



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    中国运载火箭技术研究院 哈尔滨工程大学

---

参赛队号    21832010024

---

1.徐雅涵

---

队员姓名    2.黄鑫宇

---

3.沈财旭

---

---

# 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      基于机器学习的 UWB 定位与分类技术研究

---

#### 摘                  要：

UWB 定位技术是一类室内外精确跟踪能力强，定位精度高的导航技术。近年来在各领域得到了快速发展和广泛应用。然而基于 UWB 的三维高精度定位算法仍然是一项亟待解决的技术问题。本文将基于这样的背景，采用提供的试验数据研究一类高精度的分类与定位算法，并应用提供的数据完成了方法检验和技术验证。

针对问题 1 中的数据预处理问题，本文统计了样本数据的分布特性得出了大部分样本数据符合正态分布的结论，对于样本中测量值异常的数据，采用  $3\sigma$  原则对于偏离分布中心 3 倍标准差的数据进行删除；考虑到相同相似样本会加大定位算法设计与计算负担，采用欧氏距离判定同一样本中相同相似数据并进行删除，进一步给出了面向定位算法设计的简化数据。同时应用统计分析方法检验了预处理前后数据分布特性，确定了测量数据存在有效偏差和白噪声干扰，统计了 3 附件一样本对应的 324 项标准差数据，得到了标准差符合正态分布且均值为 10，标准差为 3 的结论，确定了各组传感器性能一致且明确了测量噪声的分布规律。

针对问题二与问题三，首先分别应用传统的最小二乘法、基于 PSO 的定位算法建立了接近传统意义上的四点算法，精度分析表明，两种算法的求解精度基本一致，基本符合 x, y 轴求解精度较高而 z 轴求解精度较低的特性。为了进一步建立高精度定位算法，采用机器学习方法建立了一类高精度定位算法，以测量长度为输入、xyz 坐标数据为输出设计了三层 BP 神经网络，通过大量训练完成了定位算法设计，精度分析表明，基于机器学习的定位算法仍然表现出 x, y 轴求解精度较高而 z 轴求解精度较低的特性，但是 x、y 轴求解精度达到厘米级，而 z 轴求解精度达到了分米级，极大地提高了三点定位算法的精度特性。

针对问题四，为了进一步解决正常数据、异常数据的分类问题，首先基于传统方法出发拟寻找一类分类指标完成对正常、异常数据的二分类。然而基于统计分析建立的测距偏差分类方法和基于平面几何出发的三圆相交面积判别皆未能充分区分正常、异常数据，因此本文认为基于有限的先验信息和较大的噪声水平干扰下，无法通过有效的传统方法完成对正常数据、异常数据的识别，因此仍采用机器学习方法建立了一类有效的分类方法，基于附件 1 的测试表明识别精度达到 97.69%，并完了对附件四数据的计算。

针对问题五，首先从附件 5 中提取得到了运动过程中靶点到 4 个锚点的长度信息，再使用问题四中的分类算法验证每组数据是否存在异常干扰，对于修正后的数据采用采用前

---

面问题所建立的机器学习的方法求解坐标，最后根据坐标即可得到在空间中的运动轨迹。

本文综合了针对 UWB 技术三维坐标精确定位技术出发展开了相关研究，分析了传统方法的不足并应用机器学习方法形成了一套准确性高、适用性强的分类和定位方法，有效的解决了题目中的相关问题。

**关键词：**粒子群算法；机器学习；最小二乘法

---

## 目录

基于机器学习的 UWB 定位与分类技术研究 .....	5
1 问题重述 .....	5
1.1 研究背景 .....	5
1.2 问题描述 .....	5
1.2.1 数据预处理 .....	5
1.2.2 定位模型 .....	5
1.2.3 不同场景应用 .....	5
1.2.4 分类模型 .....	5
1.2.5 运动轨迹定位 .....	5
2 问题假设 .....	7
3 数据来源 .....	7
4 数据预处理 .....	8
4.1 附件一数据统计分析 .....	8
4.2 剔除异常数据 .....	9
4.3 对于残缺数据定义的讨论 .....	12
4.4 对于相同、相似数据的处理 .....	13
4.5 关于预处理提供数据的说明 .....	13
4.6 对于测量数据的修正策略 .....	14
5 问题二建模与分析：定位模型 .....	16
5.1 问题二分析与思路 .....	16
5.2 模型建立与求解 .....	16
5.2.1 最小二乘法 .....	16
5.2.2 粒子群算法 .....	19
5.2.3 神经网络算法 .....	24
6 问题三建模与分析：不同场景应用 .....	30
6.1 问题一分析与思路 .....	30
6.2 正常数据解算 .....	30
6.2.1 最小二乘法求解结果 .....	30
6.2.2 PSO 算法求解结果 .....	30
6.2.3 神经网络求解结果 .....	30
6.3 异常数据解算 .....	31
6.3.1 神经网络求解结果 .....	31
6.3.2 PSO 算法求解结果 .....	31
7 问题四建模与分析：分类模型 .....	32
7.1 基于测量距离的两类尝试 .....	32
7.2 基于公共区域面积大小的判别方法 .....	33
7.3 神经网络分类法 .....	34
7.4 问题四的求解 .....	34

---

8 问题五建模与分析：动态轨迹 .....	35
8.1 分析与思路 .....	35
8.2 结果 .....	35
总结与展望 .....	39
9.1 模型评价 .....	39
9.2 模型改进 .....	39

---

# 基于机器学习的 UWB 定位与分类技术研究

## 1 问题重述

### 1.1 研究背景

UWB 定位技术是一类室内外精确跟踪能力强，定位精度高的导航技术。其独特的技术优势—厘米级甚至毫米级定位精度，使得 UWB 在电力、军事、监狱、生产、医疗、特种行业等领域有着重要应用。本文着眼于一类基于飞行时间（Time of Flight, TOF）测距原理的 UWB 技术，研究一类在复杂干扰环境下仍具备高精度定位的算法，致力于推动 UWB 技术向前发展，并在一定程度上为亟待解决的技术问题提供方法参考和思路借鉴。

### 1.2 问题描述

问题通过实际场景实测提供了若干组测量数据，用于建立定位模型和分类模型，并在信号有干扰或无干扰情况下形成一类精确定位算法。

#### 1.2.1 数据预处理

题目中提供的数据源于实际测量，传感器性能的不同、实际试验引入的系统误差或人为误差、TOF 自身测量技术性能缺陷会对实际数据引入不同类型的误差。同时为了保证数据量的充足，试验时在同一状态点进行了多次测量试验，因此数据中除了可能存在异常数据，还有可能存在缺失数据、相同数据、相似数据，为了对算法进行充足的分析，有必要通过合适的手段对数据完成预处理并按照要求形成相应的数据文件。

#### 1.2.2 定位模型

基于预处理获得的相关数据，建立精确定位算法实现对靶点坐标的精确计算，并应用附件 2 中提供的数据完成算法的检验，同时完成对精确定位算法的精度分析（分别完成三维坐标精度、二维坐标精度、一维坐标精度分析）。

#### 1.2.3 不同场景应用

为了进一步检验算法对不同的试验场景和实验环境的适应能力，通过附件 3 提供的数据完成精确定位，检验不同实验场景下算法的工作能力。

#### 1.2.4 分类模型

实际测量中，难以确定信号的受干扰状态。而问题一至问题三的背景均基于已知信号是否处于异常状态。需要基于附件一提供的数据完成一类分类算法设计，区分异常数据和正常数据，并通过附件四提供的数据完成分类算法的有效性验证。

#### 1.2.5 运动轨迹定位

基于静态点的定位模型和分类算法对于静态点的精确定位和信息识别仅仅只是 UWB 技术应用的一个侧面，对于动态轨迹定位技术的实际需求，还应该结合靶点自身运动规律形成目标估计算法并结合精确定位算法和分类算法，形成一类在随机信号干扰下动态目标轨迹跟踪算法。

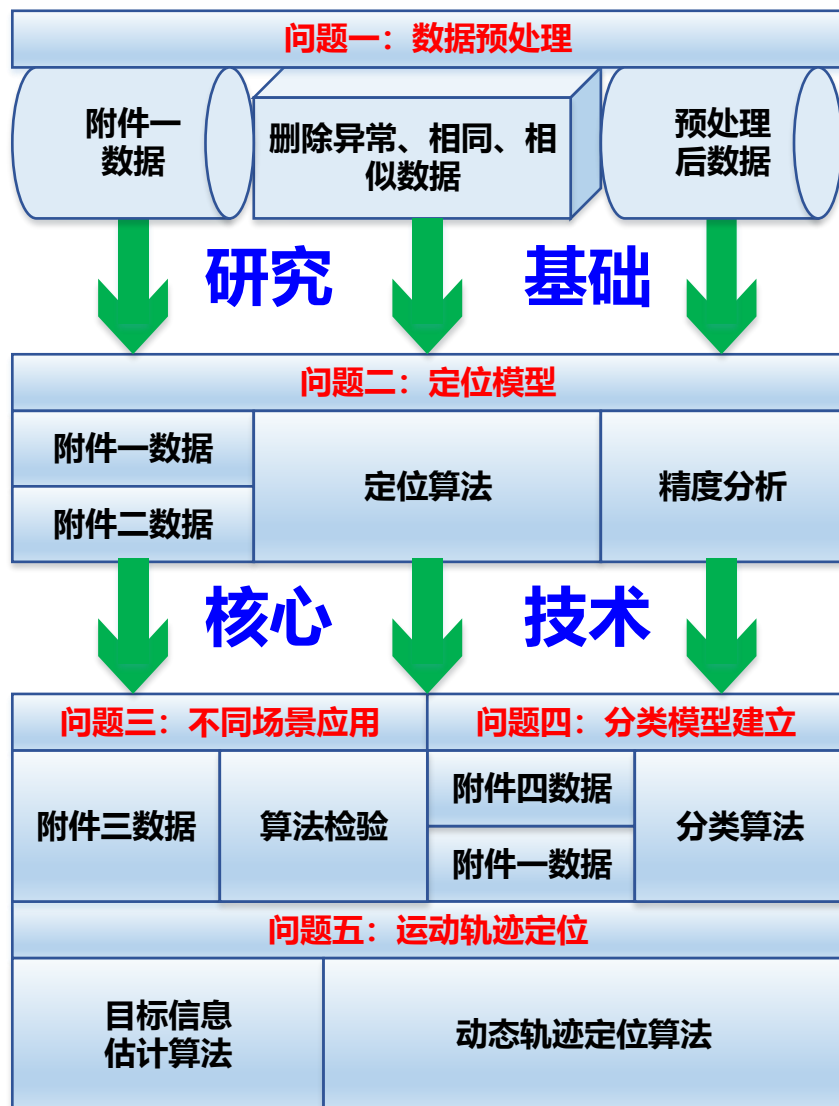


图 1.1 各问题要求子框架

---

## 2 问题假设

假设测量误差仅仅因 TOF 技术不足和传感器性能有限而引起, 误差引起的形式为一类有色偏差和一类白噪声误差的组合。

基于数据统计分析可以说明, UWO 给出的测距偏差与实际测量距离存在线性相关关系, 随着实际测量距离的增加, UWO 测量结果会被引入一类偏置误差, 同时由于传感器性能不同和实验引入的误差, 测量结果还会引入一类白噪声误差, 假定白噪声幅值与传感器性能有关, 且假设传感器的性能并不随实验次数和实验环境变化。

## 3 数据来源

本文所应用数据皆源于题目提供。



## 4 数据预处理

### 4.1 附件一数据统计分析

通过数据处理得到了附件 1 中的全数据，考虑题中规定附件 1 中数据来源于 324 个不同的位置测量获得，首先通过标准差分析检验数据的分布特点。下文为了叙述和绘图方便，约定 ZC 表示正常，YC 表示异常。

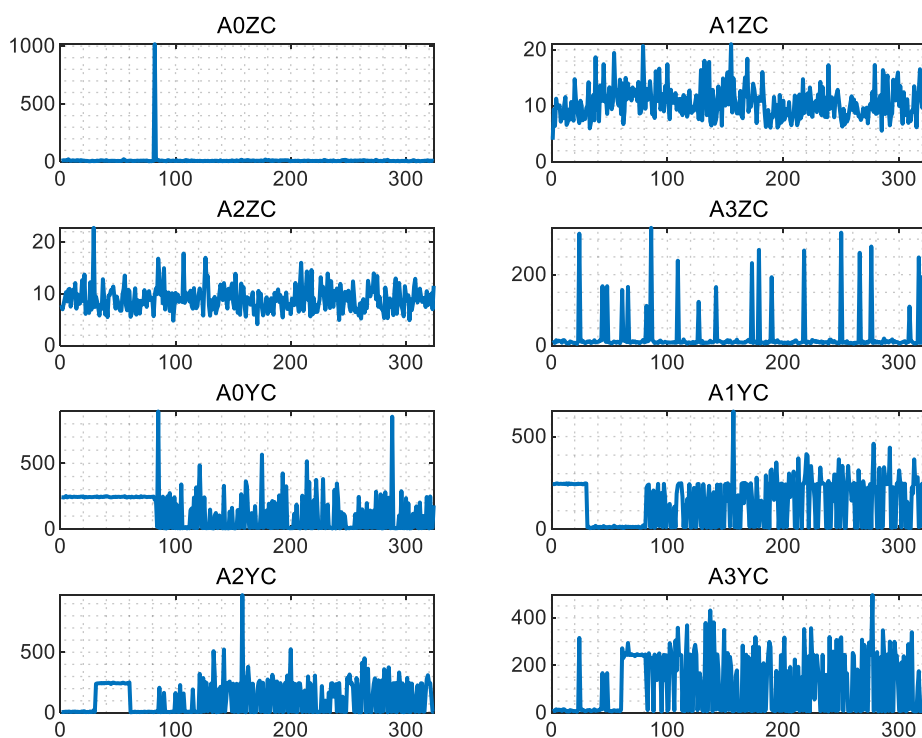


图 4.1 原始数据标准差分析

原始数据标准差分析表明，正常数据 A0ZCA3ZC 的标准差存在尖端，A1ZC、A2ZC 的标准差控制在 20 以内；异常数据中，每个锚点都存在标准差处于高水平(超过 200)的状态，选取两幅典型数据解释该问题，这是因为异常数据中，同一时间内仅有一个锚点存在异常状态，但是同一组数据内先后有两个锚点被干扰，干扰对测量数据造成的影响体现为两种情况，在原有数据基础上引入了超过 400 的测量偏差，或者引入一段均值为 400 左右的大幅值(波动水平：在 400 上下有幅值为 200 的波动)波动误差。这里要指出，对于异常数据的定义为：在正常数据水平上出现了正偏置干扰，测量数据原有的测量噪声与正常数据中的测量噪声一致看待。

题中对正常数据的定义为不存在测量干扰的数据，但是通过标准差分析发现，A0ZCA3ZC 的标准差存在尖端。提取出部分数据并做图分析可知，这是因为原始数据中存在异常点，具体表现为图像出现尖端。

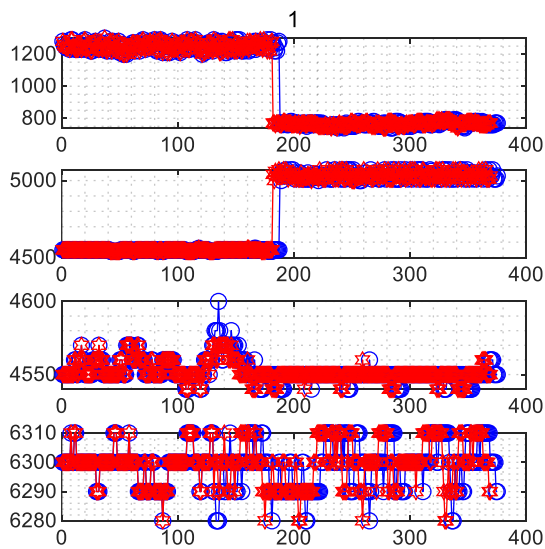


图 4.2 典型 YC 数据-第 3 组

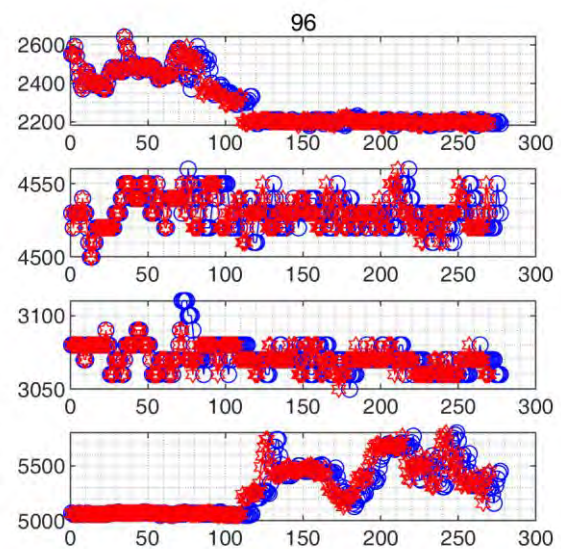


图 4.3 典型 YC 数据-第 96 组

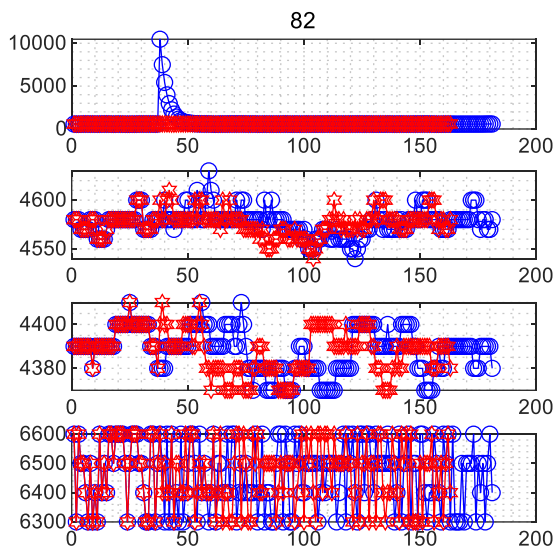


图 4.4 第 82 组 ZC 数据

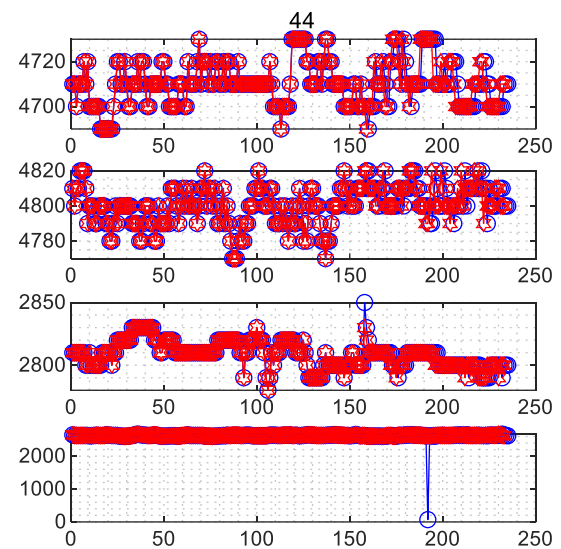


图 4.5 第 44 组 ZC 数据

## 4.2 剔除异常数据

通过对附件 1 中各个 txt 文件内测量值数据进行统计分析发现, (以正常文件夹中 1.txt 为例), A0-A3 分别到靶点的距离近似服从正态分布, 在 A2 的 365 组数据中, 仅有 1 项数据为 4600, 偏离 A2 均值较远。针对异常值处理常选用的方法有直接删除、平均值修正、视为缺失值等方法[1], 而发现异常值的方法可以通过建立统计量并设定异常点阈值、绘制散点图或箱体图分析异常值等方法, 通过实际分析发现, 样本数据基本呈现正态分布, 4

个锚点的测量数据分布情况不同，如 A0 数据方差为 117.7725，A356.8014，为但样本中的数据异常主要体现在少量数据偏离均值。假设测量数据服从正态分布，运用  $3\sigma$  原则设定异常值诊断方法，规定满足式(1.1)的数据为正常值。相对应的，不满足该原则的数据将正组被提出(某一时刻测量数据中，只要单一传感器测量数据不满足该原则，该时刻四项测量数据均被剔除)

$$|r - \text{mean}(r)| < 3\text{std}(r) \quad (1.1)$$

表 1.1 第一组数据统计信息

	A0	A1	A2	A3
均值	763.8	4548.9	4552.4	6299.2
方差	117.7725	15.6463	61.2078	56.8014

为了进一步分析 应该完成具体的统计

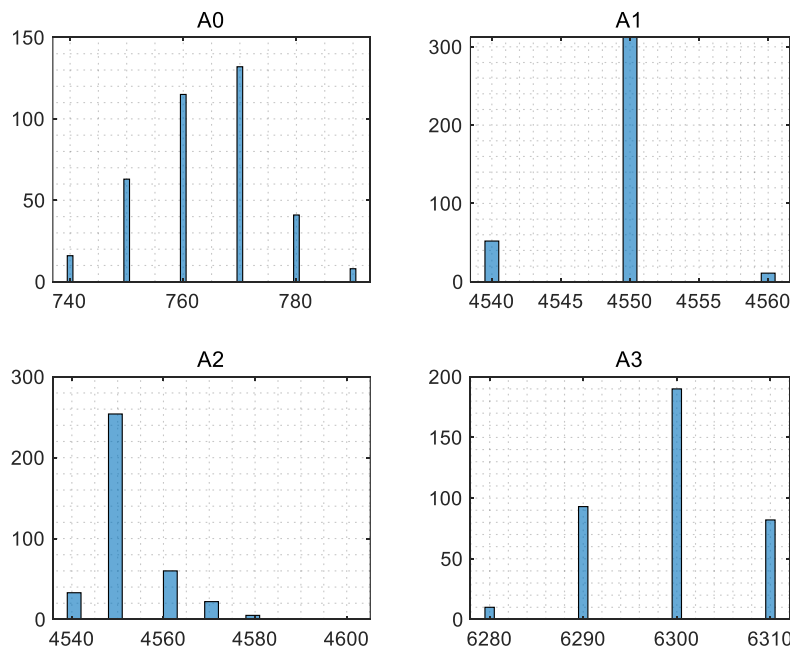


图 4.6 正常 1 直方图统计

在实际处理中，无法通过一次运算剔除所有的超出正态分布范围的异常点，这是因为尖端点增加了均值和标准差，因此部分尖端点会被划分在原始数据下的正态分布范围内。随着尖端点逐个被提出，均值和标准差逐渐缩小并向正常范围靠拢，这也使得更多尖端点被检测出进而被剔除掉，故需要通过多次迭代保证尖端点被完全剔除。可设计如下的流程消除全部的尖端点。

算法中设定了标准差门限值决定是否提前终止迭代。如图 4.8 所示，通过统计分析可发现，各个传感器的测量值标准差水平基本保持在 25 的范围内，而异常数据保持在 100 以上，可认为标准差降到 25 的水平下为正常水平数据，此时引起原始测量数据散布的原因是测量噪声带来的数据干扰。然而考虑到 A3 传感器部分情况下标准差分布偏高，将门

限值设置为 100 以上避免带来数据误删。

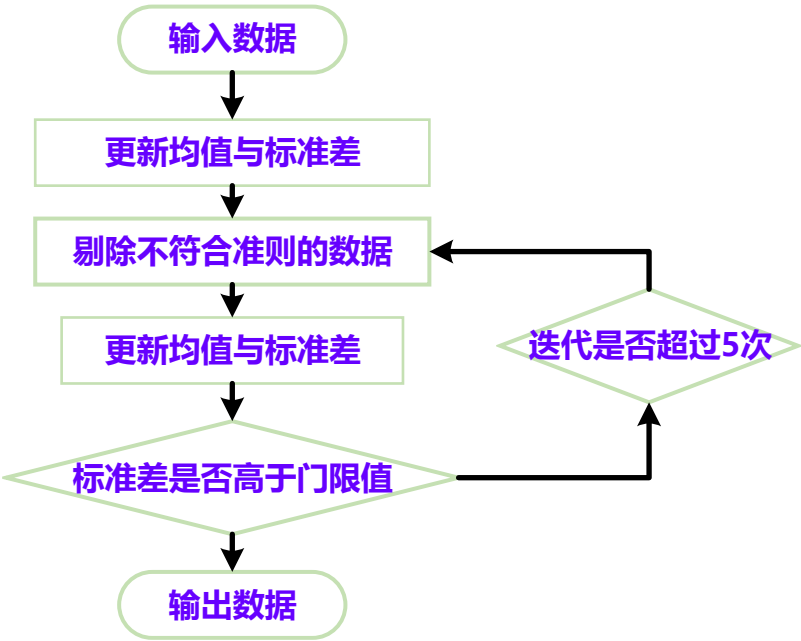


图 4.7 剔除简单数据算法流程

表 1.2 正常数据预处理前的统计信息的统计分析

	A0	A1	A2	A3
均值	10.5679	10.5187	8.9996	22.3418
标准差	3.1375	2.9178	2.4150	52.2993

表 1.3 正常数据预处理后的统计信息的统计分析

	A0	A1	A2	A3
均值	10.3933	10.3767	8.8178	9.8387
标准差	3.1728	2.9411	2.4187	3.0394

表 1.4 异常数据预处理前的统计信息的统计分析

	A0	A1	A2	A3
均值	128.3168	140.3464	116.7485	117.4819
标准差	135.1658	122.5944	134.7714	122.3155

表 1.5 异常数据预处理后的统计信息的统计分析

	A0	A1	A2	A3
均值	120.5680	138.6010	113.6125	108.8653
标准差	123.6373	122.4170	130.5125	117.5446

但是在正常数据中，第 82 组数据中 A3 数据表现出明显异常，A3 传感器测量数据偏差为 100，远远超出其他测量状态下的测量误差水平，因此第 82 组数据被认为存在问题，不予采用，但是为了保证全文顺序同一，下文仍认为数据存在 324 组。

324 组异常数据中，每一组数据可拆分为两锚点正常数据+两组锚点异常数据，两锚点

异常数据产生原因是测量值先后受到常值形式或波动形式的正偏置干扰。但是仍然有部分数据存在尖端现象。对于每组数据中的两锚点正常数据，易知可直接应用式(1.1)定义的处理规则。异常数据也可直接应用式(1.1)进行处理，这是因为异常数据中需要剔除的是尖端点，尖端点处于  $3\sigma$  之外。尽管异常锚点的测量数据中，数据按时间分为干扰和非干扰状态，但是在干扰和非干扰状态下数据先后保持在两个稳定的水平上，具体表现为均值升高，标准差增大，因此  $3\sigma$  原则可更加有效地剔除尖端点，同正常数据，仍然要多次迭代才能完全去除所有的尖端异常点。

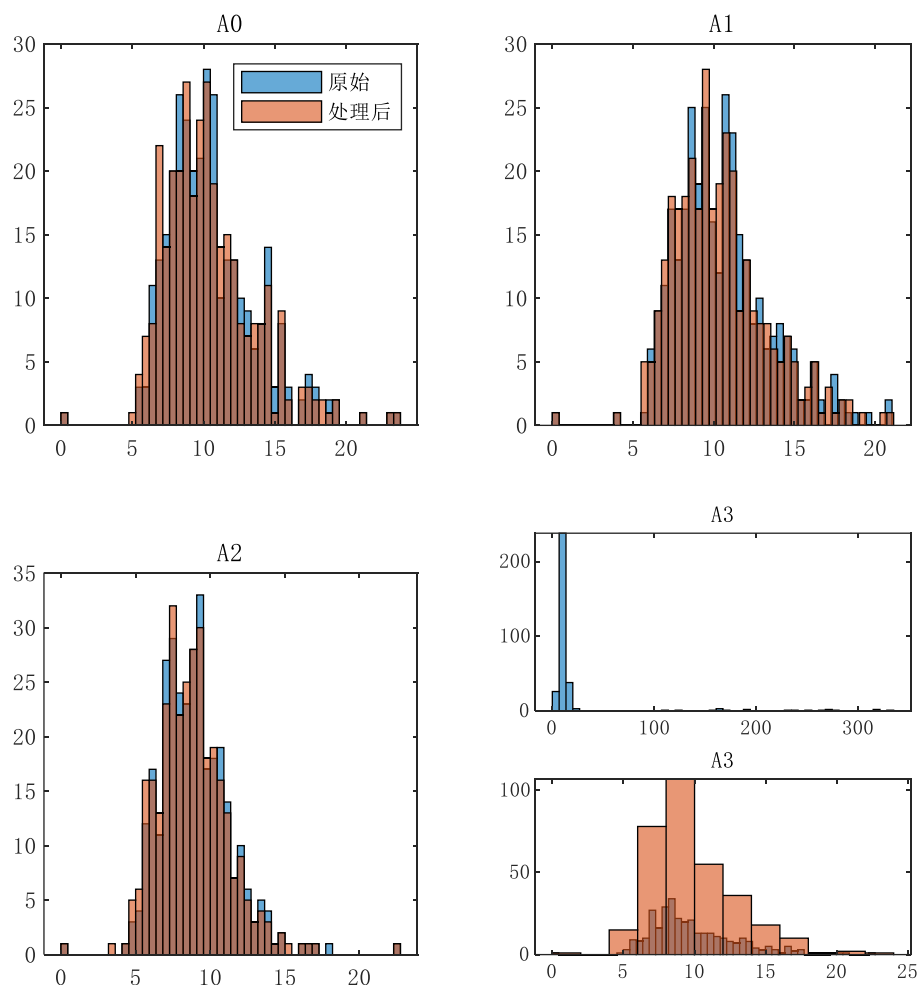


图 4.8 ZC 数据删除前后标准差统计

进一步对预处理后的正常数据进行统计分析。对于正常数据，比较有意义的信息是提取测量数据的噪声分布信息，因此需要统计分析预处理后 324 组数据对应的标准差数据 (324 项标准差)，定义 324 项标准差为测量噪声水平的分布信息，均值和标准差的统计信息见图 表。表中明确指出噪声水平数据皆低于 25，保持在均值为 10，标准差为 3 的正态分布下，该结论将在后文进行应用。

### 4.3 对于残缺数据定义的讨论

题目中强调，附件一中的数据可能在不同时间下测量，因此时间戳数据仅能由于检测

是否存在测量数据直接缺失的情况，而实际数据中并不存在该问题。同时由于各组数据测量时间不同，每个状态点下的正常数据和异常数据在数据量上存在差异，对于这一点并不按照缺失值处理，尽管所有样本不能保证在同一数量上进行统计分析，仍认为不同样本下统计结果有效且分析上占比等同，因此不会对正常、异常数据量之间的差异进行补充和修正。

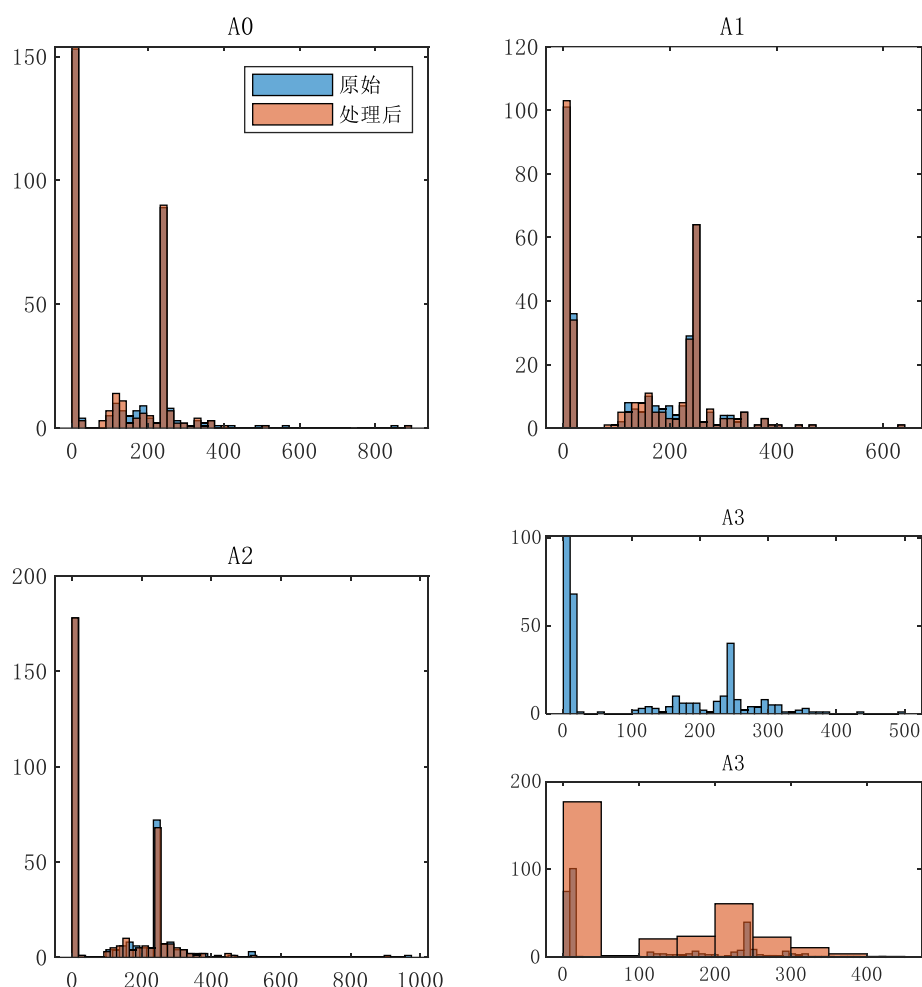


图 4.9 YC 数据删除前后标准差统计

#### 4.4 对于相同、相似数据的处理

考虑到题目中提出的具体要求应用算法针对各组数据中相同相似数据进行处理。相同数据仅仅通过判断两组数据是否相同完成消除。而对于异常数据，应用两组数据之间的欧式距离判断是否为相似数据。如果两组数据间的欧式距离低于 30 则判断两组数据为相似数据。

#### 4.5 关于预处理提供数据的说明

题目中对数据预处理提出了具体要求，因此在提交文件夹中给出了满足要求的相关数据。然而本文为了进一步对相关数据进行统计分析，仅仅使用剔除异常数据点的预处理后数据。对于剔除相同、相似数据后的简化数据，将为算法设计提供服务。



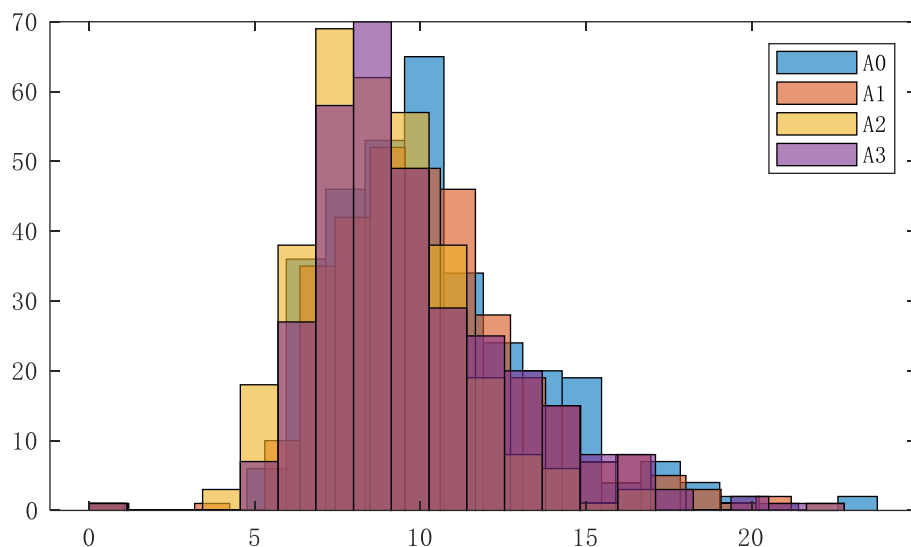


图 4.10 标准差数据正态分布图

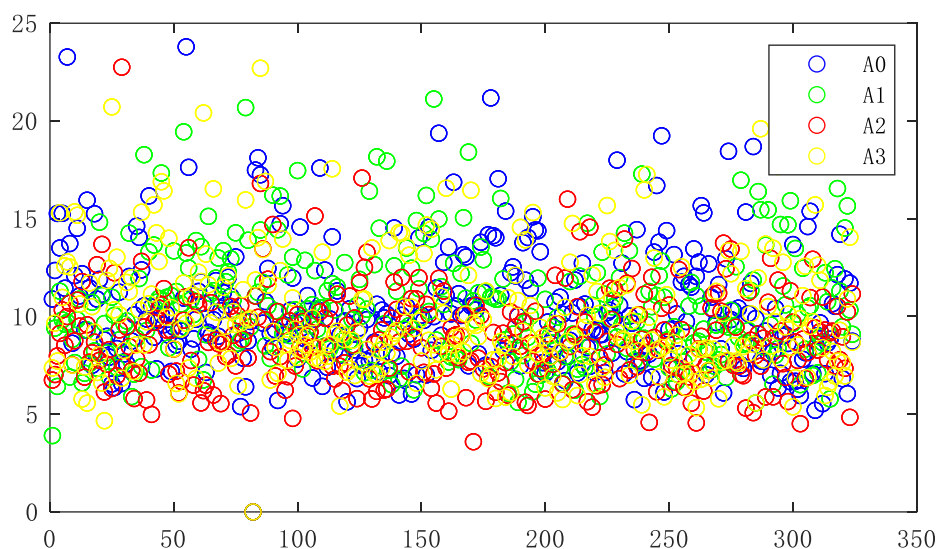


图 4.11 标准差数据散点图

#### 4.6 对于测量数据的修正策略

测量数据引起的误差原因是多方面的，考虑到题目中规定，附件一数据是基于 TOF 技术实现测量，因此本文仅仅分析给出的测量数据存在的误差组合，并不深究误差实际产生的原因，以及测量误差与 TOF 技术之间的联系，进而实现简化分析。

测量误差可被处理为白噪声偏差和有色误差，而有色误差可被处理为常值误差、正弦误差等形式。通过“附件一：UWB 数据集中的 Tag 坐标.txt”提供的靶点位置信息可以直接获得靶点到锚点的精确距离，本章简称为实际距离。将同一实际距离下不同的测量距离组成坐标，形成散点图，如图 4.12 所示。

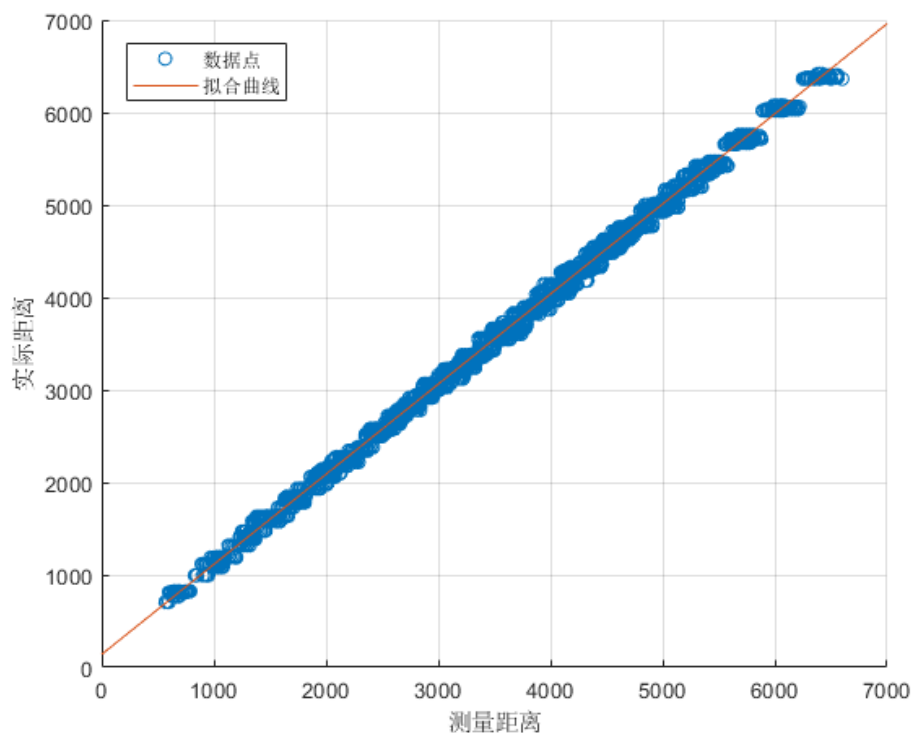


图 4.12 拟合曲线和数据点

从图中可以看出，实际距离和测量距离之间存在一定的线性关系，故对所有数据使用最小二乘法进行线性拟合，得到测量距离和实际距离存在如下的线性关系，在图 4.12 中以红色直线表示。

$$l_{\text{实际}} = 0.9743l_{\text{测量}} + 140.4$$

所有的测量长度经由上述线性变换后得到修正后的测量数据，从而在一定程度上减小了测量距离的误差，提高最终定位结果的准确性。



## 5 问题二建模与分析：定位模型

### 5.1 问题二分析与思路

首先考虑通过几何关系求解靶点的坐标。在实验场景 1 中，靶点（Tag）范围：5000mm\*5000mm\*3000mm，如图 5.1 建立空间直角坐标系，其中锚点位置如图所示，分别为 A0（0，0，1300）、A1（5000，0，1700）、A2（0，5000，1700）、A3（5000，5000，1300），而靶点可以出现在 5000mm\*5000mm\*3000mm 立方体空间中任意一处，根据靶点和锚点的空间关系，符合公式：

$$d_i^2 = (x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 \quad (5.1)$$

式中， $x, y, z$  为靶点坐标， $x_i, y_i, z_i$  为锚点的坐标， $i$  为锚点的编号（ $i=1,2,3,4$ ）。空间问题需要至少 4 个锚点来确定坐标，而上述方程为三元二次方程组。在理想情况下，三个锚点坐标信息以及这三个锚点到靶点的距离即可求解方程，得到两组解，通过第 4 个锚点的信息确定靶点的最终坐标；在实际情况下，由于测距数据存在误差，满足 4 个方程的解可能无实数解，故不能用此方法直接来求解靶点的坐标。

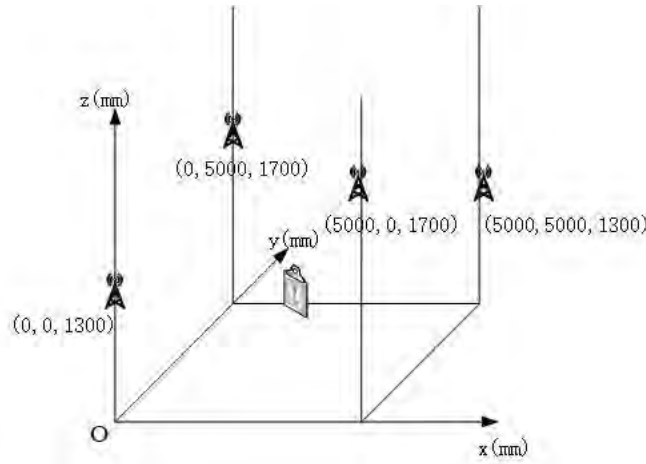


图 5.1 计算模型示意图

由于通过锚点位置信息和锚点测距值无法求解直接求解靶点坐标，故利用锚点坐标和锚点的测距值设计一个期望函数，通过函数值来反映靶点的坐标状况。

### 5.2 模型建立与求解

本题目利用最小二乘法、粒子群算法和机器学习算法进行求解并对其结果进行对比。

#### 5.2.1 最小二乘法

由于无法直接求解坐标，故可以采用最小二乘法，来求解出误差较小的结果。

##### 5.2.1.1 模型建立

假设点  $(x_i, y_i, z_i)$  是锚点坐标，标签坐标是  $(x, y, z)$ ，标签到各个锚点的距离分别为  $d_i$ ， $i=1, 2, 3, 4$ ，则由

$$d_i^2 = (x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2$$

前 3 个方程依次减去第 4 个方程，整理得到如下的矩阵关系：

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

将其写成：

$$AX = b$$

式中，

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

由于存在误差，故更合理的模型应该为：

$$AX + N = b$$

式中，N 是 3 维的随机误差向量。则由最小二乘法可以求得 X 的估计：

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

通过上式矩阵的运算即可求得靶点的估计坐标。

#### 5.2.1.2 模型求解

在 Matlab 中编写上述的算法的程序，使用已经完成预处理的数据求解 324 组正常的坐标，每一组正常数据中的 4 个靶点长度均求解一个坐标，即一组正常数据最终求解出一簇点。下图为第 1 组到第 81 组数据通过上述解法求解出的点在空间中的分布情况。

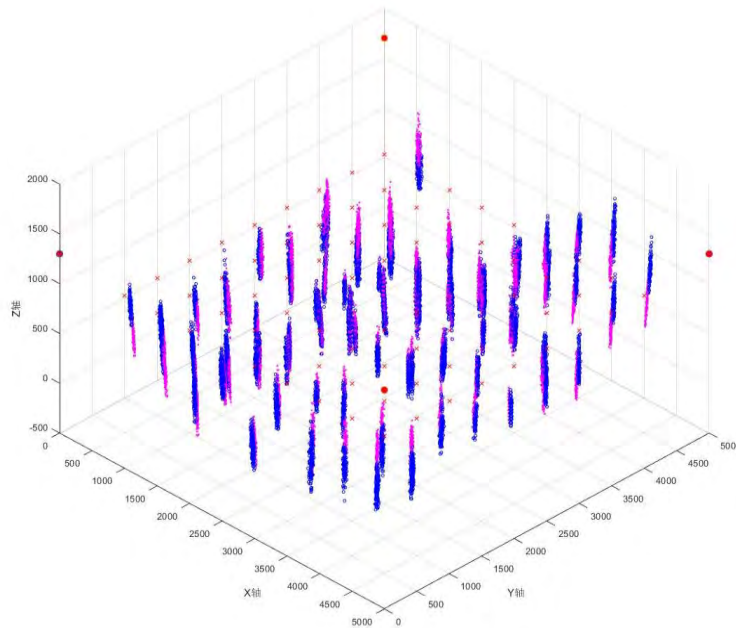


图 5.2 使用最小二乘法求解 1-81 组坐标结果

图中，红色实心圆：锚点的位置，  
 红色“X”：靶点的实际位置  
 蓝色点：修正前的测量长度求解出的靶点的位置  
 洋红色点：修正后的测量长度求解出的靶点的位置

从图中可以看出，Z 轴的求解值与实际值会有较大的偏差。事实上，采用最小二乘法求解坐标时，测量误差会逐渐积累到 Z 轴上，锚点的数量越少、锚点之间的垂直距离越小，Z 轴的偏差就会越大。

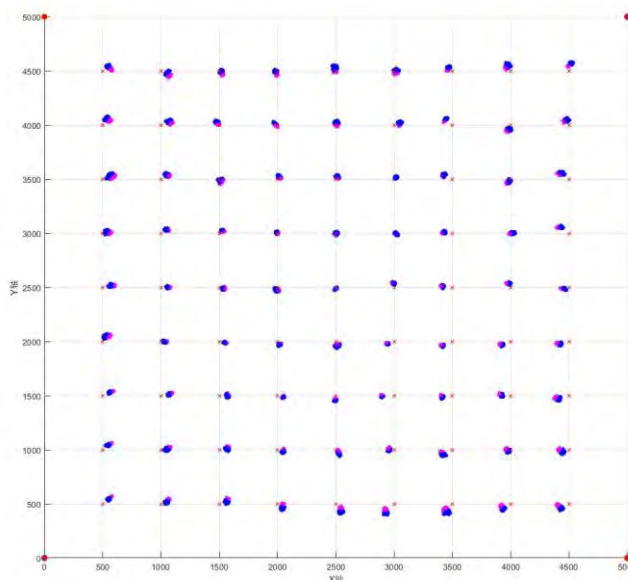


图 5.3 使用最小二乘法求解 1-81 组坐标结果在 xOy 平面内投影

上图为从 Z 轴正方向观察 1-81 组数据求解出的点簇的分布情况。可以看出，求解出的点在 (x,y) 平面中的投影与靶点的实际位置的偏差较小。

### 5.2.1.3 最小二乘法精度分析

为了更确切地分析求解坐标的精度，以均方根误差（RMSE）来表征最小二乘法求解出来的坐标的精度。分别在一维（x 方向、y 方向、z 方向）、二维（xoy 平面）、和三维空间共 5 个分别维度计算 RMSE。RMSE 的计算结果如下表：

表 5.1 正常数据修正前后 RMSE

	X 方向/mm	Y 方向/mm	Z 方向/mm	XOY 平面/mm	空间/mm
测量长度修正前	50.83	47.64	515.54	69.69	520.22
测量长度修正后	49.42	43.45	478.82	65.80	483.32

纵向比较，即判断测量长度修正前后是否会提高精度。从表中可以看出，测量长度修正后的 RMSE 与修正前相比有所减小，表明将测量长度进行变换后再计算坐标会在一定程度上提高总体的计算精度，证实了坐标变换的有效性。横向比较，即判断各个维度上的计算精度。在一维的精度上，Y 方向的精度是最高的，X 轴次之，精度最低的是 Z 方向。在三维的精度方面，由于 Z 方向的精度较低，从而导致坐标在空间的精度较低。

以上的算法和求解均是使用了在正常无干扰的情况下测量到的数据。另使用异常有干扰的数据采用以上算法求解坐标，观察部分数据发现，受异常测量数据的影响最终的计算坐标可能位于空间之外，原因是本算法仅能通过测量到的距离解算出坐标，并不能分辨距

离是否存在干扰，因此也无法较准确求解有干扰数据的坐标。求得到异常数据的 RMSE 值如下表：

表 5.2 异常数据修正前后 RMSE

	X 方向/mm	Y 方向/mm	Z 方向/mm	XOY 平面/mm	空间/mm
测量长度修正前	2503.7	2526.8	1435.8	3557.1	3836.0
测量长度修正后	200.7	204.6	2498.5	288.6	2514.9

由上表可以看出，使用修正前的测量长度计算在各个维度上都会有极大的误差，而修正后的数据在 X、Y 方向上的误差减小，Z 方向的误差增大，XOY 平面内的误差减小，空间内的误差增大。使用修正之前的长度进行求解，其 X、Y 方向的误差远远大于 Z 方向的误差，而使用修正之后的长度求解的误差恰好相反，在 XOY 平面内的精度得以提高，而空间的误差依然较大。需要注意的是，RMSE 值代表的是样本整体的误差程度，并不能代表单一样本的误差程度，因此很多样本的误差依然较大。

以上步骤基于最小二乘法完成了对正常无干扰测量数据的计算和精度分析。以本算法为基础，通过“附件二：测试集（实验场景一）”中 5 个正常无干扰的数据和 5 个正常有干扰的数据，首先计算出修正后的长度，以修正后的长度分别计算出对应的坐标，如下表：

表 5.3 使用最小二乘法求解问题二的结果

名称	坐标/mm
正常数据 1	(1193.1, 701.4, 491.1)
正常数据 2	(3166.4, 1725.4, 820.9)
正常数据 3	(2734.6, 1725.4, 820.9)
正常数据 4	(2449.9, 1028.7, 2266.2)
正常数据 5	(1492.8, 2540.2, 2130.8)
异常数据 1	(2094.2, 750.1 -247.9)
异常数据 2	(4299.6, 1599.9, -586.3)
异常数据 3	(1783.2, 1289.7, 2612.9)
异常数据 4	(3542.8, 2011.5, 2705.7)
异常数据 5	(4860.9, 1993.5, -837.0)

由上表可以看出，异常数据的求解结果可能存在较大的偏差，也就是说最小二乘法不适宜运用在异常数据的求解中，需要使用其他的算法求解。

### 5.2.2 粒子群算法

粒子群作为启发式算法的一种，对于求解多元函数问题的极值有直观，简介的特点，此算法不用对多元函数逐个未知量的求偏导，而改位多个粒子之间相互学习和运动的行为，体现向极值逼近的运动现象，仅需迭代就能使函数值逼近到期望的水平，但运用此算法需要使用适应度函数，其本质上还是四点确定空间一点的原理，只是对求解过程做了简化，从而达到方便计算的目的。

### 5.2.2.1 模型建立

假设测量距离准确，通过错误!未找到引用源。则靶点到各个锚点的距离应等于各个锚点测量的距离，则有：

$$(x-x_i)^2+(y-y_i)^2+(z-z_i)^2-r_i^2=\varepsilon,i=0,1,2,3$$

其中： $\varepsilon$  为靶点到锚点的坐标计算距离平方与锚点测量距离平方的偏差

在理想状况下， $\varepsilon = 0$  即错误!未找到引用源。等价于错误!未找到引用源。，而由于值即测量值存在偏差和误差， $\varepsilon \neq 0$ ，假设测量是准确的，那么  $\varepsilon$  应该尽可能小，对于四个锚点，都应有上述结论，那么则通过兼顾四个锚点的偏差则有：

$$E=\sum_{i=0}^3\omega_i[(x-x_i)^2+(y-y_i)^2+(z-z_i)^2-r_i^2] \quad (5.2)$$

其中： $\omega_i$  为靶点到  $i$  号锚点的坐标计算距离平方与  $i$  号锚点测量距离平方的偏差的权重， $E$  为期望函数。

在错误!未找到引用源。中假设各个锚点权重相同即  $\omega_i=1$ ，当  $E$  最小时即为靶点到各个锚点的偏差和最小，认为此时的靶点坐标值为通过锚点位置和测距信息解算出的坐标值，下简称解算值。

需要设计算法，来求取  $E$  的最小时的坐标值。

### 5.2.2.2 模型求解

粒子算法原理为在解空间中初始化一群随机粒子(随机解)，然后通过迭代找到最优解，在每一次迭代中，粒子都要向两个“最优值”进行学习。第一个最优值就是粒子自己从迭代开始但当前的一代最好的解，这个解叫做个体极值  $pBest$ ，另一个最优值是粒子群从迭代开始但当前的一代所有粒子中最好的一个解，这个最优值是全局极值  $gBest$ 。将这两个极值与自己做对比，并且乘以一个学习因子添加到粒子的运动速度中，为粒子下一次的运动方向做准备。

如图粒子群算法流程图，适应度函数即上文中计算  $\Delta V$  的函数，本问题的粒子 3 个维度分别  $x,y,z$ ，即靶点估计坐标值，而粒子群算法本身涉及的变量如表所示。

表 5.4 粒子群算法变量符号及其含义表

变量符号	变量含义
$N$	群体粒子个数
$D$	粒子维数
$T$	最大迭代次数
$c_1$	学习因子 1,学习局部最优解
$c_2$	学习因子 2,学习全局最优解
$\omega_{\max}$	惯性权重最大值

---

$\omega_{\min}$	惯性权重最小值
$x_{\max}$	粒子位置上界
$x_{\min}$	粒子位置下界
$v_{\max}$	粒子运动速度上界
$v_{\min}$	粒子运动速度下界

---

算法中粒子位置上下界由靶点范围：5000mm\*5000mm\*3000mm 确定，例子运动速度上下界粒子位置范围的 5%，群体粒子个数  $N=100$ ，最大迭代次数  $T=50$ ，学习因子  $1c_1=1.5$ ，学习因子  $2c_2=1.5$ ，惯性权重最大值  $\omega_{\max}=1.0$ ，惯性权重最小值  $\omega_{\min}=0.8$ 。

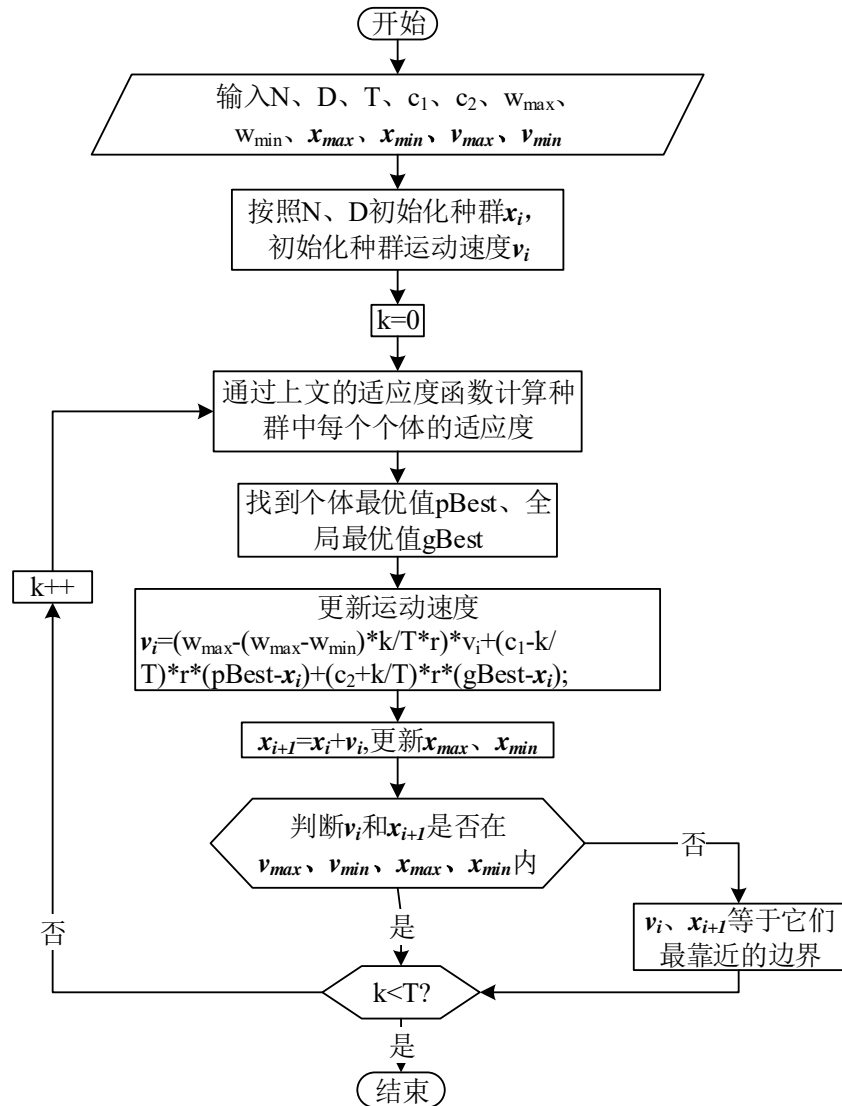


图 5.4 粒子群算法程序流程图

粒子群算法适应度函数选择(5.2)，优化适应度函数使 E 最小并记录此时的坐标值。

### 5.2.2.3 对于异常数据的处理

通过附件一的数据不难发现，异常数据只有 4 组中只有一组，且比正常值高 300mm-500mm，将一组异常数据中的，每一个异常数据依次进行减去 300mm-500mm 的随机值，然后使用粒子群算法，会得四个适应度函数和四组解算值，认为适应度函数最小的为正确解算值并作为输出。

### 5.2.2.4 实验结果分析

#### (1) 对于正常数据的学习与分析

在使用粒子群算法解算附件二的测试集之前，先对附件一的测试集进行解算，并根据测算结果进行统计学分析。

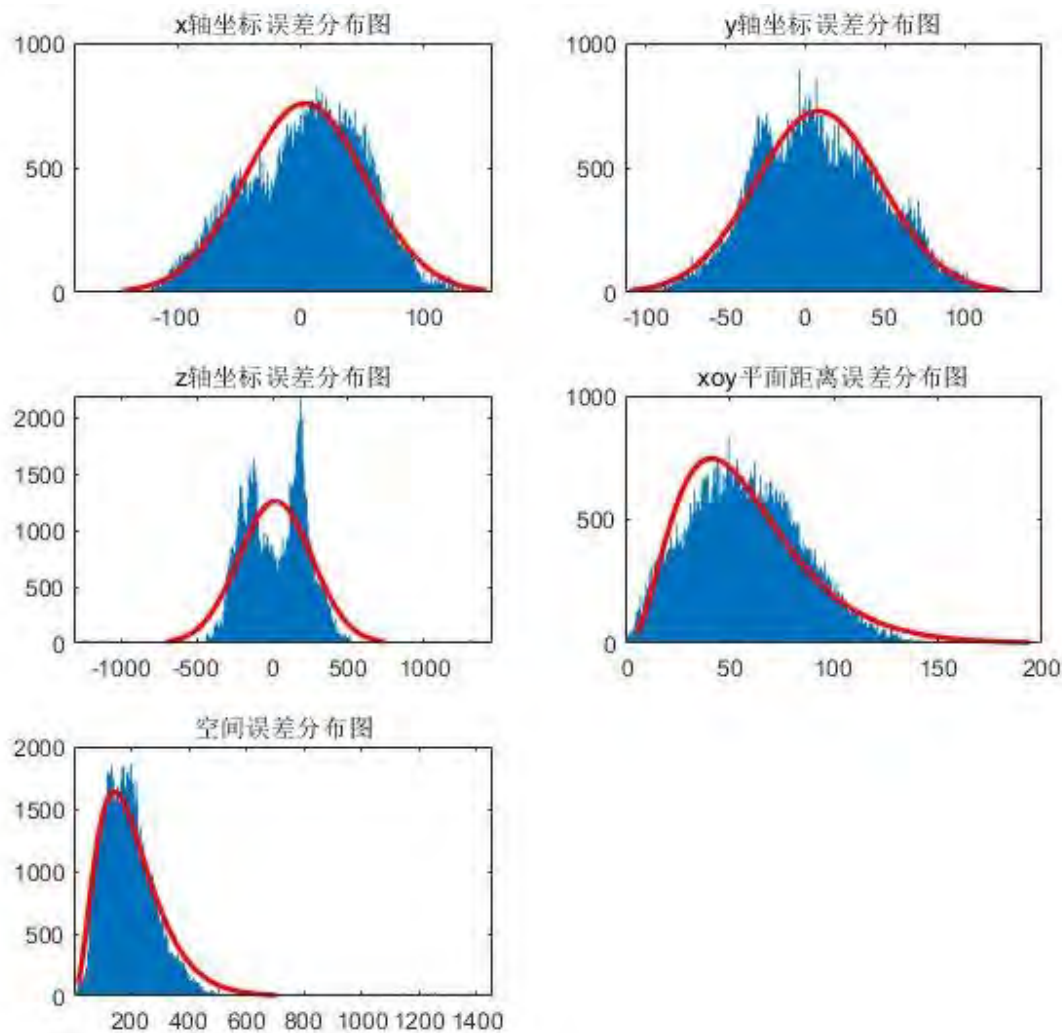


图 5.5 粒子群算法正常数据解算值直方频次分布图

表 5.5 粒子群算法各定位项目精度表

定位项目	置信区间(mm)	概率
------	----------	----

x 轴坐标	$(x-147.36, x+147.36)$	99.74%
y 轴坐标	$(y-118.42, y+118.42)$	99.74%
z 轴坐标	$(z-720.71, z+720.71)$	99.74%
xoy 平面距离	$(r, r+169.21)$	99.00%
空间距离	$(r, r+514.7)$	99.00%

通过图 5.5 和表 5.5 粒子群算法各定位项目精度表可知 PSO 算法在预测精度上 z 轴误差远远大于其它轴误差，导致其预测效果不好。

表 5.6 PSO 方法任务二正常数据解算值(单位：mm)

组数	1	2	3	4	5	精度
A0 测距值	1320	3580	2930	2740	2980	
A1 测距值	3950	2580	2600	2720	4310	
A2 测距值	4540	4610	4740	4670	2820	
A3 测距值	5760	3730	4420	4790	4320	
x 坐标解算值	1156.12	3166.72	2722.47	2462.43	1480.13	$(x-147.36, x+147.36)$
y 坐标解算值	664.37	1722.61	1171.41	1009.16	2554.50	$(y-118.42, y+118.42)$
z 坐标解算值	1143.30	1135.54	1234.15	1938.20	1731.79	$(z-720.71, z+720.71)$

## (2) 对于异常数据的学习与分析

按照 5.2.2.3 的方法求解出对附件一的异常数据进行解算并进行分析，可以看到得到的数据存在整体的漂移，且出现了多个峰值，各个解算值的方差极大，分布很散，对于准确预测坐标值已经没有意义，故不使用 PSO 算法解算异常数据以得到坐标值。



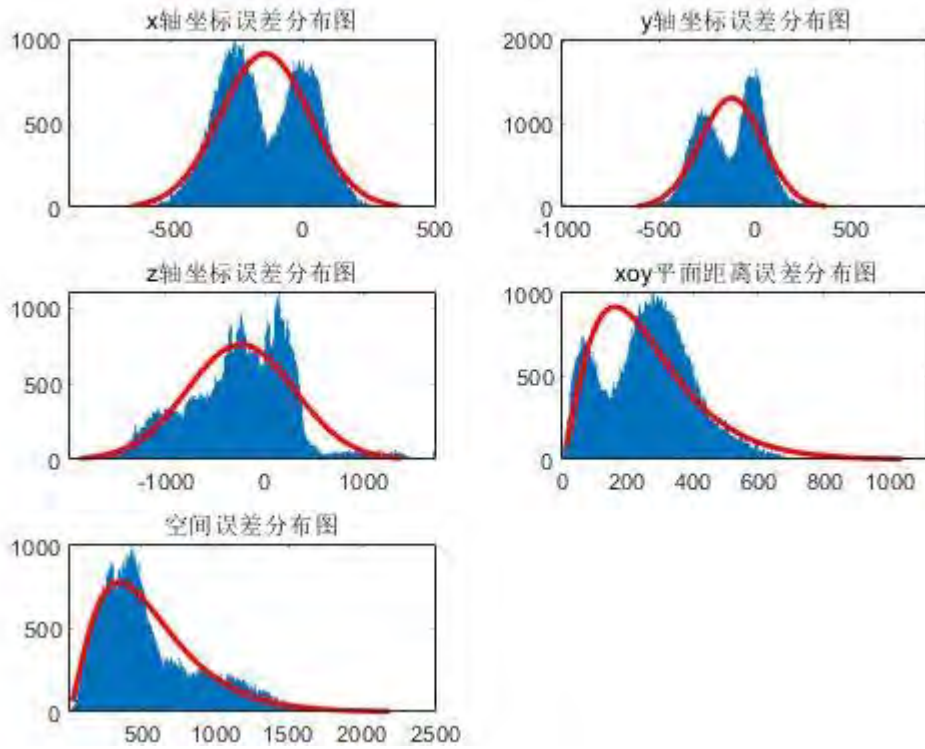


图 5.6 粒子群算法异常数据解算值直方频次分布图

### 5.2.3 神经网络算法

神经网络算法和粒子群算法有本质不同，其不关注求解方法本身也不考虑，而只是关注多层学习中的输入输出数，节点数、层数、学习方法和激活函数，也就是只关注神经网络本身的结构，而在学习过程中也仅仅需要提供学习可以使用是实验数据，端对端式的进行学习。本算法需要大量训练数据才可以使用，且速度较慢，但是训练后的结果，偏差大幅减小，拟合程度大幅度提高，在任务二的模型中，是没有数据可以进行训练的，假定任务一和任务二使用的仪器相同，即认为系统特性不发生改变，使用任务一的数据进行训练再将模型用于任务二。

#### 5.2.3.1 模型建立

由于神经网络算法是一个“端对端”的智能算法，故仅需设计算法需要的输入模型和输出模型即可。

本任务的输出模型非常直观，设置为靶点位置的坐标即：

$$[x \ y \ z]$$

在任务一和任务二的情景下，仅有四个锚点的测量值发生变化，故输出模型为：

$$[r_0 \ r_1 \ r_2 \ r_3]$$

但是由于在后续任务场景中，场地和锚点位置都发生了变化，这种变化也会对靶点坐标的解算造成巨大的影响，故将四个锚点坐标纳入输入模型中，变化后的输入模型为：

$$[x_0 \ y_0 \ z_0 \ x_1 \ y_1 \ z_1 \ x_2 \ y_2 \ z_2 \ x_3 \ y_3 \ z_3 \ r_0 \ r_1 \ r_2 \ r_3]$$

#### 5.2.3.2 模型求解

BP(back propagation)神经网络是 1986 年由 Rumelhart 和 McClelland 为首的科学家提出的概念，是一种按照误差逆向传播算法训练的多层前馈神经网络，是应用最广泛的神经网络。

如图 5.7 所示，神经网络的基本构造包括输入层，中间隐含层，输出层，输入层即为问题的多个输入，隐含层与输入层输出层进行全连接，并且隐含层不一定只含有一层，可以有多个隐含层，隐含层直接也采取全连接的方式进行链接，输出层则是输出期望的输出函数。

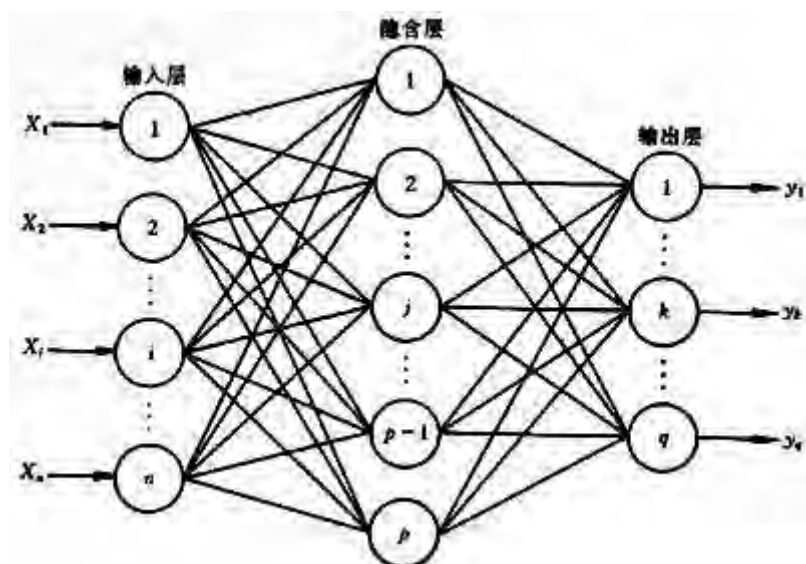


图 5.7 神经网络基本构造

如图所示神经网络算法流程为，将输入的学习样本进行归一化处理，再设定训练最大次数，学习精度等神经网络参数，然后初始化各层间的权值、阈值、初始学习速度等参数，然后计算各层输入和输出，计算输出层的误差  $E$ ，并判断  $R$  是否符合学习精度要求  $\varepsilon$ ，以及是否到达最大迭代次数  $n$ ，如果是则直接输出训练得到的各个层间的参数，如果不是则修正权值和阈值，然后重新进行计算各层的输入和输出，循环迭代。

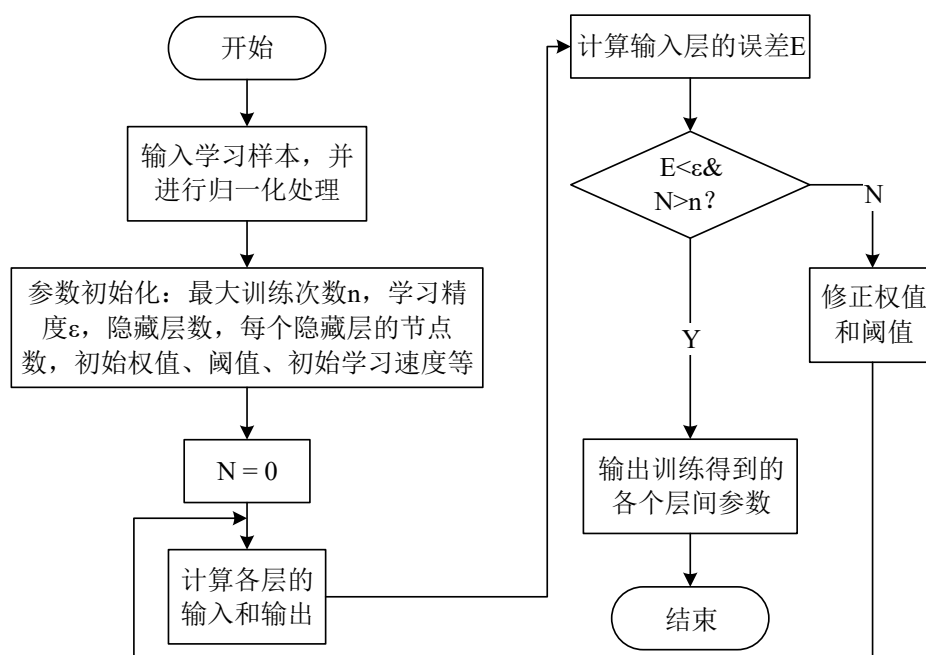


图 5.8 神经网络算法程序流程图

### 5.2.3.3 实验结果分析

#### (1) 正常数据学习与分析

使用附件一正常数据的训练一个 3 层, 节点分别为 30、25、20 的神经网络, 并利用神经网络预测附件一的正常数据, 预测数据和真实数据误差频数直方分布如图所示, 通过图片很容易看出解算出的  $x, y, z$  轴的坐标值呈正态分布,  $xoy$  平面距离和空间距离误差都呈  $\chi^2$  分布, 这也间接再次证明了  $x, y, z$  轴的坐标值呈正态分布。

根据误差范围的置信区间来定义各个测量值的精度, 对于  $x, y, z$  轴的坐标值采用  $3\sigma$  定律, 而对于  $\chi^2$  分布采用置信区间内的分布概率为 99% 的方法计算其精度, 其精度及置信区间如表所示。

表 5.7 各定位项目精度表

定位项目	置信区间(mm)	概率
x 轴坐标	$(x - 86.95, x + 86.95)$	99.74%
y 轴坐标	$(y - 76.61, y + 76.61)$	99.74%
z 轴坐标	$(z - 111.83, z + 111.83)$	99.74%
$xoy$ 平面距离	$(r, r + 101.20)$	99.00%

空间距离

 $(r, r+140.90)$ 

99.00%

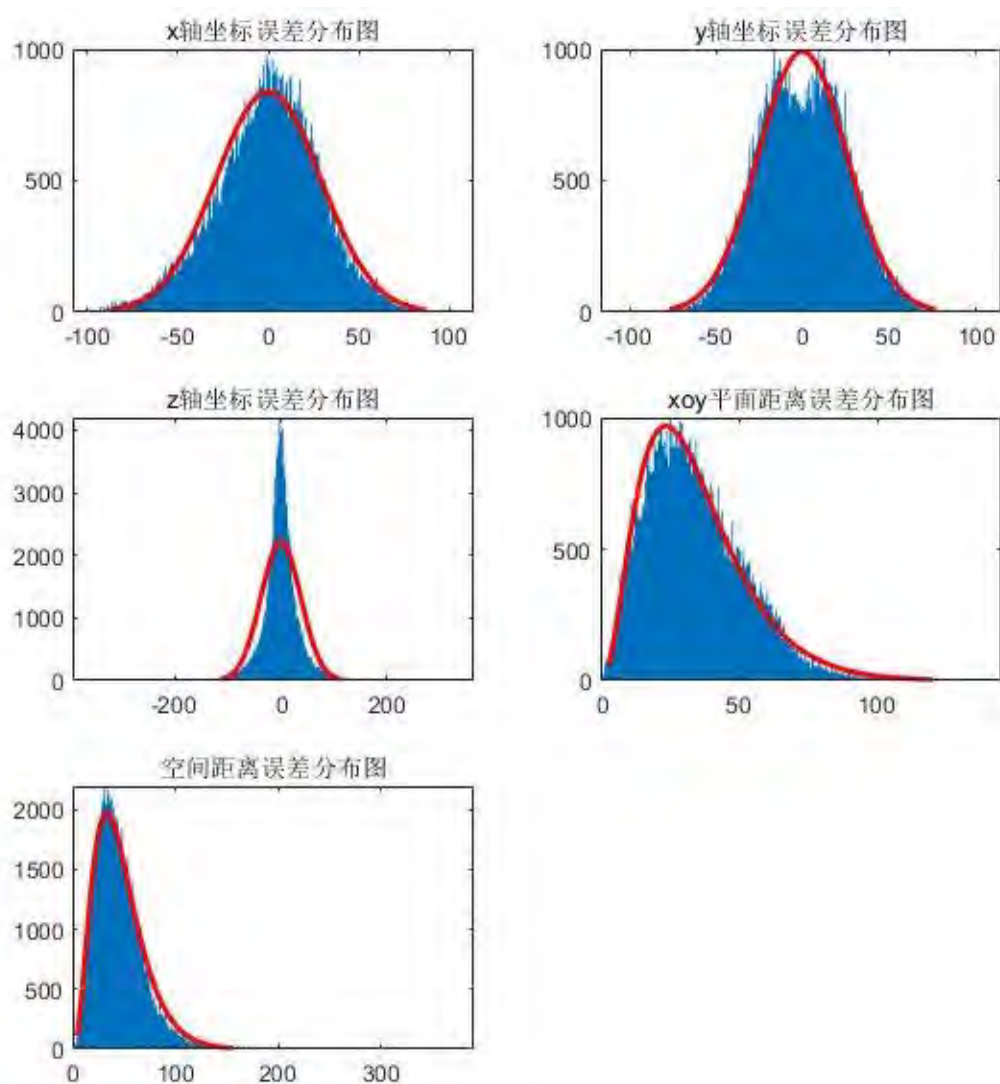


图 5.9 正常数据解算误差频数直方分布图

将附件二中场景一下的正常数据利用机器学习得到的结果进行解算，得到的结果如表所示

表 5.8 任务二正常数据解算值(单位: mm)

组数	1	2	3	4	5	精度
A0 测距值	1320	3580	2930	2740	2980	
A1 测距值	3950	2580	2600	2720	4310	
A2 测距值	4540	4610	4740	4670	2820	
A3 测距值	5760	3730	4420	4790	4320	
x 坐标解算值	1112.34	3216.36	2828.75	2482.48	1508.89	$(x-86.95, x+86.95)$

y 坐标解算值	604.12	1725.85	1174.65	1072.72	2510.90	$(y - 76.61, y + 76.61)$
z 坐标解算值	888.10	786.08	768.10	2007.36	2063.89	$(z - 111.83, z + 111.83)$

## (2) 异常数据学习与分析

使用附件一异常数据的训练一个 3 层，节点分别为 40、30、20 的神经网络，并利用神经网络预测附件一的异常数据，预测数据和真实数据误差频数直方分布如图所示，通过图片很容易看出解算出的  $x, y, z$  轴的坐标值呈正态分布，xoy 平面距离和空间距离误差都呈  $\chi^2$  分布，这也间接再次证明了  $x, y, z$  轴的坐标值呈正态分布。但同时，可以看到解算出的坐标值分布很不稳定，存在多个尖峰，训练出的系统存在“谐波”，可以通过增加隐藏层数和每层节点数来减少这样的谐波，但由于能够使用的算力不足等原因，无法承载更大规模的学习。

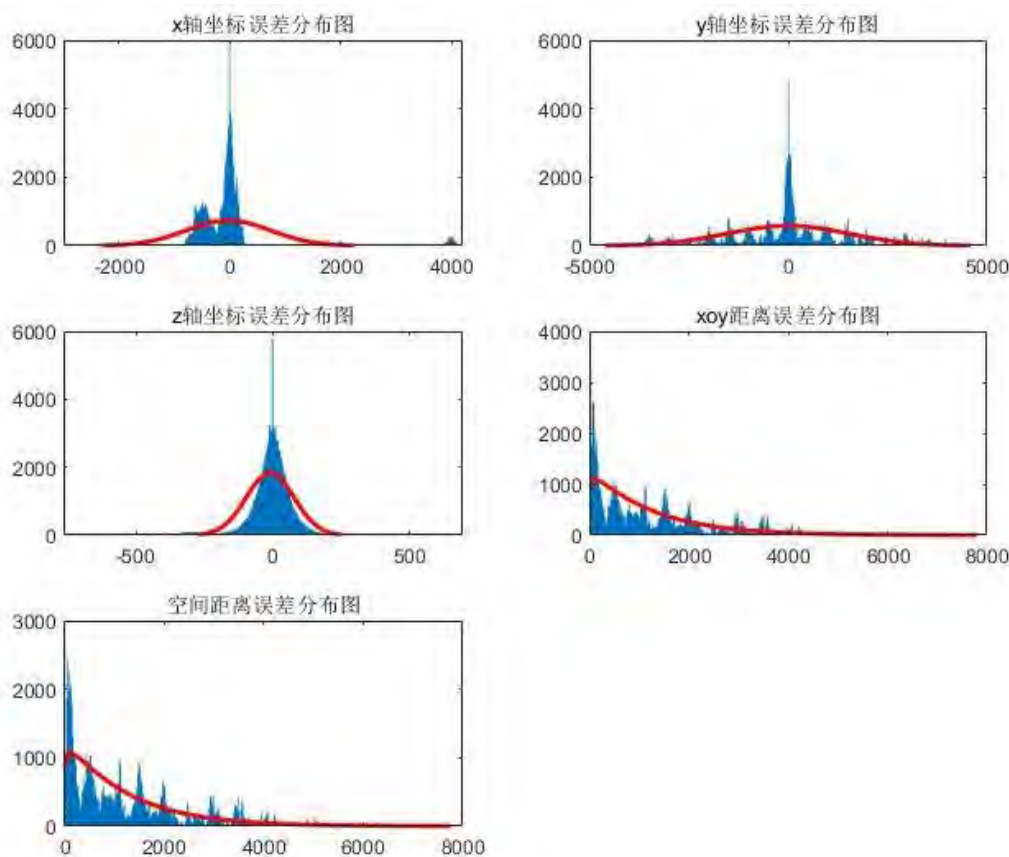


图 5.10 异常数据解算误差频数直方分布图

根据误差范围的置信区间来定义各个测量值的精度，由于训练出的模型存在而导致正态分布曲线和  $\chi^2$  分布曲线背拉宽已经没有意义，故认为中间峰为精确值，给出中间峰在所有峰中面积的占比，表示此模型解算出的解是正确解的概率，并将这个概率定义为精度。

对于异常数据，可以先进行分类再使用分类后的数据训练神经网络，或者通过增加神

---

经网络的层数和节点数，来提高模型的精度，以消除谐波，解算出合理的解。

表 5.9 任务二异常数据解算值(单位：mm)

组数	1	2	3	4	5	精度
A0 测距值	2230	4520	2480	4220	5150	
A1 测距值	3230	1990	3530	2510	2120	
A2 测距值	4910	5600	4180	4670	5800	
A3 测距值	5180	3360	5070	3492	2770	
x 坐标解算值	2007.652	4168.847	1637.887	3584.506	4386.972	65.40%
y 坐标解算值	755.2196	1672.897	1234.318	1838.467	2257.694	53.20%
z 坐标解算值	1276.208	1708.154	1198.482	1556.597	1616.428	100%

## 6 问题三建模与分析：不同场景应用

### 6.1 问题一分析与思路

在第 5 章中对于第 6 章的不同应用场景下，通过四个锚点坐标和锚点距离测量值的问题已经做了考量，故直接使用第 5 章的方法进行求解。

### 6.2 正常数据解算

第 5 章利用了最小二乘法、PSO 算法和神经网络算法两个算法对正常数据进行了求解，本章也利用这三个算法对正常数据进行求解。

#### 6.2.1 最小二乘法求解结果

组数	1	2	3	4	5
A0 测距值	4220	4500	3550	3300	720
A1 测距值	2580	1940	2510	3130	4520
A2 测距值	3730	4420	3410	2900	3050
A3 测距值	1450	1460	2140	2790	5380
x 坐标解算值	3643.5	4206.2	3165.3	2585.4	538.8
y 坐标解算值	2205.3	1735.3	1724.8	1872.3	58.9
z 坐标解算值	899.9	1063.7	918.0	1687.4	988.4

#### 6.2.2 PSO 算法求解结果

组数	1	2	3	4	5
A0 测距值	4220	4500	3550	3300	720
A1 测距值	2580	1940	2510	3130	4520
A2 测距值	3730	4420	3410	2900	3050
A3 测距值	1450	1460	2140	2790	5380
x 坐标解算值	3670.439	4199.144	3177.178	2590.015	535.6363
y 坐标解算值	2216.251	1735.828	1737.702	1879.034	99.7232
z 坐标解算值	1279.471	1213.5	1293.563	2098.014	747.1726

#### 6.2.3 神经网络求解结果

组数	1	2	3	4	5
A0 测距值	4220	4500	3550	3300	720
A1 测距值	2580	1940	2510	3130	4520
A2 测距值	3730	4420	3410	2900	3050
A3 测距值	1450	1460	2140	2790	5380
x 坐标解算值	3765.21	4656.48	2858.94	1745.15	214.85

y 坐标解算值	3257.90	2783.47	2880.07	2650.22	1641.65
z 坐标解算值	227.00	-256.59	90.36	187.68	2350.60

### 6.3 异常数据解算

第 5 章只使用了神经网络算法对数据进行求解，故此处也利用神经网络算法对数据进行求解，并且增加视所有异常数据都是正常的，利用 PSO 算法求解。

#### 6.3.1 神经网络求解结果

组数	1	2	3	4	5
A0 测距值	5100	2900	2380	2150	1620
A1 测距值	2220	3210	3530	3220	3950
A2 测距值	4970	3140	2320	3140	2580
A3 测距值	800	2890	3760	3640	4440
x 坐标解算值	6042.76	2278.14	2273.60	2150.89	1109.31
y 坐标解算值	4489.35	2395.42	3274.90	2093.57	2053.31
z 坐标解算值	190.78	-832.14	2589.43	-666.55	-2152.27

#### 6.3.2 PSO 算法求解结果

组数	1	2	3	4	5
A0 测距值	5100	2900	2380	2150	1620
A1 测距值	2220	3210	3530	3220	3950
A2 测距值	4970	3140	2320	3140	2580
A3 测距值	800	2890	3760	3640	4440
x 坐标解算值	4768.50	2481.97	1722.16	2025.43	1191.68
y 坐标解算值	2156.54	1539.72	1420.36	803.04	777.77
z 坐标解算值	1115.32	518.03	1974.62	795.02	1983.34



## 7 问题四建模与分析：分类模型

### 7.1 基于测量距离的两类尝试

采用四点算法解算数据后，需要对求解精度进行分析，检验策略是求解得到的靶点坐标是否为真实解。其中一类检验思路是检验解算前后的测距是否一致，如图 7.1 所示。分析解算前、后测距  $R_0-R_3$  与  $r_0-r_3$  间偏差，检验结果有效性。对于存在噪声干扰的正常数据，即使四点算法求解得到靶点真实坐标，测距偏差仍然是存在的。基于测距进行偏差分析的方法有效利用了测量值，可在不知道靶点真实坐标的情况下检验结果有效性。通过进一步分析发现，该算法除了可以用于检验算法精度，还可以用于异常数据的分类识别。

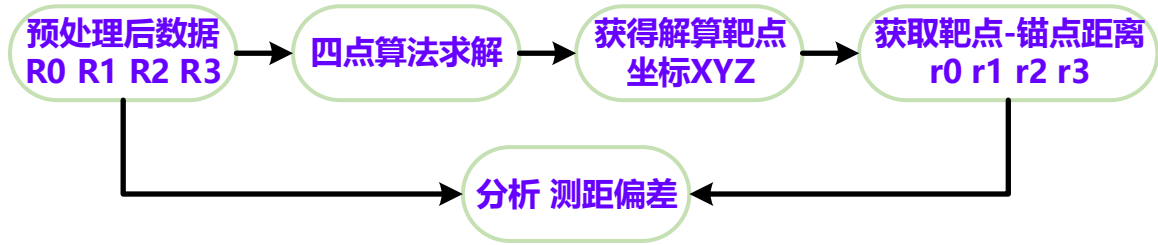


图 7.1 分析算法

分别应用三种方法完成结果讨论和分析。如图 7.2 所示，预处理后的测距用符号  $R_i, i=1,2,3,4$  表示，解算后获取的测距用符号  $r_i, i=1,2,3,4$  表示，定义测距偏差为预处理后的测距  $R_i$  与解算后获取的测距  $r_i$  之差，并用符号  $E_r^i, i=1,2,3,4$  表示测距偏差，

$$E_r^i = |R_i - r_i|, i=1,2,3,4$$

基于偏差定义判别函数  $\phi$

$$\phi = \sqrt{\sum_{i=1}^4 (E_r^i)^2}$$

基于全部样本进行统计存在不公平问题，尽管误差是随机分布的，但是每个测量点的数据可被判别的能力所需的统计样本并不是全部样本，而是仅仅包含典型样本的抽样样本，通过典型数据可以直接分析针对不同状态异常数据的区分能力，进而得到确切判别方法识别精度与应用环境，而不是以概率为指标评价判别能力。因此这种情况，除了要考虑各种情况的分布情况，也要关注误差的分布范围，然而，基于误差准则的判别方法并不能区分两类数据，基于 PSO 的算法基于给定的数据可以寻找到一组稳定的解，这组靶点坐标到锚点的测距与给定的测距，较为接近，这种情况下基于测距偏差的判别方法将认定这组数据为正常数据，这也侧面说明了没有先验信息情况下，基于几何关系出发的判别方法在仅仅利用测距信息时判别精度有限，如果能尽可能提高正常数据下的求解精度，降低正常数据测距偏差的散布，也许可以这一类判别方法可以获得更好的效果。

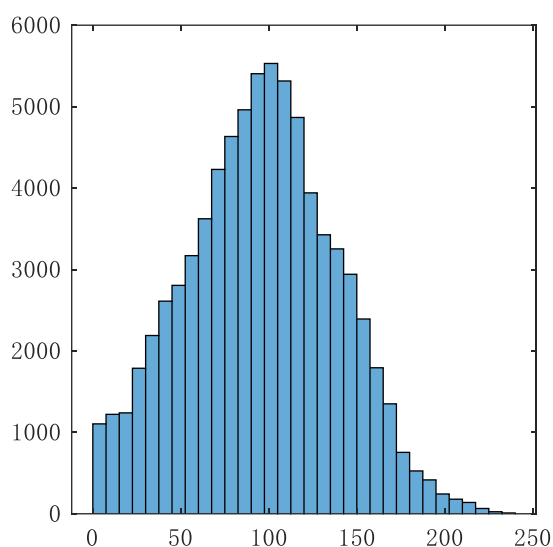


图 7.2 正常数据判别函数 $\phi$ 分布水平

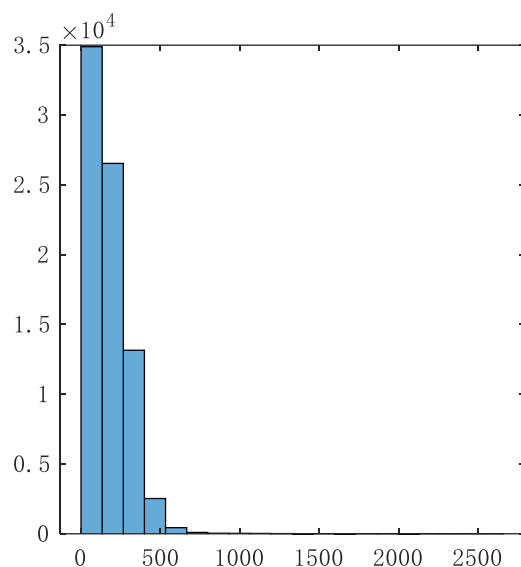


图 7.3 异常数据判别函数 $\phi$ 分布水平

表 7.1 统计分析

	ZC 数据	YC 数据
样本量	76141	77762
误差<250	76141	57011
占比	100%	73.31%

## 7.2 基于公共区域面积大小的判别方法

基于平面几何出发，三圆交于一点时可以完成平面定位，然而由于测距的扰动三圆往往难以汇成一点，可能会形成一类封闭区域，但是在不知道那一组测距信息出现显著误差的情况下，无法通过调整测距令三圆交于一点形成唯一解。但是在小幅噪声扰动情况下，和在大幅常值扰动情况下，封闭区域将表现不一致的性质，其中小幅扰动产生的封闭区域面积较小，而大幅常值扰动产生的封闭区域面积较大，因此可以猜想，能否通过封闭区域面积来区分异常情况数据和正常情况数据。对于平面几何，三圆关系可以分为彼此不相交、两圆相交并与另一圆、三圆相交等情况。异常数据是在三圆相交一点的基础上将某一数据加入了正偏执，因此三圆关系容易从三圆相交于一点转移至三圆形成封闭区域，噪声扰动可能形成小幅值的封闭区域或者未形成封闭区域，而常值绕回会形成大范围的封闭区域。因此正常数据和异常数据，在平面关系上来看应该处于显著差异的状态。从这一点出发进行猜想，空间中四球关系是否会得到类似的结论，能否通过分析四球相交空间体积的大小来识别正常数据和异常数据。

但是空间情况较为复杂且不直观，因此引入了一个大幅简化的想法，能否将测距直接应用到平面上，从平面几何出发讨论区域相交的情况。但是综合来看，这一类做法并不能获得较好的效果。对于第一组数据，正常数据和异常数据被显著分开，而对于其他数据则出现了各类复杂情况，因此这一类分类方法也是无效的。综合来看，仅仅基于测量距离出发的判别方法从统计和几何的角度并不能直接实现具体分类，这两类的方法是企图寻找一

项有效的单一指标对数据进行分类，但是结果表明需要进一步挖掘有效的信息或者使用学习类的方法来识别正常数据和异常数据之间的显著特征，从而实现分类。

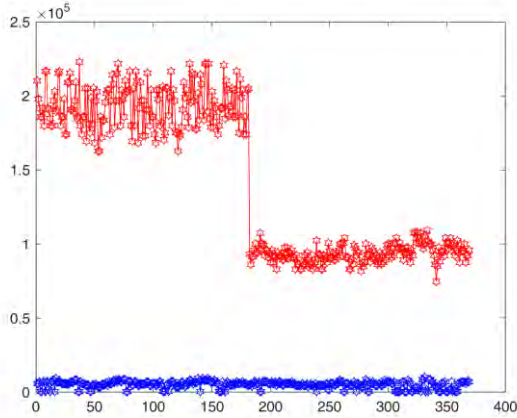


图 7.4 第 1 组数据-判据有效

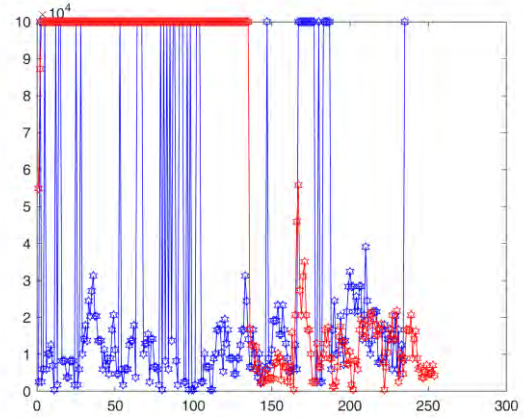


图 7.5 第 108 组数据—判据无效

### 7.3 神经网络分类法

由于本问题使用和任务一、任务二同样的场景，故任务一、任务二的数据皆可利用，且由于场景相同，故无需再考虑场地的变化以及锚点位置的变化，问题得到降维，只需要考虑对于不同靶标位置，四个锚点的测量距离中的特征，来判断四个测定的数据是否存在异常，本问题同样使用传统的启发式算法结合分类的方法，以及训练神经网络的办法来建立分类模型，从而判断数据是否存在异常。

神经网络的原理已经在 5.2.3 中介绍，在此不再重复介绍，对于“端对端的”模型，只需建立它的输入模型和输出模型即可，与 5.2.3 中的输入模型不同，此处的场景不会发生变化，无需考虑锚点的坐标变化，只需要将四个锚点距离测量值作为输入即可，即

对于输出模型，由于本任务的目标是判断锚点测量数据是否有异常，则输出只有两个状态：有异常和无异常，则将输出设置为：

$$sign = \begin{cases} 0, & \text{数据不存在异常} \\ 1, & \text{数据存在异常} \end{cases}$$

然后通过对神经网络模型进行训练，得到模型参数，来判断附件四中的数据是否存在异常。

### 7.4 问题四的求解

使用附件一和附件二数据的训练一个 3 层，节点分别为 40、30、20 的神经网络，然后使用附件一中 324 组样本点对样本点进行检验，得到的识别成功率：

$$rate_{success} = 97.69\%$$

对附件四的样本使用机器学习算法求解，得到的结果如表。其中仅 1、3、6、10 组数据为正常值，其余数据存在异常。

表 7.2 任务四数据是否异常判断表

组数	1	2	3	4	5	6	7	8	9	10
r0	2940	5240	4800	5010	2840	5010	5050	5050	4840	2740
r1	4290	5360	2610	4120	4490	5320	3740	4110	2600	2720
r2	2840	2040	4750	3810	2860	1990	3710	3710	4960	4670
r3	4190	2940	2550	2020	4190	2930	2070	2110	2700	4790
预测结果	0	1	0	1	1	0	1	1	1	0

## 8 问题五建模与分析：动态轨迹

### 8.1 分析与思路

问题五为问题一至问题四的综合运用。首先将“附件 5：动态轨迹”中每个位置下靶点到四个锚点的距离提取出来，然后通过问题 4 中神经网络的算法进行分类，判断数据是否存在异常干扰，再使用神经网络算法求解出对应的坐标，生成靶点的动态运动轨迹。另外给出了不判断是否存在干扰，直接对提取出的数据分别使用神经网络算法、粒子群算法解算坐标，并生成轨迹图，作为对结果的对比。

### 8.2 结果

采用 7.3 中训练得到的模型参数对提取到的数据进行异常性判断，再使用 5.2.3 中的神经网络求解方法完成坐标的计算，将结果绘制成图，得到结果如图 8.1。

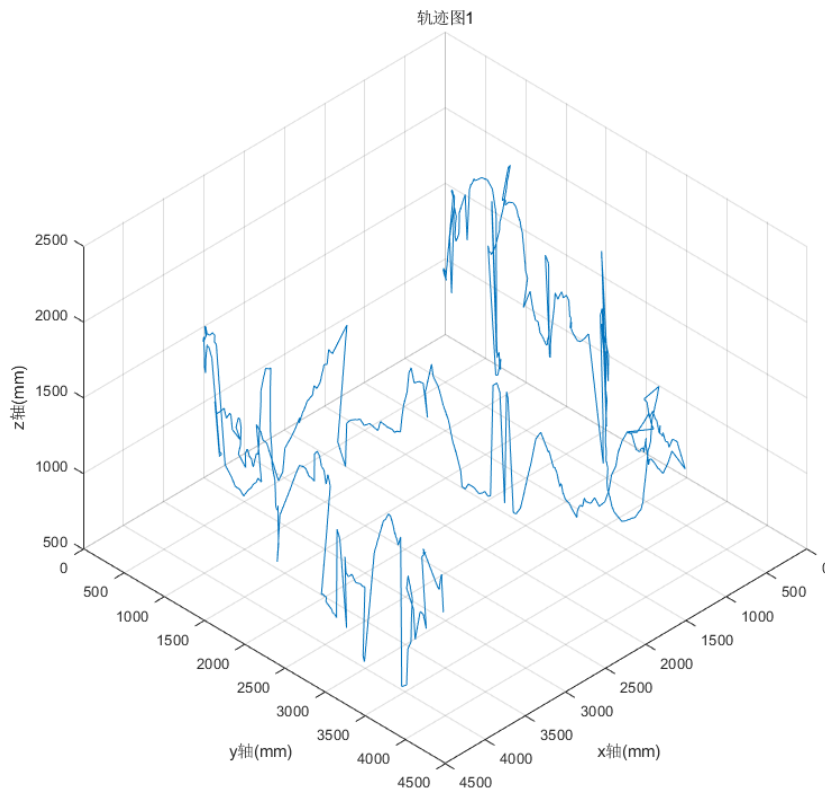


图 8.1 判断异常数据后神经网络算法求解结果

可以看到，受算法的精度影响，求解坐标在 Z 方向有较大的波动。图 8.2 为轨迹在 xOy 平面内投影的轨迹。

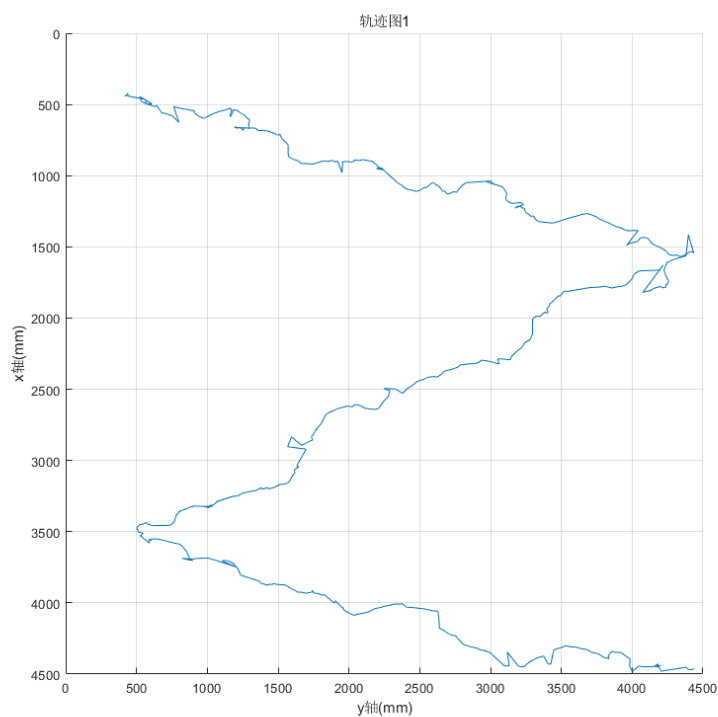


图 8.2 “图 8.1” 轨迹的 xOy 投影

图 8.3 为仅使用神经网络算法求解坐标后得到的结果，没有判断附件的数据是否异常，即将所有数据视为正常。图 8.4 为该求解轨迹在 xOy 平面内的投影。

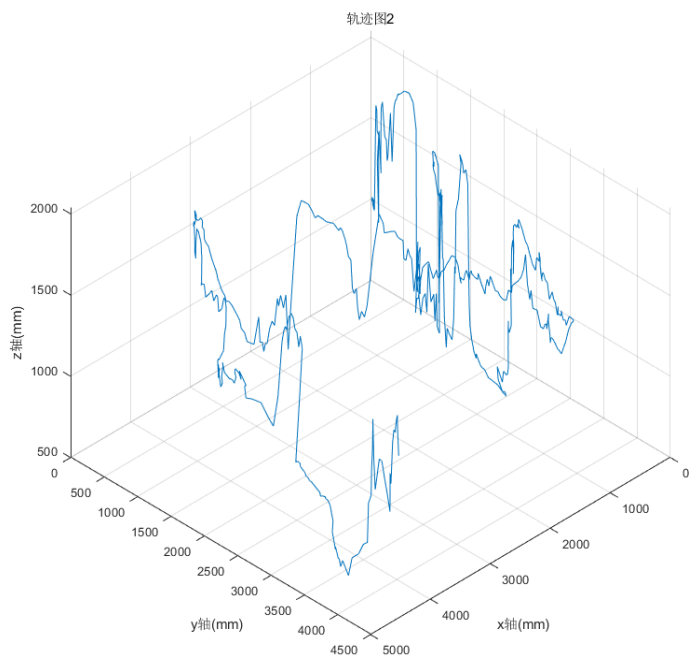


图 8.3 仅使用神经网络算法求解结果

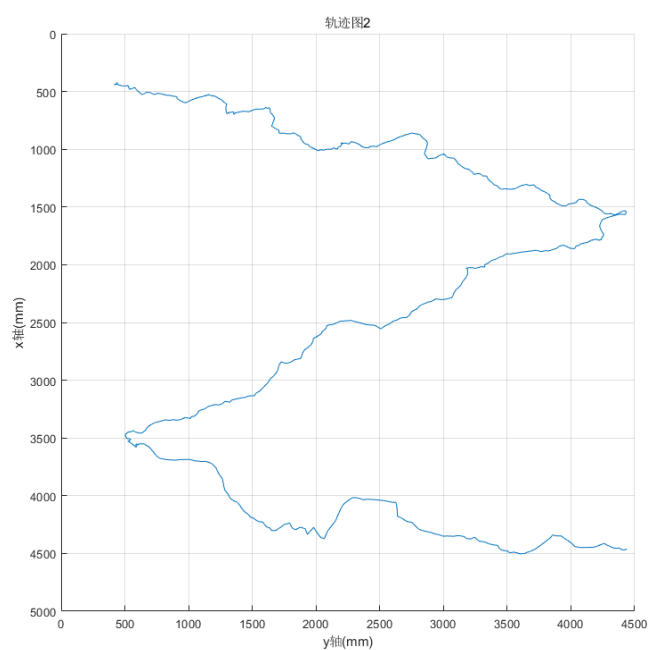


图 8.4 “图 8.3” 轨迹的 xOy 投影

图 8.5 为仅使用粒子群算法求解坐标后得到的结果，同样没有判断数据是否异常，即也将所有数据视为正常。图 8.6 为该求解轨迹在 xOy 平面内的投影。

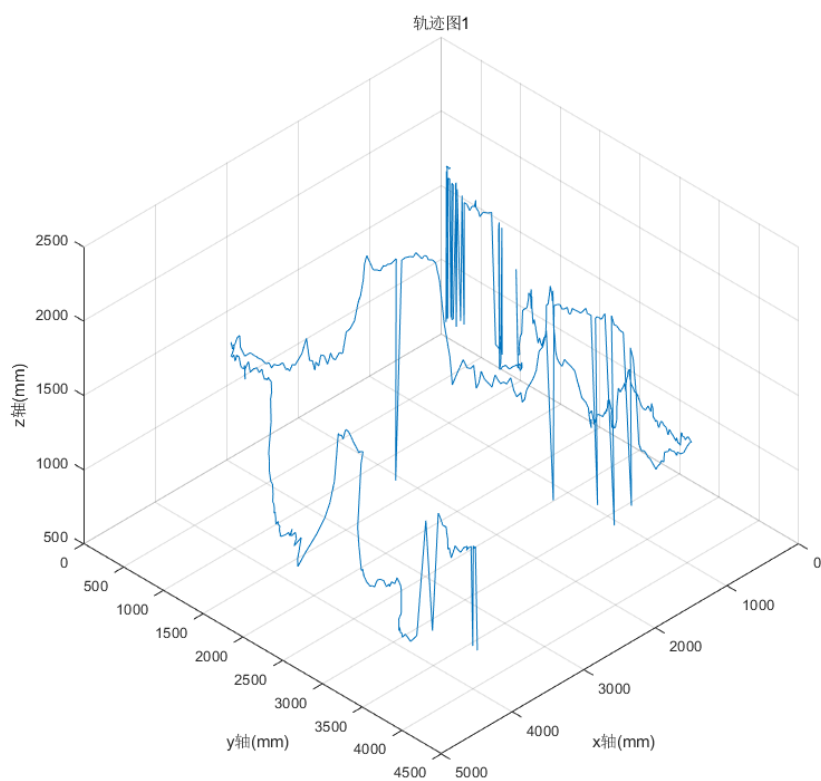


图 8.5 仅使用 PSO 算法求解结果

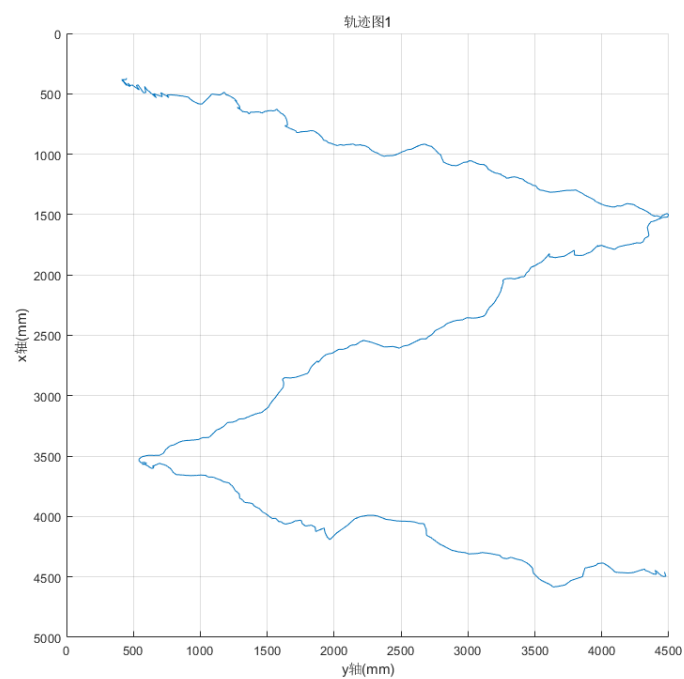


图 8.4 “图 8.5” 轨迹的 xOy 投影  
以上三组结果分别是附件中的“track1.fig”、“track2.fig”、“track3.fig”。

---

## 总结与展望

### 9.1 模型评价

本文完成了针对不同测量数据的定位算法与分类算法设计，提出了一类基于机器学习的定位与分类方法，并结合最小二乘法的传统方法以及基于 PSO 的四点算法完成了算法性能的对比，给出了算法精度的评价。同时应用机器学习方法结合几何法、统计分析方法对比验证了基于机器学习分类方法的有效性，为复杂信号干扰环境下高精度 UWB 定位算法与分类算法建立了一类有效的算法架构，并与传统方法进行了充分对比分析验证了所提出方法的有效性。

### 9.2 模型改进

针对问题五没有完成充分解答，采用了统计分析方法充分讨论了数据分布特性，但是再结合算法设计时与先验的统计信息结合不够紧密，应进一步结合可获取的先验信息改进算法，使得所提出的改进算法更加完备。



---

## 参考文献

- [1] <https://zhuanlan.zhihu.com/p/358944859>
- [2] 许佩佩, 基于TOA方案的超宽带室内定位技术研究, 东南大学, 15—16.
- [3] 蔡红, 基于 UWB 的定位方法研究, 北京邮电大学.