



中国研究生创新实践系列大赛
“华为杯”第十八届中国研究生
数学建模竞赛

学 校 北京航空航天大学

参赛队号 21100060056

1.郑耀威

队员姓名 2.陈俊旭

3.皮翔宇

中国研究生创新实践系列大赛
“华为杯”第十八届中国研究生
数学建模竞赛

题 目 基于神经网络噪声预测和混合高斯降噪的超宽带精确定位

摘 要：

近些年来，室内精确定位技术在军事和民用领域得到了越来越广泛的应用，提高室内定位精度成为了通信及信息科学领域的前沿课题之一。超宽带（UMB）定位技术由于具有高定位精度以及良好的抗多径干扰能力和穿透能力，受到了学术界和业界的广泛关注。在室内定位的应用中，复杂多变的室内环境会对 UWB 通信信号造成较大的干扰，严重影响定位精度。为此，我们通过采集实际场景的数据，利用 UWB 技术中基于飞行时间（TOF）的测距原理建立数学模型，实现信号干扰下目标物的三维精确定位。

本文的贡献主要分为以下三点：（1）我们采用最小二乘法和牛顿迭代法对靶点位置进行估计，将定位精度提高到了厘米级；（2）我们采用混合高斯模型对测量距离的误差进行建模，并且根据噪声的统计信息建立了相应的降噪模型，将误差平均减少了 23.4 厘米；（3）我们提出了基于神经网络的噪声分类方法，结合降噪模型和锚点坐标估计算法生成输入特征，将实验数据作为训练数据对神经网络模型进行拟合，得到了 90% 的分类准确率。最终定位算法对于正常数据的三维定位精度为 14.96 厘米，对于异常数据的三维定位精度为 29.84 厘米。

针对问题一，需要对测量数据进行预处理。先利用字符串处理方法将数据文件中每组数据四个锚点的测距值提取出来并转换为二维表形式，然后删除同一时间戳下未包含全部四个锚点的测量距离的数据。接着，通过统计分析发现正常数据符合高斯分布，异常数据符合混合高斯分布，分别利用高斯 σ 原则和 k-means 聚类算法删除偏离均值超过单倍标准差的离群数据。结果表明，经过删除缺失数据和离群数据两步数据处理，正常数据和异常数据保留率分别为 18.14% 和 15.88%。

针对问题二，需要分别对正常数据和异常数据建立定位模型。首先对实验数据进行分布统计，得到实验数据的统计特征，从而对测量技术自身的误差和受干扰时产生的误差建立噪声模型，进而使用神经网络预测噪声的类别。接着我们根据最大似然估计得到对不同噪声的降噪方法，使用该方法对测量距离进行降噪。然后，我们由锚点坐标和测量距离建立定位方程组，通过最小二乘法求解得到靶点的估测坐标，再利用牛顿迭代法对该坐标进行进一步的优化，使之更加靠近真实的靶点坐标。结果表明，我们提出的降噪模型有效提升了定位的准确性，定位模型精度较为理想，对测试数据均能求解得到合理的靶点坐标。

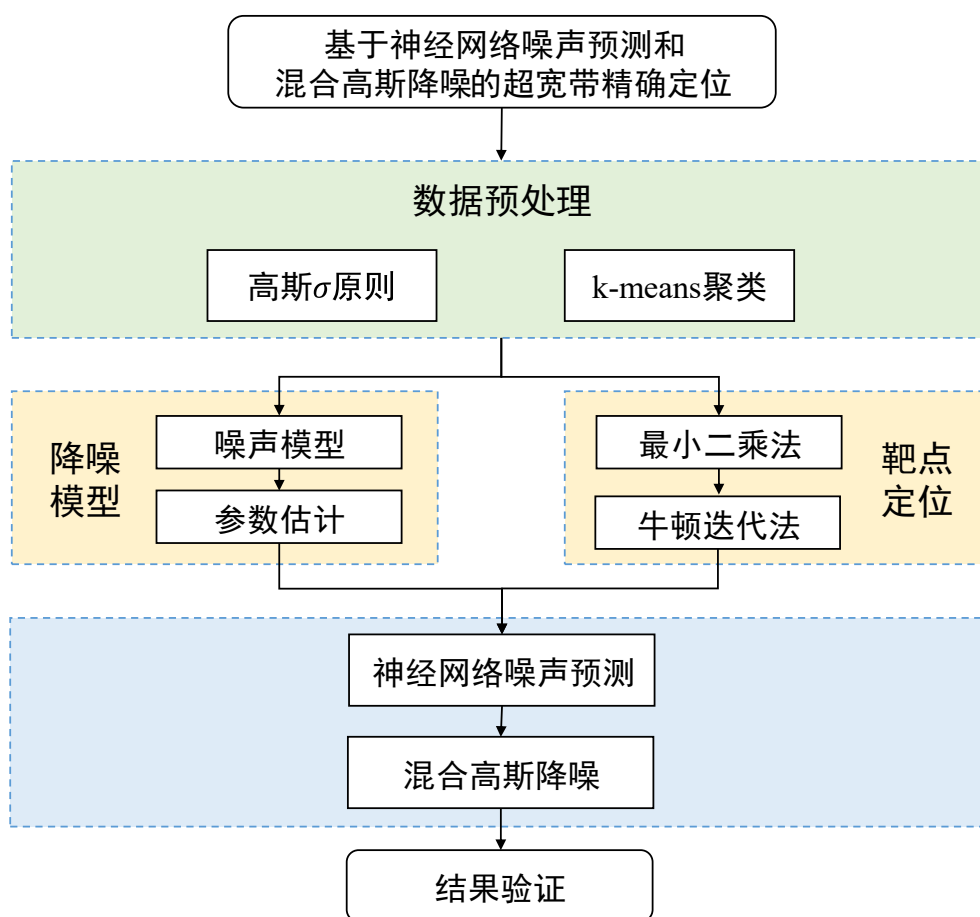
针对问题三，需要将定位模型应用于不同场景。我们基于锚点坐标信息建立的定位模型当锚点坐标改变时，仍然能利用降噪模型对测量数据进行降噪，再利用新的锚点坐标重

新建立定位方程组，求解得到靶点的坐标。结果表明，更换场景之后我们所设计的定位模型依然可以合理估计出靶点的空间位置。

针对问题四，需要建立能够根据测量数据判断采集数据时信号有无干扰的分类模型。在问题二中所建立的神经网络可以对噪声类型进行预测，得出在采集数据时是否存在干扰，因此能够直接对测量数据进行分类。结果表明，我们建立的神经网络对测量数据的分类准确率较高。

针对问题五，需要根据运动靶点的测量数据定位靶点的运动轨迹。我们利用定位模型，对运动靶点在每一时刻的坐标进行静态定位，对得到的所有靶点坐标绘制散点图，再利用最小二乘法多项式拟合得到靶点在空间中的运动轨迹。结果表明，靶点在空间中的动态运动轨迹近似一条平滑曲线。除此之外，静态估计中得出的靶点坐标与运动轨迹之间的偏离较小，因此可认为该运动轨迹定位方法较为有效。

关键词：UWB 定位；混合高斯降噪；神经网络；信号处理



目 录

一、问题阐述	4
1.1 问题背景	4
1.2 问题分析	4
二、假设和符号	5
2.1 模型假设	5
2.2 符号说明	5
三、模型建立	5
3.1 数据预处理	5
3.2 三维定位模型	7
3.2.1 建立定位方程组	7
3.2.2 最小二乘法求解方程组	8
3.2.3 牛顿迭代法优化近似解	9
3.3 噪声模型	10
3.3.1 测量误差模型	10
3.3.2 干扰误差模型	11
3.3.3 最大似然估计	13
3.3.4 降噪方法	14
3.4 神经网络噪声预测	14
3.4.1 生成训练数据特征	14
3.4.2 生成训练数据标签	15
3.4.3 训练神经网络	15
3.5 靶点位置计算	16
3.6 运动定位	16
4 模型求解	17
4.1 问题一	17
4.2 问题二	17
4.3 问题三	18
4.4 问题四	18
4.5 问题五	18
结论	19
参考文献	20
附录	20

一、问题阐述

1.1 问题背景

近些年来，室内定位技术服务迅速发展，在特殊人群监护、建筑管理及物联网等领域得到了广泛的应用^[1]。以 GPS 为代表的传统定位技术由于受到技术原理的限制，在室内环境的定位效果较不理想。而 UWB (Ultra-Wideband) 的脉冲宽度仅为纳秒或亚纳秒级，理论上可以获得厘米级甚至更高的定位精度，因此在精确定位的应用场景中极具潜力。同时由于其时间分辨率高，具有抗多径干扰能力和一定的穿透能力，能够在室内环境下完成定位。因此 UWB 精确定位技术将会对导航系统起到一个极好的补充作用，在军用、商用及民用领域具有广阔的发展前景。

UWB 定位技术主要包括三种方法，即基于到达角度、基于接收信号强度和基于到达时间进行估计^[2]。本文仅考虑基于飞行时间 (Time of Flight, TOF) 的测距技术，该技术通过计算信号在两个模块的飞行时间，再乘以光速来求得两个模块之间的距离。这个距离虽然会受信道噪声、波形失真、时钟异步等因素的影响存在一定程度的误差，但仍然具有较高的精度。若已知 4 个锚点的位置信息和目标物体离对应锚点的距离，我们利用 UWB 定位技术便可求解出目标物体的位置。然而，室内环境复杂多变，常见的墙体和材料的电磁特性会严重影响 UWB 信号的传播，使信号受到较强的干扰，从而测得的距离数据会有较大误差以致无法完成室内定位^[3]。因此，如何在信号受到干扰的条件下实现超宽带精确定位成为了业界学者十分重视的问题。

事实上，UWB 信号受到的干扰大致可总结为信号衰减、目标移动、多径效应等类型，具有一定的规律。通过对这些干扰因素进行统计分析建模并修正测距误差，能够减轻信号干扰的影响，提高室内定位精度，这将对室内定位的发展具有里程碑式的意义。

1.2 问题分析

信号干扰下的超宽带精确定位共有五个问题需要进行研究。首先通过实际场景实测，在信号无干扰和有干扰两种情况下，利用 UWB 定位技术采集锚点 (Anchor) 与靶点 (Tag) 之间的距离，分别作为“正常数据”和“异常数据”记录。接着需要对数据进行预处理，删除部分异常或缺失的数据，分别对正常数据和异常数据设计合适的数学模型，估测靶点的精确位置，并验证定位模型的有效性。同时，该定位模型还需要能够在不同的实际场景下也能进行有效定位。除此之外，我们还需要对在未知场景下采集到的数据进行分类，以判断信号在采集时是否受到了干扰；最后，我们要对动态靶点进行定位并给出其运动中的动态轨迹。接下来，我们对这五个问题分别进行详细的分析。

问题一：在信号受到干扰和无干扰两种情况下，靶点在同一坐标点上都会采集到多组数据，这些数据不能直接用于模型的建立，我们需要对这些数据进行预处理，将其转化为便于使用程序处理的格式，删除掉异常或缺失的数据，最后以矩阵的形式输出保留的所有“正常数据”和“异常数据”。

问题二：针对处理后的正常数据和异常数据分别建立数学模型，估计靶点的精确位置，给出定位模型的 3 维精度、2 维精度以及 1 维精度，说明定位模型的有效性，并利用 5 组信号无干扰的数据和 5 组信号有干扰的数据进行检验。

问题三：虽然训练数据采集于同一实验场景，但建立的定位模型需要能够应用于不同的实际场景。因此通过对不同实验场景下 5 组信号无干扰的数据和 5 组信号有干扰的数据进行精确定位，检验模型对不同场景的适用性。

问题四：定位模型是在已知信号有无干扰的条件下建立的，然而在实际应用时，UWB

无法在采集数据时得知信号是否受到干扰，因此我们需要建立一个分类模型，判断采集到的信号有无干扰并利用采集于实验场景下的 10 组数据检验模型的有效性。

问题五：实际应用时，靶点可能会移动，并且信号干扰是随机出现的，需要利用分类模型和降噪模型对靶点的对动态靶点的运动轨迹进行精确定位并画出运动轨迹图。

二、假设和符号

2.1 模型假设

为了方便模型建立，我们在本文中考虑了以下假设条件：

- (1) 假设四个锚点在测距时不会互相影响，即四个测距值相互独立；
- (2) 假设锚点空间位置的变化对测距不会产生影响；
- (3) 假设靶点的运动状态不影响测距过程；
- (4) 假设每次测量时锚点的状态均不变。

2.2 符号说明

表 1 符号说明

符号	说明
A	锚点
T	靶点
d	测量距离
d^*	真实距离
g	测量误差
u	干扰误差
e	总误差
q	估测靶点
f	分类特征
y	类别概率分布
p	挡板状态

三、模型建立

3.1 数据预处理

对于每个数据文件，我们首先使用 Python 编写字符串处理算法，从每组数据中提取对于四个锚点的测量距离。接着我们对每组数据进行预处理，去除缺失和离群数据。

在预处理步骤中，我们先删除存在缺失的数据。我们将当前时间下测量距离不足四个的数据视为缺失数据，表 2 呈现了正常数据和异常数据中的一个缺失样例。

表 2 缺失数据样例

存在缺失的正常数据	存在缺失的异常数据
T:090531711:RR:0:0:780:780:232:3304	T:091216784:RR:0:0:3460:3460:131:5251
T:090531711:RR:0:1:4550:4550:232:3304	T:091216785:RR:0:1:5420:5420:131:5251
T:090531712:RR:0:2:4550:4550:232:3304	T:091216785:RR:0:2:2150:2150:131:5251
T:090531712:RR:0:3:6300:6300:232:3304	T:091216785:RR:0:3:4870:4870:131:5251

我们对经过处理后的数据进行进一步的分析，首先绘制正常数据的直方图，观察四个

测量距离的频数统计数据，如图 1 所示。

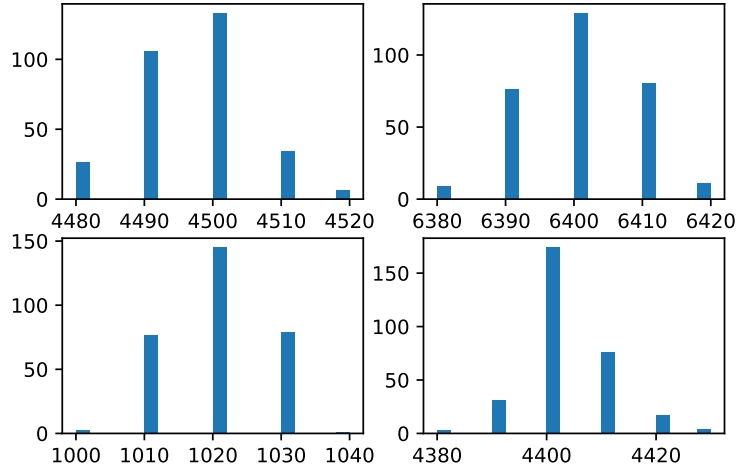


图 1 正常数据中四个锚点测量距离的频数分布

由图 1 可知正常数据大致呈高斯分布，因此我们假设正常数据相对第 i 个锚点的测量距离满足：

$$D_i \sim \mathcal{N}(\mu_i, \sigma_i^2), \quad i = 1, 2, 3, 4$$

我们从正常数据中随机抽取足够多个样本，利用样本均值和样本方差分别估计 μ_i 和 σ_i ，删除处于 $(\mu_i - \sigma_i, \mu_i + \sigma_i)$ 之外的数据。

我们再绘制异常数据的直方图，观察四个测量距离的频数统计数据，如图 2 所示。

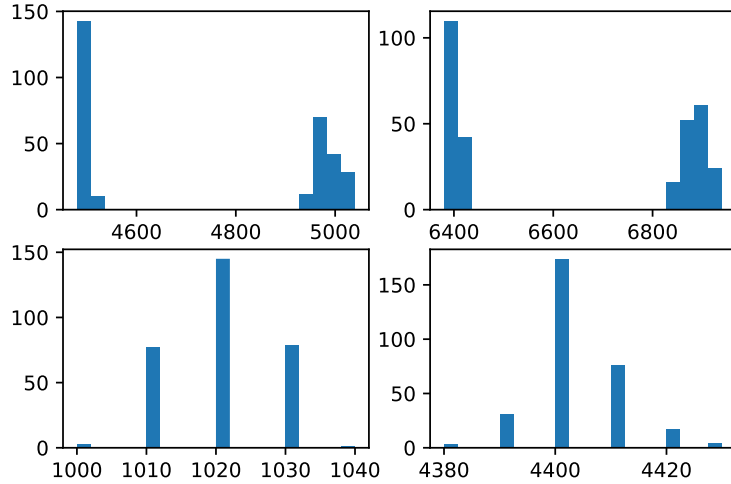


图 2 异常数据中四个锚点测量距离的频数分布

由图 2 可知异常数据大致呈混合高斯分布。对于异常数据，我们将每组数据中的四个测量距离看作四维向量，采用 k-means 聚类算法对某个文件的全部数据进行分类，以寻找混合高斯分布中每个子分布的中心，并删除离聚类中心过远的数据。

k-means 聚类算法是一种通过迭代方法求解的聚类分析算法^[4]，其基本思想是通过不断迭代改变 K 个类的中心，使各个样本距离所属类中心点的误差平方和（SSE）最小。

通常来讲，误差平方和 SSE 随着类数 K 的增加而单调减少。当 K 刚开始增加时，SSE 会呈现迅速减小的趋势；当 K 增加到一定值时，K 值减小的速度会显著放缓；而当 K 的值增加到等于总样本数时，误差平方和 SSE 将会等于零，即每个样本都是当前聚类的中心。因此 K 值的选择将会直接影响最终的聚类效果，当 K 过小时，算法无法将特征差异明显的的数据分成不同类别；当 K 过大时，将会产生冗余的类别并影响聚类效果。

为了找到合适的 K 值，我们绘制误差平方和 SSE 随 K 的变化关系曲线，并且选择

曲线的拐点作为最优 K 值。

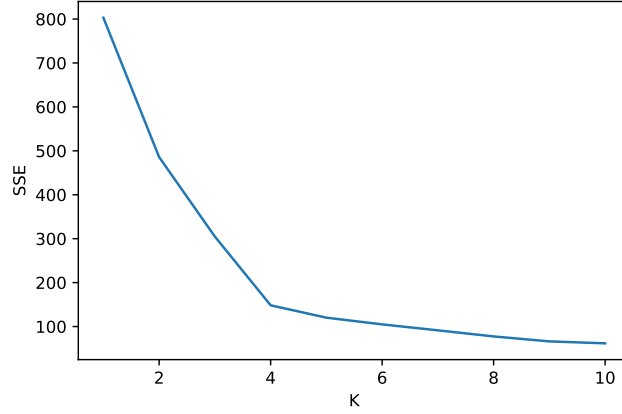


图 3 误差平方和 SSE 随 K 的变化曲线

由图 3 可见，当 K 值为 4 时曲线出现了明显的拐点，因此我们将数据分为四簇。经过 k -means 算法聚类后，我们对每个类中的所有数据点求均值和方差，类似地删除离均值超过单倍标准差的数据点。

最后，我们将每个文件中经过预处理后的所有数据转换为二维表形式，使用 `numpy` 的库函数存储数据，以备后续使用。

3.2 三维定位模型

3.2.1 建立定位方程组

对于给定的应用场景，我们假设四个锚点的坐标已知，将其分别记为 $A_1(x_1, y_1, z_1)$ 、 $A_2(x_2, y_2, z_2)$ 、 $A_3(x_3, y_3, z_3)$ 和 $A_4(x_4, y_4, z_4)$ 。接着将通过 UWB 定位技术测量得到的靶点与每个锚点之间的距离记为 d_1, d_2, d_3, d_4 。同时，记靶点的坐标为 $T(x, y, z)$ 。

根据 UWB 定位技术的原理，靶点与四个锚点在三维空间中的几何关系如图 4 所示。若排除误差的影响，靶点应当同时处于四个以锚点为球心，半径为测量距离的球面上，即四球的交点所在处。

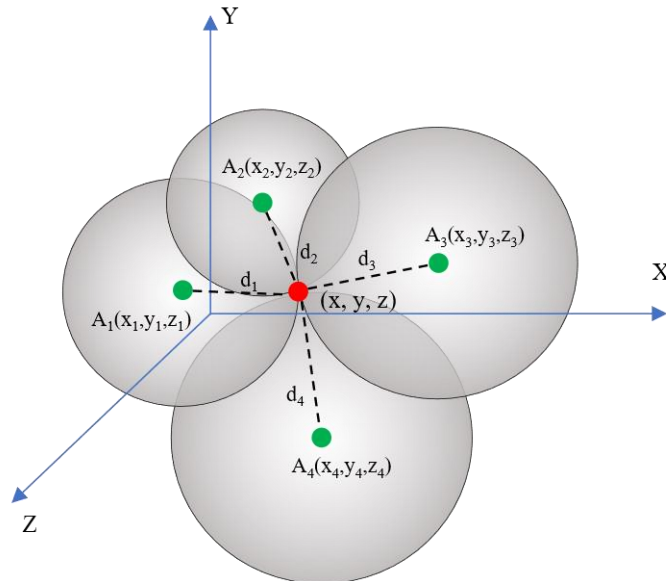


图 4 靶点求解示意图

通过距离公式，我们可以计算靶点 T 到锚点 A_i 的距离为：

$$d_i = \text{Dist}(A_i, T) \quad (1)$$

其中：

$$\text{Dist}(A_i, T) = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} \quad (2)$$

为了求解靶点的坐标，我们利用锚点的坐标和测量距离构建方程组：

$$\begin{cases} d_1^2 = (x_1 - x)^2 + (y_1 - y)^2 + (z_1 - z)^2 \\ d_2^2 = (x_2 - x)^2 + (y_2 - y)^2 + (z_2 - z)^2 \\ d_3^2 = (x_3 - x)^2 + (y_3 - y)^2 + (z_3 - z)^2 \\ d_4^2 = (x_4 - x)^2 + (y_4 - y)^2 + (z_4 - z)^2 \end{cases} \quad (3)$$

通过求解上述方程组，我们即可估算出靶点坐标。

3.2.2 最小二乘法求解方程组

在实际应用场景中，我们在测量距离时会存在信号干扰，使用 UWB 技术测得的靶点到锚点的距离与真实距离存在偏差，因此该方程组是一个没有解的非线性矛盾方程组。首先我们可以通过一些线性变换，将非线性方程组转化为线性方程组，这时利用最小二乘法即可求得线性矛盾方程组的近似解。

对非线性方程组进行线性变换得到线性方程组：

$$AX = b \quad (4)$$

其中：

$$A = \begin{bmatrix} 2(x_1 - x_4) & 2(y_1 - y_4) & 2(z_1 - z_4) \\ 2(x_2 - x_4) & 2(y_2 - y_4) & 2(z_2 - z_4) \\ 2(x_3 - x_4) & 2(y_3 - y_4) & 2(z_3 - z_4) \end{bmatrix} \quad (5)$$

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (6)$$

$$b = \begin{bmatrix} (x_1^2 + y_1^2 + z_1^2) - (x_4^2 + y_4^2 + z_4^2) - (d_1^2 - d_4^2) \\ (x_2^2 + y_2^2 + z_2^2) - (x_4^2 + y_4^2 + z_4^2) - (d_2^2 - d_4^2) \\ (x_3^2 + y_3^2 + z_3^2) - (x_4^2 + y_4^2 + z_4^2) - (d_3^2 - d_4^2) \end{bmatrix} \quad (7)$$

利用最小二乘法，可以求得：

$$X = (A^T A)^{-1} A^T b \quad (8)$$

其中X为该方法估算的靶点坐标。

我们使用已有的正常数据，对该方法估算出的靶点坐标进行评估，并将估算的靶点坐标与真实靶点坐标之间的距离进行统计，结果如图 5 所示。

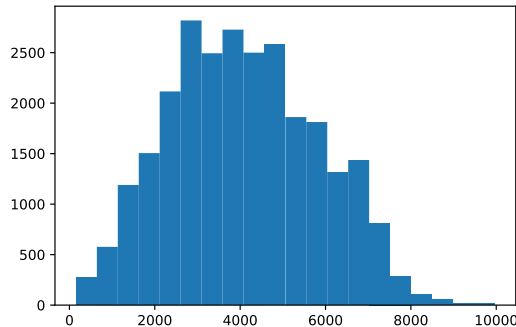


图 5 最小二乘法的误差直方图

在图 5 中，我们可以发现使用最小二乘法估算出的靶点坐标误差非常大，甚至导致靶点超出了实验场景的范围。因此，我们必须对该方法估算出的靶点坐标做进一步的优化，减小估算的误差。

3.2.3 牛顿迭代法优化近似解

为了进一步优化近似解，我们将利用最小二乘法估算的近似解作为初始解，接着使用牛顿迭代方法继续逼近矛盾方程组的最佳解。牛顿迭代法通过多次计算，利用梯度信息迭代地优化近似解，因此可以取得相比最小二乘法更佳的效果。

在牛顿迭代法中，首先需要建立用于优化的目标函数，我们选取当前估算的靶点与锚点之间的距离和测量距离的误差平方和作为优化目标函数：

$$\phi(T) = \frac{1}{2} \sum_{i=0}^4 r_i^2(T) \quad (9)$$

其中

$$r_i(T) = d_i - \text{Dist}(A_i, T), \quad i = 1, 2, 3, 4 \quad (10)$$

牛顿迭代法使用梯度信息对近似解进行优化，其最优条件定义为目标函数取极小值，也就是使梯度为零的点：

$$\nabla \phi(T) = J_r(T)^T r(T) = 0 \quad (11)$$

其中

$$J_r = \begin{bmatrix} \frac{\partial r_1}{\partial x} & \frac{\partial r_1}{\partial y} & \frac{\partial r_1}{\partial z} \\ \frac{\partial r_2}{\partial x} & \frac{\partial r_2}{\partial y} & \frac{\partial r_2}{\partial z} \\ \frac{\partial r_3}{\partial x} & \frac{\partial r_3}{\partial y} & \frac{\partial r_3}{\partial z} \\ \frac{\partial r_4}{\partial x} & \frac{\partial r_4}{\partial y} & \frac{\partial r_4}{\partial z} \end{bmatrix} \quad (12)$$

$$r(T) = \begin{bmatrix} r_1(T) \\ r_2(T) \\ r_3(T) \\ r_4(T) \end{bmatrix} \quad (13)$$

接着，该方法通过泰勒展开进行迭代计算，其优化目标为求解 T 满足该方程。

不失一般性地，我们讨论一元方程的情况。设一元方程为 $f(x) = 0$ ，首先令初始解为 x_0 ，将初始解与真实解之间的距离设为 $\Delta x = x - x_0$ 。

对函数 $f(x)$ 进行泰勒展开：

$$f(x_0 + \Delta x) = f(x_0) + \Delta x f'(x_0) + O((\Delta x)^2) \quad (14)$$

在本文中，我们使用最小二乘法得到初始解。由于该初始解与真实解的距离相对较近，因此我们可以忽略高阶项 $O((\Delta x)^2)$ 得到：

$$\Delta x \approx -\frac{f(x_0)}{f'(x_0)} \quad (15)$$

在迭代计算时，令 $x_n = x_{n-1} + \Delta x$ ，当 Δx 小于某个值时停止迭代，并将最后得到的 x_n 作为优化后的解。

在多元函数的情形下，我们可以直接利用多元函数的泰勒展开，其余推导与一元函数

情况类似：

$$\Delta x = -J_F^{-1}(x_0)F(x_0) \quad (16)$$

其中 $F(x)$ 为多元函数， $J_F(x)$ 为 F 对 x 的雅可比矩阵。

我们将牛顿法引入到问题的求解中，设 $F(T) = \nabla \phi(T) = J_r(T)^T r(T)$ ，利用牛顿法进行迭代求解，获得对于靶点坐标的一个最优估计。

牛顿法中使用的初始解对于最终的估计结果具有较大的影响，若初始解与真实解的差距较大，该方法便可能陷入局部最优解导致求解失败，所以使用最小二乘法得到近似解作为初始解可以避免陷入局部最优解，从而最大程度地逼近方程组的真实解。

类似地，我们使用已有的正常数据，使用牛顿迭代法对最小二乘法的结果进行优化，对该方法估算出的靶点坐标进行评估，并将估算的靶点坐标与真实靶点坐标之间的距离进行统计，结果如图 6 所示。

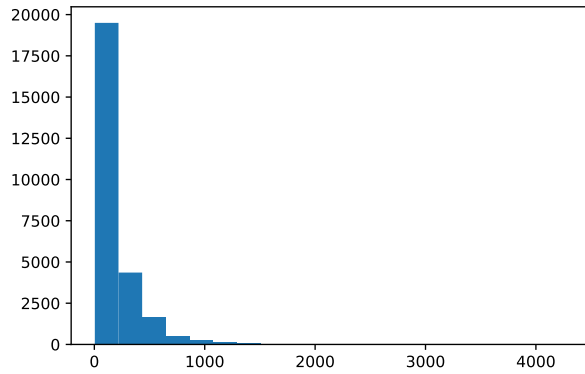


图 6 牛顿迭代法的误差直方图

对比图 5 和图 6 的优化结果，我们可以发现使用牛顿迭代法进行优化后，估测的靶点坐标与真实靶点坐标的距离显著减小，因此牛顿迭代法进一步缩小了估算的误差。

3.3 噪声模型

受到时钟异步、波形失真等多种因素的影响，我们的测量数据会不可避免地包含噪声。如果使用带噪声的测量数据计算靶点坐标，会导致估算的靶点与真实靶点坐标偏离较远。因此我们还需要进一步建立噪声模型，利用噪声的统计信息对测量数据进行去噪，从而缓解噪声导致的靶点估算误差。具体地讲，我们首先利用已有的测量数据，对测量距离进行统计，确定噪声的统计特性，由此建立对应的噪声模型。然后再对正常数据和异常数据分别建立与噪声相关的数学模型，得到对应的降噪方法，进而对数据进行降噪处理。

3.3.1 测量误差模型

在无信号干扰情况下进行测量时，所得的距离会存在一定的测量误差。我们设测量误差为 g ，真实距离为 d^* ，则测量距离 d 可表示为：

$$d = d^* + g \quad (17)$$

通过对实验数据的分析和观察，我们计算测量误差的统计特征。由于每个靶点的真实坐标以及锚点坐标均已给出，我们利用距离公式求出靶点与锚点之间的真实距离，并用测量距离与真实距离作差，结果如图 7 所示。

四个图分别代表四个锚点对应的测量距离和真实距离之间的差值，由图可知，四个锚点对应的测量距离的误差都具有类似的统计特性。由统计分析可得，测量误差大致服从均值为 -45 ，方差为 60 的高斯分布。

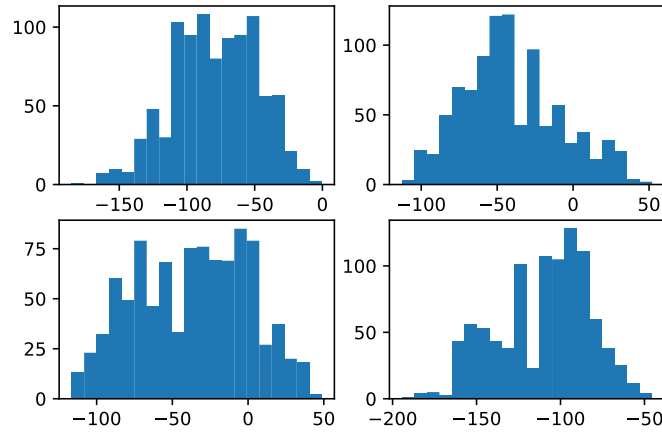


图 7 测量误差导致测量距离与真实距离的差值

3.3.2 干扰误差模型

使用 UWB 技术定位目标时，如果目标与锚点之间存在遮挡，UWB 定位信号会通过穿透和绕射的方式到达定位目标，此时就会产生非视距传播（NLOS）误差。在 NLOS 误差存在的情况下，我们得到的测量距离就会大于真实距离^[5]。

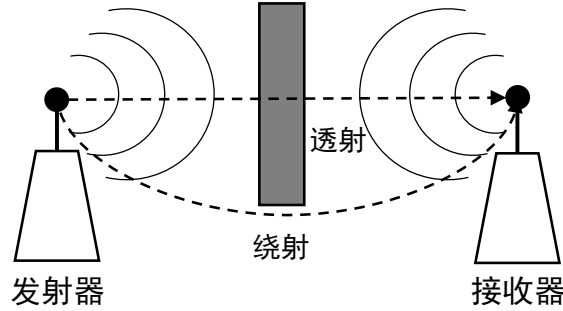


图 8 非视距传播误差的产生原理

非视距传播带来的噪声来自于环境中存在的干扰，因此该噪声可被称作干扰误差。我们设测量误差为 g ，干扰误差为 u ，真实距离为 d^* ，则测量距离 d 可表示为：

$$d = d^* + u + g \quad (18)$$

通过对实验数据的分析和观察，我们计算干扰误差的统计特征。由于每个靶点的真实坐标以及锚点坐标均已给出，我们利用距离公式求出靶点与锚点之间的真实距离，并用测量距离与真实距离作差，结果如图 9 所示：

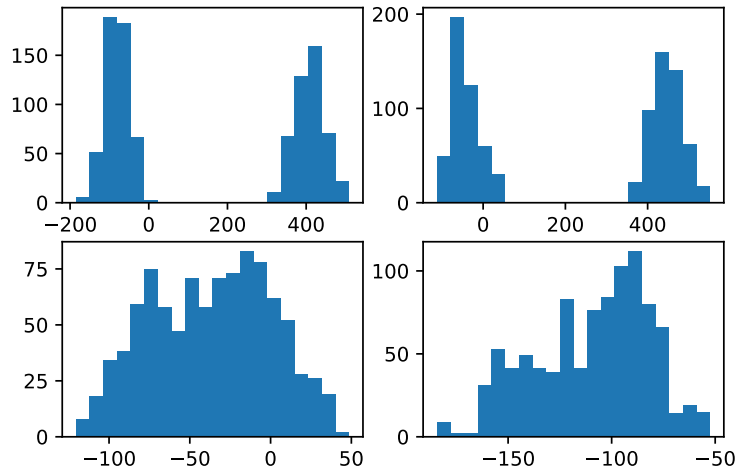


图 9 干扰误差导致测量距离与真实距离的差值

从图 9 可以看出，有两个锚点对应的干扰误差分布明显与另外两个锚点不同。并且具有相同的特征，即为混合高斯。

通过直方图不足以看出干扰信号随时间的变化趋势。我们绘制某个靶点对应的测量距离随时间变化的曲线，在无干扰信号和有信号干扰时的曲线分别如图 10 和图 11 所示。

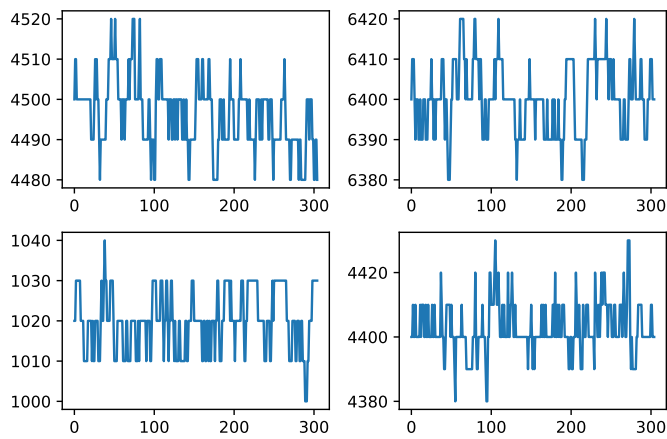


图 10 无干扰情况下测量距离随时间的变化曲线

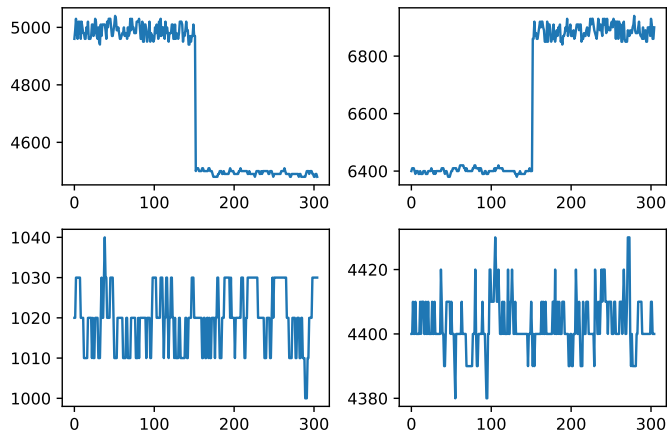


图 11 有干扰情况下测量距离随时间的变化曲线

由图 11 可以明显看出，当存在信号干扰时，在时间标志为 150 的时刻，第一、第二个锚点对应的测量距离发生了明显的跳变，而另外两个锚点对应的测量距离则以很小的幅度围绕均值上下波动。这说明在该时刻，实验场景中的干扰发生了改变，干扰误差发生了变化，即该跳变是由干扰误差的变化产生的。

为了进一步判断某个锚点对应的测量距离的误差与其他锚点的测量距离的误差是否具有一定的相关性，我们求解它们的相关系数，如图 12 所示：

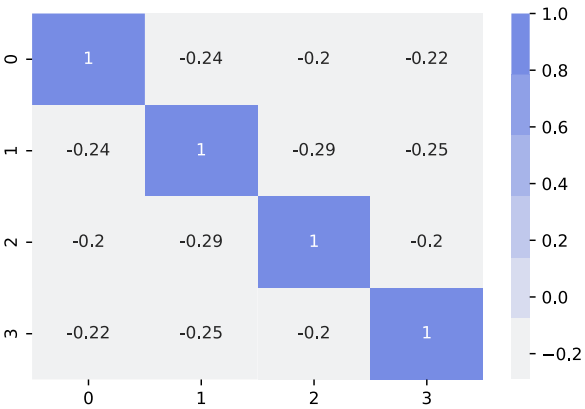


图 12 不同锚点对应测量距离的误差之间的相关系数

由图 12 可见，不同锚点对应的测量距离的误差之间的相关系数都非常小，说明不同锚点对应的测量误差的相关性很小。根据以上对测量距离的统计分析，我们可以使用高斯分布来建模干扰误差，干扰误差大致服从均值为445，方差为220的高斯分布，并且当前锚点对应的干扰误差不会对其他锚点产生影响。

3.3.3 最大似然估计

建立好噪声模型以后，我们采用最大似然估计减少测量距离的误差。

在信号无干扰条件下，对于某个锚点，将测量误差的随机变量记为 G ，真实距离为 d^* ，进行了 M 次测量，第 j 次测量的测量距离 d_j 可表示为：

$$d_j = d^* + g_j \quad (19)$$

测量误差服从均值为 μ_g ，方差为 σ_g^2 的高斯分布，即 $G \sim \mathcal{N}(\mu_g, \sigma_g^2)$ 。由于 G 的分布已知，我们可以使用最大似然估计得到对真实距离的估计，令似然函数为：

$$L(d^*|d_1, \dots, d_M) = \prod_{j=1}^M f(d_j|d^*) = \prod_{j=1}^M \frac{1}{\sqrt{2\pi}\sigma_g} \exp\left(-\frac{(d_j - d^* - \mu_g)^2}{2\sigma_g^2}\right) \quad (20)$$

经过变换得到：

$$L(d^*|d_1, \dots, d_M) = \left(\frac{1}{\sqrt{2\pi}\sigma_g}\right)^M \exp\left(-\frac{1}{2\sigma_g^2} \sum_{j=1}^M (d_j - d^* - \mu_g)^2\right) \quad (21)$$

对 d^* 进行最大似然估计：

$$\widehat{d^*} = \arg \max_{d^*} \prod_{j=1}^M f(d_j|d^*) \quad (22)$$

通过导数可以求得 L 的极大值点：

$$\frac{\partial L}{\partial d^*} = 0 \quad (23)$$

再得到对于 d^* 的最大似然估计为：

$$\widehat{d^*} = \frac{1}{M} \sum_{j=1}^M d_j - \mu_g \quad (24)$$

在信号受到遮挡的情况下，对于某个锚点，将干扰误差的随机变量记为 U ，干扰误差服从均值为 μ_u ，方差为 σ_u^2 的高斯分布，即 $U \sim \mathcal{N}(\mu_u, \sigma_u^2)$ 。当测量距离存在干扰误差时，第 j 次测量的测量距离 d_j 可表示为：

$$d_j = d^* + u_j + g_j \quad (25)$$

令总误差 $E = U + G$ ，由于干扰误差 U 和测量误差 G 均为高斯分布且相互独立，因此总误差 E 也服从高斯分布，满足 $E \sim \mathcal{N}(\mu_u + \mu_g, \sigma_u^2 + \sigma_g^2)$ 。

类似地，我们得到信号有干扰条件下的 d^* 的最大似然估计为：

$$\widehat{d^*} = \frac{1}{M} \sum_{j=1}^M d_j - \mu_u - \mu_g \quad (26)$$

3.3.4 降噪方法

基于之前建立的噪声模型，我们在对测量距离进行降噪时，可以对每个锚点进行单独讨论。首先无论信号是否受到干扰，测量误差在每个锚点的测距过程中都是存在的，同时若锚点发送的信号受到干扰，则测量距离产生干扰误差。

在我们设计的降噪模型中，为了能够对干扰误差进行精准去除，需要先找出受到了干扰的锚点。假设锚点发送的信号受到干扰视为在锚点前随机放置了挡板，并且该挡板不会对其他锚点发送的信号产生影响。

用 i 表示锚点序号，将锚点 i 前有挡板记为 $p = i$ ，而 $p = 0$ 表示所有锚点前都没有挡板的情况，对其测得的数据进行降噪的方法如下所示：

$$\text{Denoise}(d_i, p) = \begin{cases} d_i - \mu_u - \mu_g, & i = p \\ d_i - \mu_g, & i \neq p \end{cases} \quad (27)$$

3.4 神经网络噪声预测

我们设计了一个神经网络来帮助我们判断当前哪个锚点发送的信号是否受到了干扰，该神经网络可以利用四个测量距离找出被放置了挡板的锚点。利用现有的数据能够训练出具有找出挡板锚点能力的神经网络，即神经网络分类器。

3.4.1 生成训练数据特征

算法 1 特征生成算法

```

for  $p \leftarrow 0$  to 4
  for  $i \leftarrow 1$  to 4
     $\hat{d}_i = \text{Denoise}(d_i, p)$ 
  end

   $q_0 = \text{Estimate}(\hat{d}_1, \hat{d}_2, \hat{d}_3, \hat{d}_4)$ 

  for  $i \leftarrow 1$  to 4
     $q_i = \text{Estimate}(\{\hat{d}_j | j \neq i\})$ 
  end

   $f_p = [\phi(q_0), \dots, \phi(q_4), \text{Dist}(q_0, q_1), \dots, \text{Dist}(q_0, q_4), \mathbb{D}(q_1, q_2, q_3, q_4)]$ 
end

 $f = [f_0, f_1, f_2, f_3, f_4]$ 

```

我们首先循环遍历 $p = 0, 1, 2, 3, 4$ ，对于不同的 p 生成相应特征。举例来说，假设 $p = 2$ ，即第二个锚点 A_2 前有挡板，我们首先对测量距离进行降噪处理：

$$\hat{d}_i = \text{Denoise}(d_i, p) \quad (28)$$

使用四个测量距离估算靶点 q_0 ：

$$q_0 = \text{Estimate}(\hat{d}_1, \hat{d}_2, \hat{d}_3, \hat{d}_4) \quad (29)$$

接着从中选择三个测量距离估算靶点，共有 $C_4^3 = 4$ 中选法，得到 q_1, q_2, q_3, q_4 ：

$$q_1 = \text{Estimate}(\hat{d}_2, \hat{d}_3, \hat{d}_4) \quad (30)$$

$$q_2 = \text{Estimate}(\hat{d}_1, \hat{d}_3, \hat{d}_4) \quad (31)$$

$$q_3 = \text{Estimate}(\hat{d}_1, \hat{d}_2, \hat{d}_4) \quad (32)$$

$$q_4 = \text{Estimate}(\hat{d}_1, \hat{d}_2, \hat{d}_3) \quad (33)$$

然后计算 $p = 2$ 对应的特征向量 $f_2 \in \mathbb{R}^{12}$:

$$f_2 = [\phi(q_0), \dots, \phi(q_4), \text{Dist}(q_0, q_1), \dots, \text{Dist}(q_0, q_4), \mathbb{D}(q_1, q_2, q_3, q_4)] \quad (34)$$

其中 $\mathbb{D}(q_1, q_2, q_3, q_4) \in \mathbb{R}^3$ 是指对 q_1, q_2, q_3, q_4 四个样本分别计算沿三个坐标轴上的方差, $\text{Dist}(q_i, q_j)$ 为 q_i 和 q_j 两点的距离, 而 $\phi(q)$ 为在定位方程组求解过程中使用的优化目标函数。

其余情况同理, 最终得到当前数据的特征 $f \in \mathbb{R}^{60}$:

$$f = [f_0, f_1, f_2, f_3, f_4] \quad (35)$$

我们将训练数据特征生成算法展现在算法 1 中。

3.4.2 生成训练数据标签

对于异常数据, 我们使用每个锚点对应的测量距离减去真实距离得到误差, 假设误差最大的数据对应的是被挡板所阻挡的锚点所测出的数据, 其中

$$d_i^* = \text{Dist}(A_i, T) \quad (36)$$

$$p = \arg \max_i (d_i - d_i^*) \quad (37)$$

对于正常数据, 令 $p = 0$, 代表没有挡板存在于任何锚点前。

3.4.3 训练神经网络

我们采用 z-score 方法对输入特征进行归一化, 首先计算特征的均值 μ_f 和标准差 σ_f , 然后施加 z-score 方法:

$$\tilde{f} = \frac{f - \mu_f}{\sigma_f} \quad (38)$$

我们使用全连接网络FFN(\cdot)对特征进行分类得到预测标签 y 。

$$y = \text{FFN}(\tilde{f}) \quad (39)$$

具体地讲, 该网络由 3 个全连接层构成, 其计算公式为:

$$h_1 = \text{ReLU}(W_1 f + b_1) \quad (40)$$

$$h_2 = \text{ReLU}(W_2 f + b_2) \quad (41)$$

$$y = \text{ReLU}(W_3 f + b_3) \quad (42)$$

该神经网络的结构示意图如图 13 所示。

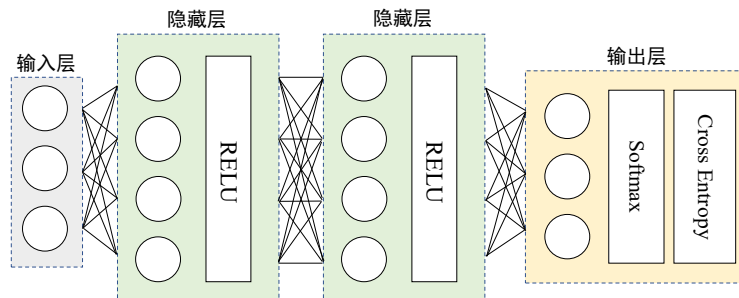


图 13 神经网络示意图

我们使用交叉熵损失函数和 Adam 优化器对参数进行优化，其中损失函数定义为：

$$\mathcal{L} = - \sum_{k=1}^N \sum_{j=1}^5 p_k^j \log(y_k^j) + \beta \sum_{\theta} \|\theta\|_2^2 \quad (43)$$

其中 β 控制 L_2 正则权重。

我们将已有的数据按照10:1的比例分为训练集和验证集，使用训练集优化神经网络参数，并在验证集上进行测试和调参，选择验证集上准确率最高的模型作为最优模型，保存记录。

得到训练好的网络后，我们即可使用神经网络计算预测标签：

$$\hat{p}_k = \arg \max_j y_k^j \quad (44)$$

3.5 靶点位置计算

通过四个锚点对应的测量距离求解靶点位置的计算过程如图 14 所示。首先将四个测量距离输入到神经网络噪声预测模型中来判断锚点发送的信号有无受到干扰，如果有干扰则去除测量噪声和干扰噪声，否则只需去除测量噪声。利用去噪后的四个测量距离通过最小二乘法和牛顿迭代法估算出靶点三维坐标。

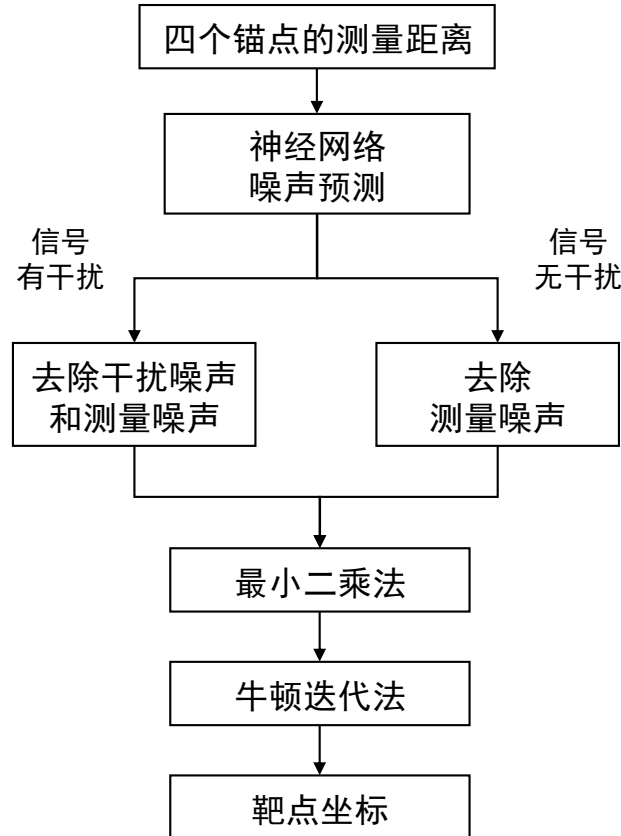


图 14 靶点位置计算流程图

3.6 运动定位

UWB 在现实场景中应用时通常需要对移动的靶点进行定位，而其应用场景一般为小范围内的定位，靶点的运动速度和运动的范围都非常有限，可以将每个时刻的靶点近似为静止状态，使用静态定位模型求解。

我们根据题目所给的运动靶点的测量数据对每一个时刻的靶点进行定位，得到在 x, y, z

轴上的坐标随时间的变化曲线，由此得到了靶点在不同时刻的位置。我们可以发现靶点在空间位置基本围绕着某一曲线，只有在少数几个时刻上有比较大的偏离，这些时刻点可以视为测量出现问题的时刻点。

结合靶点运动规律，我们使用最小二乘法多项式拟合得到靶点在三维空间内的运动轨迹。

4 模型求解

4.1 问题一

为了对数据进行预处理，我们首先通过检测时间戳删除掉缺失数据，然后用高斯分布的特性删除掉离均值点过远的离群数据。经过这两步数据处理，保留的数据个数如表 3 所示。

表 3 数据清洗后保留的数据个数及保留率

	正常数据个数	异常数据个数	正常数据保留率	异常数据保留率
未处理	77122	79176	100.00%	100.00%
删除掉缺失数据	61062	62886	79.18%	79.43%
删除掉异常数据	13987	12570	18.14%	15.88%

经过数据预处理后，正常数据文件夹中“24.正常.txt”保留的数据如附录中表 8 所示，“109.正常.txt”保留的数据如附录中表 9 所示。异常数据文件夹中“1.异常.txt”保留的数据如附录中表 10 所示，“100.异常.txt”保留的数据如附录中表 11 所示。

4.2 问题二

我们首先对实验场景一进行讨论，构建大小为 $5000mm * 5000mm * 3000mm$ 的三维空间，划定靶点的范围，然后标注了四个锚点的位置（单位：mm）分别为 $A_0(0,0,1300)$ 、 $A_1(5000,0,1700)$ 、 $A_2(0,5000,1700)$ 、 $A_3(5000,5000,1300)$ 。为了检验模型的有效性，我们计算模型的精度如表 4 所示。除此之外，在实验场景一中我们将附件二中提供的前 5 组（信号无干扰）数据依次输入到信号无干扰定位模型中得出靶点的三维坐标，后 5 组（信号有干扰）数据依次输入到信号有干扰定位模型中同样得出靶点的三维坐标，结果如表 5 所示。

表 4 精度数据

精度 (单位 cm)	3 维 (x,y,z)精度	2 维 (x,y)精度	1 维 (x)精度	1 维 (y)精度	1 维 (z)精度
正常数据	14.96	6.71	4.59	4.01	12.24
异常数据	29.84	11.34	7.62	7.33	25.74
全部数据	20.20	8.90	6.02	5.58	18.63

表 5 场景一预测数据对应的靶点坐标

测试数据类型	测试数据编号	靶点的三维坐标(cm)
信号无干扰	数据 1	(114.80, 64.65, 90.74)
	数据 2	(318.07, 170.80, 89.23)
	数据 3	(273.47, 115.22, 97.52)
	数据 4	(245.31, 99.01, 216.33)
	数据 5	(146.70, 254.70, 200.33)
信号有干扰	数据 6	(221.85, 57.15, 146.41)

	数据 7	(411.06, 180.75, 205.74)
	数据 8	(166.56, 115.32, 144.89)
	数据 9	(337.97, 183.28, 98.65)
	数据 10	(466.49, 219.92, 165.18)

4.3 问题三

为了检验模型能够在不同场景下应用，我们再对实验场景二进行讨论，构建一个 $5000mm * 3000mm * 3000mm$ 的三维空间，将四个锚点位置（单位：mm）分别设置为 $A_0(0,0,1200)$ 、 $A_1(5000,0,1600)$ 、 $A_2(0,3000,1600)$ 、 $A_3(5000,3000,1200)$ 。在实验场景二下，我们将附件 3 中提供的前 5 组（信号无干扰）数据依次输入到信号无干扰定位模型中得出靶点的三维坐标，后 5 组（信号有干扰）数据同样依次输入到信号有干扰定位模型中得出靶点的三维坐标，结果如表 6 所示。

表 6 场景二预测数据对应的靶点坐标

测试数据类型	测试数据编号	靶点的三维坐标(cm)
信号无干扰	数据 1	(367.58, 223.16, 120.37)
	数据 2	(422.07, 171.83, 98.88)
	数据 3	(318.14, 173.92, 116.27)
	数据 4	(258.11, 191.53, 136.41)
	数据 5	(58.22, 6.72, 121.63)
信号有干扰	数据 6	(456.88, 227.11, 142.44)
	数据 7	(234.88, 176.16, 92.86)
	数据 8	(171.10, 141.02, 214.30)
	数据 9	(190.43, 103.20, 155.34)
	数据 10	(107.85, 106.67, 187.36)

4.4 问题四

为了判断 UWB 采集数据时候信号有无干扰，我们建立了分类模型来进行区分。将附件 4 中提供的采集于实验场景一的 10 组数据输入到分类模型中，判断结果如表 7 所示。

表 7 对未知有无干扰的数据进行分类

测试数据编号	信号是否受到干扰
数据 1	否
数据 2	是
数据 3	否
数据 4	是
数据 5	是
数据 6	否
数据 7	否
数据 8	是
数据 9	是
数据 10	否

4.5 问题五

为了实现对动态靶点的精确定位，我们利用之前建立的静态点数据模型加上靶点自身

运动规律，得到动态靶点的运动轨迹。首先对附件 5 中动态靶点采集的数据输入到分类模型中进行分类，然后将信号有干扰的数据和信号无干扰的数据分别输入到对应的定位模型，得出动态靶点的三维坐标，并绘制坐标在 x, y, z 轴维度上的变化散点图，然后使用最小二乘法多项式拟合这些散点得到曲线图如图 15 所示。

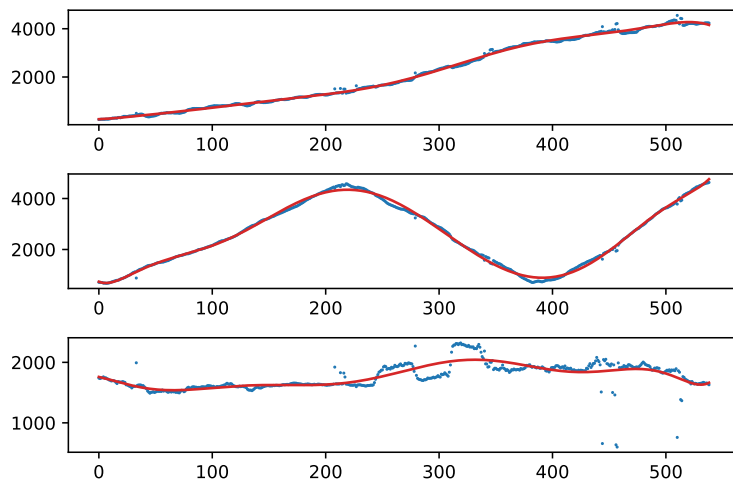


图 15 动态靶点坐标在 x, y, z 轴上的变化曲线拟合结果

最后在三维坐标系内，依据时间顺序依次将所有的三维坐标相连得到运动轨迹图如图 16 所示。

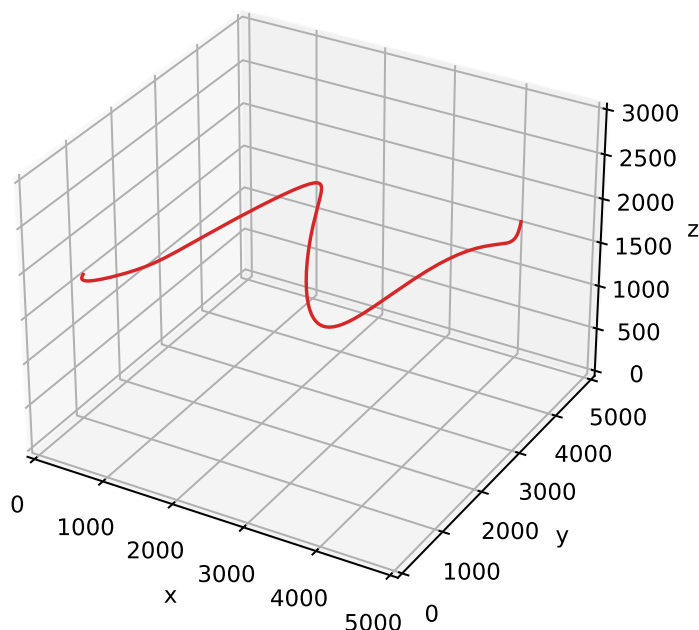


图 16 动态靶点的三维运动轨迹

结论

本文基于题目给出的测量数据，采用了混合高斯模型对噪声进行建模，并且利用最大似然估计理论设计了降噪模型，建立了能够进行噪声预测的神经网络，对经过降噪后的测量数据使用最小二乘法得到靶点坐标的估测值，再利用牛顿迭代法对该估测值进行优化，得到对于靶点坐标的精确估测。本文所建立的神经网络能够对给定的测量数据进行分类，

分辨数据是否为在信号干扰条件下得到的。

针对问题一，设计了数据预处理方法，先删减掉不含全部锚点测距值的缺失数据，再通过 k-means 聚类算法和高斯 σ 原则删除掉离中心点较远的离群数据。处理后正常数据和异常数据分别保留 18.24%，15.88%。

针对问题二，我们分别对正常数据和异常数据建立了定位模型。首先对噪声进行统计分析，搭建了神经网络进行噪声预测，建立了降噪模型，使测量距离降噪后更加接近于真实距离，对降噪后的测量距离使用最小二乘法和高斯迭代法便可以求出靶点位置。正常数据定位模型的三维定位精度为 14.96 厘米，异常数据定位模型的三维定位精度为 29.84 厘米，定位精度较高。

针对问题三，我们建立的定位模型是包含锚点坐标信息的，锚点坐标的改变并不会影响降噪模型对测量数据的降噪效果，依然能够正确估计出靶点的空间位置，说明我们建立的定位模型能够应用于不同场景。

针对问题四，我们基于数据特征建立了基于神经网络的分类模型，判断是否存在干扰的准确率为 90%，说明我们的分类准确有效。

针对问题五，结合静态点定位模型和靶点运动规律设计了运动靶点的定位方法，利用最小二乘多项式成功拟合了运动靶点在空间中的运动轨迹。

本文也存在一些不足，仍有改进的空间。对于干扰误差只粗略地使用一种分布进行估计，不够全面。

参考文献

- [1] 邓中亮,余彦培,袁协,等. 室内定位现状与发展趋势研究[J]. 中国通信,2013,10(3):42-55.
- [2] 肖竹,王勇超,田斌,等. 超宽带定位研究与应用:回顾和展望[J]. 电子学报, 39(1):133-141. 2011.
- [3] Safaai-Jazi A, Riad S M, Muqaibel A, et al. Ultra-wideband propagation measurements and channel modeling[J]. Report on Through-the-Wall Propagation and Material Characterization, 2002.
- [4] 王千,王成,冯振元,等. K-means 聚类算法研究综述[J]. 电子设计工程, 20(7):21-24. 2012.
- [5] 姚健,刘送永,崔玉明,顾聪聪. 基于 KF/PF 测距滤波的 UWB 精确定位试验研究[J]. 煤炭工程. 53(09):122-126. 2021.

附录

表 8 “24.正常.txt”数据内容

3280 4660 2600 3910
3280 4660 2600 3900
3280 4660 2600 3920
3280 4660 2600 3920
3280 4660 2600 3900
3280 4660 2600 3900
3280 4660 2600 3910
3280 4660 2600 3890
3280 4660 2600 3880

3280 4660 2600 3910

表 9 “109.正常.txt” 数据内容

4910 5320 2020 2860
4890 5330 2030 2880
4890 5330 2030 2880
4910 5320 2030 2880
4910 5320 2040 2880
4890 5320 2020 2880
4890 5320 2020 2880
4910 5320 2030 2880
4890 5320 2020 2880
4890 5320 2020 2880
4890 5330 2020 2880
4910 5330 2030 2880
4910 5330 2020 2880
4890 5330 2020 2880
4890 5330 2020 2880
4890 5330 2020 2880
4890 5330 2020 2880
4890 5320 2040 2890
4910 5320 2030 2880
4890 5320 2020 2880
4890 5330 2040 2880
4910 5330 2040 2880
4910 5330 2040 2880
4910 5320 2040 2880
4890 5320 2040 2880
4890 5320 2030 2880
4890 5320 2030 2870
4890 5320 2020 2880
4890 5320 2020 2880
4890 5320 2020 2880
4910 5330 2020 2890
4890 5330 2020 2870
4890 5330 2020 2870
4890 5330 2040 2870
4910 5330 2040 2880
4910 5330 2040 2880
4910 5330 2040 2880
4890 5320 2040 2880
4890 5330 2040 2880
4890 5330 2040 2880
4890 5330 2040 2880
4910 5320 2040 2880

4910 5320 2040 2880
4910 5320 2030 2870
4910 5330 2040 2870
4890 5330 2040 2870
4910 5330 2040 2880
4910 5330 2040 2900
4910 5320 2040 2880
4910 5330 2040 2880
4890 5320 2040 2880
4910 5330 2040 2880
4910 5330 2030 2890
4910 5330 2040 2890
4910 5330 2040 2890
4910 5330 2040 2890
4890 5330 2040 2890
4890 5330 2040 2890
4890 5330 2040 2890
4910 5320 2040 2880
4910 5320 2040 2880
4910 5320 2030 2890
4910 5320 2030 2900
4910 5320 2040 2890
4910 5320 2030 2890
4910 5320 2030 2900
4910 5330 2030 2900
4910 5320 2030 2900
4890 5330 2020 2890
4910 5330 2030 2880
4910 5330 2020 2880
4890 5330 2020 2880
4890 5330 2030 2880
4890 5330 2040 2880
4890 5330 2040 2880
4910 5330 2030 2880
4910 5330 2040 2880

表 10 “1.异常.txt” 数据内容

1280 4550 4550 6300
1270 4550 4560 6300
1280 4550 4560 6300
1270 4550 4560 6300
1270 4550 4550 6300
1280 4550 4560 6300
1270 4550 4550 6300
1270 4550 4550 6300

1270 4550 4550 6300
1270 4550 4550 6300
1270 4550 4560 6300
1280 4550 4560 6300
1280 4550 4550 6300
1270 4550 4550 6300
1270 4550 4550 6300
770 5030 4550 6300
770 5030 4550 6300
760 5010 4550 6300
760 5030 4550 6300
770 5030 4550 6300
770 5020 4550 6300
760 5020 4550 6300
760 5020 4550 6300
770 5010 4550 6300
760 5030 4550 6300
760 5030 4550 6300
760 5030 4550 6300
760 5010 4550 6300
760 5010 4550 6300
770 5020 4550 6300
760 5020 4550 6300
770 5010 4550 6300
760 5030 4550 6300
770 5010 4550 6300
770 5030 4550 6300
770 5010 4550 6300
770 5030 4550 6300
760 5010 4550 6300
770 5010 4550 6300
1240 4550 4550 6300
1230 4550 4550 6300
1230 4550 4550 6300
1230 4550 4550 6300
1240 4550 4560 6300
1230 4550 4550 6300
1240 4550 4550 6300
1240 4550 4550 6300
1220 4550 4550 6300
1220 4550 4550 6300
1230 4550 4560 6300
1240 4550 4550 6300
1230 4550 4550 6300

1220 4550 4550 6300
1240 4550 4550 6300
1240 4550 4550 6300
1240 4550 4550 6300
1230 4550 4550 6300
1240 4550 4560 6300
1240 4550 4560 6300
1230 4550 4550 6300
1240 4550 4550 6300
1240 4550 4550 6300
1220 4550 4550 6300
1230 4550 4550 6300
1230 4550 4550 6300
1220 4550 4550 6300
1220 4550 4560 6300
1240 4550 4560 6300
1230 4550 4560 6300
1220 4550 4560 6300
1240 4550 4550 6300
1240 4550 4550 6300
760 5070 4550 6300
770 5070 4550 6300
770 5070 4550 6300
760 5070 4550 6300
770 5070 4550 6300

表 11 “100.异常.txt” 数据内容

1510 3730 4680 5710
1510 3730 4680 5700
1510 3760 4680 5710
1510 3790 4680 5700
1510 3770 4680 5710
1510 3800 4680 5710
1510 3730 4680 5710
1510 3740 4680 5700
1510 3750 4680 5700
1510 3750 4680 5710
1510 3740 4680 5710
1510 3740 4680 5710
1510 3730 4680 5700
1510 3730 4680 5700
1510 3730 4680 5700
1510 3760 4680 5700
1510 3750 4680 5710
1510 3740 4680 5710

1510 3560 4930 5700
1510 3560 4910 5700
1520 3560 4890 5700
1520 3560 4870 5700
1510 3560 4930 5700
1520 3560 4870 5700
1520 3560 4860 5700
1510 3560 4880 5700
1510 3560 4890 5700
1510 3560 4890 5700
1510 3560 4890 5700
1520 3560 4880 5700
1510 3560 4880 5700
1510 3560 4870 5700
1510 3560 4930 5700
1500 3570 5880 5690
1510 3560 5660 5700
1510 3560 5190 5700
1510 3560 5110 5700