适应性。BP 神经网络的过程主要分为两个阶段: 第一阶段是信号的前向传播, 从输入层经过隐藏层, 最后到输出层; 第二阶段是误差的反向传播, 从输出层到隐藏层, 最后到输入层, 调整隐藏层到输出层的权重和偏置, 输入层的权重和偏置为隐藏层[8]。

2.2.2. 遗传算法

遗传算法(GA)是一种通过模拟生物进化机制进行全局优化的启发式搜索方法, 它从一组初始的随机解开始, 称为种群, 种群中的每一个个体称为染色体。根据目标函数评价各染色体的适应度。经过选择、交叉、变异等一系列操作, 剔除适应度较低的染色体, 得到一个新的种群。这些操作重复进行, 直到染色体能够满足一定的条件, 即问题的近似最优解。BP 神经网络对初始网络权值敏感, 容易陷入平坦区域、局部极小化等问题。

2.3. 分类模型

2.3.1. 支持向量机

支持向量机(SVM), 是解决一种二分类问题的模型, 它的基本模型是定义在特征空间上的间隔最大

的线性分类器[9]。SVM 的学习算法就是求解凸二次规划的最优化算法，其基本想法是求解能够正确划分训练数据集并且几何间隔最大的分离超平面，如图 2 中的 $w \cdot x + b = 0$ 即为分离超平面。

支持向量一般用于分析线性可分的情况，在遇到线性不可分问题时，也可将低维数据空间映射到高维的特征空间，转化成线性可分的问题。因为 SVM 目标函数和分类决策函数都只涉及实例和实例之间的内积，所以不需要显式地指定非线性变换，而是用核函数替换当中的内积。使用 $\varnothing(x)$ 表示 x 经过映射后的向量，得到在特征空间中划分超平面后对应的模型：

$$f(x) = w^T \varnothing(x) + b \quad (1)$$

最终模型可使用线性核函数 $k(x_i, y_i) = x_i^T x_j$ 。

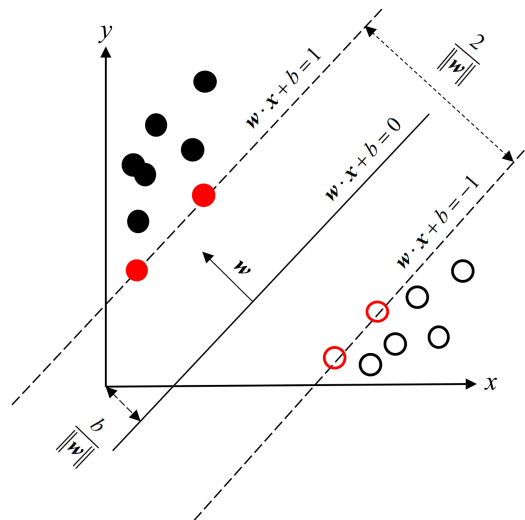


Figure 2. The principle of support vector machine
图 2. 支持向量机原理

2.3.2. 随机森林

随机森林是一种以决策树为基础的集成学习模型，其包含多个 CART 树分类器。随机森林模型的适应性较强，对于复杂的数据也能较好地处理，不容易产生过拟合问题。

随机森林的生成步骤如下：

1) 为了降低森林中每棵树之间的相关性，用 Bootstrap 从样本容量为 N 的训练集中有放回地随机抽取等量的样本，作为一棵决策树的训练样本。而未被 Bootstrap 方法抽中的样本被称为这棵树的袋外样本 (out-of-bag samples)，也称为 OOB 样本。

2) 对于单棵决策树，给定样本后，决策树由根节点开始自上而下递归分裂，即从根节点开始依次对训练样本进行分类。

3) 决策树的每个节点均按照步骤 2 进行分裂，直到节点无法继续分裂时停止。

4) 按照上述 3 个步骤生成多棵决策树，将生成的多棵决策树组成随机森林。在分类问题中，森林中的每一棵决策树分别进行判别和分类并输出一个分类结果，然后对所有决策树的分类结果进行投票，根据少数服从多数的原则，确定随机森林的最终分类结果。

2.3.3. XGBoost 模型

XGBoost 是一种改进的梯度提升决策树算法 (GBDT)，本质上仍然是基于 boosting 集成思想的加法模

型。它的模型训练时采用前向分布算法进行贪婪的学习，每次迭代都学习一棵 CART 树来拟合之前 t-1 棵树的预测结果与训练样本真实值的残差，具有计算复杂度低、运行速度快、准确度高等特点[10]。通过 3 个步骤优化 XGBoost 的目标函数：

- 1) 二阶泰勒展开，去除常数项，优化损失函数项；
- 2) 正则化项展开，去除常数项，优化正则化项；
- 3) 合并一次项系数、二次项系数，得到最终目标函数。

具体做法就是分裂后的目标函数值比单子叶子节点的目标函数的增益，同时为了限制树生长过深添加一个阈值，只有当增益大于该阈值才进行分裂，不断形成新的树，每次在上一次的预测基础上取最优进一步建树。为防止过拟合，通过设置树的最大深度且当样本权重和小于设定阈值时停止生长：

- 1) 当引入的分裂带来的增益小于设定阈值的时候，我们可以忽略掉这个分裂，所以并不是每一次分裂 lossfunction 整体都会增加的，有点预剪枝的意思，阈值参数为(即正则项里叶子节点数 T 的系数)；
- 2) 当树达到最大深度时则停止建立决策树，设置一个超参数 max_depth，避免树太深导致学习局部样本，从而过拟合；
- 3) 样本权重和小于设定阈值时则停止建树。

通过网格搜索法对 XGBoost 算法进行参数优化，即将各个参数可能的取值进行排列组合，列出所有可能的组合结果生成“网格”。然后将各组合用于 SVM 训练，并使用交叉验证对表现进行评估。在拟合函数尝试了所有的参数组合后，返回一个合适的分类器，自动调整至最佳参数组合。在原始数据集划分成训练集和测试集以后，其中测试集除了用作调整参数，也用来测量模型的好坏，这样做导致最终的评分结果比实际效果要好，而我们的目的是将训练模型应用在未知数据上。因此如图 3 本文对原训练集进行二次划分，分成训练集和验证集，划分的最终结果为训练集、验证集和测试集，其中训练集用来模型训练，验证集用来调整参数，而测试集用来衡量模型表现好坏。

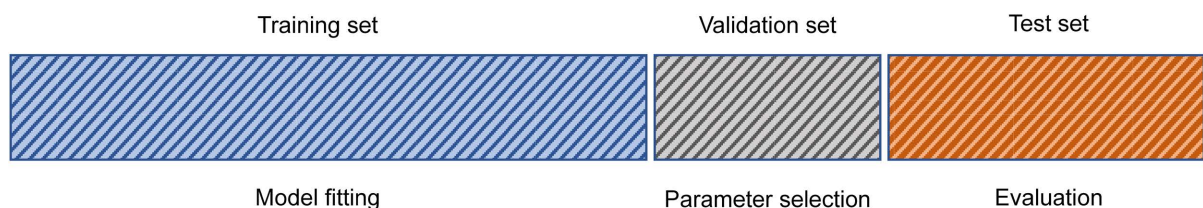


Figure 3. Dataset secondary division

图 3. 数据集二次划分

3. 数据获取

3.1. 实验场景

如图 4 所示，在 5000 mm*5000 mm*3000 mm 的测试环境中，分别在 4 个角落 A0 (0, 0, 1300), A1 (5000, 0, 1700), A2 (0, 5000, 1700), A3 (5000, 5000, 1300) 放置 UWB 锚点(anchor)，锚点向所有方向发送信号。Tag 是 UWB 标签(靶点)，即需要定位的目标(只在测试环境范围内)。Tag 接收到 4 个 UWB 锚点(anchor) 的信号(无论信号是否干扰，Tag 一般都可以接收到信号)，利用 TOF 技术，分别解算出对应的 4 个距离数据。

实验在实验场景中采集了 Tag 在 324 个不同位置，在信号无干扰和信号干扰下的 UWB 数据，即每个位置采集 2 次，一次信号无干扰，另一次信号有干扰(锚点与靶点间有遮挡)。

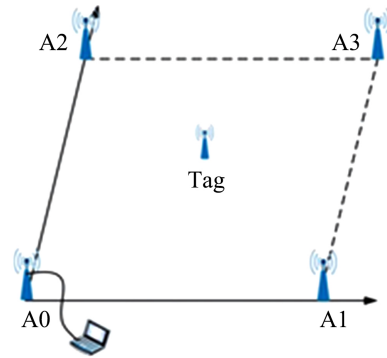


Figure 4. Schematic diagram of the measured environment
图 4. 实测环境示意图

3.2. 数据预处理

3.2.1. 异常值处理

在异常值处理前, 首先通过箱线图对异常值进行识别。在箱线图中, 异常值被定义为小于 $Q_L - 1.5IQR$ 或大于 $Q_U + 1.5IQR$ 的值, 其中 Q_L 是下四分位数, Q_U 是上四分位数, IQR 为四分位距。下图 5 中(a)、(b) 分别表示根据无干扰数据和有干扰数据画出的箱线图, 并剔除异常数据。

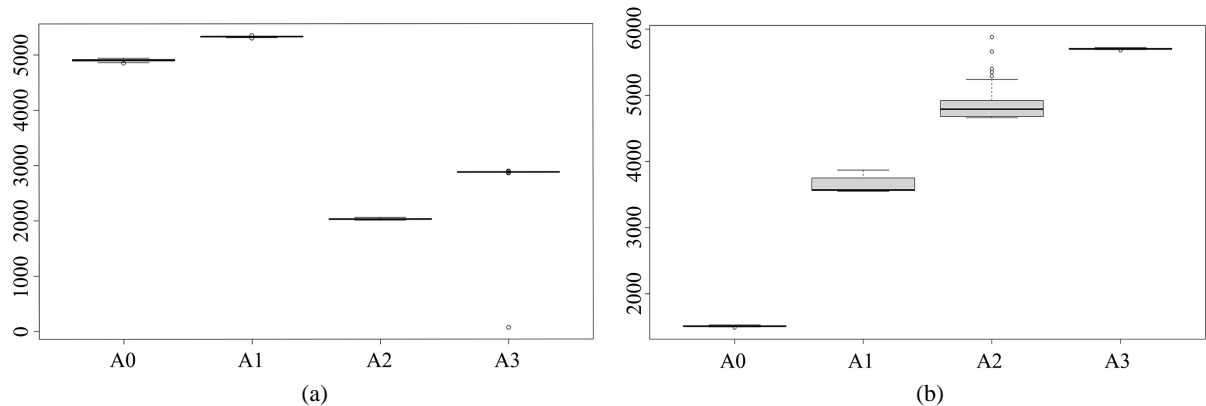


Figure 5. Box plot
图 5. 箱线图

3.2.2. K-means 聚类处理重复值和相似值

K 均值聚类也是一种判断数据相似性的方法, 其步骤如下:

- 1) 给定类别 K , 并随机选取 K 个数据点作为质心;
- 2) 比较数据集中每个数据到质心的欧式距离, 并将其分配到距离最近的一类。
- 3) 分类完成后, 对每类分别计算平均值, 作为新的质心;
- 4) 重复上述步骤直到分类不再变化为止。

在进行聚类分析之前, 应该先确定要分成的类别数量 K 。通常情况下, 随着 K 增多, 每一个类别中的数据会减少, 组内平方和(WSS)也就越来越小。当 WSS 减少得缓慢时, 说明进一步增大类别数量对聚类效果影响较小, 此时对应的 K 值就是最合适的类别数量。本文以无干扰数据为例进行聚类分析, 当 $K = 2$ 时, K 增大对 WSS 的影响程度明显减小, 则最合理的 K 取值为 2。然后, 根据 $K = 2$ 对数据进行聚类, 聚类结果可视化如图 6 所示。

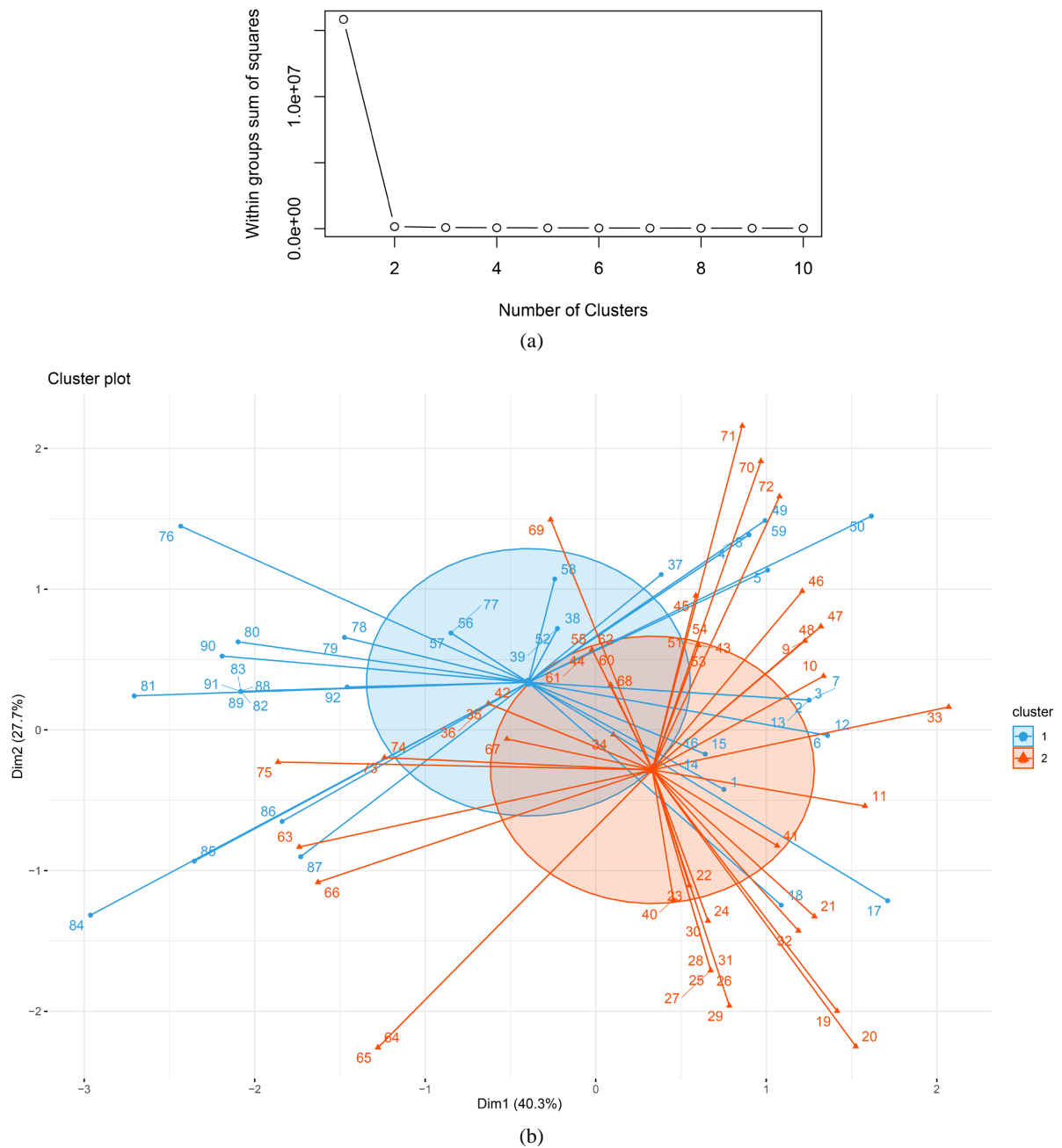


Figure 6. Cluster visualization
图 6. 聚类可视化

4. 定位预测

4.1. 三维几何模型

在三维坐标系中，令 A_0 的坐标为 (x_0, y_0, z_0) ， A_1 的坐标为 (x_1, y_1, z_1) ， A_2 的坐标为 (x_2, y_2, z_2) ， A_3 的坐标为 (x_3, y_3, z_3) ，同时，令 Tag 的坐标为 (x, y, z) 。 l_i 表示锚点 A_i 到 Tag 的测量距离，其中 $i=0,1,2,3$ 。根据根据立体几何原理，分别以 A_0 、 A_1 、 A_2 、 A_3 为球心， l_i 为半径画球，为半径画球，靶点所在的位置即为各球的交点，如图 7 所示。

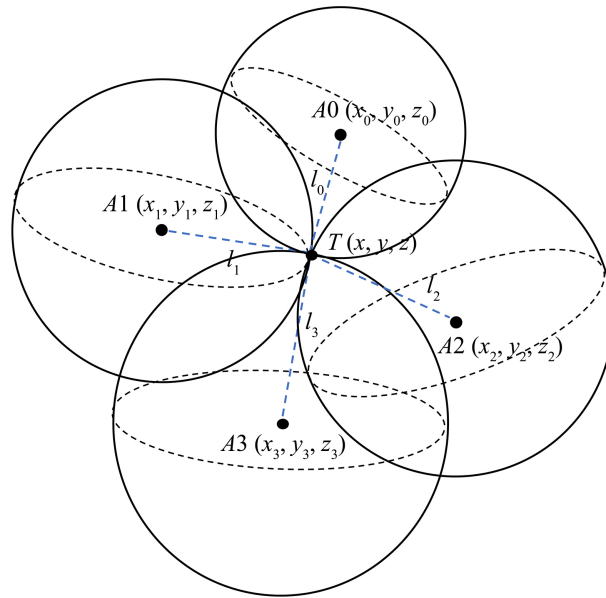


Figure 7. 3D geometric model diagram
图 7. 三维几何模型图

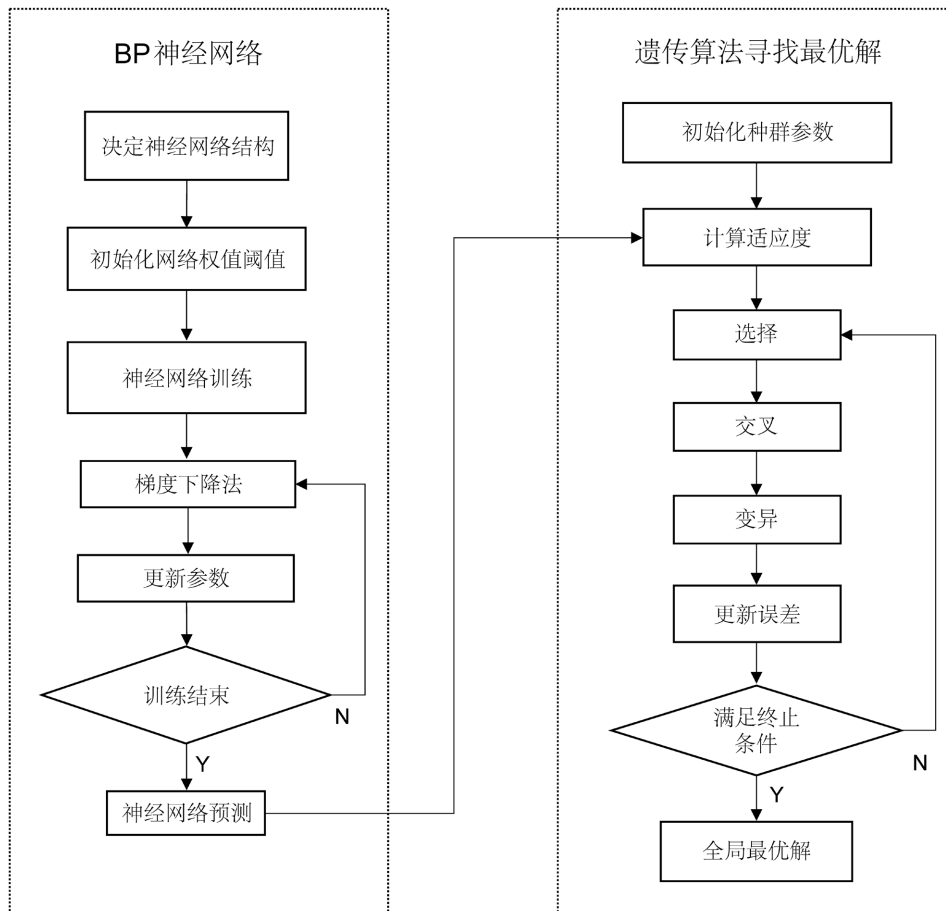


Figure 8. GA-BP flowchart
图 8. GA-BP 流程图

4.2. GA-BP 神经网络搭建

4.2.1. 模型框架

对于非线性方程组，仅通过数值的代入求解可能会有多个解，且往往难以找到最优解，因此本文使用 GA-BP 算法，将 BP 神经网络和遗传算法相结合，利用神经网络的非线性拟合能力和遗传算法的寻优能力寻找非线性方程组的最优解。

GA-BP 算法求最优解主要分为 BP 神经网络训练拟合和遗传算法极值寻优两步，算法流程如图 8，神经网络首先根据方程组的特点构建合适的 BP 神经网络，并用数据进行训练以设置合适的权值，然后按照一定的规则获得输出数据。遗传算法把训练后的 BP 神经网络预测结果作为个体适应度，经过反复选择、交叉和变异，剔除适应度较低的个体，逼近全局最优解。

4.2.2. 三维坐标定位预测

本文选用均方误差(MSE)对模型的有效性(精度)进行评价。GA-BP 神经网络寻找最优解过程中，其模型的 MSE 值在不断变化，最后趋于收敛。本文将基于 GA-BP 的 MSE 变化过程进行可视化，如图 9 所示。

图 9(a)是根据“无干扰数据”计算得到的 MSE，图 9(b)是根据“有干扰数据”得到的 MSE 值。

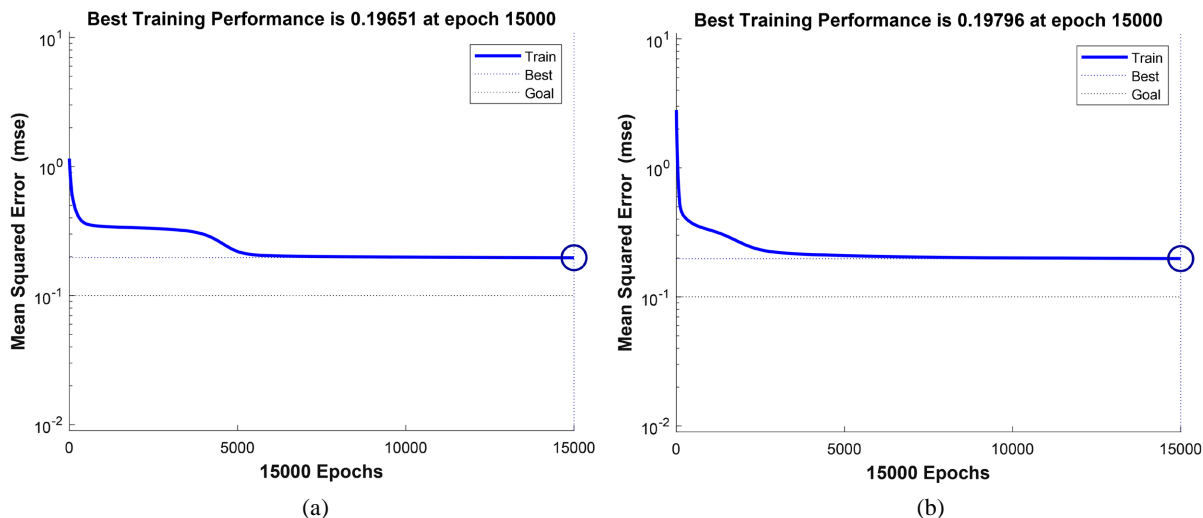


Figure 9. MSE of 3D coordinate localization of GA-BP model

图 9. GA-BP 模型三维坐标定位的 MSE

由此可见，GA-BP 神经网络的预测效果较好，故本文利用该算法对前 5 组(信号无干扰)数据和后 5 组(信号有干扰)的测试数据进行精确定位，结果如表 1:

Table 1. Coordinate estimation
表 1. 坐标估计

	x	y	z
数据 1	779.8537	1967.2588	1506.1116
数据 2	3412.2455	2450.9619	1467.5003
数据 3	2644.4258	2396.8730	1485.4639

Continued

数据 4	2240.5695	2359.5642	1493.4047
数据 5	1800.3975	2225.9610	1493.9124
数据 6	1883.2266	762.8736	1417.1968
数据 7	4425.8317	1337.5702	1485.0029
数据 8	1669.7740	1537.3152	1460.7829
数据 9	3944.0055	1802.8743	1497.6125
数据 10	4604.7814	1505.9533	1485.8438

5. 信号干扰判断

5.1. 问题分析

室内环境复杂,会产生较大的定位误差,而误差的主要来源是遮挡造成时延引起的误差,故需要提前进行有无信号干扰的判断。但 UWB 在采集数据时并不知道信号有无干扰,所以判断信号有无干扰是 UWB 精确定位问题的重点和难点[11]。我们根据本文,从经过数据处理后的已知数据信息中进行特征提取,选取合理特征参量集进行信号分类,使用支持向量机、随机森林、K 近邻和 XGBoost 提升算法分别对已有数据进行训练和测试,然后通过比较模型的分类准确率和 F-score 确定最终分类模型,以构建一个识别准确率较高、环境适用性较强、计算复杂度较低的实时信号干扰识别方法。

5.2. 多模型分类预测

5.2.1. SVM 和 KNN 分类

根据 SVM 和 KNN 原理,在本文的 SVM、KNN 分类器中,对 10 组测试数据分类结果如表 2 所示。

Table 2. Discriminant result
表 2. 信号有无干扰判别结果

数据	SVM 分类结果	KNN 分类结果
1	无干扰	无干扰
2	有干扰	有干扰
3	无干扰	无干扰
4	无干扰	无干扰
5	无干扰	有干扰
6	有干扰	有干扰
7	无干扰	无干扰
8	有干扰	有干扰
9	有干扰	有干扰
10	无干扰	无干扰

5.2.2. 随机森林分类

在建立随机森林的模型之前, 需要寻找模型中参数 $mtry$ 和 $ntree$ 的最优值, $mtry$ 是指树节点用于分裂的变量个数, $ntree$ 是指随机森林所包含的决策树数目。首先根据 OOB 误判率确定最佳 $mtry$ 值。本文自变量有 5 个, 将 $mtry$ 参数设置为 1 至 5 进行 5 次建模, 计算不同 $mtry$ 取值下基于 OOB 数据的模型误判率, 可以发现 OOB 误判率最小时对应的 $mtry$ 值为 5, 因此本文选择 5 作为最优 $mtry$ 值。其次, 根据模型误差确定最佳参数 $ntree$ 。绘制模型误差与决策树数量的关系图, 具体结果如图 10。模型误差大概在 $ntree$ 为 400 时趋于平稳, 故本文选择 400 作为 $ntree$ 参数的取值。根据以上结果, 以 $mtry = 5$, $ntree = 400$ 建立随机森林模型。

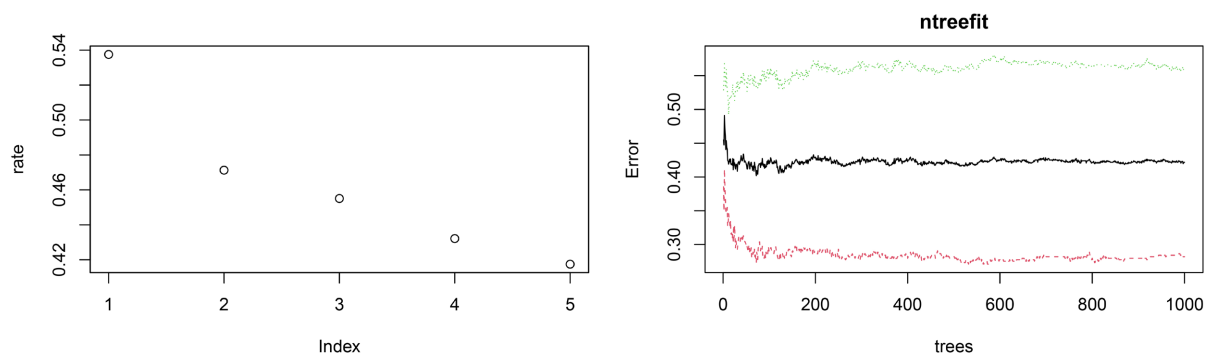


Figure 10. Random forest parameter selection

图 10. 随机森林参数选择

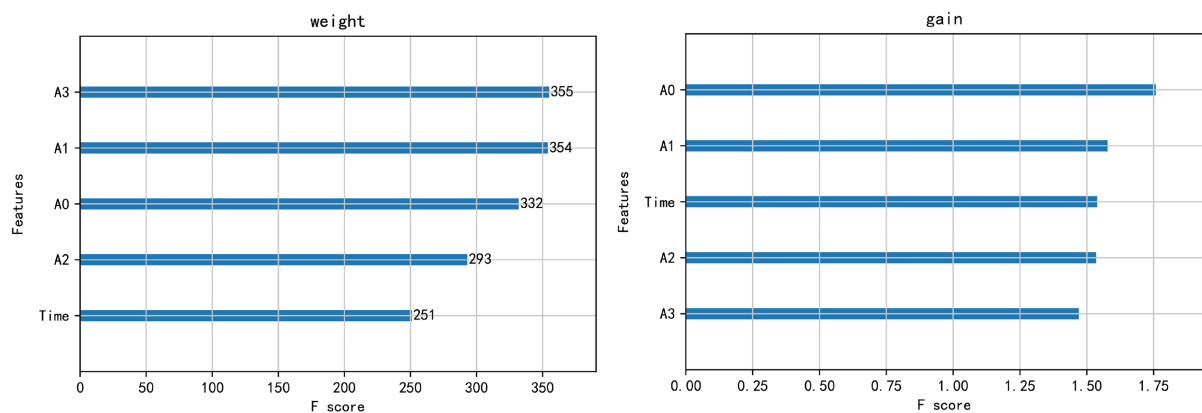


Figure 11. Weight and gain

图 11. 特征使用次数及其平均信息增益

5.2.3. XGboost 分类

本文通过特征用到的次数(weight)、平均信息增益(gain)来评估特征的重要性, 选出特征参量。图 11 中 A0 代表 A0 锚点到靶点的测距值, A2、A3 和 A4 同理; Time 表示上一次测量的时间与当前时刻的时间差。Weight 表示在所有的树中, 一个特征总共多少次被用于分割数据, A2 锚点到靶点的测距值是用到最多的特征, A3、A1 的 F-score 也高达 0.9 以上, 相比而言最低的是时间差 Time, 仅有 0.52。这个指标是一个变量被用于分割的绝对次数, 这就造成了 weight 的缺点, 对于定类变量和定序变量, 由于他们天生能用于分割的点就比较少, 很容易被 weight 指标所忽略。gain 表中的 F-score 值代表使用该特征的所有分割的平均信息增益(average gain), 也就是节点分裂时, 该特征带来信息增益(目标函数)优化的平均值,

是通过取每个特征对模型中每棵树的贡献来计算对应特征对模型的相对贡献， $gain$ 指标的值越高，意味着它对生成预测更重要，从图中可以看到距离特征要优于时间差。

在特征分析后，我们选取了 A0、A1、A2、A3 四个特征与标签组合绘制散点图和直方图，寻找特征解释。将类别定义为两类，0 代表无信号干扰，1 代表受到信号干扰，0 和 1 在不同锚点和不同距离上近似服从高斯分布，且 0 和 1 的散点几乎重合分布，但是因为受定位和信号干扰，相近坐标的点的实际靶点位置可能差别很大，因此需要通过合适的分类模型，通过学习数据特征和训练神经网络来构建分类器判断无标签的输入数据是否受到信号干扰。

5.3. 模型评估与比较

本问题对模型的有效性评估主要通过衡量分类问题的评价指标，包括混淆矩阵、准确率、精确率、召回率和 F1-Score。在评价分类模型时，我们通常将关注的类别作为积极类别，其他类别作为消极类别，计算分类器对数据集的预测正确或错误的概率。图 12(a)~(d) 分别表示 SVM、随机森林、KNN 和 XGBoost 得到的混淆矩阵：

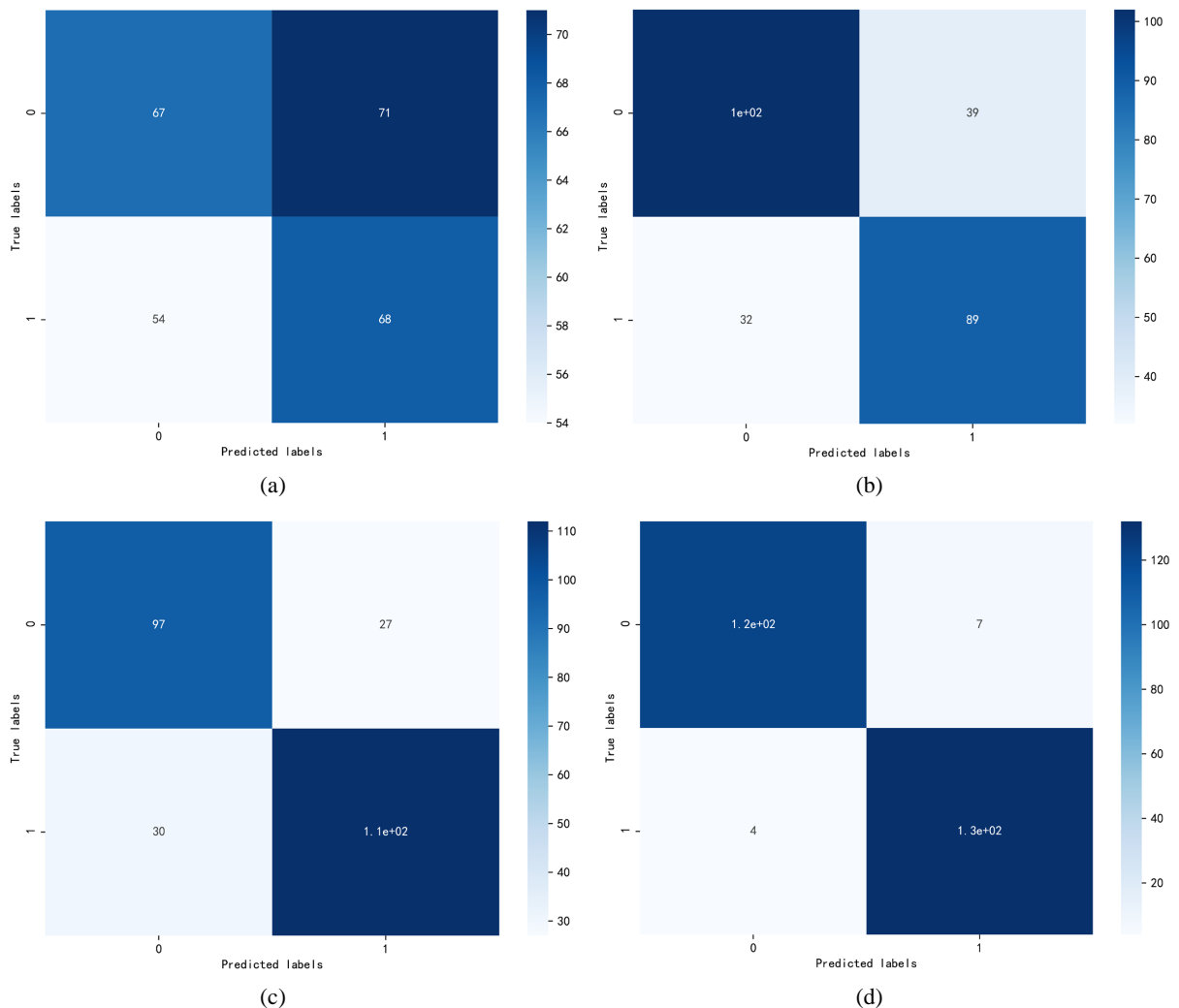


Figure 12. Confusion matrix diagram for each classification method

图 12. 各分类方法的混淆矩阵图

根据各方法的混淆矩阵可以计算出各方法的四个评价指标如表 3 所示, 其中准确率是预测正确的样本数量占总量的百分比; 精准率是在模型预测为正样本的结果中真正是正样本所占的百分比; 召回率是在实际为正样本中被预测为正样本所占的百分比; F1 指数精准率和召回率的调和平均数。

Table 3. The evaluation index of each classification model

表 3. 各分类模型评价指标

评价指标	SVM	随机森林	KNN	XGBoost
准确率	0.52	0.73	0.78	0.86
精确率	0.55	0.76	0.76	0.86
召回率	0.49	0.72	0.78	0.88
F1 指数	0.52	0.74	0.77	0.87

6. 总结

超宽带信号以其时间分辨能力强和抗多径等性能, 可以达到很好的定位效果。本文在简要介绍超宽带信号、信道的基础上, 对超宽带定位技术进行了深入研究, 使其实现高精度定位。现将主要工作总结如下:

1) 本文介绍了超宽带无线通信理论, 研究分析了超宽带技术在定位方面的优势。同时分析了 UWB 信号在室内传播时容易受多径和噪声等因素的干扰, 造成接收端接收信号发生相当程度的畸变, 相关接收机难以实现准确的相关接收, 导致时间测量值出现偏差, 进而影响测距值精度。

2) 对实验数据进行清洗和分析后, 构建 GA-BP 神经网络模型, 将其应用于预测三维坐标的精确定位问题, 分别对无信号干扰数据和有信号干扰数据进行预测, 预测的均方误差皆小于 0.1, 证明了该定位模型的有效性。

3) 为了解决室内环境复杂会产生较大的定位误差的问题, 对定位数据进行有无信号干扰的分类判断。本文从已知的数据信息中进行特征提取, 选取合理特征参量集进行信号分类, 引入 SVM、KNN、随机森林和 XGboost 四种算法分别建立二分类模型, 然后通过评估指标比较模型的性能, 确定 XGboost 模型效果最优。

参考文献

- [1] Fontana, R.J. (2004) Recent System Applications of Short-Pulse Ultra-Wideband (UWB) Technology. *IEEE Transactions on Microwave Theory and Techniques*, **52**, 2087-2104. <https://doi.org/10.1109/TMTT.2004.834186>
- [2] 李楠, 梁冰. 基于 UWB 的三维定位和优化滤波方法[J]. 河北大学学报(自然科学版), 2021, 41(3): 329-336.
- [3] 沈国杰, 周德胜, 郑雪娜, 李敏. 基于 UWB 的矿井巷道精准三维定位系统设计[J]. 矿业安全与环保, 2020, 47(6): 80-84
- [4] 朱自强. 超宽带定位的关键技术研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2019.
- [5] Ni, D., Postolache, O.A., Mi, C., et al. (2019) UWB Indoor Positioning Application Based on Kalman Filter and 3-D TOA Localization Algorithm. 2019 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE), Bucharest, 28-30 March 2019, 1-6. <https://doi.org/10.1109/ATEE.2019.8724907>
- [6] 李鲁昭. 基于 UWB 的停车场汽车定位系统的研究与应用[D]: [硕士学位论文]. 大连: 大连理工大学, 2019.
- [7] 焦海龙. 复杂环境下 UWB 定位准确度的测试及优化研究[D]: [硕士学位论文]. 海口: 海南大学, 2018.
- [8] 沈冬冬, 李晓伟, 宋旭文, 等. 基于多层神经网络的超宽带室内精确定位算法[J]. 电子科技, 2014, 27(5): 161-163.

- [9] Poulou, A., Emeršič, Ž., Eyobu, O.S., *et al.* (2020) An Accurate Indoor User Position Estimator for Multiple Anchor UWB Localization. 2020 *International Conference on Information and Communication Technology Convergence (ICTC)*, 21-23 October 2020, 478-482. <https://doi.org/10.1109/ICTC49870.2020.9289338>
- [10] Sahin, E.K. (2020) Assessing the Predictive Capability of Ensemble Tree Methods for Landslide Susceptibility Mapping Using XGBoost, Gradient Boosting Machine, and Random Forest. *SN Applied Sciences*, **2**, 1-17. <https://doi.org/10.1007/s42452-020-3060-1>
- [11] 张少辉, 亓玉浩, 翟方文, 吕洪波, 宋亦旭. 基于冗余距离筛选的 UWB 定位优化方法[J/OL]. 清华大学学报(自然科学版), 2021: 1-9. <https://doi.org/10.16511/j.cnki.qhdxxb.2021.21.040>, 2021-09-10.