

HF 和波动参数辅助的优化 XGBoost 室内定位方法^{*}

刘高辉, 凌凤智

(西安理工大学 自动化与信息工程学院, 陕西 西安 710048)

摘要: 针对复杂室内环境下接收信号强度测量数据中包含噪声使其呈现波动性导致定位精度低的问题, 提出一种基于混合滤波 (HF)、波动参数辅助的优化极限梯度提升 (XGBoost) 室内定位方法。首先采用 HF 的方法对数据子集进行优化, 降低噪声的影响, 得到初始数据库; 另外, 考虑到波动不能完全消除, 引入能够反映数据变化程度的波动参数; 其次, 针对 XGBoost 算法性能易受初始参数的影响, 采用粒子群 (PSO) 算法对其进行寻优, 并将波动参数与优化后的数据作为算法输入训练生成定位模型; 最后, 将目标点处信息输入到模型中完成位置估计, 同时将该点数据保存到数据库中完成更新。实验结果表明, 与传统算法相比, 所提算法具有良好的定位效果, 在 1 m、2 m 和 3 m 范围内, 定位准确率分别提升 9.2%、14.1% 和 18.45%。

关键词: 室内定位; 混合滤波; 波动参数; 粒子群算法; XGBoost

中图分类号: TN911

文献标识码: A

DOI: 10.19358/j.issn.2097-1788.2025.09.008

引用格式: 刘高辉, 凌凤智. HF 和波动参数辅助的优化 XGBoost 室内定位方法 [J]. 网络安全与数据治理, 2025, 44 (9): 51–58, 65.

Optimization of XGBoost indoor positioning method with HF and fluctuation parameter assistance

Liu Gaohui, Ling Fengzhi

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: Aiming at the problem that the received signal strength measurement data in complex indoor environments contain noise which makes them show fluctuation and leads to low positioning accuracy, an optimized extreme gradient boosting (XGBoost) indoor positioning method based on hybrid filtering (HF) and fluctuation parameters assistance is proposed. Firstly, the HF method is used to optimize the data subset, reduce the influence of noise, and obtain the initial database; in addition, considering that the fluctuation can't be completely eliminated, the fluctuation parameter that can reflect the degree of data change is introduced; secondly, for the performance of the XGBoost algorithm is susceptible to the influence of the initial parameter, the particle swarm (PSO) algorithm is used for optimization of the parameter, and the fluctuation parameter and optimized data are used as inputs to train the algorithm to generate the localization model; finally, the target point information is input into the model to complete the position estimation, and the point data is saved into the database to complete the update. The experimental results show that compared with the traditional algorithms, the algorithm in this paper has a good localization effect, and the localization accuracy is improved by 9.2%, 14.1%, and 18.45% in the range of 1 m, 2 m, and 3 m, respectively.

Key words: indoor positioning; hybrid filtering; fluctuation parameters; PSO algorithm; XGBoost

0 引言

随着无线局域网、智能终端的普及与发展, 基于位置信息的需求不断增加。传统全球定位系统 (Global Positioning System, GPS)^[1] 和基站定位技术能够满足室外场

景的定位需求, 然而室内环境复杂多变且信号弱, 使得传统技术在此环境下定位精度低, 无法满足人们对定位服务精度的需求^[2], 因此室内定位吸引了众多学者的研究。

常见的室内定位技术有多种, 如超宽带^[3]、蓝牙^[4]、射频识别和 Wi-Fi^[5] 等。因具有覆盖面积广、成本低等

^{*} 基金项目: 国家自然科学基金 (61671375)

特点, Wi-Fi 设备已广泛部署在超市、医院、车站、教室等室内场景且随着移动智能设备的普及提升了数据测量和采集效率, Wi-Fi 指纹定位已成为该领域的研究热点^[6]。

在当前的研究中, Wi-Fi 指纹定位是最普遍的室内定位方法。该方法利用接收信号强度与物理位置之间的某种映射关系进行定位, 它包括离线数据采集和在线匹配两个阶段^[7]。此外, 在复杂室内环境如障碍物遮挡、多径效应、噪声^[8]等影响下, 接收信号存在不稳定性, 导致采集的接收信号强度数据呈现波动性, 其中的异常数据容易导致模型性能不稳定、精度低等问题, 所以近年来相关算法被陆续提出。卡尔曼滤波 (Kalman Filter, KF) 算法^[9-10]在处理含有高斯噪声的信号时性能较好, 而粒子滤波 (Particle Filter, PF) 算法^[11-12]可以滤除非高斯噪声, 单一滤波算法不能完全消除干扰。文献 [13] 为优化测量数据值, 采用均值滤波、高斯滤波、KF 等混合滤波 (Hybrid Filtering, HF) 算法对其进行处理。文献 [14] 提出一种双滤波器集成算法, 先利用 KF 获得平滑约束指纹, 然后采用扩展卡尔曼滤波将处理后的指纹与运动传感器相结合完成定位。此外, 在有 Wi-Fi 设备覆盖的室内场景中, 定位区域内采集的数据和特征越多, 定位精度越高。文献 [15] 将静态环境下采集的原始数据和时频域波动参数共同作为输入训练决策树模型提升了定位精度, 但在复杂动态环境 (噪声、障碍物阻挡) 下性能不理想, 且未处理的异常数据会导致计算量增加。文献 [16] 首先将接收的信号强度转换成用于指纹定位的五个统计特征, 然后利用聚类 and 行人航位推算算法结合完成目标的定位, 但该方法需要高精度传感器从而增加了成本。随着机器学习算法的不断成熟, 很多学者将机器学习算法与传统定位方法相结合, 把坐标求解转化为分类预测问题。文献 [17] 提出改进接入点 (Access Point, AP) 选择策略并结合利用 Bagging 方法生成用于分类和匹配的模型来预测目标的位置, 该方法虽然提升了精度, 但传统 Bagging 方法构建的决策树模型容易产生过拟合且复杂度高的问题。文献 [18] 设计新型的迁移学习网络得到新的指纹库, 完成对待测点位置预测, 提高了室内定位的鲁棒性。文献 [19] 首先使用主成分分析对采集数据进行降维, 再利用核岭回归算法对目标位置进行预测, 但在此过程中会产生冗余信息从而增加计算量并影响算法性能。

基于以上研究, 本文提出一种基于混合滤波、波动参数辅助的粒子群算法优化 XGBoost 的室内定位方法。该方法共分为两个阶段, 在离线阶段, 混合滤波是指利用粒子-卡尔曼滤波算法对每个采样点处数据优化的过程,

首先采用对非高斯噪声效果较好的粒子滤波算法进行处理, 然后将其输出作为卡尔曼滤波算法的输入进一步滤除非高斯噪声, 以最后的输出构建初始数据库; 此外, 计算可以体现数据细微波动的波动参数。其次, 由于初始化参数的选取严重影响 XGBoost 的性能, 采用粒子群算法 (Particle Swarm Optimization, PSO) 获取最优的参数提升 XGBoost 算法的预测精度, 将波动参数和处理后的数据输入到算法中构建定位模型。在线匹配阶段, 将目标点的样本经过混合滤波后送入训练好的定位模型中预测目标位置, 同时将滤波后的待定位点的数据保存到初始数据库中完成数据库的更新。

1 数据预处理与混合滤波算法

1.1 数据预处理

在 Wi-Fi 室内定位中有基于信道状态信息 (Channel State Information, CSI) 和基于接收信号强度 (Receive Signal Strength, RSS) 两种定位方案。后者已被众多学者广泛应用, 该方案的无线信号衰减模型为:

$$r(d) = r(d_0) + 10\alpha \lg\left(\frac{d}{d_0}\right) + \xi \quad (1)$$

式中, $r(d)$ 、 $r(d_0)$ 分别表示在距离 d 和 d_0 处测量得到的 RSS 值, α 为路径损耗系数, ξ 为环境噪声。

式 (1) 表明在复杂的室内环境中, 测量数据易受噪声和距离 d 的影响使采集样本呈现波动性, 导致定位性能不稳定、精度低等问题。某个采样点的测量数据可用向量 $\mathbf{R}$ 表示:

$$\mathbf{R}_i = [r_i^1, r_i^2, \dots, r_i^n] \quad (2)$$

式中, r_i^n 表示第 i 个采样点处的第 n 个数据。

将前后两次采集数据 r 最大允许波动范围设置为 M , 其主要目的是对接收信号前后两次的范围进行限制, 减少因干扰产生的 r 突变, 提高采集数据的可靠性。采集样本可以表示为:

$$r_i^n = \begin{cases} r_i^n, & |r_i^n - r_i^{n-1}| \leq M \\ r_i^{n-1}, & |r_i^n - r_i^{n-1}| > M \end{cases} \quad (3)$$

为防止首次采集的数据 r 为异常值, 在预处理操作之前将该值与 n 个采集数据的中值进行比较, 如果两者差值小于 $2M-1$, 则保留本数据, 反之则用采集数据的中值替换。

1.2 PF-KF 混合滤波算法

采集的数据样本中包含高斯和非高斯噪声, 为了减少两种噪声对定位精度的影响, 采用 PF 和 KF 相结合的方法处理采集样本。该方法分为两个步骤: 首先对预处理后的数据进行粒子滤波, 降低非高斯噪声的影响; 然后将其输出作为 KF 算法的输入进一步处理高斯噪声, 以

最后的输出建立初始数据库。

粒子滤波是通过寻找一组在状态空间中传播的随机样本，积分运算用均值来替换，得到状态最小方差分布的近似概率密度函数的过程。将预处理后的数据采用粒子滤波处理，其状态空间模型如下：

$$\begin{cases} x_j = F(x_{j-1}, v_{j-1}) \\ z_j = H(x_j, \mu_j) \end{cases} \quad (4)$$

其中， F 、 H 为系统函数， x_j 表示 j 时刻状态， z_j 表示 j 时刻的观测值， v_{j-1} 、 μ_j 分别为过程噪声和观测噪声。

首先生成 L 个粒子 $\{p_l, \omega_l\}_{l=1}^L$ ； p_l 、 ω_l 分别表示第 l 个粒子及其对应的权重，粒子 p_l 服从式 (5) 的正态分布。

$$p_l \sim N(r_j, \sigma^2) \quad (5)$$

采用式 (6) 进行权值更新，并对粒子权重进行归一化操作：

$$\begin{cases} \omega_l = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(r_{j-1}-p_l)^2}{2\sigma^2}} \\ \hat{\omega}_l = \frac{\omega_l}{\sum_{l=1}^L \omega_l} \end{cases} \quad (6)$$

其次，为了解决重采样可能导致粒子多样性降低的问题，采用 Neff 准则（式 (7)）计算有效粒子数，判断是否进行重采样。

$$\text{Neff} = \frac{1}{\sum_{l=1}^L (\hat{\omega}_l)^2} \quad (7)$$

当 $\text{Neff} < N_{\text{threshold}}$ 时，进行重采样得到新的粒子，由 Neff 准则可知， $N_{\text{threshold}} = \frac{2L}{3}$ 时效果最优。

最后，将 L 个粒子的加权平均值作为第 j 个数据 r 的估计值：

$$\hat{r}_j = \sum_{l=1}^L p_l \omega_l \quad (8)$$

对预处理数据进行处理，得到的结果表示为：

$$\hat{\mathbf{R}}_i = [\hat{r}_1, \hat{r}_2, \dots, \hat{r}_j, \dots, \hat{r}_n] \quad (9)$$

采用卡尔曼滤波对上述结果进一步优化，KF 的观测方程和测量方程为：

$$\begin{cases} x_j = Ax_{j-1} + \mu_j \\ z_j = H \cdot \hat{r}_j + v_j \\ j=1, 2, \dots, n \end{cases} \quad (10)$$

其中， μ_j 、 v_j 分别为过程噪声和测量噪声，分别服从均值为 0，方差为 Q 、 R 的正态分布。

$$\begin{cases} \rho(\mu) \sim N(0, Q) \\ \rho(v) \sim N(0, R) \end{cases} \quad (11)$$

将经过 PF 处理后的数据作为测量值，计算其方差：

$$\sigma^2 = \text{var}(\hat{r}_j - \hat{r}_{\text{mean}}) \quad (12)$$

则第 j 个数据 r 的最优估计值为 x_j ，计算公式为：

$$x_j = x_{j-1} + \text{Kg}(\hat{r}_j - Ax_{j-1}) \quad (13)$$

式中卡尔曼增益 Kg 可由式 (14) 计算得出：

$$\text{Kg} = \rho_j^- \mathbf{H}^T (\mathbf{H} \rho_j^- \mathbf{H}^T + R)^{-1} \quad (14)$$

式中 $\rho_j^- = \rho_{j-1} + Q$ ，更新后的 ρ_i 可由式 (15) 计算得出：

$$\rho_j = (\mathbf{I} - \text{Kg} \mathbf{H}) \rho_j^- \quad (15)$$

最后通过式 (16) 得到优化后的向量 $\mathbf{R}$ ：

$$\mathbf{R}_i = \frac{1}{n} \sum_{j=1}^n x_j \quad (16)$$

综上所述，经混合滤波后得到的初始数据库如下：

$$\mathbf{D} = \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \\ \vdots \\ \mathbf{R}_k \end{bmatrix} = \begin{bmatrix} r_1^1, r_1^2, r_1^3, \dots, r_1^n, (x_1, y_1) \\ r_2^1, r_2^2, r_2^3, \dots, r_2^n, (x_2, y_2) \\ \vdots \\ r_k^1, r_k^2, r_k^3, \dots, r_k^n, (x_k, y_k) \end{bmatrix} \quad (17)$$

2 加入波动参数辅助的 PSO-XGBoost 模型构建

2.1 波动参数

针对室内采集的数据 r 时变波动性无法完全消除的问题，采用波动特征参数表示样本的变化情况，将它们作为特征输入到模型中训练，从而提高模型的性能。

将在每个 AP 处采集的数据样本看成一个数学序列 $s(n)$ ，然后计算其波动特征。在样本波动参数的选取方面，选择可以描述数据分布情况和平均变化的偏度 (Skewness)、峰谷差值 (Yp2p) 等参数，此外还加入方差 (Var) 作为波动参数，将它们与训练数据共同作为算法的输入，构建用于定位的模型。波动参数定义如表 1 所示。

表 1 波动参数

符号	计算公式	含义
Skewness	$E\left[\left(\frac{s(n) - \overline{s(n)}}{\sigma}\right)^3\right]$	偏度
Mean	$\frac{1}{N} \sum_{n=1}^N s(n)$	均值
Yp2p	$\{s(n)\} \big _{\max} - \{s(n)\} \big _{\min}$	峰谷差值
Var	$\frac{1}{N} \sum_{n=1}^N (s(n) - \overline{s(n)})^2$	方差
Kurtosis	$E\left[\left(\frac{s(n) - \overline{s(n)}}{\sigma}\right)^4\right]$	峰度

2.2 粒子群算法

粒子群算法是一种智能优化算法，因操作简单、收敛速度快等优点被广泛应用于各领域。该算法原理为在 n

维向量空间中, 每个粒子 m 对应的位置和速度分别为 X , V , 对应的更新公式如下:

$$\begin{cases} V_{mn}^{t+1} = \beta V_{mn}^t + c_1 r_1 (P_{mn}^t - X_{mn}^t) + c_2 r_2 (G_{mn}^t - X_{mn}^t) \\ X_{mn}^{t+1} = X_{mn}^t + V_{mn}^{t+1} \end{cases} \quad (18)$$

其中, β 为惯性权重, r_1 和 r_2 表示常数, c_1 和 c_2 为随机数; V_{mn}^{t+1} 表示第 $t+1$ 次迭代中粒子 m 在第 n 维的速度, X_{mn}^{t+1} 表示第 $t+1$ 次迭代中粒子 m 在第 n 维的位置, G_{mn}^t 表示第 t 次迭代中全局最优位置。

2.3 XGBoost 算法

XGBoost 是一种极限梯度提升模型, 具有效率高和灵活性强等优点, 模型框架如图 1 所示。它将新模型拟合到先前模型的残差并对代价函数在某点处进行二阶泰勒展开, 加入正则化项控制模型复杂度^[20]; 将数据划分为 k 个数据子集, 然后分别加入波动参数作为模型的输入, 最后通过 k 个加性函数集成来预测输出。

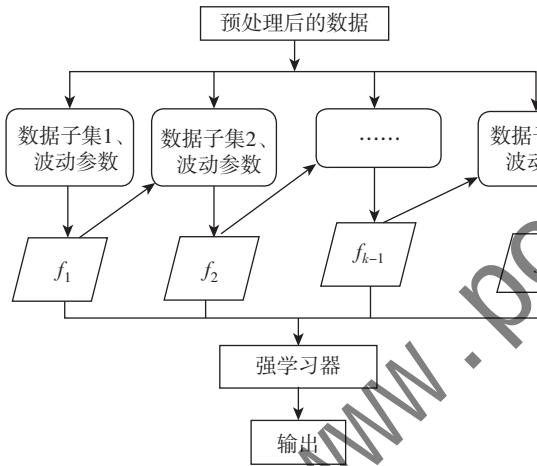


图 1 XGBoost 基本原理

XGBoost 目标函数定义如下 (包括损失函数和正则化两部分):

$$\text{Obj}^{(t)} = \sum_{j=1}^n \text{loss} (y_j, \hat{y}_j^{(t-1)} + f_t(r_j)) + \Omega(f_k) \quad (19)$$

式中, n 表示样本数量, y_j 、 $\hat{y}_j$ 分别为第 j 个样本的真实值、 $t-1$ 时刻的预测值, $\Omega(f_k)$ 、 f_t 分别表示正则化项、加性函数。正则化项也就是树的复杂度定义为:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\tau_j\|^2 \quad (20)$$

式中, T 为树中叶子节点的个数, γ 和 λ 为正则化参数, τ_j 为对应叶子节点的权重。

利用泰勒公式对式 (19) 中的损失函数部分进行展开 (忽略高次项) 得到:

$$\text{Obj}^{(t)} \approx \sum_{j=1}^n [\text{loss}(y_j, \hat{y}_j^{(t-1)}) + g_j f_t(r_j) + \frac{1}{2} h_j f_t^2(r_j)] + \Omega(f_t) \quad (21)$$

式中, g_j 、 h_j 分别为一阶、二阶梯度; 由于损失函数为常数项, 故式 (21) 可进一步化简为:

$$\widehat{\text{Obj}}^{(t)} = \sum_{j=1}^T \left[\left(\sum g_j \right) \tau_j + \frac{1}{2} \left(\sum h_j + \lambda \right) \tau_j^2 \right] + \lambda T \quad (22)$$

对式 (22) 中 τ_j 求导可得最优叶子权重和最小损失。详细推导过程可参见文献 [20]。

2.4 粒子群算法改进的 XGBoost

与传统的机器学习算法相比, XGBoost 复杂度低且精度比较高, 但初始参数的取值对精度的影响较大, 所以本文采用粒子群算法对 XGBoost 某些重要参数的取值进行优化。PSO-XGBoost 模型构建流程如图 2 所示, 优化参数包括学习率 (learning_rate)、最大迭代次数 (max_iter)、最大深度 (max_dep) 和正则化权重 (Re_w), 适应度函数设置为模型预测的准确率、均方误差函数。

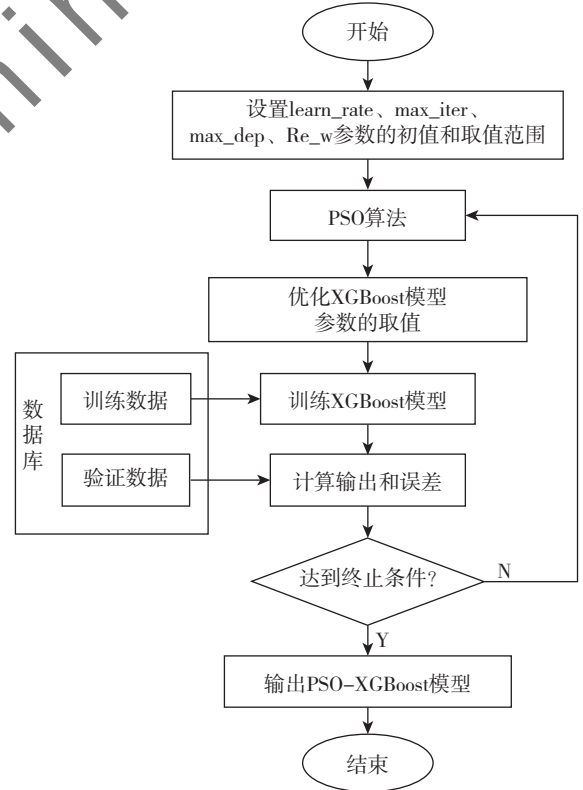


图 2 PSO-XGBoost 模型构建流程

设数据为: $D_i = \{(\mathbf{R}_1, p_1), \dots, (\mathbf{R}_k, p_k)\}$, $\mathbf{R}_i$ 表示第 i 个采样点接收到的 n 个 RSS 测量数据 r 组成的向量, $p_i = (x_i, y_i)$ 为采样点的位置坐标。通过 PSO-XGBoost 定位模型得到的预测结果为:

$$\hat{p}_i = \sum_k^k f_k(R_i) \tag{23}$$

3 基于混合滤波、波动参数辅助的优化 XGBoost 定位方法

本文提出的基于混合滤波、波动参数的优化 XGBoost 定位方法流程如图 3 所示，该方法分为离线模型训练和在线匹配两个阶段。

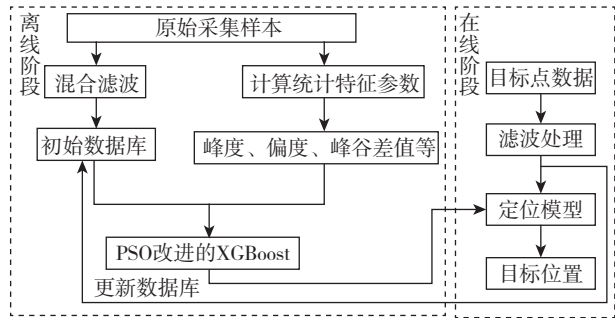


图3 基于混合滤波、波动参数辅助的 PSO-XGBoost 室内定位方法

离线训练阶段：首先利用 PF-KF 混合滤波的方法对预处理后的采集数据进行优化，降低噪声的影响，并以此构建初始数据库；其次，计算预处理后样本序列的波动参数，采用 PSO 优化的 XGBoost 算法对输入的数据和波动参数进行训练，得到最优的定位模型。

在线匹配阶段：根据目标点实时采集的数据 r 利用滤波算法进行处理，然后送入已有的 PSO-XGBoost 定位模型得到目标点的位置坐标；最后将处理后的目标点数据存入数据库中，完成数据库的更新。

4 实验与结果分析

4.1 实验环境及参数设置

为了验证本文所提方法的性能，采用 Zenodo 存储库中的 UJIndoorLoc 数据集，它是由西班牙某大学专业人员测量得到的，每一层包含 96 个采样点在 25 个月内收集的不同网络设备的数据及对应位置信息^[21]；采集区域结构如图 4、图 5 所示。

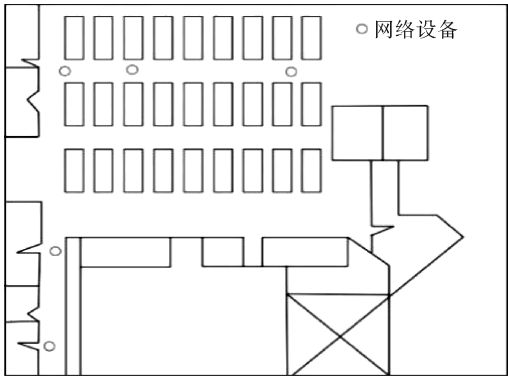


图4 网络设备分布结构图

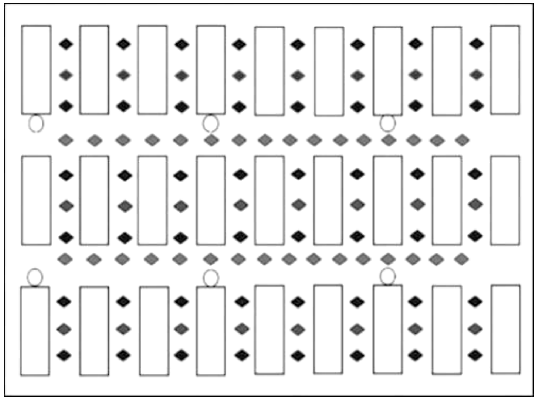


图5 数据采样点位置分布图

图 4 中的圆形表示采集区域的网络设备，图 5 中的矩形表示书架、菱形表示数据采样点的位置，采样点间的距离在 1.5 m ~ 4.5 m 之间，采集区域的覆盖面积约为 154 m²。算法的初始参数设置如表 2 所示。

表 2 仿真参数设置

参数名称	参数值
种群规模	300
最大迭代次数	200
速度范围	[-50, 50]
位置范围	[-100, 100]

图 6 显示了三种优化算法适应度变化情况，PSO 算法在迭代次数达到 20 次时趋于稳定且优于 SSA、FPA 两种算法，收敛速度更快，因此选择采用 PSO 对初始参数进行优化。

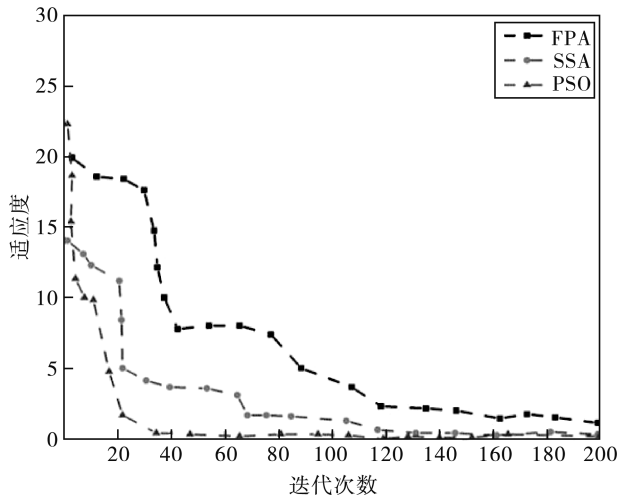


图6 不同优化算法适应度迭代曲线

4.2 不同滤波方法的处理时间对比

为验证不同滤波方法对数据的处理效率，选择对来

自某一参考点数据滤波处理的时间作为评价指标,如表3所示。可以看出 PF-KF 混合滤波算法的平均滤波时间比均值滤波,中位值滤波,偏度-峰度正态检验和卡尔曼滤波(Kurtosis-Skewness and Kalman, KSK)法,高斯、中值滤波、狄克逊检测法(Gaussian, Median and Dixon, GMD)融合法,狄克逊检测法和高斯滤波(Dixon Test and Gaussian, DTG)混合算法的时间分别减少 1.06 s、1.1 s、0.91 s、0.97 s、0.33 s,数据处理效率明显优于其他几种滤波算法。

表3 不同算法的滤波时间对比

算法	滤波时间/s
均值滤波	3.07
中位值滤波	3.11
KSK	2.92
GMD	2.98
DTG	2.34
PF-KF	2.01

4.3 不同滤波方法处理后的定位效果对比

表4展示了室内场景下不同滤波方法处理后的数据对定位性能的影响。可以看出,PF-KF混合滤波方法处理后数据的最大、最小定位误差分别为2.63、0.34,明显优于其他五种滤波方法;相较于均值滤波算法、中位值滤波算法、KSK混合滤波、GMD混合滤波、DTG混合滤波的平均定位误差,PF-KF混合滤波的平均定位误差分别下降44.81%、37.2%、35.67%、33.9%、32.21%。

表4 室内场景下不同滤波方法处理后的定位效果 (m)

滤波方法	最大误差	最小误差	平均误差	均方根误差
均值滤波	3.86	1.02	1.83	2.19
中位值滤波	3.99	1.26	1.61	1.93
KSK	3.67	0.95	1.57	1.65
GMD	3.94	0.43	1.49	1.78
DTG	3.55	0.69	1.53	1.66
PF-KF	2.63	0.34	1.01	1.02

图7给出了室内场景下均值滤波、中位值滤波、KSK、GMD、DTG、PF-KF六种算法预处理后算法的累计分布概率。由图可知,经PF-KF混合滤波处理后的定位误差在1.5 m范围内的概率约为50%,其他五种算法处理后定位误差在1.5 m范围内的概率分别约为30%、31%、39%、41%、45%。

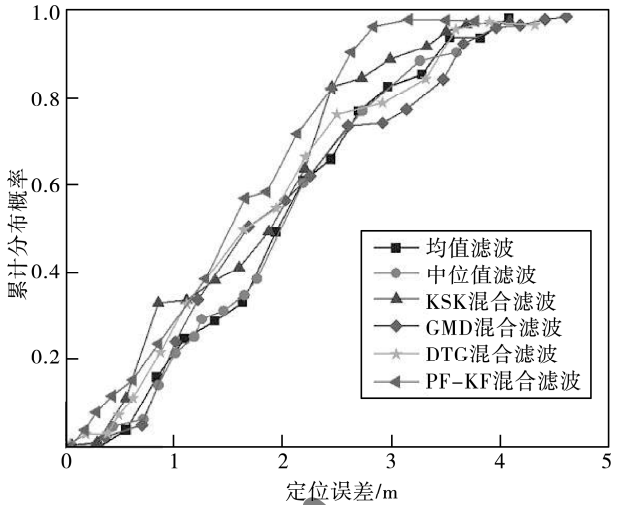


图7 室内场景下误差累计分布概率

表5展示了室外(走廊)场景下不同滤波方法处理后的数据对定位性能的影响。可以看出,PF-KF混合滤波方法处理后数据的最大、最小定位误差分别为2.43、0.23,明显优于其他五种滤波方法;相较于均值滤波算法、中位值滤波算法、KSK、GMD、DTG混合滤波的平均定位误差,PF-KF混合滤波的平均定位误差分别下降58.81%、52.4%、50.3%、30.7%、42.5%;室外环境下的最大、最小、平均定位误差均略低于室内场景,主要是因为室外场景的信号受干扰小,采集的数据受噪声干扰弱。由图8可知,经PF-KF混合滤波处理后的定位误差在2 m范围内的概率为82.9%,其他五种算法处理后定位误差在2 m范围内的概率分别为79.3%、78.1%、66.4%、58.6%、42.7%。

表5 走廊内不同滤波方法处理后的定位效果 (m)

滤波方法	最大误差	最小误差	平均误差	均方根误差
均值滤波	4.02	0.92	2.33	2.01
中位值滤波	4.11	0.98	1.89	1.84
KSK	3.86	0.86	1.77	1.89
GMD	3.73	0.36	1.49	1.62
DTG	3.65	0.48	1.53	1.67
PF-KF	2.43	0.23	0.88	1.13

4.4 算法参数对定位效果的影响

虽然通过滤波方法可以有效处理数据大幅度变化的问题,但仍然存在小幅波动,因此引入可以反映数据微弱变化的波动参数并将它当作一种特征用来训练定位模型。机器学习算法中提供了衡量特征重要性的函数 importance(),表6显示了五种波动特征参数在算法中的重要程度,前四种特征对模型的贡献较大。

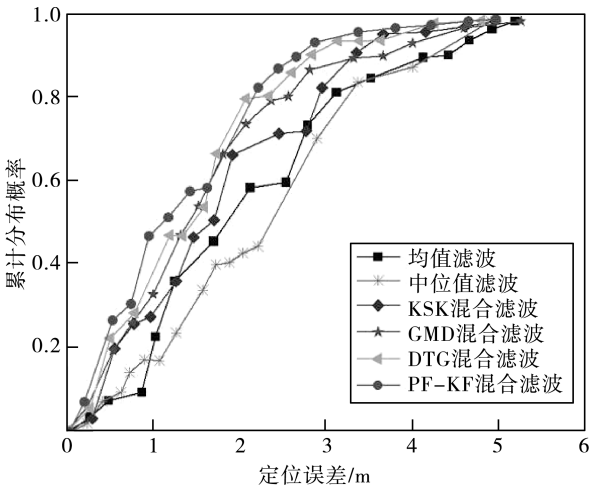


图8 走廊内定位误差累计分布概率

表6 统计特征参数在算法中的重要程度

特征参数	均值	峰度	峰谷差值	偏度	方差
重要程度	0.42	0.28	0.22	0.31	0.18

为了更直观地显示五种统计参数对算法定位性能的影响,首先对表6中不同参数对算法的重要程度进行降序排列,得到0.42(均值)>0.31(偏度)>0.28(峰度)>0.22(峰谷差值)>0.18(方差),然后对排名前五的参数进行逐个添加,观察算法的准确率,结果如图9所示。其中,四种算法分别为加权K近邻(Weighted K-Nearest Neighbors, WKNN)算法、改进随机森林(Improved Random Forest, IRF)算法、XGBoost算法及本文方法。

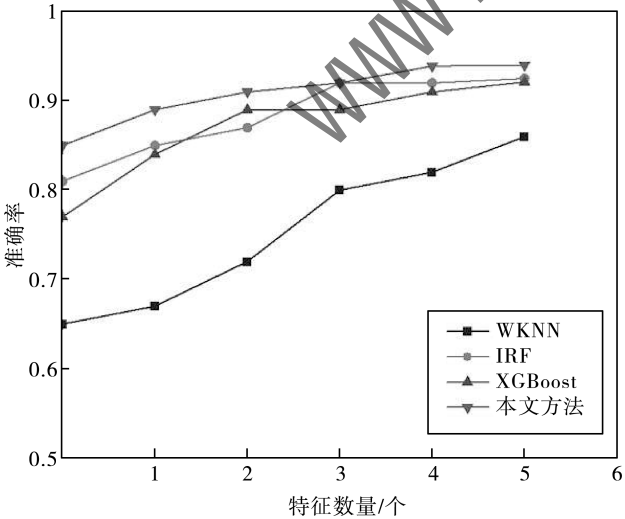


图9 特征数量对算法性能的影响

可以看出,随着特征数目的增加这四种算法准确率并没有明显降低,说明合适的特征数量有助于提升算法性能,均值、偏度、峰度这三个特征对算法性能的贡献

比较大;但随着特征数量的增多,对应的计算成本也会增加。

XGBoost算法是基于分类/回归树模型的一种机器学习算法,图10显示了优化参数(以树深度、迭代次数)对定位效果的影响。由图可以看出,随着迭代次数的增加,平均定位误差逐渐减小;但超过500次迭代时,部分深度的树模型出现误差异常的情况,这主要是因为出现了过拟合现象。一般情况下,树越深模型结构越复杂,定位效果越好,然而模型越复杂导致计算成本增加,随着树深度的增加,算法的收敛速度也有一定的提升。

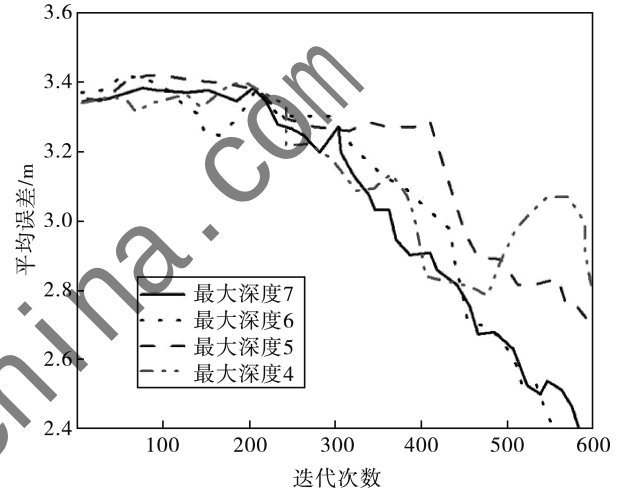


图10 优化参数对定位效果的影响

4.5 本文方法与其他方法的对比

为验证改进方法的有效性,选择XGBoost、PSO改进的XGBoost和FCM改进的XGBoost三种方法与本文方法进行比较。由图11可知,本文方法有81%的样本定位误差在2m范围内,而FCM-XGBoost算法误差在2m范围内的累计分布概率为64.5%、PSO-XGBoost算法误差在2m内的累计分布概率为61.7%;进一步观察可知,本文方法在各个误差范围内的累计分布概率均优于其他三种算法。

为验证本文方法的有效性,将本文方法与其他方法进行对比分析。图12展示了本文方法与K近邻算法(K-Nearest Neighbors, KNN)、AP选择的WKNN、IRF、麻雀搜索算法(Sparrow Search Algorithm, SSA)改进的XGBoost四种算法的对比结果。可以得出,当误差为1m、1.5m、2m时本文方法的累计分布概率分别为48%、71%、83%;高于KNN的27%、32%、46%,AP选择的WKNN的34%、45%、58%,IRF的38%、53%、72%以及SSA-XGBoost的40%、63%、68%,本文算法相比其他算法有明显优势。

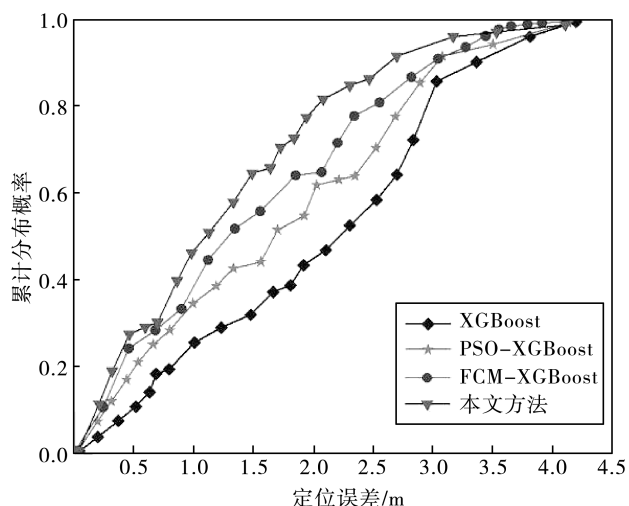


图 11 不同改进策略的累计概率分布

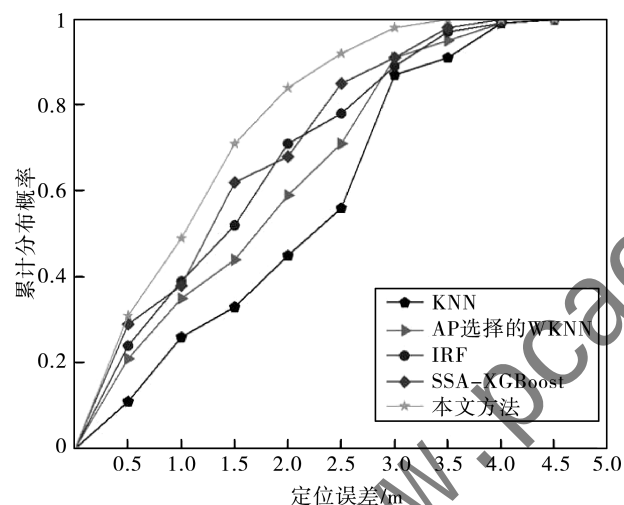


图 12 不同算法的定位效果对比

5 结论

本文针对障碍物阻挡、环境噪声等使测量数据呈现明显波动而影响定位效果的情况,提出一种融合混合滤波与波动参数辅助的 XGBoost 定位方法。首先采用 PF-KF 结合的方法对原始 RSS 样本处理,效果优于单一滤波方法,更有效地抑制了高斯与非高斯噪声干扰;同时将优化后的 RSS 和波动参数双特征结合,防止单一 RSS 特征提供信息有限、不稳定而降低算法预测精度;其次,采用 PSO 对 XGBoost 的初始参数进行动态调整,提升算法的运行效率和预测性能;最后根据目标点的信息估计其位置坐标,同时将该点的信息保存到数据库中完成更新。实验表明,本文方法可以实现较好的定位效果,有效提升了定位精度。后续工作可以融合 AOA、CSI 等信息,进而扩展到多层空间、多类型 AP、多模态传感融合(如

UWB + Wi-Fi) 环境下的室内定位中。

参考文献

- [1] HE S, CHAN S H G. Tilejunction: mitigating signal noise for fingerprint-based indoor localization [J]. IEEE Transactions on Mobile Computing, 2016, 15 (6): 1554 - 1568.
- [2] JO K. Erratum for "Interacting multiple model filter-based sensor fusion of GPS with in-vehicle sensors for real-time vehicle positioning" [J]. IEEE Transactions on Intelligent Transportation Systems, 2012, 13 (2): 973 - 973.
- [3] YOON P K, ZIHAJEZHDADEH S, KANG B S, et al. Robust bio-mechanical model-based 3-D indoor localization and tracking method using UWB and IMU [J]. IEEE Sensors Journal, 2017, 17 (4): 1084 - 1096.
- [4] 王海英, 梁鹏, 阮祺, 等. 基于 ZigBee 的隧道人员定位系统设计 [J]. 兰州交通大学学报, 2016, 35 (3): 94 - 99.
- [5] SHANG S, WANG L X. Overview of WiFi fingerprinting-based indoor positioning [J]. IET Communications, 2022, 16 (7): 725 - 733.
- [6] 陈锐志, 叶锋. 基于 Wi-Fi 信道状态信息的室内定位技术现状综述 [J]. 武汉大学学报 (信息科学版), 2018, 43 (12): 2064 - 2070.
- [7] 崔丽珍, 王巧利, 郭倩倩, 等. 井下基于动态指纹更新的指纹定位算法研究 [J]. 系统仿真学报, 2021, 33 (4): 818 - 824.
- [8] XUE W, QIU W, HUA X, et al. Improved Wi-Fi RSSI measurement for indoor localization [J]. IEEE Sensors Journal, 2017 (7): 2224 - 2230.
- [9] DIVERSI R, GUIDORZI R, SOVERINI U. Kalman filtering in extended noise environments [J]. IEEE Transactions on Automatic Control, 2005, 50 (9): 1396 - 1402.
- [10] CHEN Z H, ZOU H, JIANG H, et al. Fusion of WiFi, smart-phone sensors and landmarks using the Kalman filter for indoor localization [J]. Sensors, 2015, 15 (1): 715 - 732.
- [11] ARULAMPALAM M S, MASKELL S, GORDON N, et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking [J]. IEEE Transactions on Signal Processing: A Publication of the IEEE Signal Processing Society, 2002, 50 (2): 174 - 188.
- [12] WANG F S, ZHANG J X, LIN B W, et al. Two stage particle filter for nonlinear Bayesian estimation [J]. IEEE Access, 2018, 6: 13803 - 13809.
- [13] 董思宇. 基于混合滤波算法的无线传感网络定位系统的研究 [D]. 呼和浩特: 内蒙古大学, 2016.
- [14] ZHUANG Y, LI Y, QI L, et al. A two-filter integration of MEMS sensors and Wi-Fi fingerprinting for indoor positioning [J]. IEEE Sensors Journal, 2016: 5125 - 5126.

(下转第 65 页)

- [12] 孙跃元. 算法决策应用的外部风险及其公共治理路径研究 [J]. 河北法学, 2023, 41 (4): 156-175.
- [13] 关保英. 行政法教科书之总论行政法 [M]. 北京: 中国政法大学出版社, 2005.
- [14] 王明敏. 行政机关利用个人信息进行自动化决策的合宪性检视 [J]. 法学, 2025 (2): 36-50.
- [15] 张凌寒. 算法自动化决策与行政正当程序制度的冲突与调和 [J]. 东方法学, 2020 (6): 4-17.
- [16] 郭春镇, 勇琪. 算法的程序正义 [J]. 中国政法大学学报, 2023 (1): 164-180.
- [17] 王怀勇, 邓若翰. 算法行政: 现实挑战与法律应对 [J]. 行政法学研究, 2022 (4): 104-118.
- [18] 汪庆华. 算法透明的多维维度和算法问责 [J]. 比较法研究, 2020 (6): 163-173.
- [19] 解志勇, 吴梦玉. 发展与共享: 数字社会的权利鸿沟及其平权治理 [J]. 法学评论, 2025, 43 (1): 13-25.
- [20] 查云飞. 自动化行政中的事实认定——以《行政处罚法》第41条为中心 [J]. 行政法学研究, 2024 (2): 33-44.
- [21] 卢克·多梅尔. 算法时代: 新经济的新引擎 [M]. 胡小锐, 钟毅, 译. 北京: 中信出版集团, 2016.
- [22] 马辉. 公共性视角下平台协议的法律规制研究 [J]. 法学家, 2025 (2): 98-115, 194.
- [23] 张恩典. 通过合同实现算法行政问责 [J]. 南昌大学学报 (人文社会科学版), 2024, 55 (3): 76-87.
- [24] 王锡锌. 数治与法治: 数字行政的法治约束 [J]. 中国人民大学学报, 2022, 36 (6): 17-34.
- [25] 邓明峰. 算法行政对正当程序原则的挑战与法律规制研究 [J]. 河北法学, 2024, 42 (6): 112-126.
- [26] 陶锋. 人工智能模拟时代的艺术品 [J]. 南开学报 (哲学社会科学版), 2024 (5): 54-65.
- [27] 唐林垚. 公共治理视域下自动化应用的法律规制 [J]. 交大法学, 2022 (2): 17-27.
- [28] 张恩典. 人工智能算法决策对行政法治的挑战及制度因应 [J]. 行政法学研究, 2020 (4): 34-45.
- [29] 李丹. 数字权利的生成基础及法治化保障 [J]. 求是学刊, 2024, 51 (2): 124-137.
- [30] 斜晓东. 论生成式人工智能推进应用中的能源数据安全“分类分级治理体系”建构——基于国家大安全的审视 [J]. 法学论坛, 2025, 40 (2): 114-127.
- [31] 雷刚. 论算法行政的程序法治: 挑战与回应 [J]. 电子政务, 2024 (6): 81-95.
- [32] 张涛. 个人信用评分的地方实践与法律控制——以福州等7个城市为分析样本 [J]. 行政法学研究, 2020 (1): 116-129.
- [33] 赵精武. 人工智能科技伦理审查制度的体系化建构 [J]. 当代法学, 2025, 39 (1): 84-96.
- [34] BRENNAN T, DIETERICH W, EHRET B. Evaluating the predictive validity of the COMPAS risk and needs assessment system [J]. Criminal Justice and Behavior, 2009, 36 (1): 21-40.
- [35] 张凌寒, 杨建军, 文禹衡, 等. 续章与新篇: 中国人工智能法治研究 2024 年度观察 [J]. 法律适用, 2025 (3): 146-166.
- [36] 姚显森, 王森. 算法行政相对人的权利及其双重保护 [J]. 中国软科学, 2024 (10): 68-83.
- [37] HEISKANEN E. Taming the golem-an experiment in participatory and constructive technology assessment [J]. Science & Technology Studies, 2005, 18 (1): 37-44.
- [38] 丁晓东. 基于信任的自动化决策: 算法解释权的原理反思与制度重构 [J]. 中国法学, 2022 (1): 99-118.
- (收稿日期: 2025-03-18)
-
- 作者简介:**
- 王潘 (2001-), 女, 硕士研究生, 主要研究方向: 宪法与行政法、数字法学。

(上接第 58 页)

- [15] SU Z, PAHLAVAN K, AGU E. Performance evaluation of COVID-19 proximity detection using bluetooth LE signal [J]. IEEE Access, 2021: 938891-38906.
- [16] KUMAR R, TORRES-SOSPEDRA J, CHAURASIYA K V. H2LWRF-PDR: an efficient indoor positioning algorithm using a single Wi-Fi access point and Pedestrian Dead Reckoning [J]. Internet of Things, 2024: 27101271-101271.
- [17] 牟平, 凌铭, 胡锐. 基于改进 AP 选择的融合随机森林室内定位算法 [J]. 全球定位系统, 2021, 46 (5): 33-38.
- [18] 李玉柏, 孙迅. 基于迁移学习提高 WiFi 室内定位中信道状态信息指纹库的鲁棒性 [J]. 电子与信息学报, 2023, 45 (10): 3657-3666.
- [19] 张贺娜, 乐燕芬, 施伟斌. 基于特征降维的核岭回归室内定位算法 [J]. 仪器仪表学报, 2020, 41 (10): 83-91.
- [20] 钱宇, 何益丰, 谢斌鑫, 等. 基于特征工程的 XGBoost 风速短期预测 [J]. 武汉大学学报 (工学版), 2022, 55 (10): 1057-1064.
- [21] GERNÁN M S, PHILIPP R, JOAQUÍN T S, et al. Long-Term Wi-Fi fingerprinting dataset for research on robust indoor positioning [J]. Data, 2018, 3 (1): 3-3.
- (收稿日期: 2025-05-09)
-
- 作者简介:**
- 刘高辉 (1968-), 男, 博士, 教授, 主要研究方向: 通信信号处理、通信辐射源识别、随机共振、无源定位等。
- 凌凤智 (2000-), 通信作者, 男, 硕士研究生, 主要研究方向: 室内定位. E-mail: 2826259302@qq.com。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部

www.pcachina.com