

基于 REINFORCE 算法的组网测向选站方法

朱晨晨, 张 静

(信息工程大学, 河南 郑州 450001)

摘要: 针对传统组网测向方法应对高速移动、短时出联目标信号时, 存在时效性差、难以实时跟踪的问题, 提出一种基于 REINFORCE 算法的组网测向选站方法。采用马尔可夫决策过程 (Markov Decision Process, MDP) 框架, 将站点组合过程分解为多步决策过程, 通过逐次选取站点来构建组合, 实现组合优化问题到序贯决策问题的转化; 同时, 以站点与目标的位置关系作为状态空间, 用选站组合数量约束智能体的学习步数, 对已选择的站点进行动作屏蔽, 并利用几何精度因子 (Geometrical Dilution of Precision, GDOP) 作为奖惩函数评价选站方案的优劣。实验结果表明, 该方法在时间复杂度上显著优于遍历方法与遗传算法, 在定位精度上与遍历方法基本持平, 在泛化能力上, 能够学习到测向站空间构型与 GDOP 之间的映射关系, 对于未训练的目标场景也可以完成组网方案决策。

关键词: 组网测向; 序贯决策; 强化学习; 动作屏蔽; 几何精度因子

中图分类号: TN953

文献标识码: A

文章编号: 1671-0673(2026)03-0291-06

A Station Selection Method for Networked Direction-Finding Based on the REINFORCE Algorithm

ZHU Chenchen, ZHANG Jing

(Information Engineering University, Zhengzhou 450001, China)

Abstract: To address the issue of the poor timeliness and difficulty of real-time tracking in traditional networked direction-finding methods when handling fast-moving targets with short signal burst durations, a station selection method based on the REINFORCE algorithm is proposed. A Markov decision process (MDP) framework is employed to decompose the station combination process into a multi-step sequential decision process. By selecting stations one by one, the combinatorial optimization problem is transformed into a sequential decision problem. The positional relationship between stations and the target is used as the state space, the number of stations in the combination constrains the agent's learning steps, action masking is applied to the selected stations, and the geometrical dilution of precision (GDOP) serves as the reward function to evaluate the quality of station selection schemes. Experimental results demonstrate that the proposed method significantly outperforms both the exhaustive method and the genetic algorithm in terms of time complexity, while achieving comparable localization accuracy to the exhaustive method. Furthermore, it exhibits strong generalization ability by learning the mapping between station spatial configuration and GDOP, enabling effective network deployment decisions for unseen target scenarios.

Key words: networked direction-finding; sequential decision; reinforcement learning; action masking; geometric dilution of precision

随着无线电技术的发展,对无人机、飞机和轮船等目标信号进行测向定位监控在保证航空、航海安全方面发挥着重要作用。传统目标信号一出就调用全网测向站的组测模式,不仅加重了站点任务负荷,更易在多目标并发场景下导致漏测与失踪。为此,亟需一种自主协同组网测向模式,通过优化选站,在保障测向精度的前提下减轻站点压力,提升多目标实时监控能力。

组网测向的本质是站点优化选择问题,即在现有的测向站分布中选择合适的布局对目标进行定位。测向站布局优化是提高测量精度的有效手段,在设备精度与定位算法确定的前提下,通过选择最优布局来最大化测量精度^[1]。

现有研究主要分为两类:一类是对布局模型的研究。这类研究主要对常见的布局形式如矩形^[2]、菱形^[3]进行建模,仿真分析不同的几何布局对定位精度的影响,为布站提供参考。文献[4]分析了三站不同布局、站点数量以及测向误差对定位误差的影响。文献[5]分析了双站测向交叉定位的布站方案。然而,这类研究仅适用于初期建站选址,无法用于从已有测向站中选站,且局限于规则区域布局^[6],难以适应实际应用中需考虑地理条件和信号质量的不规则布站场景。另一类是对模型求解算法的研究。在站点位置已知、存在多种组合方案的情况下,根据约束条件建立模型并求解最优几何布局,主要算法包括遍历方法和启发式算法。遍历方法在小数量级的站点组合下可以快速得到全局最优解,但随着测向站规模的扩大,组网方案呈非线性增长,求解速度显著下降,甚至无法在有限时间内求解;启发式算法则以几何精度因子(Geometrical Dilution of Precision, GDOP)、克拉美罗下界(Cramer-Rao Lower Bound, CRLB)等作为适应度函数,模拟生物界觅食、遗传进行解空间搜索,在非均匀解空间中寻找近似最优解。文献[7]分析了自适应差分进化算法在TDOA(Time Difference of Arrival)定位布站优化中的优势,并在时间复杂度上与穷举法进行对比。文献[8]使用改进粒子群算法求解最优布站问题。文献[1]采用自适应调整策略改进遗传算法来解决测控设备布站优化问题。基于启发式算法的优化布站方法具有算法简单、收敛速度快等优点,但主要面向静态站点优化问题,面对目标位置移动、信号时变的动态场景,需要重新建模求解,通常只能获得近似优解。

上述算法各有其适用场景,不存在普适的最

优算法。近年来,强化学习(Reinforcement Learning, RL)在人工智能领域表现突出,已广泛应用于充电桩位置规划^[9]、应急避难所选址^[10]等实际选站(址)问题,为本文采用强化学习方法开展选站研究提供了技术思路。

针对目标信号出联时间窗口短的特点,结合强化学习的自主决策能力,本文提出了一种基于REINFORCE算法^[11]的组网测向选站方法,该方法采用了马尔可夫决策过程(Markov Decision Process, MDP)框架,将组合选站过程分解为多步决策过程,以站点与目标的位置关系作为状态空间,用选站组合数量约束智能体学习步数。针对站点不可重复选择的规则,引入动作屏蔽机制加以约束,以GDOP作为奖励函数引导智能体向更优站点组合方向学习。为验证所提方法的性能,本文设计了仿真实验,将该方法与遗传算法及遍历方法进行了对比分析。

1 问题模型

1.1 AOA测向定位模型

在基于目标信号到达角度值(Angle of Arrival, AOA)的球体测向定位模型中(如图1所示),将地球建模为半径为1的标准球体,N表示北极,目标信号S的位置为(B, L),其中 B 表示纬度, L 表示经度;第 i 个测向站的坐标为(B_i, L_i), $i = 1, 2, \dots, N$,第 i 个测向站观测目标的方位角为 θ_i 。

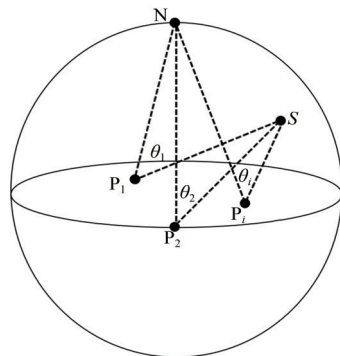


图1 AOA球体测向定位模型

根据示向线与目标信号源的关系可得^[12]:

$$\theta_i = \arctan\left(\frac{x}{y}\right) \quad (1)$$

$$\begin{cases} x = \cos B \sin(L - L_i); \\ y = \cos B_i \sin B - \sin B_i \cos B \cos(L - L_i). \end{cases} \quad (2)$$

式中, $\arctan(\cdot)$ 为四象限反正切函数。

将式(1)展开成可求解的非线性方程组,并进行牛顿迭代法求解,可得目标信号的位置(B, L)。

1.2 定位精度指标

在AOA测向定位系统中,定位精度主要受测量设备误差、定位算法、测向站几何布局的影响。测量设备误差和定位算法属于系统固有误差,受客观因素制约,而测向站几何布局可通过优化加以改善,从而提高定位精度^[12]。常用GDOP衡量测向误差对定位误差的放大程度,反映几何构型对定位精度的影响,其值越小,定位精度越高。多站测向定位中,GDOP的计算如下^[13]:

$$N_{\text{GDOP}} = \sqrt{\text{trace}\left(\left(\mathbf{H}^T \mathbf{H}\right)^{-1}\right)} \quad (3)$$

$$\mathbf{H} = \begin{bmatrix} \frac{Y - Y_1}{r_1^2} & \frac{-X + X_1}{r_1^2} \\ \vdots & \vdots \\ \frac{Y - Y_N}{r_N^2} & \frac{-X + X_N}{r_N^2} \end{bmatrix} \quad (4)$$

$$\begin{cases} X_i = \frac{\pi R}{180} L_i; \\ Y_i = R \ln \left(\tan \left(\frac{\pi}{4} + \frac{B_i}{2} \right) \right). \end{cases} \quad (5)$$

式中: $\mathbf{H}$ 为测向几何矩阵,反映站点与目标间的空间布局特性; r_i 为平面坐标系中第*i*个测向站与目标的几何距离; X_i, Y_i 分别为地理坐标系下的纬度、经度投影到平面直角体系下的横坐标、纵坐标的值; R 为地球半径。

1.3 限定站数的测向选站问题

基于前述测向定位系统模型与定位精度评价指标,本文研究如下组合优化问题:从*N*个测向站中选择*M*个构成一个组网方案,以最大化该方案下的定位精度为目标。

1.3.1 自变量

在二维平面中,第*i*个测向站的坐标为 (X_i, Y_i) , $i = 1, 2, \dots, N$ 。该问题的自变量 c_i 为

$$c_i = \begin{cases} 1, & \text{第 } i \text{ 个测向站参与组网;} \\ 0, & \text{其他.} \end{cases} \quad (6)$$

1.3.2 约束条件

参与组网的站点数量约束:s.t. $\sum_{i=1}^N c_i = M, c_i \in \{0, 1\}$ 。

1.3.3 目标函数

GDOP可反映组网方案的定位精度,选择GDOP最小的测向站布局,可有效提高定位精度。因此目标函数为: $\min N_{\text{GDOP}}(\mathbf{l}_s), \mathbf{l}_s = (c_1, c_2, \dots, c_N)$ 。其中: $\mathbf{l}_s$ 代表一个组网方案, $c_i = 1$ 时,表示组网方案包含编号*i*的测向站;反之, $c_i = 0$ 时,表示不包含编号为*i*的测向站。

2 模型构建

2.1 REINFORCE算法原理

强化学习是一种机器学习范式,旨在让智能体以马尔可夫决策过程为框架与环境进行交互。其交互过程为感知状态、作出决策、改变状态、得到反馈等。其中:学习、决策的机器称为智能体;智能体所处的空间称为环境;当前所处的环境称为状态;每步所作的决定称为动作;作用于环境的动作得到的反馈称为奖励;在每个状态下选择动作的依据称为策略 π 。根据强化学习训练过程中动作选取方式的不同,可以分为基于值函数算法和基于策略梯度算法^[14]。

基于值函数算法是智能体在学习过程中策略保持不变,通过优化值函数去适应策略,进而找到最优策略。常用的策略有贪婪法、汤普森采样等。

值函数包括状态值函数 $V^\pi(s)$ 和动作状态值函数 $Q^\pi(s, a)$ 。 $V^\pi(s)$ 表示在*t*时刻处于状态*s*时,遵循当前策略 π 所获得的未来累积回报的期望; $Q^\pi(s, a)$ 表示在*t*时刻处于状态*s*并执行动作*a*,遵循当前策略 π 所获得的未来累积回报的期望^[15]。

$$V^\pi(s) = E[G_t | S_t = s] \quad (7)$$

$$Q^\pi(s, a) = E[G_t | S_t = s, A_t = a] \quad (8)$$

$$G_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots = \sum_{k=t}^T r_k \gamma^{k-t} \quad (9)$$

式中: $E[\cdot]$ 为期望函数; G_t 表示从*t*时刻到终止时刻*T*获得的累计回报之和; γ 是衰减因子。

值函数算法的最优策略可以通过最优值函数来表示,对于任意状态*s*,最优策略 π^* 为

$$\pi^*(s) = \underset{a}{\operatorname{argmax}} Q^*(s, a) \quad (10)$$

式中, $Q^*(s, a)$ 为最优动作状态值函数,表示在状态*s*下采取动作*a*后,遵循最优策略所能获得的期望累积回报。

最优状态值函数 $V^*(s)$ 与最优动作状态值函数满足以下关系:

$$V^*(s) = \max_a Q^*(s, a) \quad (11)$$

式中, $V^*(s)$ 表示在状态*s*下,选取最优动作所能达到的最大期望累积回报。

最优值函数是所有策略对应一般值函数的上界,满足以下关系:

$$\begin{cases} V^*(s) = \max_{\pi} V^\pi(s); \\ Q^*(s, a) = \max_{\pi} Q^\pi(s, a). \end{cases} \quad (12)$$

一般值函数可通过贝尔曼方程迭代求解得

到,该方程反映了固定策略下当前状态与下一状态价值的递推规律。具体形式如下:

$$V^\pi(s) = E[r + \gamma V^\pi(s') | s] \quad (13)$$

式中: r 表示状态 s 下即时奖励; $V^\pi(s')$ 表示下一状态 s' 的值函数。

基于策略梯度算法是通过直接优化策略函数,使其朝着累积奖励的期望 $J(\theta)$ 增大的方向更新,从而实现最优策略求解。该方法通常构造一个以 θ 为参数、以状态 s 为输入、输出动作概率分布的策略 $\pi_\theta(a|s)$,以累积奖励的期望 $J(\theta)$ 作为目标函数,策略 $\pi_\theta(a|s)$ 是参数 θ 的函数,目标函数 $J(\theta)$ 是策略 $\pi_\theta(a|s)$ 的函数,则目标函数为参数 θ 的函数,原问题变成基于 θ 的最优化问题^[16]。用梯度上升法优化参数 θ 使目标函数取得最大值,从而得到最优策略。

REINFORCE算法是Williams^[11]等提出的一种典型的策略梯度算法,是深度强化学习中策略梯度类方法的基础。算法采用回合制更新,在一个完整回合结束后,利用整条轨迹的累积回报来更新策略,无需价值函数辅助,具体步骤如下。

在给定的策略 π_θ 下,智能体与环境交互得到一条轨迹 τ :

$$\tau = \{s_1, a_1, r_1, \dots, s_T, a_T, r_T\} \quad (14)$$

该轨迹发生的概率是:

$$\pi_\theta(\tau) = P(s_1) \prod_{i=1}^T (\pi_\theta(a_i | s_i) P(s_{i+1} | s_i, a_i)) \quad (15)$$

式中: $\pi_\theta(a|s)$ 表示在状态 s 下选择动作 a 的概率; $P(s_{i+1} | s_i, a_i)$ 表示在状态 s_i 下执行动作 a_i 后,转移到状态 s_{i+1} 的概率。

所有轨迹的奖励期望 $J(\theta)$ 为:

$$J(\theta) = \sum_{\tau} G_{\tau} \pi_{\theta}(\tau) \quad (16)$$

$$G_{\tau} = \sum_{t=1}^T r_t \gamma^{t-1} \quad (17)$$

式中, G_{τ} 表示一条轨迹的累积奖励。

通过不断优化参数 θ ,使得 $J(\theta)$ 最大化,常用梯度上升法优化参数 $\theta: \theta \leftarrow \theta + \alpha \nabla J(\theta)$,其中, α 是学习率。

最优策略对应参数 θ^* 应满足:

$$\theta^* = \underset{\theta}{\operatorname{argmax}} J(\theta) \quad (18)$$

对奖励期望 $J(\theta)$ 进行求导、变换可得:

$$\nabla J(\theta) = E_{\tau \sim \pi_{\theta}(\tau)} (G_{\tau} \nabla \log_2 \pi_{\theta}(\tau)) \quad (19)$$

式中, $E_{\tau \sim \pi_{\theta}(\tau)}$ 表示对策略 $\pi_{\theta}(\tau)$ 生成的所有可能轨迹 τ 的期望。

由于期望难以直接计算,通过采样 N 条轨迹,可近似得到 $\nabla J(\theta)$:

$$\nabla J(\theta) = E_{\tau \sim \pi_{\theta}(\tau)} (G_{\tau} \nabla \log_2 \pi_{\theta}(\tau)) \approx \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^{T_n} (\nabla \log_2 \pi_{\theta}(a_t^n | s_t^n) \cdot G_{\tau}^n) \quad (20)$$

针对策略梯度方差大的问题,REINFORCE算法对累积奖励 G_{τ} 进行修改,用 t 时刻后的回报函数 R_t 代替 G_{τ} :

$$R_t = \sum_{k=t}^T r_k \gamma^{k-t} \quad (21)$$

$$\nabla J(\theta) = \frac{1}{N} \sum_{n=1}^N \sum_{t=1}^{T_n} (\nabla \log_2 \pi_{\theta}(a_t^n | s_t^n) \cdot R_t^n) \quad (22)$$

2.2 基于REINFORCE算法的求解模型

将目标视为智能体,智能体依据含有测向站和目标信号等信息的当前状态 S_t ,按照某种策略作出动作 A_t ,即选择 i 号测向站参与组网,动作作用于环境得到反馈 R_t ,并更新自己的状态为 S_{t+1} ,持续与环境交互,直到得到一个选站方案。组网测向选站MDP模型如图2所示。

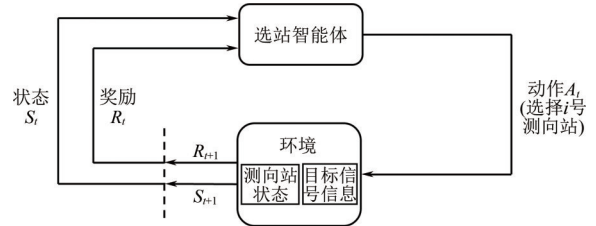


图2 组网测向选站MDP模型

针对选站问题,重点从状态空间、动作空间与奖励函数3个方面展开设计。

1) 状态空间设计。在强化学习模型中,状态反映智能体对所处环境的感知。在组网测向定位模型中,第 i 个测向站与目标的相对位置关系 s_i 为 $(X - X_i, Y - Y_i)$, $i = 1, 2, \dots, N$ 。则 T 时刻的状态 S_T 为所有测向站与目标的相对位置关系:

$$S_T = [s_1, s_2, \dots, s_N] \quad (23)$$

T 时刻的动作为 n ,表示第 n 个测向站已经被选择,则 $(T+1)$ 时刻模型的状态为

$$S_{T+1} = [s_1, s_2, \dots, s_{n-1}, s_n, s_{n+1}, \dots, s_N] \quad (24)$$

式中, s_n 为第 n 个测向站与目标的相对位置关系,且 $s_n = (0, 0)$ 。

2) 动作空间设计。动作是智能体在当前状态下可执行的决策。在测向站选站问题中,每个动作对应选择一个未被选过的测向站。智能体需要从 N 个测向站中选出 M 个站点,系统原始动作空间大小为 N 。由于测向站选择具有无重复性,即在同一选站过程中已选站点不可被重复选择,实际可

选动作空间会随着学习过程逐步缩小。设智能体需要 M 步来完成一组站点的选择:第1步,智能体的可选动作空间大小是 N ,第2步是 $(N-1)$,第 M 步是 $(N-M+1)$,当 M 步结束后,所选的动作集合 $A=\{A_1, A_2, \dots, A_M\}$ 即构成一个完整的选站方案。为避免智能体重复选择站点,引入动作屏蔽机制,以应对模型动作空间缩减。

3)奖励函数设计。奖励是对智能体所做动作的反馈,因GDOP能够衡量站点布局对定位误差的影响,因此选择其作为奖励函数的设计依据。由于GDOP值越小表示定位精度越高,为使奖励值与优化目标方向一致,本文将奖励函数定义为GDOP的相反数:

$$R_t = -N_{\text{GDOP}}(l_s) \quad (25)$$

本文算法如下所示。

输入:测向站坐标集,目标坐标,限定数量 M

输出:选站结果 $A=\{a_1, a_2, \dots, a_M\}$

1. 初始化状态空间 S 、策略参数 θ 、目标位置;
2. for each episode do
3. for $t=1$ to M do
4. 获取网络输出 logits; 根据状态 S_t 取值为0的维度, 将对应位置动作的 logits 设为 $-\infty$; 施加 Softmax, 得到概率分布 $\pi_\theta(a|s)$, 根据当前分布采样选择动作 a_t ;
5. 执行动作 a_t , 获得奖励, 进入下一状态 S_{t+1} ;
6. end for
7. 将轨迹 $\tau = \{s_1, a_1, r_1, \dots, s_T, a_T, r_T\}$ 存入缓冲区;
8. for $t=1$ to M do
9. 计算 R_t ;
10. $\nabla J(\theta) \leftarrow \nabla J(\theta) + \nabla_\theta \log_2 \pi_\theta(s_t, a_t) \cdot R_t$
11. end for
12. $\theta_{\text{new}} \leftarrow \theta_{\text{old}} + \alpha \nabla_\theta J(\theta)$
13. $\nabla J(\theta) \leftarrow 0$
14. 丢弃轨迹 τ ;
15. end for

3 实验验证

3.1 仿真环境设置

为验证本文算法的有效性,在相同条件下,将其与遍历方法、遗传算法进行对比实验。基于Python 3.8.19 环境与Tensorflow 2.11.0 框架搭建仿真平台,环境配置如下:在东经 $[110^\circ, 150^\circ]$ 、北纬 $[15^\circ, 55^\circ]$ 范围内随机生成60个测向站,假设各测向站的测量误差相互独立且均服从零均值高斯分布。

3.2 性能评价标准及参数设置

一个性能良好的组测算法,即便无法保证始

终获得全局最优解,通常也能以较高的概率找到近似最优解。解的质量排名占比 D 为全部解数量和优于算法解的数量之差与全部解数量的比值,即:

$$D = \frac{C - C_R}{C} \quad (26)$$

式中: C_R 表示解的值小于算法求得GDOP的值的解的数量; C 表示全部解的数量。 D 值越大,表明本文算法解在全部解中的排名越靠前,解的质量越高; D 值越小,表明解的质量越低。强化学习方法网络仿真参数及遗传算法参数设置见表1和表2。

表1 强化学习方法网络仿真参数设置

参数名称	参数值
输入层	$2N$
连接层神经元个数	$(128, 64, 32)$
输出层神经元个数	M
学习率	0.000 05
贪婪系数	初始 0.95, 随后递减
折扣系数	0.9
更新步数	100

表2 遗传算法参数设置

参数名称	参数值
种群规模	50
编码长度	N
迭代次数	200
交叉概率	0.8
变异概率	0.05
选择方式	锦标赛选择

3.3 仿真结果对比分析

3.3.1 算法时间复杂度分析

在 $M=5$ 的条件下,利用Python语言下time模块仿真出遍历方法、遗传算法以及强化学习方法从不同测向站数量中选取组网方案所使用的时间。图3是3种方法所用的平均时间对比。

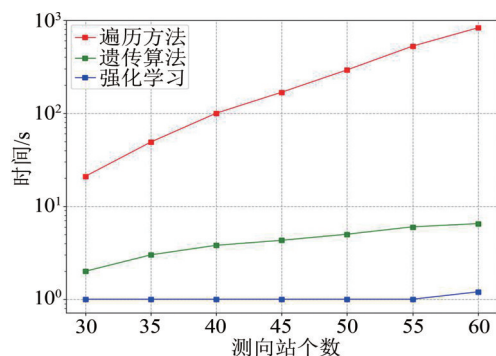


图3 3种方法求解所用平均时间对比

由图3可得,在相同测向站数量下,遍历方法所需时间最多,强化学习方法所需时间最少,遍历方法与遗传算法、强化学习方法所需时间数量级相差较大。随着测向站数量增多,遗传算法和遍历方法用时逐渐增大,且遍历方法时间增加幅度显著,强化学习方法运行时间保持相对平稳,消耗时间受测向站数量影响不大。。

3.3.2 算法定位性能分析

$M=5$ 时,仿真出3种方法在不同测向站数量的场景下,选取组网方案解的质量排名占比情况。图4对比了3种方法求得的解的质量排名占比。

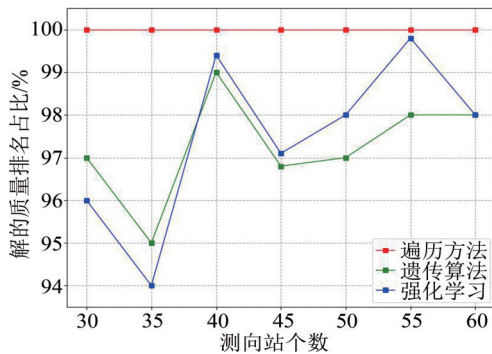


图4 解的质量排名占比对比

排名占比越大,越接近最优解,其中遍历方法代表全局最优解,占比为100%。由图4可得,在测向站数量较少时,强化学习方法与遗传算法得到的解的排名占比相近,无显著差异。随着测向站数量增多,解的空间增大,强化学习方法解的质量排名占比略优于遗传算法。在不同测向站数量的场景下,强化学习方法解的质量排名占比均在94%以上,具备良好的鲁棒性。

3.3.3 移动目标定位性能分析

在 $N=40$ 、 $M=5$ 的场景下,当算法收敛后,仿真研究了移动目标的定位性能的变化情况。表3分析了随着目标位置机动变化,3种算法的选站方案GDOP情况。

表3 3种算法组网方案的GDOP值对比

算法	目标 位置1	目标 位置2	目标 位置3	目标 位置4
遍历方法	10.40	7.80	6.30	10.50
强化学习	10.50	7.90	6.40	10.80
遗传算法	10.50	8.40	6.70	11.40

由表3可得,强化学习推荐的选站方案计算出的GDOP值与遍历算法无显著差异,整体略优于遗传算法。表明智能体已经在训练中学习到了测向

站空间构型与GDOP之间的映射关系,对于未训练的目标场景也可以完成测向站点组网方案决策,具备良好的泛化能力。

4 结束语

本文针对组网测向场景下的站点选择问题,提出一种基于REINFORCE算法的组网测向选站方法。该方法将站点组合过程分解为多步决策过程,实现组合优化问题到序贯决策问题的转化。仿真结果表明,本文方法在定位精度与遍历法持平的前提下,时间复杂度显著优于遍历法和遗传算法;同时,该方法能够有效学习测向站空间构型与GDOP之间的映射关系,对未训练的目标场景仍具备良好的组网决策泛化能力。

参考文献:

- [1] 刘宏建,王明孝,蔡中祥,等.一种基于自适应遗传算法的测控设备布站优化方法[J].测绘科学技术学报,2020,37(3):325-330.
- [2] 张世航.基于TDOA的对地目标定位技术研究[D].成都:电子科技大学,2023:49.
- [3] 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,2022:15.
- [4] 曹文焕,韩升明,陈超.多站测向交叉定位的定位精度研究[J].现代电子技术,2026,49(5):51-55.
- [5] 孟圣波,李鹏飞,黄建廷.对空观察哨双站测向交叉定位误差分析与验证[J].舰船电子工程,2023,43(7):36-39.
- [6] 窦雪倩,李华,许博.基于几何稀释度的多站外源时差定位最优布站方法[J].电子信息对抗技术,2018,33(5):37-40.
- [7] 曾凤,蔡方凯,龙宁.基于自适应DE算法的TDOA定位布站优化策略[J].电子信息对抗技术,2025,40(1):53-62.
- [8] 丁磊,任璐,侯轩,等.面向高精度AOA定位的布站优化算法[J].计算机科学,2025,52(增刊2):679-687.
- [9] 高铭壑.基于多目标融合的充电桩位置规划方法研究[D].北京:华北电力大学,2020:3.
- [10] 刘雅婷.基于多智能体强化学习的应急避难所动态选址算法研究[D].北京:北京交通大学,2023:7.
- [11] WILLIAMS R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning [J]. Machine Learning, 1992,8(3/4):229-256.
- [12] 张天予,滕鹏晓,吕君,等.结合球面伪线性估计器和传播噪声估计的广域次声源定位算法[J].声学学报,2025,50(2):395-411.