

基于局部线性嵌入和梯度提升决策树的信道状态信息室内指纹定位算法研究

李新春¹, 赵忠婷^{2*}, 于洪仕¹

¹辽宁工程技术大学电子与信息工程学院, 辽宁 葫芦岛 125105;

²辽宁工程技术大学研究生院, 辽宁 葫芦岛 125105

摘要 针对室内定位中相邻定位点之间信道状态信息(CSI)指纹特征易模糊和定位算法鲁棒性低等问题,提出了一种基于局部线性嵌入(LLE)和梯度提升决策树(GBDT)的定位算法。离线阶段,首先将经过预处理的幅值和相位作为CSI联合指纹,然后再用LLE降维前用弹性网络(EN)对联合指纹的不同子载波进行加权,既保证了降维后CSI指纹的真实性又增强了其独有的特征,最后用基于果蝇优化算法(FOA)改进的GBDT算法训练降维后的数据以提高CSI指纹的可靠性和稳定性,并建立指纹库。在线阶段,将待测点的联合指纹代入LLE+GBDT算法训练出指纹信息,然后通过与指纹库匹配预测出待测点的物理位置。室内定位实验结果表明,所提算法相较于对比算法具有较高的定位精度和鲁棒性,具有一定的应用价值。

关键词 室内定位; 信道状态信息; 梯度提升决策树; 局部线性嵌入

中图分类号 TN925+.93

文献标志码 A

doi: 10.3788/LOP202259.0215008

Channel State Information Indoor Fingerprint Localization Algorithm Based on Locally Linear Embedding and Gradient Boosting Decision Tree

Li Xinchun¹, Zhao Zhongting^{2*}, Yu Hongshi¹

¹School of Electronics and Information Engineering, Liaoning Technical University, Huludao, Liaoning 125105, China;

²Graduate School, Liaoning Technical University, Huludao, Liaoning 125105, China

Abstract Aiming at the problems such as the low algorithm robustness and easily blurred fingerprint of channel state information (CSI) in indoor localization, a location algorithm based on locally linear embedding (LLE) and gradient boosting decision tree (GBDT) is proposed. In the offline stage, first, the preprocessed amplitude and phase are regarded as joint CSI fingerprints, and then the individual subcarriers of the joint fingerprint are weighted using the elastic network (EN) algorithm before dimensionality reduction by LLE, which not only ensures the authenticity of CSI fingerprint after dimensionality reduction, but also enhances its unique characteristics. Finally, GBDT algorithm based on fruit fly optimization algorithm (FOA) is used to train the reduced dimension data to improve the reliability and stability of CSI fingerprint, and the fingerprint database is established. In the online stage, the LLE+GBDT algorithm is adopted to find the fingerprint information of the test point, so that actual physical locations can be predicted by matching with the fingerprint library. The indoor localization experiments results show that the proposed algorithm has higher localization accuracy and robustness compared with the comparison algorithm, and has certain application value.

Key words indoor localization; channel state information; gradient boosting decision tree; locally linear embedding

收稿日期: 2021-07-12; 修回日期: 2021-08-10; 录用日期: 2021-09-08

基金项目: 国家自然科学基金(61372058)、辽宁省教育厅青年项目(NOLJ2019QL024)

通信作者: *1213259067@qq.com

1 引言

现如今,各种大型建筑物内部已经成为了人类活动的主要聚集地,如机场、商场、火车站等,因此高精度的室内定位成为了基于位置的定位服务(LBS)^[1]的一个热门研究方向。室内定位技术主要包括几何定位和指纹定位^[2-3],其中指纹定位需要的设备少、操作简单且定位精度高,被诸多学者当作研究重点。目前可以作为指纹特征的数据主要有接收信号强度指示(RSSI)^[4-5]和信道状态信息(CSI),其中RSSI是多径幅度叠加的结果,易受多径效应的影响,从而导致指纹特征不完整;CSI反映子信道在各个空间的频率特性,因此可将CSI的幅度与相位作为联合指纹,来保证指纹特征的完整性。

华盛顿大学的Halperin等^[6-7]于2010年通过改写驱动,实现了从商业Intel 5300网卡中获取Wi-Fi环境下的CSI,使得在Wi-Fi环境下获取CSI成为可能。到目前为止,已有多篇文献对CSI指纹定位进行了深入研究,文献[8]利用深度神经网络(DNN)算法进行指纹位置预估定位,但只用了CSI的幅度信息;文献[9]利用改进的加权K近邻(WKNN)算法进行定位,但只用了CSI的相位信息。这两种方法虽然有一定的定位精度,但都不能完整地表达定位点的指纹特征。文献[10]把CSI的幅值和相位结合作为联合指纹,然后分别用Bayes算法和KNN算法实现室内定位,保证了指纹特征的完整性,但联合指纹增加了数据维度,从而导致数据冗余,定位存在延时。为了避免数据冗余、缩短定位时长,文献[11]和文献[12]分别利用主成分分析(PCA)和卷积自编码器(CAE)对指纹特征降维,但降维后的指纹特征不能准确地表达指纹特征之间的非线性关系,从而导致定位点模糊,最终影响定位效果。文献[13]指出在室内定位算法中,合适的机器学习算法比Bayes算法、KNN算法和神经网络算法具有更好的鲁棒性,能提高定位精度。

综上所述,本文提出了一种基于局部线性嵌入(LLE)和梯度提升决策树(GBDT)的定位算法。首先采用弹性网络(EN)算法分析每个参考点(RP)上CSI联合指纹不同子载波的特征,并进行适当的加权处理,加大相邻参考点之间指纹特征的区别。然后参照文献[14-15],用GBDT算法建立室内定位

模型,同时考虑到GBDT算法对低维数据具有更高的预测精度,在数据训练之前用局部线性嵌入(LLE)算法对每个参考点的CSI联合指纹进行降维处理。整个定位过程需要提前在Wi-Fi环境下提取所有参考点的CSI原始数据,离线阶段将每个参考点采集到的CSI数据进行异常值处理和相位矫正,处理后的幅值和相位作为联合指纹,代入EN算法加权,加权后的指纹特征输入LLE算法对数据进行局部划分,保留局部最具代表性的数据作为降维后的数据,避免了数据冗余。将这两种算法组成elastic net locally linear embedding (ENLLE)算法,最后用GBDT算法训练降维后的数据,并通过果蝇优化算法(FOA)的迭代,不断寻求GBDT算法中所有超参数的全局最优值,从而得到每个参考点的指纹信息并联合对应的物理位置信息构建指纹库,将这两种算法组成FOA-GBDT算法。在整个室内定位算法中,连接ENLLE算法和FOA-GBDT算法,并简化命名为LLE+GBDT算法。在线阶段对随机选取的测试点(TP)进行幅值、相位预处理后,用LLE+GBDT算法对其进行物理位置估计。

2 CSI数据的采集与预处理

2.1 CSI介绍

CSI是描述在利用正交频分复用(OFDM)方式传输信号时发射机与接收机之间信道属性的物理量。与RSSI相比,CSI具有更细粒度的物理信息。采集CSI信息时,发送端选用对障碍物具有较好穿透力的5 GHz Wi-Fi信号,接收端选用终端装有的无线网卡的天线来收集CSI数据。Wi-Fi所用的协议是IEEE 802.11n,此协议的无线通信标准是基于OFDM和多输入多输出(MIMO)的^[16],故OFDM系统的信道模型可以描述为

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}, \quad (1)$$

式中: \mathbf{Y} 表示接收信号; \mathbf{X} 表示发送信号; \mathbf{N} 表示高斯白噪声; \mathbf{H} 表示信道频率响应。

选取OFDM系统中的30个子载波信号作为CSI原始数据^[17],即 $\mathbf{Y} = [H_1, H_2, \dots, H_{30}]$,第*i*个子载波的信道频率响应为

$$H_i = |H_i| \exp(j\angle H_i), \quad (2)$$

式中: $|H_i|$ 和 $\angle H_i$ 分别表示第*i*个子载波的幅度响应和相位响应, $i = 1, 2, \dots, 30$ 。

假设有*m*根发送天线*n*根接收天线,则某一位置点采集的CSI数据集可以表示为

$$\mathbf{Y}_{\text{CSI}} = \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1m} \\ H_{21} & H_{22} & \cdots & H_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ H_{n1} & H_{n2} & \cdots & H_{nm} \end{bmatrix}, \quad (3)$$

式中： H_{nm} 表示第 m 个发射天线和第 n 个接收天线之间 CSI 的幅度和相位。

所提算法通过 Ubuntu 操作系统中提供的函数将采集到的 CSI 幅值数据和相位数据保存成 .dat 格式,供后续定位使用。

2.2 幅值滤波

实验人员的走动、噪声和室外环境的变化等都可能使 CSI 幅值产生较大波动而出现异常值。为了在去除异常值的同时保证数据的完整,所提算法将同一天线中不同子载波的中值作为参考值,将该中值三倍及以上的幅值视为异常值,并用中值将其替换。图 1(a)表示滤波前的幅值,图 1(b)表示滤波后的幅值。从图中可以看出,滤波后无异常值,表明中值替换法能够有效地去除异常值。

2.3 相位矫正

在采集 CSI 数据时,由于收发设备不稳定,信号的接入点(AP)与采样点之间信道中心频率不完全同步会造成载波频移(CFO),AP与采样点之间采样频率时钟不同步会造成采样频移(SFO),因此实际获得的第 i 个 CSI 的相位信息为

$$\theta_i = \angle H(f_i) + 2\pi\omega \frac{m_i}{N} \Delta t + \beta + \delta_i, \quad (4)$$

式中： $\angle H(f_i)$ 表示真实的 CSI 相位； $2\pi\omega \frac{m_i}{N}$ 和 β 分别表示由 SFO 和 CFO 引起的相位偏移^[18], m_i 表示第 i 个子载波的索引号, N 表示快速傅里叶变换

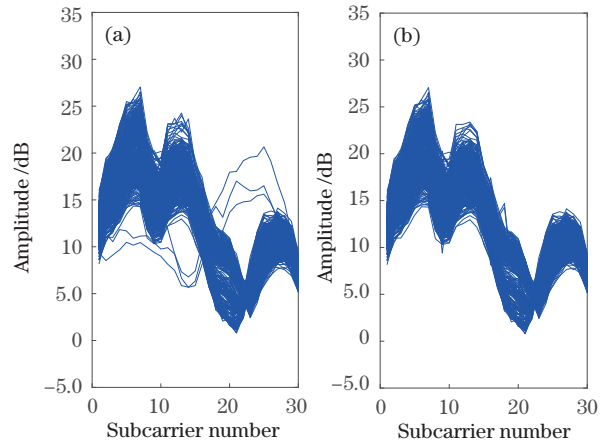


图 1 滤波处理前后对比。(a)滤波处理前;(b)滤波处理后
Fig. 1 Comparison before and after filtering. (a) Before filtering; (b) after filtering

(FFT) 的长度； Δt 表示非同步时钟引起的时延； δ_i 表示测量噪声,可忽略。

结合(4)式,采用线性拟合的方法进行相位矫正^[19],先选取两个中间量 k 与截距 b ,表达式为

$$k = \frac{\theta_{30} - \theta_1}{m_{30} - m_1}, \quad (5)$$

$$b = \frac{1}{30} \sum_{i=1}^{30} \theta_i. \quad (6)$$

再使用 θ_i 减去 $(km_i + b)$ 得到校正后的 CSI 相位,表达式为

$$\angle H(f_i)' = \theta_i - (km_i + b). \quad (7)$$

CSI 相位通过上述算法矫正后可抵消 CFO 与 SFO 引起的偏差,图 2(a)、(b)分别表示线性拟合前后的相位。

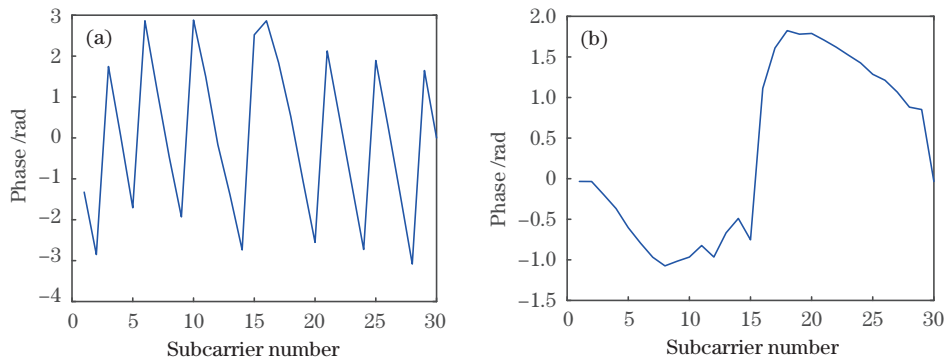


图 2 相位矫正前后对比。(a)矫正前;(b)矫正后

Fig. 2 Comparison before and after phase correction. (a) Before correction; (b) after correction

3 基于 LLE+GBDT 的定位算法研究

所提定位算法需要提前规划好每个 RP 的位

置,整个定位过程分为离线阶段和在线阶段两部分,这两部分的理论原理一致。离线阶段是为了训练所有 RP,并建立离线指纹库;在线阶段则是为了

训练出 TP 的指纹信息,方便与离线指纹库匹配实现室内定位。首先采集所需采样点的原始 CSI 数据,并进行异常值处理和相位矫正,以便作为联合

指纹;然后用 ENLLE 算法降维,最后将降维后的联合指纹特征用 FOA-GBDT 算法训练生成指纹信息。室内定位流程如图 3 所示。

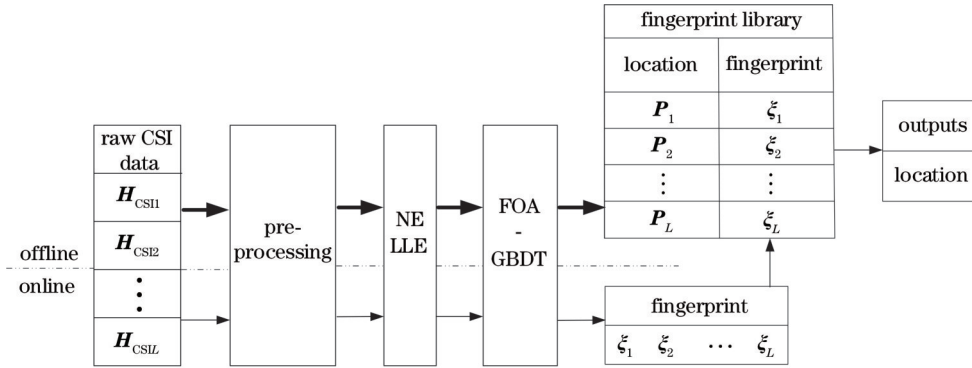


图 3 室内定位流程

Fig. 3 Process of indoor positioning

3.1 ENLLE 算法

在采集数据时,实际采集到的 CSI 数据是一个 $2 \times 3 \times 30$ 的复数矩阵,其中 30 代表子载波数,每个子载波至少采集 1000 个数据包。为了避免数据冗余同时减少计算量,所提算法对 CSI 数据进行降维,但降维会带来另一个突出问题:原始 CSI 数据中同一位置点的不同子载波会产生不同的波动,即一些子载波幅度会有较大的起伏,一些子载波幅度仍保持稳定,这是区分不同位置点的重要特征,降维后的数据会削弱这一特征。为了解决这一问题,提出了先用 EN 回归^[20]对同一天线的不同子载波进行适当的加权,然后用 LLE^[21-22]对 CSI 数据进行降维,以此提高降维后的数据对复杂环境的表现力。

LLE 算法属于流形学习,将高维数据划分为不同局部数据的组合,每个局部都用权值最小的邻域数据代表整体数据,从而实现降维。与其他降维算法相比,LLE 算法能在降维时保持数据局部的线性特征,最大程度地保证了原数据的真实性。

假设有 L 个位置点,每个点包含 W 个 CSI 数据,并用 ξ 来表示每个点的 CSI 联合指纹数据,则 ENLLE 算法的主要步骤如下。

1) 定义 $\xi_{l1}, \xi_{l2}, \dots, \xi_{lw}$ 是第 l 个点的所有原变量特征,用 P_l 来标记不同位置的点,有

$$\begin{cases} P_1 = \omega_{11}\xi_{11} + \omega_{12}\xi_{12} + \dots + \omega_{1W}\xi_{1W} \\ P_2 = \omega_{21}\xi_{21} + \omega_{22}\xi_{22} + \dots + \omega_{2W}\xi_{2W} \\ \vdots \\ P_L = \omega_{L1}\xi_{L1} + \omega_{L2}\xi_{L2} + \dots + \omega_{LW}\xi_{LW} \end{cases}, \quad (8)$$

式中:系数 ω_{lw} 为原变量特征的属性权重系数, $l = 1, 2, \dots, L, w = 1, 2, \dots, W$ 。

2) 用 EN 回归求解(8)式中的属性权重系数的最优解,即 ω_{lw} ,得出新的 CSI 数据,其中 λ_2, λ_1 为给定常数,可用果蝇优化算法寻求其最优值。

$$\hat{\beta} = \min \left(\|P_l - \xi_l \beta^T\|_2^2 + \lambda_2 \|\beta\|_2^2 + \lambda_1 \|\beta\|_1 \right). \quad (9)$$

3) 假设一个点的 W 维中所有数据属于同一流形 $A_e, e = 1, 2, \dots, a + 1$, 计算每个数据的 $H (H \geq 3)$ 个近邻点,其中 V_{eh} 为 A_e 与它的 H 个近邻点之间的权重。

$$A_e = \sum_{h=1}^H V_{eh} A_{eh}. \quad (10)$$

4) 通过差平方和取最小值来计算 V_{eh} , 即

$$\begin{cases} \min \left(\sum_{e=1}^{a+1} \left\| A_e - \sum_{h=1}^H V_{eh} A_{eh} \right\|^2 \right) \\ \sum_{h=1}^H V_{eh} = 1 \end{cases}. \quad (11)$$

5) 保持 V_{eh} 固定不变化,在低维空间 $W_{low} (W_{low} < W)$ 中对原来的数据点进行重构,用特征值求解得到的 B 作为降维后的数据,即

$$\min \left(\sum_{e=1}^{a+1} \left\| B_e - \sum_{h=1}^H \omega_{eh} B_{eh} \right\|^2 \right). \quad (12)$$

3.2 改进的 GBDT 算法

GBDT 算法^[23]将分类与回归树(CART)中的回归树作为弱学习器,通过多个弱学习器的“集合”,组成一个新的强学习器。GBDT 采用向前分步法,利用损失函数的负梯度值作为残差向前拟合构建回归

树,最后通过求和每棵回归树误差最小的预测值,从而得到最终预测结果。上述过程如图 4 所示。

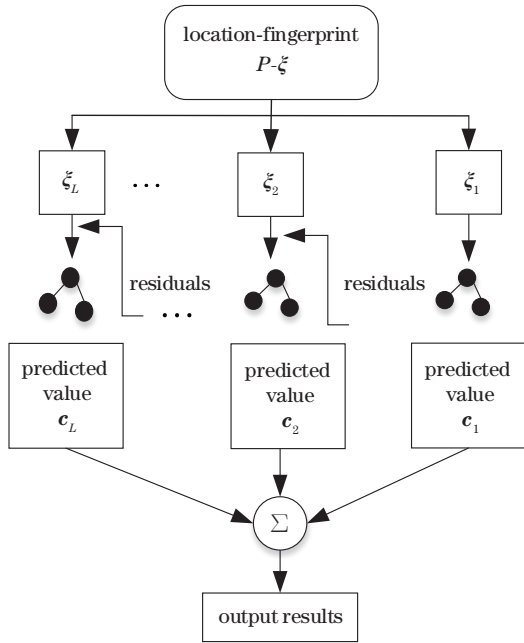


图 4 GBDT 原理示意图
Fig. 4 Schematic of GBDT

假设 $\xi_l = [\xi_{l1}, \xi_{l2}, \dots, \xi_{lW_{low}}]^T$ 和 $P_l = (x_l, y_l)$ 分别表示经过 ENLLE 降维后第 l 个位置点的 CSI 指纹特征和位置坐标, $l = 1, 2, \dots, L$ 。初始化第 1 棵回归树 $f_0(\xi)$, 有

$$f_0(\xi) = \arg \min_c \sum_{l=1}^L L_o(\xi_l, c_l), \quad (13)$$

式中: c_l 为第 l 个位置点的指纹信息估计值; L_o 为模型的损失函数。

为了提高定位模型位置预测值的精度, 选择具有高鲁棒性的 Huber 作为损失函数, 函数中的 δ 为超参数, δ 用 FOA 寻求最优值。

$$L_\delta[\xi_l, f(\xi_l)] = \begin{cases} \frac{1}{2} [f(\xi_l) - \xi_l]^2, & |f(\xi_l) - \xi_l| \leq \delta \\ \delta \left(|f(\xi_l) - \xi_l| - \frac{\delta}{2} \right), & |f(\xi_l) - \xi_l| > \delta \end{cases}. \quad (14)$$

然后计算(13)式的残差, 拟合下一棵回归树。第 $t-1$ 棵回归树的残差可以表示为

$$r_u = -\frac{\partial L_\delta[\xi_l, f_{t-1}(\xi_l)]}{\partial f_{t-1}(\xi_l)}, \quad l = 1, 2, \dots, L. \quad (15)$$

将 (ξ_l, r_u) 作为第 t 棵回归树的训练数据集, 将 (ξ_l, r_u) 划分出两个子区域 R_1, R_2 , 每次划分都依照

使得残差与预测值之间差的平方最小来确定最优切分量 j 和切分点 s 。

$$\min_{j^s} \left[\min_{c_1} \sum_{\xi_l \in R_1(j,s)} (r_{il} - c_1)^2 + \min_{c_2} \sum_{\xi_l \in R_2(j,s)} (r_{il} - c_2)^2 \right], \quad (16)$$

通过(16)式算出的 (j, s) 来确定子区域的输出值,

$$\begin{cases} R_{1l}(j, s) = \{\xi_l | \xi_l^{(j)} \leq s\} \\ R_{2l}(j, s) = \{\xi_l | \xi_l^{(j)} > s\} \end{cases}. \quad (17)$$

重复(16)式和(17)式的划分过程, 将第 t 棵回归树划分为 J 个互不相干的子区域 $R_{jt} (j = 1, 2, 3, \dots, J)$, 线性扫描各个子区域的输出值, 找到误差最小的输出值, 通过求和得到回归树的精确预测值。

$$c_{jt} = \arg \min_c \sum_{\xi \in R_{jt}} L_\delta[\xi_l, f_{t-1}(\xi_l) + c]. \quad (18)$$

更新第 t 棵回归树, 有

$$f_t(\xi) = f_{t-1}(\xi) + \sum_{j=1}^J c_{jt} I, \quad \xi \in R_{jt}, \quad (19)$$

式中: $\sum_{j=1}^J c_{jt} I$ 是每个回归树的模型, 且 $I = \begin{cases} 1, & \xi \in R_{jt} \\ 0, & \xi \notin R_{jt} \end{cases}$ 。

将 T 次迭代生成的回归树进行叠加, 有

$$f_T(\xi) = \sum_{t=1}^T \sum_{j=1}^J c_{jt} I, \quad \xi \in R_{jt}. \quad (20)$$

GBDT 通过最小化损失函数来更新模型, 为防止训练模型发生过度拟合并提升预测精度, 又在回归树模型中加入一个压缩因子 $v, 0 < v \leq 1$ 。最终 GBDT 的定位模型为

$$f(\xi) = \sum_{t=1}^T \sum_{j=1}^J v c_{jt} I, \quad \xi \in R_{jt}. \quad (21)$$

综上所述, GBDT 中有很多超参数, 如用于构建 CART 的迭代次数 T 、压缩因子 v , 还有损失函数的超参数 δ 。其中弱学习器的 T 过小易造成欠拟合, 过大又会形成过拟合; v 过小会导致寻优时间增加, 过大又容易跳过最优值; δ 的大小直接影响了损失函数的性能, 因此 δ 的选择至关重要; 它们共同决定了 GBDT 算法的好坏。为了提高 GBDT 的泛化能力和位置预测的准确性, 采用 FOA 优化算法对相关参数进行选择, 并命名为 FOA-GBDT 算法。

FOA 算法仿照果蝇强大的视觉能力及对气味的敏感性来找食物的特性, 通过不断迭代寻找参数的最优值^[24], 搜寻过程如图 5 所示。

图 5 中 $F_{ly1}, F_{ly2}, F_{ly3}$ 共属一个果蝇群, 它们在经过上一次搜索寻优之后分别到达了现在的位置

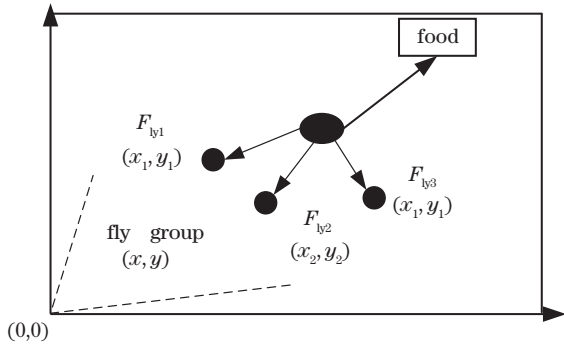


图 5 果蝇群体觅食示意图

Fig. 5 Schematic diagram of fruit fly population foraging $(x_1, y_1), (x_2, y_2), (x_3, y_3)$.

设定果蝇种群的数量为 q , GBDT 的超参数群可表示为

$$\Gamma = \{ \sigma_s | \sigma_s = (\sigma_{1s}, \sigma_{2s}, \dots, \sigma_{zs}), s = 1, 2, \dots, q \}, \quad (22)$$

式中: σ_s 为果蝇种群向量; 优化的超参数个数为 z 。

基于 FOA-GBDT 算法步骤如下。

1) 初始化 FOA 参数, 随机给定待优化参数的初始种群位置 (x_{u0}, y_{u0}) 其中 $u = 1, 2, \dots, z$, 搜索范围 $[x_{u \min}, x_{u \max}]$ 和最大迭代次数 T_F 。

2) 确定果蝇个体位置坐标, 随机赋予果蝇个体搜索食物的位置和方向:

$$\begin{cases} x_{ut} = x_{u0} + r_{\text{and}}(\cdot) \\ y_{ut} = y_{u0} + r_{\text{and}}(\cdot) \end{cases}, \quad (23)$$

式中: $r_{\text{and}}(\cdot)$ 为随机搜索的距离。

3) 确定果蝇坐标与坐标原点的距离 D_{ut} , 计算浓度判定值 S_{ut} :

$$D_{ut} = \sqrt{x_{ut}^2 + y_{ut}^2}, \quad (24)$$

$$S_{ut} = \frac{1}{D_{ut}}. \quad (25)$$

4) 将 GBDT 模型预测的准确率作为浓度判定函数 f_s , 带入判定值 S_{ut} 判定果蝇个体的味道浓度。

5) 寻找最优值, 求种群中离食物位置最近的果蝇 $F_{\text{ly max}}$, 即求浓度判定函数的极大值 $f_{s \max}$ 对应的果蝇个体 I_{best} , 有

$$\begin{cases} f_{s \max} = \arg \max f_s(\sigma_1, \sigma_2, \dots, \sigma_z) \\ I_{\text{best}} = F_{\text{ly max}} \end{cases}, \quad (26)$$

6) 确定果蝇最佳气味浓度和对应的最佳位置, 该位置为下一次迭代的最优位置:

$$\begin{cases} f_{s_{l+1}} = f_{s \max} \\ x_{l+1} = x_{I_{\text{best}}}, \\ y_{l+1} = y_{I_{\text{best}}} \end{cases}, \quad (27)$$

7) 重复运行步骤 2~6 直到满足最大迭代次数或收敛条件。

3.3 LLE+GBDT 算法

因为 GBDT 算法对低维数据的预测精度更高, 所以需要先用 ENLLE 对预处理后的数据进行降维, 同时使指纹特征更准确地表达复杂环境的变化。在后续的数据处理阶段中, 为了使 GBDT 算法处于最优状态, 又利用 FOA 算法优化 GBDT 算法中的超参数的全局最优值, 最终建立了高精度的 LLE+GBDT 室内定位算法, 工作原理如图 6 所示。

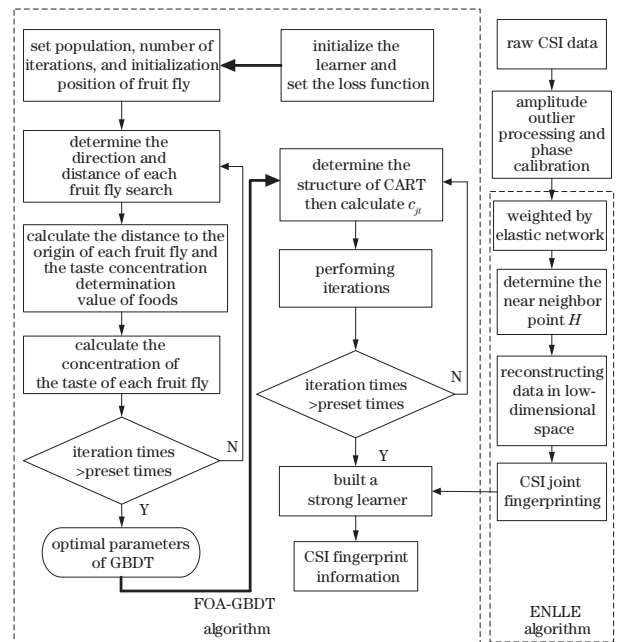


图 6 LLE+GBDT 模型流程图

Fig. 6 Flow chart of LLE+GBDT model

4 实验验证

4.1 实验方案

为了研究 LLE+GBDT 算法在实际环境中的定位性能, 在辽宁工程技术大学的实验楼东 302 和南 411 分别模拟复杂办公室环境和理想空旷环境采集数据, 实际环境如图 7(a)、(b) 所示。选用一台两天线的 TP-Link 路由器作为 AP, 发射信号, 一台装有三天线 Intel 5300 网卡的笔记本电脑接收信号, 采集各个位置的 CSI 数据, 设置每根天线采集 30 个子载波, 每个子载波至少包含 1000 个数据包, 最后通过 CSI tool 工具提取无线网卡里 $2 \times 3 \times 30$

的 CSI 原始数据。采集到的数据分别在 MATLAB R2019b 软件和 Spyder 软件上完成数据处理和最终定位仿真实验。所提算法的定位性能

分别用定位准确度、真实位置和定位结果的估计误差和累积分布函数(CDF)来判定。

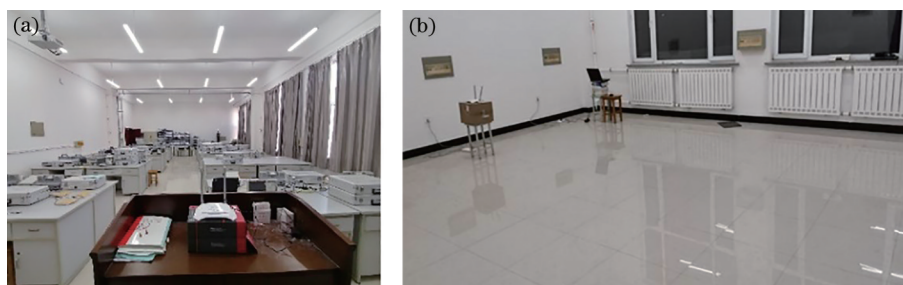


图7 实验环境。(a)东 302;(a)南 411

Fig. 7 Experimental environment. (a) East 302; (a) south 411

4.2 采样间隔

为了研究所提定位算法在不同采样间隔下的定位性能,在辽宁工程技术大学实验楼东 302 选取面积为 $4.8\text{ m} \times 4.8\text{ m}$ 的实验区域,并设置了 3 类参考点,它们的间隔分别是 1.2, 0.9, 0.6 m, 然后分别用等间隔采样法采集 20、30、81 个参考点。本次实验的采样点分布如图 8 所示。

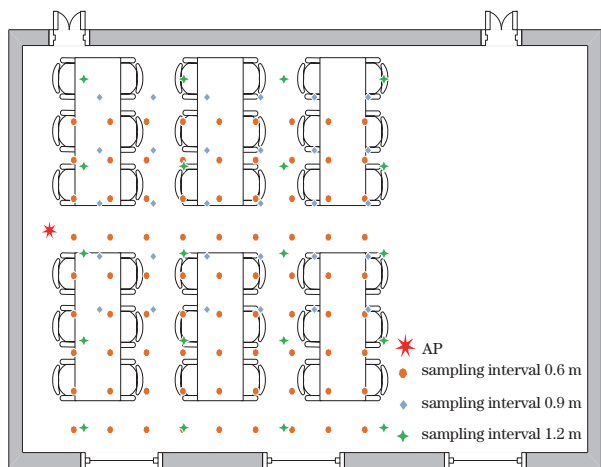


图 8 不同采样间隔实验场景图

Fig. 8 Experimental scene diagrams for different sampling intervals

表 1 为不同采样间隔的定位性能比较。从表中可以看出:1.2 m, 0.9 m 和 0.6 m 这 3 种采样间隔的定位准确度始终保持在 97% 以上,说明所提 LLE+GBDT 定位算法即使在分米级也能很好地区分不同的位置;LLE+GBDT 算法的训练时间会随着采样间隔的缩小呈指数型增长,且定位性能也会降低,采样间隔为 1.2 m 和 0.9 m 的训练时间相差不大且远远小于 0.6 m 的训练时间。故为了保证能在实验楼东 302 采集足够数量的采样点,同时保持良好的定位

性能,在后续实验中将 0.9 m 作为采样间隔。

表 1 不同采样间隔的定位性能比较

Table 1 Comparison of positioning performances at different sampling intervals

Positioning performance	Sampling interval /m		
	1.2	0.9	0.6
Positioning accuracy /%	97.8	97.6	97.3
Average positioning error /m	0.035	0.047	0.063
Training time /s	1.71	15.33	345.98

4.3 ENLLE 算法对定位性能的影响

ENLLE 算法不但减少了模型的训练时间、避免了信息冗余,还使降维后的 CSI 联合指纹能更准确地表达复杂环境的特征。为了验证 ENLLE 算法的定位性能,选用第 4.2 节中采样间隔为 0.9 m 的测试集和训练集作为本次实验的实验数据,用第 2.2 节和 2.3 节中的方法对原始 CSI 数据进行异常值处理和相位矫正,将处理后的 CSI 的幅值和相位作为联合指纹,然后分别用 ENLLE+GBDT 算法、LLE+GBDT 算法、PCA+GBDT 算法进行室内定位,定位性能比较如表 2 和图 9 所示。

表 2 不同降维算法的定位性能

Table 2 Localization performance of different dimensionality reduction algorithms

Algorithm	Average positioning error /m
PCA+GBDT	0.262
LLE+GBDT	0.155
ENLLE+GBDT	0.068

从表 2 中可以看出,在平均定位误差上,ENLLE+GBDT 算法比 PCA+GBDT 算法降低了 74.05%,比 LLE+GBDT 算法降低了 56.13%。从

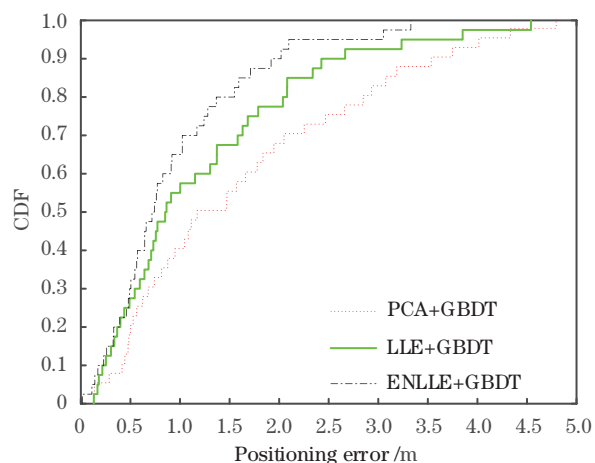


图 9 不同降维算法的 CDF

Fig. 9 CDF of different dimensionality reduction algorithms

图 9 可以看出, ENLLE+GBDT、LLE+GBDT、PCA+GBDT 在定位误差小于 2.5 m 时的累积概率分布 (CDF) 分别达到了 95.86%, 90.01%, 76.21%, 说明通过 ENLLE 算法处理后, 每一个采样点的指纹特征会明显区别于其他点的指纹特征, 因此提升了定位准确率, 从而表明了 ENLLE 算法的有效性。

4.4 果蝇优化算法对定位性能的影响

FOA 算法能够快速且准确地找到超参数的全局最优值, 为了验证 FOA 算法的优化效果, 选用第 4.2 节中采样间隔为 0.9 m 的训练集和测试集数据作为本次实验的实验数据, 用第 2.2 节和 2.3 节中的方法对原始 CSI 数据进行异常值处理和相位矫正, 将处理后的 CSI 的幅值和相位作为联合指纹, 然后分别用 FOA, 粒子群算法 (PSO)^[25] 和 LightGBM 算法^[26] 优化 GBDT, 设置最大迭代次数为 200。图 10 为这 3 种算法的定位准确度分布情况。

从图 10 可以看出: GBDT 算法在分别经 FOA、PSO 和 LightGBM 这 3 种算法优化后, 定位准确率分别可以达到 96.11%, 94.01%, 92.02%; 且 FOA 算法在迭代 50 次左右时, 就使得 GBDT 算法达到了最优, 定位准确度也始终优于与其他对比算法的优化结果。这表明了用 FOA 算法能快速准确地搜寻到 GBDT 中超参数的全局最优值, 是这 3 种优化算法中的最优选择。

4.5 经典定位算法的鲁棒性能比较

考虑到在实际环境中, 复杂环境的影响会导致数据在采集过程中出现丢失从而导致数据集不完整, 也就是说存在缺失值。经过多次采集实验数据, 并分析统计实验数据发现, 缺失值在本次实验

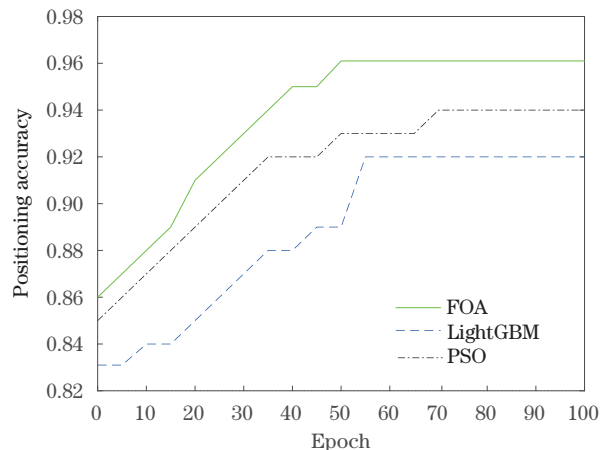


图 10 不同优化算法的定位准确度

Fig. 10 Positioning accuracy of different optimization algorithms

的占比为 2% 左右。

缺失值的存在非常考验算法的鲁棒性, 为了验证 LLE+GBDT 定位算法的鲁棒性与定位性能, 对所提算法与文献[8]中将 CSI 的幅度作为指纹特征, 先用 density-based spatial clustering of applications with noise (DBSCAN) 算法对幅度指纹特征降噪, 然后用 PCA 加 DNN 算法定位的方法、文献[9]中将通过线性变换矫正后的 CSI 相位作为指纹特征, 然后用改进的 WKNN 算法定位的方法、文献[10]中将 CSI 的幅度和相位作为联合指纹特征, 先用拉依达准则去除幅值异常值, 线性变换矫正相位, 然后用 KNN 算法定位的方法进行对比实验。本次实验在第 4.1 节理想空旷环境中划分出 4.5 m × 8.1 m 的区域作为实验场地, 采样间隔取 0.9 m, 等间隔采样 60 个点, 然后对采集到的实验数据集进行修改, 使之组成含有 2% 缺失值的对照数据, 将这两组数据分别用以上 4 种算法进行定位。表 3 为不同算法在没有缺失值和有缺失值情况下的定位性能比较, 图 11 为在有缺失值的情况下, 4 种算法的累积概率分布。

从表 3 可以看出, 缺失值对 LLE+GBDT 算法的影响要小于其他算法。在数据完整的情况下, LLE+GBDT 算法的定位准确率和平均定位误差略优于联合指纹+KNN 和幅度指纹+PCA-DNN 算法, 而相位指纹+WKNN 的平均定位误差较大。当数据中存在缺失值时, LLE+GBDT 算法的定位准确率降低了 1.4 个百分点, 平均定位误差增加了 2.28 倍; 联合指纹+KNN 的定位准确度降低了约 2 个百分点, 平均定位误差增加了 2.83 倍; 幅度指纹+PCA-DNN 算法的定位准确度降低了约 3 个百

表 3 缺失值对不同算法定位性能影响

Table 3 Effect of missing values on positioning performance of different algorithms

Algorithm	Positioning performance	No missing value	With missing value
LLE+GBDT	Positioning Accuracy / %	98.6	97.2
	Average positioning error / m	0.039	0.128
Joint fingerprint+KNN	Positioning Accuracy / %	97.8	95.6
	Average positioning error / m	0.048	0.184
Amplitude fingerprint+PCA-DNN	Positioning Accuracy / %	97.3	93.8
	Average positioning error / m	0.056	0.235
Phase fingerprint + WKNN	Average positioning error / m	1.014	4.877

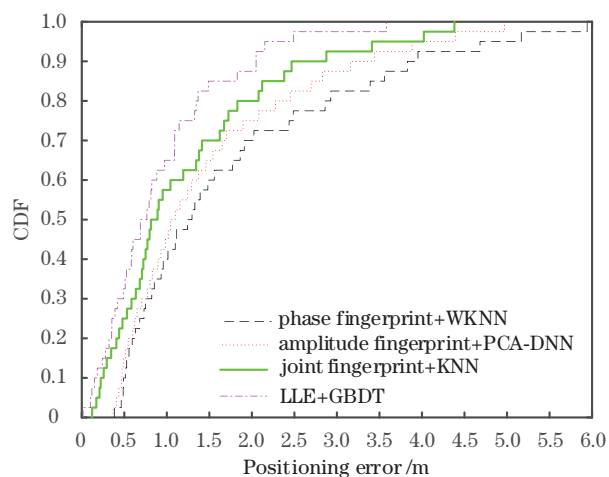


图 11 有缺失值的定位结果 CDF

Fig. 11 Localization result CDF with missing values

分点,平均定位误差增加了 3.20 倍;相位指纹+WKNN 算法的平均定位误差增加了 3.81 倍。

从图 11 可以看出,相位指纹+WKNN、幅度指纹+PCA-DNN、联合指纹+KNN 和 LLE+GBDT 算法在数据集存在 2% 缺失值的情况下,定位误差在 2.5 m 以内时的累积概率分布分别达到了 78.11%, 82.52%, 90.15%, 98.21%, 表明所提 LLE+GBDT 算法的鲁棒性要高于相位指纹+WKNN、幅度指纹+PCA-DNN 和联合指纹+KNN 算法。故 LLE+GBDT 算法在室内复杂环境的定位效果最佳,要远优于另外 3 种对比算法。

5 结 论

提出了一种基于 LLE+GBDT 的 CSI 室内定位算法。该算法将经过幅值滤波和相位矫正后的 CSI 数据作为联合指纹特征,通过 ENLLE 算法对 CSI 数据进行降维的同时突出指纹特征对环境的敏感性,最后通过 FOA-GBDT 算法实现高精度指纹定位。实验结果表明:LLE+GBDT 模型能很好地

适应环境变化,解决了因子载波特征不明显而导致的“误匹配”问题;当采集的 CSI 数据存在缺失值时,该算法具有良好的鲁棒性,定位性能明显优于联合指纹+KNN 算法和幅度指纹+PCA-DNN 算法,定位准确度始终保持在 95% 以上。下一步将对室内环境下,人员的实时运动轨迹进行研究。

参 考 文 献

- [1] Yan D Y, Song W, Wang X D, et al. Review of development status of indoor location technology in China[J]. Journal of Navigation and Positioning, 2019, 7(4): 5-12.
闫大禹, 宋伟, 王旭丹, 等. 国内室内定位技术发展现状综述[J]. 导航定位学报, 2019, 7(4): 5-12.
- [2] Yan M. Device-free indoor localization using CSI fingerprint[D]. Nanjing: Nanjing University, 2017.
黎明. 基于 CSI 指纹信息的室内定位技术研究[D]. 南京: 南京大学, 2017.
- [3] Xu S W, Wu Y, Wang X F. Visible light positioning algorithm based on particle swarm optimization compressed sensing[J]. Chinese Journal of Lasers, 2021, 48(3): 0306004.
徐世武, 吴怡, 王徐芳. 基于粒子群优化压缩感知的可见光定位算法[J]. 中国激光, 2021, 48(3): 0306004.
- [4] Mahfouz S, Nader P, Abi-Char P E. RSSI -based classification for indoor localization in wireless sensor networks[C]//2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), February 2-5, 2020, Doha, Qatar. New York: IEEE Press, 2020: 323-328.
- [5] Xu H, Wang X D, Wu N. Indoor visible light fingerprint positioning scheme based on convolution neural network[J]. Laser & Optoelectronics Progress, 2021, 58(16): 1606001.
许浩, 王旭东, 吴楠. 基于卷积神经网络的室内可见

- 光指纹定位方法[J]. 激光与光电子学进展, 2021, 58(16): 1606001.
- [6] Halperin D, Hu W J, Sheth A, et al. Tool release: gathering 802.11n traces with channel state information [J]. ACM SIGCOMM Computer Communication Review, 2011, 41(1): 53.
- [7] Halperin D, Hu W J, Sheth A, et al. Predictable 802.11 packet delivery from wireless channel measurements[C]// Proceedings of the ACM SIGCOMM 2010 conference, August 30-September 3, 2010, New Delhi, India. New York: ACM Press, 2010: 159-170.
- [8] Zhou R, Hao M, Lu X, et al. Device-free localization based on CSI fingerprints and deep neural networks[C]//2018 15th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), June 11-13, 2018, Hong Kong, China. New York: IEEE Press, 2018: 1-9.
- [9] Liu Z Y, Chen L W, Huang L. Research on indoor fingerprint positioning technology based on CSI phase correction[J]. Radio Engineering, 2020, 50(2): 102-107. 刘兆岩, 陈立伟, 黄璐. 基于 CSI 相位矫正的室内指纹定位技术研究[J]. 无线电工程, 2020, 50(2): 102-107.
- [10] Zhang L, Ding E J. Indoor fingerprint localization method using channel state information[J]. Journal of Henan University of Science and Technology (Natural Science), 2019, 40(3): 7, 46-52. 张雷, 丁恩杰. 一种利用信道状态信息的室内指纹定位方法[J]. 河南科技大学学报(自然科学版), 2019, 40(3): 7, 46-52.
- [11] Yao Z. Research on positioning technology based on RSSI and CSI[D]. Beijing: Beijing University of Posts and Telecom, 2019. 姚喆. 基于 RSSI 与 CSI 的定位技术研究[D]. 北京: 北京邮电大学, 2019.
- [12] Li X C, Ji X L, Wei W, et al. Research on fingerprint location algorithm based on OCAE-SOM [J]. Laser & Optoelectronics Progress, 2021, 58(8): 0812001. 李新春, 纪小璐, 魏武, 等. 基于 OCAE-SOM 的室内指纹定位算法研究[J]. 激光与光电子学进展, 2021, 58(8): 0812001.
- [13] Xu L Y. A survey of indoor positioning methods based on machine learning[J]. Computer Knowledge and Technology, 2018, 14(1): 217-219. 徐龙阳. 基于机器学习的室内定位方法综述[J]. 电脑知识与技术, 2018, 14(1): 217-219.
- [14] Kang X F, Li M M, Qiao W. High-precision indoor localization fusion algorithm based on Wi-Fi fingerprint [J]. Journal of Xi'an University of Science and Technology, 2020, 40(3): 470-476. 康晓非, 李梦梦, 乔威. 基于 Wi-Fi 指纹的高精度室内定位融合算法[J]. 西安科技大学学报, 2020, 40(3): 470-476.
- [15] Zhang X L, Xiu C D, Wang Y Z, et al. High-precision Wi-Fi indoor localization algorithm based on CSI-XGBoost[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(12): 2536-2544. 张玄黎, 修春娣, 王延昭, 等. 基于 CSI-XGBoost 的高精度 Wi-Fi 室内定位算法[J]. 北京航空航天大学学报, 2018, 44(12): 2536-2544.
- [16] Chapre Y, Ignjatovic A, Seneviratne A, et al. CSI-MIMO: indoor Wi-Fi fingerprinting system[C]//39th Annual IEEE Conference on Local Computer Networks, September 8-11, 2014, Edmonton, AB, Canada. New York: IEEE Press, 2014: 202-209.
- [17] Yan D. Research on gesture recognition method of unknown position based on Wi-Fi signal[D]. Kaifeng: Henan University, 2020. 闫迪. 基于 Wi-Fi 信号的位置未知的手势识别方法研究[D]. 开封: 河南大学, 2020.
- [18] Zhuo Y W, Zhu H Z, Xue H, et al. Perceiving accurate CSI phases with commodity Wi-Fi devices [C]//IEEE INFOCOM 2017 - IEEE Conference on Computer Communications, May 1-4, 2017, Atlanta, GA, USA. New York: IEEE Press, 2017: 1-9.
- [19] Wang X Y, Yang C, Mao S W. TensorBeat[J]. ACM Transactions on Intelligent Systems and Technology, 2017, 9(1): 1-27.
- [20] Zou H, Hastie T. Regularization and variable selection via the elastic net[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2005, 67(2): 301-320.
- [21] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.
- [22] Li Y C, Guo R Z, Cui D L. Foundations of differentiable manifolds[M]. Beijing: Science Press, 2011. 李养成, 郭瑞芝, 崔登兰. 微分流形基础[M]. 北京: 科学出版社, 2011.
- [23] Friedman J H. Stochastic gradient boosting[J]. Computational Statistics & Data Analysis, 2002, 38(4): 367-378.
- [24] Pan W T. A new fruit fly optimization algorithm: taking the financial distress model as an example[J].

Knowledge-Based Systems, 2012, 26: 69-74.

- [25] Gu Y F, Zhang D Y, Bao Z D. Lithology prediction of tight sandstone reservoirs using the PSO-GBDT: a case study of the Chang 4+5 members in the western Jiyuan oilfield[J]. Bulletin of Mineralogy, Petrology and Geochemistry, 2021, 40(3): 624-634.
谷宇峰, 张道勇, 鲍志东. PSO-GBDT 识别致密砂岩储集层岩性研究: 以姬塬油田西部长 4+5 段为例

[J]. 矿物岩石地球化学通报, 2021, 40(3): 624-634.

- [26] Liu Y C, An J. A method for predicting blade cracking of GBDT wind turbine based on improved LightGBM framework[J]. Journal of Technology, 2020, 20(1): 63-70.
刘钰宸, 安静. 一种基于 LightGBM 框架改进的 GBDT 风力发电机叶片开裂预测方法[J]. 应用技术学报, 2020, 20(1): 63-70.