

基于随机森林的无源时差定位误差估计与误差修正算法

宋定宇¹, 张君毅^{1,2}

(1. 中国电子科技集团公司 第 54 研究所, 石家庄 050081;

2. 河北省电磁频谱认知与管控重点实验室, 石家庄 050081)

摘要: 针对无源时差定位算法在时差估计误差未知条件下, 无法通过计算几何稀释因子准确估计定位误差的问题, 提出基于随机森林的无源时差定位误差估计与误差修正算法, 通过学习信号特征参数和定位信息与定位误差之间的映射关系, 准确估计定位误差并对定位结果进行修正, 实现高精度定位; 通过匹配时差定位结果与参考源信息制作了测试数据集, 并使用该数据集验证了所提算法的有效性; 对各特征参数的重要性进行了定量分析, 并验证了随机森林模型在时间上的泛化能力; 实验结果表明, 所提算法实现了在时差估计误差未知条件下对时差定位误差的准确估计, 提高无源时差定位的精度, 具有较高的工程应用价值。

关键词: 无源定位; 时差定位; 随机森林; 误差估计; 误差修正

Estimation and Correction Algorithm for Passive TDOA Localization Based on Random Forest

SONG Dingyu¹, ZHANG Junyi^{1,2}

(1. The 54th Research Institute of China Electronics Technology Group Corporation, Shijiazhuang 050081, China;

2. Hebei Province Key Laboratory of Electromagnetic Spectrum Cognition and Control, Shijiazhuang 050081, China)

Abstract: Under the condition of unknown TDOA estimation error, passive time difference of arrival (TDOA) positioning algorithm can not accurately estimate positioning error by calculating geometric dilution of precision (GDOP), this paper proposes a TDOA positioning error estimation and correction algorithm based on random forest. By learning the mapping relationship between signal feature parameters, positioning information, and positioning error, it accurately estimates the positioning error and corrects the positioning result to achieve high-precision positioning. A test dataset was created by matching the TDOA positioning result with reference source information, and the effectiveness of the proposed algorithm is verified by this dataset. The importance of each feature parameter is quantitatively analyzed, and the temporal generalization ability of the random forest model is validated. Experimental results show that the proposed algorithm achieves accurate estimation of TDOA positioning error under the condition of unknown TDOA estimation error, improves the accuracy of passive TDOA positioning, and has high engineering application value.

Keywords: passive localization; TDOA; random forest; error estimation; error correction

0 引言

无源定位技术是一种探测器无需向定位目标发射电磁信号, 仅通过截获信号对目标辐射源进行定位的技术。与有源定位相比, 无源定位具有探测距离远、隐蔽性强、抗电磁干扰能力强等优点, 在电子侦查领域发挥着重要作用。根据定位所需参数, 无源定位可以分为:

基于到达时差的定位 (TDOA, time difference of arrival)^[1]、基于到达频差的定位 (FDOA, frequency difference of arrival)^[2]、基于信号到达角方向的定位方法 (DOA, direction of arrival)^[3] 以及基于接收信号强度的定位 (RSS, received signal strength)^[4] 等。其中 TDOA 是指通过辐射源信号到达各个观测站的时差进行定位的方法。该方法具有观测站组网便捷、定位原理

收稿日期: 2024-08-28; 修回日期: 2024-10-08。

基金项目: 国家自然科学基金(U20B2071)。

作者简介: 宋定宇(2000-), 男, 硕士研究生。

引用格式: 宋定宇, 张君毅. 基于随机森林的无源时差定位误差估计与误差修正算法[J]. 计算机测量与控制, 2025, 33(1): 269-275, 284.

简单易实现、定位精度高、成本低、实时性强等优点^[5]，是无源定位的重要方法。

随着无源时差定位技术的发展，如何提高时差定位的定位精度成为研究热点。文献 [6-9] 通过优化部站方式提高了定位精度，但部站时需要设置较长的基线长度。由于无源时差定位需要多个观测站同时截获辐射源信号才能完成定位，基线长度过长会降低多个观测站同时截获信号的概率，部站时基线长度受到限制^[10]。因此，通过优化部站方式提高定位精度存在困难。

在时差估计误差标准差和站址误差标准差已知条件下，可以通过计算几何稀释因子 (GDOP, geometric dilution of precision) 估计时差定位的定位精度，但在实际定位中，时差估计误差标准差和站址误差标准差往往难以获得^[13]。因此，难以通过 GDOP 方法准确估计并修正定位误差。

在有源时差定位研究中，Ashwin 等人提出了一种基于神经网络模型的全球导航卫星系统 (GNSS, global navigation satellite system) 定位误差修正方法，通过训练模型学习伪距残差和视距无线传输 (LOS, line-of-sight) 向量与定位误差之间的关系，对初始定位进行修正，降低了 GNSS 定位误差^[11]。Adyasha 等人提出使用图卷积神经网络 (GCNN, graph convolution neural network) 学习具有多星座和多频率信号的不同图结构，降低了 GNSS 对智能手机的定位误差^[12]。以上两种方法适用于导航定位，而在无源时差定位中，由于观测站仅通过截获信号对目标辐射源进行定位，且观测站数量较少，不具备导航定位中充足的定位信息，因此这两种方法不适用于无源时差定位。

目前，针对无源时差定位的误差修正方法研究较少，为解决无源时差定位误差估计和误差修正困难的问题，提出一种基于随机森林模型的误差估计和误差修正算法，并制作数据集验证算法有效性。使用随机森林模型学习信号特征参数和定位信息与定位误差之间的映射关系，将误差估计转化为回归拟合问题，通过寻找最优拟合，准确估计定位误差，从而对定位误差进行修正。

针对目标辐射源真实位置难以获取的问题，使用无源时差定位数据和参考源制作数据集，并使用数据集验证无源时差定位误差估计与修正算法的有效性。无源时差定位误差估计与修正示意图如图 1 所示，观测站截获目标辐射源信号后，通过两步最小二乘法 (TSWLS, two-step weighted least squares) 获得目标辐射源的初始定位结果，通过计算定位结果和对应参考源位置的距离和方位角获得真实的定位误差标签。训练随机森林模型估计定位误差，并使用获得的误差估计值修正定位结果。

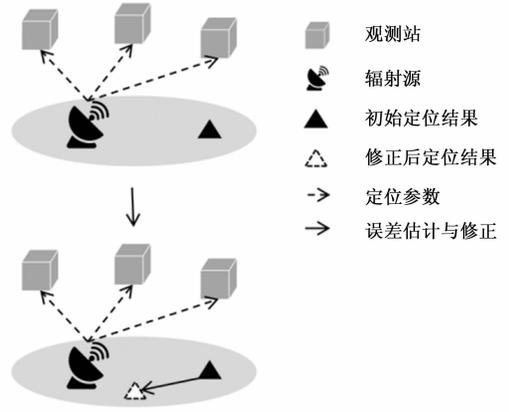


图 1 误差估计与修正示意图

1 时差定位与 GDOP 原理

1.1 三站时差定位

如图 2 所示，三站无源时差定位系统由一个主观测站和两个辅观测站构成。地固坐标系中，主观测站位置为 $S_0(x_0, y_0, z_0)$ ，辅观测站 1 位置为 $S_1(x_1, y_1, z_1)$ ，辅观测站 2 位置为 $S_2(x_2, y_2, z_2)$ ，目标辐射源位置为 $T(x, y, z)$ 。

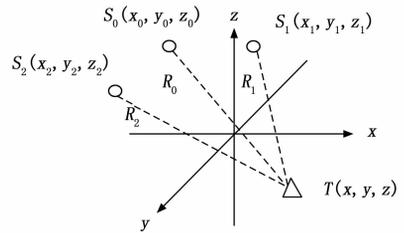


图 2 三站时差定位

设目标辐射源发射的脉冲到达主观测站时间为 t_{TOA0} ，到达辅观测站 1 时间为 t_{TOA1} ，到达辅观测站 2 时间为 t_{TOA2} ，则脉冲到达主观测站与辅观测站 1 的时间差为：

$$t_{TDOA1} = t_{TOA0} - t_{TOA1} \quad (1)$$

脉冲到达主观测站与辅观测站 2 的时间差为：

$$t_{TDOA2} = t_{TOA0} - t_{TOA2} \quad (2)$$

设目标辐射源与各观测站距离为 R_i ，则：

$$R_i = \sqrt{(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2} \quad (i = 0, 1, 2) \quad (3)$$

时差方程为：

$$R_0 - R_1 = c \cdot t_{TDOA1} \quad (4)$$

$$R_0 - R_2 = c \cdot t_{TDOA2} \quad (5)$$

其中： c 为电磁波传播速度。

WGS-84 地球椭球模型下，目标辐射源位置为：

$$\begin{cases} x = (N+H) \cos B \cos L \\ y = (N+H) \cos B \sin L \\ z = (N(1-e^2) + H) \sin B \end{cases} \quad (6)$$

其中: L 和 B 分别为目标辐射源的经度和纬度, H 为目标辐射源的高程, N 为目标辐射源位置的卯酉圈半径, e 为第一偏心率。

假设地球表面辐射源目标高程为 0, 则辐射源目标所在椭球方程为:

$$\frac{x^2}{a^2} + \frac{y^2}{a^2} + \frac{z^2}{a^2(1-e^2)} = 1 \quad (7)$$

其中: a 为地球长半轴, 通过 (1) 至 (5) 式和 (7) 式可以解算出目标辐射源位置。

1.2 GDOP 原理

设目标辐射源定位误差为 dx, dy, dz , 各观测站站址误差以及误差分量 dx_i, dy_i, dz_i 之间不相关, 根据误差传递原理, 对 $\Delta R_i = R_0 - R_i, i=1, 2$ 等号两边求微分并化简得:

$$d\Delta R_i = (C_{ix} - C_{0x})dx + (C_{iy} - C_{0y})dy + (C_{iz} - C_{0z})dz + (k_i - k_0) \quad (8)$$

其中:

$$\begin{cases} C_{ix} = \frac{\partial R_i}{\partial x} = \frac{x - x_i}{R_i} \\ C_{iy} = \frac{\partial R_i}{\partial y} = \frac{y - y_i}{R_i} \\ C_{iz} = \frac{\partial R_i}{\partial z} = \frac{z - z_i}{R_i} \\ k_i = C_{ix}dx_i + C_{iy}dy_i + C_{iz}dz_i \end{cases} \quad (i = 0, 1, 2) \quad (9)$$

其矢量形式为:

$$C \cdot dR = d\Delta R - dS \quad (10)$$

$$dR = \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix}, \text{ 为所求目标辐射源定位误差。}$$

$$d\Delta R = \begin{bmatrix} d\Delta R_1 \\ d\Delta R_2 \end{bmatrix}, \text{ 为 TDOA 估计引入误差。}$$

$$dS = \begin{bmatrix} \frac{x-x_0}{R_0} - \frac{x-x_1}{R_1} & \frac{y-y_0}{R_0} - \frac{y-y_1}{R_1} & \frac{z-z_0}{R_0} - \frac{z-z_1}{R_1} \\ \frac{x-x_0}{R_0} - \frac{x-x_2}{R_2} & \frac{y-y_0}{R_0} - \frac{y-y_2}{R_2} & \frac{z-z_0}{R_0} - \frac{z-z_2}{R_2} \end{bmatrix},$$

表示目标辐射源相对各站的方向余弦, 由目标辐射源和各站的相对位置决定, 与布站方式有关。

由伪逆法得到定位误差估计值:

$$dR = (C^T C)^{-1} C^T (d\Delta R - dS) \quad (11)$$

设误差服从均值为零的正态分布, 则定位误差协方差矩阵可表示如下:

$$P_{dR} = E[dR \cdot dR^T] = B\{E[d\Delta R \cdot d\Delta R^T] + E[dS \cdot dS^T]\} B^T \quad (12)$$

$$B = (C^T C)^{-1} C^T \quad (13)$$

$$E[d\Delta R \cdot d\Delta R^T] = \begin{bmatrix} \sigma_{\Delta R_1}^2 & \eta_{12} \sigma_{\Delta R_1, \Delta R_2}^2 \\ \eta_{12} \sigma_{\Delta R_1, \Delta R_2}^2 & \sigma_{\Delta R_2}^2 \end{bmatrix} \quad (14)$$

其中: $\sigma_{\Delta R_i}$ 为误差 ΔR_i 的标准差, η_{12} 为 ΔR_1 与 ΔR_2

的相关系数:

$$\eta_{12} = \frac{\text{cov}(\Delta R_1, \Delta R_2)}{\sigma_{\Delta R_1} \sigma_{\Delta R_2}} \quad (15)$$

设站址误差矢量的各分量标准差大小相同, 即:

$$\sigma_{x_i}^2 = \sigma_{y_i}^2 = \sigma_{z_i}^2 = \sigma_s^2 \quad (i = 0, 1, 2) \quad (16)$$

因为:

$$C_{ix}^2 + C_{iy}^2 + C_{iz}^2 = 1 \quad (i = 0, 1, 2) \quad (17)$$

代入上两式得:

$$E[dS \cdot dS^T] = \begin{bmatrix} 2\sigma_s^2 & \sigma_s^2 \\ \sigma_s^2 & 2\sigma_s^2 \end{bmatrix} \quad (18)$$

则:

$$P_{dR} = B\{E[d\Delta R \cdot d\Delta R^T] + E[dS \cdot dS^T]\} B^T = B \begin{bmatrix} \sigma_{\Delta R_1}^2 + 2\sigma_s^2 & \eta_{12} \sigma_{\Delta R_1, \Delta R_2}^2 + \sigma_s^2 \\ \eta_{12} \sigma_{\Delta R_1, \Delta R_2}^2 + \sigma_s^2 & \sigma_{\Delta R_2}^2 + 2\sigma_s^2 \end{bmatrix} B^T \quad (19)$$

定位的几何稀释因子为:

$$\text{GDOP} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} = \sqrt{\text{trace}(P_{dR})} \quad (20)$$

式中, $\text{trace}(P_{dR})$ 表示矩阵 P_{dR} 的迹, 即矩阵 P_{dR} 主对角线上所有元素之和。由推导可知, 定位误差与时差测量误差、观测站站址误差、目标辐射源与各观测站的几何位置相关^[13]。

2 基于随机森林的定位误差估计和修正

针对 GDOP 方法在时差估计误差标准差和站址误差标准差未知条件下, 无法准确估计定位误差的问题, 提出使用随机森林模型, 通过学习信号特征参数和定位信息与定位误差之间的映射关系, 拟合无源时差定位系统的误差模型, 准确估计定位误差。

2.1 随机森林模型

随机森林 (RF, random forest)^[14] 是一种基于 bagging 方法集成多棵决策树完成分类或回归任务的集成学习方法。随机森林模型由于在集成多棵决策树预测结果的同时引入了随机性, 所以具有较高的预测准确度和较强的抗过拟合能力。因此, 选择随机森林回归模型估计时差定位误差, 并基于误差估计结果进行定位误差修正。模型结构如图 3 所示。

2.1.1 Bagging

Bagging (自助聚合) 的基本原理是通过有放回的抽样方法从原始数据集中生成多个子数据集, 每个子数据集用于训练一个基模型。这些基模型在训练过程中是相互独立的, 最终通过投票或平均的方式将它们的预测结果结合起来, 形成最终的预测^[15]。Bagging 的主要优点包括减少模型的方差, 从而提高模型的泛化能力, 防止过拟合, 并提高模型的鲁棒性。

2.1.2 决策树

决策树通过选择最佳特征和相应的阈值将数据递归地分割成不同的分支, 直到达到决策的终点^[16]。决策

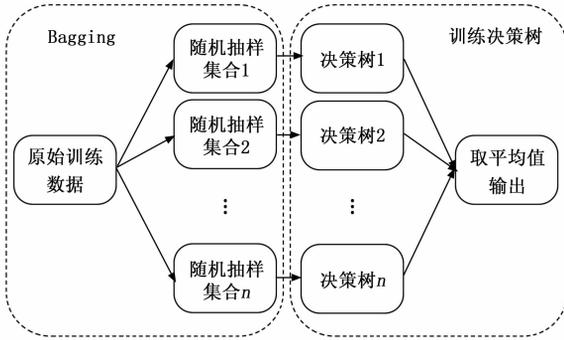


图 3 随机森林回归模型图

树具有计算效率高、无需数据预处理等优点，但存在易于过拟合、对噪声数据敏感等问题^[17]。

随机森林模型通过随机抽取决定每棵决策树的样本集合；决策树在每个节点分裂时，从所有特征中随机选择一部分特征，从部分特征中选择最佳特征进行分裂。随机森林模型通过样本随机采样和特征随机选择引入随机性，克服了易于过拟合、对噪声数据敏感等问题^[18]，可以获得更高的预测精度和更强的泛化能力。

2.2 基于随机森林的定位误差估计和修正

将定位误差估计视为回归问题，训练随机森林回归模型估计定位误差。

对于一棵回归树，设输入参数为 x ，输出目标为 d ，给定训练数据集：

$$D = \{(x_1, d_1), (x_2, d_2), \dots, (x_N, d_N)\} \quad (21)$$

将输入空间划分为 M 个单元 R_1, R_2, \dots, R_m ，回归树模型可表示为：

$$f(x) = \sum_{m=1}^M c_m I(x \in R_m) \quad (22)$$

其中： c_m 为每个单元 R_m 上的固定的输出值。

使用均方误差 $MSE = \frac{1}{N} \sum_{i=1}^N (d_i - \hat{d}_i)^2$ 表示回归树预测误差，其中， d_i 代表样本的真实值， \hat{d}_i 代表样本的预测值， N 为样本数。使用最小均方误差准则 (MMSE, minimum mean-square error) 求解每个单元上的最优输出值。易知，单元 R_m 上的 c_m 的最优值 \hat{c}_m 是 R_m 上所有输入实例 x_i 对于的输出 d_i 的均值，即：

$$\hat{c}_m = ave(d_i | x_i \in R_m) \quad (23)$$

采用启发式对输入空间进行划分，选择第 j 个变量 x_j 和它的取值 s ，作为切分变量和切分点，并定义两个区域：

$$R_1(j, s) = \{x | x_j \leq s\} \quad (24)$$

$$R_2(j, s) = \{x | x_j > s\} \quad (25)$$

然后寻找最优切分变量 j 和最优切分点 s 。即求解：

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (d_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (d_i - c_2)^2 \right] \quad (26)$$

对固定输入变量 j 可以找到最优切分点 s ：

$$\hat{c}_1 = ave(d_i | x_i \in R_1(j, s)) \quad (27)$$

$$\hat{c}_2 = ave(d_i | x_i \in R_2(j, s)) \quad (28)$$

遍历所有输入变量，找到最优切分变量 j ，构成一个对 (j, s) ，依次将输入空间划分为两个区域。接着，每个对每个区域重复上述划分过程，直到满足停止条件，得到一棵回归树。

对于回归森林模型，模型输出的预测值可以表示为：

$$\hat{d} = \frac{1}{K} \sum_{i=1}^K f_i(x) \quad (29)$$

其中： K 为随机森林模型中回归决策树棵数， $f_i(x)$ 为第 i 棵决策树模型，每棵决策树的训练集为随机森林模型训练集随机抽取的真子集，特征为随机森林模型特征随机抽取的真子集。

由已知定位结果和估计得到的定位误差估计值 \hat{d} 和方位角 α 可以计算修正后的定位结果：

$$\theta_c = \theta_2 + \frac{\hat{d} \cdot \cos\alpha}{R \frac{2\pi}{360}} \quad (30)$$

$$\phi_c = \phi_2 + \frac{\hat{d} \cdot \sin\alpha}{R \cos(\theta_c) \frac{2\pi}{360}} \quad (31)$$

其中： θ_2 和 ϕ_2 为定位结果的纬度和经度， θ_c 和 ϕ_c 为修正后的纬度和经度， R 为地球平均半径， $R = 6371.393 \text{ km}$ 。

综上所述，基于随机森林的定位误差估计和修正算法流程如下：

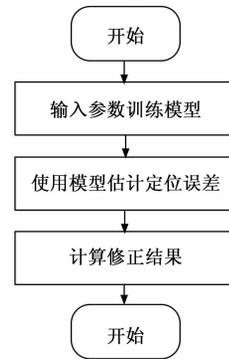


图 4 定位误差估计和修正算法流程图

1) 使用训练集定位参数和定位误差标签训练随机森林模型估计定位误差。

2) 将测试集定位参数输入训练好的随机森林模型，模型输出定位误差估计值 \hat{d} 。

3) 由初始定位结果的纬度 θ_2 和经度 ϕ_2 、模型输出的定位误差估计值 \hat{d} 以及方位角 α 计算修正后的定位结果。

3 实验与分析

为验证随机森林模型估计定位误差和修正定位误差的有效性, 建立无源时差定位数据集并设计以下实验进行验证。实验使用的操作系统为 Windows10, 内存为 64 GB, GPU 为 NVIDIA GeForce RTX3060。使用 python 库中的 scikit-learn 库实现随机森林回归, 随机森林回归模型参数设置如表 1 所示。

表 1 模型参数设置

模型参数名称	参数设置
n_estimators	192
max_depth	None
Bootstrap	Ture
max_features	Log2
Obb_score	Ture

其中: n_estimators 为随机森林模型中树的数量, 增加数量通常会提高模型的性能, 但同时也会增加计算成本。max_depth 为每棵树的深度, 限制深度可以防止过拟合, 但太浅可能导致欠拟合。Bootstrap 为是否对样本进行随机抽样。max_features 为寻找最佳分割时要考虑的特征数量, 适当选择可以提高模型的多样性。obb_score 为随机抽样条件下, 是否使用袋外数据作为验证集。各参数最优值通过网格搜索获得。

3.1 数据集

针对无源时差定位, 使用无源时差定位数据和参考源建立数据集。数据集包含 121 个参考源坐标和 191 574 条对应的定位结果。参考源的分布跨东西 5 897.1 km, 南北 4 842.2 km, 跨度较大, 具有普遍意义。参考源包括导航雷达、机场雷达等辐射源。定位结果由无源观测站截获的参考源信号信息, 经两步最小二乘法解算获得。

数据集建立步骤为: 由大量无源时差定位系统截获的信号, 经两步最小二乘法解算获得大量定位结果。将大量定位结果与参考源进行初步匹配, 匹配规则为参考源与定位结果距离小于 10 km。根据参考源的频率、脉冲宽度、脉冲重复周期等信息和定位信息对初步匹配结果进行数据清洗, 最终得到包含 121 个参考源和 191 574 条对应的定位结果的数据集。

由于定位结果由两步最小二乘法解算获得, 模型学习到的误差模型同时包含无源时差定位系统的误差模型和定位算法的误差模型。

定位误差计算公式如下:

$$Error = R \cdot \arccos(\sin\theta_1 \sin\theta_2 + \cos\theta_1 \cos\theta_2 \cos(\phi_2 - \phi_1)) \quad (32)$$

其中: θ_1 和 ϕ_1 为参考源位置的纬度和经度, θ_2 和 ϕ_2 为定位结果的纬度和经度, R 为地球平均半径, $R =$

6 371.393 km。

数据集中每条定位数据包含的特征参数如表 2 所示。

表 2 数据集特征参数

特征参数名称	参数的含义	数据类型
Sensor_x	主观测站在地心地固坐标系下的 x 坐标	Double
Sensor_y	主观测站在地心地固坐标系下的 y 坐标	Double
Sensor_z	主观测站在地心地固坐标系下的 z 坐标	Double
Longitude	定位结果的经度	Double
Latitude	定位结果的纬度	Double
Freq_mean	用于定位的信号的频率的平均值	Double
PW_mean	用于定位的信号的脉冲宽度的平均值	Double
PRI_mean	用于定位的信号的脉冲重复周期的平均值	int
PA	用于定位的信号的脉冲幅度的平均值	int
Tri	1 秒内 3 个观测站同时接收到该辐射源信号的次数	int
Dis	主观测站在地球表面投影点到定位结果的距离	Double
Az	定位结果到主观测站在地球表面投影点的方位角	Double
Error	定位误差绝对值	Double

其中: Sensor_x 为主观测站在地心地固坐标系下的 x 坐标, Sensor_y 为主观测站在地心地固坐标系下的 y 坐标, Sensor_z 为主观测站在地心地固坐标系下的 z 坐标, Longitude 和 Latitude 为初始定位结果的经纬度, 初始定位结果由 1 秒内单个辐射源被 3 个观测站同时截获的信号经两步最小二乘法解算获得, Tri 为 1 秒内单个辐射源被 3 个观测站同时截获的次数, Freq_mean 为 1 秒内被用于定位信号的频率平均值, PW_mean 为 1 秒内被用于定位信号的脉冲宽度平均值, PRI_mean 为 1 秒内被用于定位信号的重复周期平均值, PA 为 1 秒内被用于定位信号的脉冲幅度平均值, Dis 为主观测站在地球表面投影点到定位结果的距离, Az 为定位结果到主观测站在地球表面投影点的方位角, Error 为真实定位误差标签。

地心地固坐标系 (ECEF, earth-centered, earth-fixed) 是一种三维直角坐标系, 原点位于地球质心, Z 轴与地球自转轴重合指向北极, X 轴指向赤道平面和本初子午线的交点, Y 轴垂直于 XZ 平面, 与 X 轴和 Z 轴构成右手坐标系。这种坐标系固定在地球上, 随地球

一起旋转，因此也称为固定坐标系。在地心地固坐标系中，地球上任何一个点的位置可以通过 3 个坐标值 (X, Y, Z) 来确定。地心地固坐标系与地理坐标系 (经纬度) 之间可以相互转换。数据集中观测站坐标采用 WGS84 (World Geodetic System 1984) 坐标系。经纬高坐标系 (LLA, longitude, latitude, altitude)，其中，Altitude 表示地球表面上某一点相对于参考椭球体或大地水准面的垂直距离。数据集中，辐射源定位结果坐标采用经纬高坐标系。地心地固坐标系和经纬高坐标系可以相互转换，在数据集中，主观测站坐标和定位结果坐标仅通过一种坐标系表示。

将数据划分，80% 作为训练集，20% 作为测试集。由于随机森林模型可以使用“袋外数据”进行验证，故不需要设置验证集^[19]。

3.2 实验结果

3.2.1 评价指标

为评价所提方法误差估计和误差修正的效果，选取平均绝对误差 (MAE, mean absolute error)、均方根误差 (RMSE, root mean square error)、决定系数 (Coefficient of determination, R squared) 以及圆概率误差 (CEP, circular error probability) 4 个参数作为评价指标，MAE 和 RMSE 越小，误差估计越准确； R^2 反映的是模型拟合的准确程度，取值范围为 [0, 1]， R^2 越接近 1 表示模型对数据的拟合程度越好，CEP 越小，定位误差越小。模型评价指标的具体计算公式如下所示：

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (33)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (34)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (35)$$

$$CEP = k \times \sigma \quad (36)$$

其中： y_i 代表样本的真实值， \hat{y}_i 代表样本的预测值， \bar{y} 代表真实值的平均值， N 代表样本个数， k 表示标准正态分布表格中对应概率系数， σ 表示误差的标准差。 P 取 50%，即定位的 CEP 值为定位结果以 50% 概率落在圆内时圆的半径值。

3.2.2 误差估计结果

由于时差估计误差标准差和站址误差未知，无法通过计算 GDOP 估计定位误差。为验证随机森林模型估计时差定位误差的有效性，将随机森林回归模型与 XGBoost (XGB) 模型^[20]、神经网络模型 (ANN, artificial neural network) 和支持向量回归 (SVR, support vector regression) 模型^[21] 进行对比。各算法模型的误差估计精度如表 3 所示。实验结果表明，随机森林模型具有最高的误差估计精度，误差估计的均方根误差为

0.609 0 km，优于极端梯度提升模型的 0.891 0 km、人工神经网络的 0.955 5 km 和支持向量回归的 1.223 5 km。同时，随机森林模型在测试集上的 R^2 系数为 0.902，说明模型拟合效果较好。

表 3 各算法模型的误差估计精度

模型名称	均方根误差/km
RF	0.609 0
XGB	0.891 0
ANN	0.955 5
SVR	1.223 5

3.2.3 误差修正结果

为验证随机森林模型修正时差定位误差的有效性，将使用随机森林模型、神经网络模型和 XGBoost 模型估计误差和修正误差的结果和使用两步最小二乘法^[22] 定位算法获得的未经修正的定位误差进行对比。不同模型修正定位误差的结果如表 4 所示。实验结果表明，随机森林模型效果最佳，以平均绝对误差统计，随机森林模型将误差由 1.742 2 km 修正为 0.452 1 km，优于人工神经网络的 0.751 5 km 和极端梯度提升的 0.660 6 km。以圆概率误差统计，随机森林模型将误差由 0.954 7 km 修正为 0.149 6 km，优于人工神经网络的 0.426 2 km 和极端梯度提升的 0.302 4 km。随机森林模型能够较好的估计并修正误差，以圆概率误差统计，定位误差被修正了 84.33%。

表 4 不同模型修正定位误差的结果

模型名称	平均绝对误差/km	圆概率误差/km
未修正	1.742 2	0.954 7
ANN	0.751 5	0.426 2
XGB	0.660 6	0.302 4
RF	0.452 1	0.149 6

3.3 模型在不同时间的泛化能力

为验证随机森林模型在不同时间的泛化能力，使用 2019 年数据中的测试集和 2021 年数据进行测试，测试结果如表 5 所示。以平均绝对误差统计，测试集 (与训练集时间相同) 将定位误差由 1.742 2 km 修正至 0.452 1 km，2021 年的数据 (2 年之后的数据) 将定位误差由 2.010 4 km 修正至 1.351 5 km，误差修正率由 74.05% 下降到 32.77%。以圆概率误差统计，测试集 (与训练集时间相同) 将定位误差由 0.954 7 km 修正至 0.149 8 km，2021 年的数据 (2 年之后的数据) 将定位误差由 1.151 7 km 修正至 0.307 7 km，误差修正率由 84.33% 下降到 73.28%。说明模型具有一定时间迁移能力，但存在随时间迁移退化的情况。误差修正率 C 的计算公式为：

$$C = \frac{E_0 - E_1}{E_0} \quad (37)$$

其中: E_0 是未修正的误差, E_1 是修正后的误差。

根据工程经验, 无源时差定位系统因时钟同步误差随时间增加等原因, 存在定位误差随时间增加的情况。说明无源时差定位系统的误差模型是随时间推移轻微变化的。在后续工作中, 可以通过增量学习的方法, 使用不同时间的数据, 增量训练模型适应无源时差定位系统随时间推移的轻微变化, 从而保持较高的误差估计和修正精度。

表 5 不同时间数据测试结果

数据时间	是否修正	平均绝对误差/km	圆概率误差/km
2019	未修正	1.742 2	0.954 7
	RF	0.452 1	0.149 6
2021	未修正	2.010 4	1.151 7
	RF	1.351 5	0.307 7

3.4 特征重要性分析

通过训练模型得到输入特征参数的特征重要性如图 5 所示。对定位误差影像最大的特征参数为主观测站在地球表面投影点到定位结果的距离, 主观测站在地球表面投影点到定位结果的距离越远, 定位误差越大。除此之外, 定位误差与定位结果的经纬度、截获的目标辐射源信号的频率、脉宽等特征相关性较强。定位结果的纬度越高, 观测站与目标辐射源位置的几何构型越差, 定位误差越大。截获的目标辐射源信号的频率越高、脉宽越窄, 信号到达时间的估计误差越大, 定位结果误差越大。模型学习到的特征重要性与工程经验一致, 说明模型学习到了正确的无源时差定位误差模型。

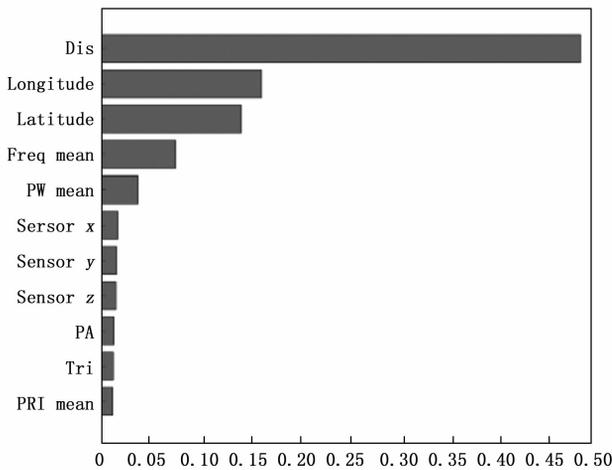


图 5 特征重要性

图 5 各特征参数的含义见表 2 数据集特征参数。

4 结束语

针对三站时差定位中, 定位精度需要提高, 且

GDOP 方法在时差测量误差标准差和站址误差标准差未知条件下无法准确估计定位误差的问题, 提出将误差估计视为回归问题, 基于数据驱动, 使用随机森林模型估计误差, 并基于模型估计获得的误差和初始定位结果对误差进行修正。实验表明, 随机森林模型可以较为准确的估计误差; 以圆概率误差统计, 相较于初始定位结果, 定位误差被修正了 84.33%。并验证了随机森林模型在时间上的泛化能力。在后续工作中, 可以通过增量学习的方法, 使用不同时间的数据, 增量训练模型适应无源时差定位系统随时间推移的轻微变化, 从而保持较高的误差估计和修正精度。

参考文献:

- [1] 杨林, 周一宇, 孙仲康. TDOA 被动定位方法及精度分析 [J]. 国防科技大学学报, 1998 (2): 52-56.
- [2] 刘洋, 杨乐, 郭福成, 等. 基于定位误差修正的运动目标 TDOA/FDOA 无源定位方法 [J]. 航空学报, 2015, 36 (5): 1617-1626.
- [3] 杨萃元, 郑思海. 基于运动辐射体 TOA 和 DOA 测量的单站被动定位算法 [J]. 电子学报, 1996 (12): 66-69.
- [4] 刘栩楠, 石荣, 何彬彬, 等. 基于 RSS 信息的测向交叉虚假定位点排除方法 [J]. 电子信息对抗技术, 2020, 35 (4): 28-33.
- [5] 李文臣, 赵会宁, 李宏, 等. 时差定位系统仿真与评估 [M]. 西安: 西安电子科技大学出版社, 2022.
- [6] LEI Y P, GONG F X, MA Y Q. Optimal distribution for four-station TDOA location system [C] // Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on, IEEE, 2010, 7: 2858-2862.
- [7] BISHOP A N, FIDAN B, ANDERSON B D O, et al. Optimality analysis of sensor-target geometries in passive localization: Part 1-Bearing-only localization [C] // 2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information, IEEE, 2007: 7-12.
- [8] 汪波, 薛磊. 基于遗传算法的 TDOA 定位系统的最优布站算法 [J]. 系统工程与电子技术, 2010, 31 (9): 2125-2128.
- [9] 赵家凯. 基于群智能算法的无源时差定位最优布站研究 [D]. 西安: 西安电子科技大学, 2022.
- [10] 安文琦. 基于到达时差测量的短基线测向优化布站方法 [D]. 西安: 西安电子科技大学, 2022.
- [11] KANHERE, ASHWIN V, et al. Improving gnss positioning using neural-network-based corrections [J]. NAVIGATION: Journal of the Institute of Navigation 69. 4 (2022).

(下转第 284 页)