

## 基于生成对抗网络的高精度室内无线定位方法

王福展, 朱晓荣, 陈美娟, 朱洪波

(南京邮电大学通信与信息工程学院, 江苏 南京 210003)

**摘要:** 无线信号在传播过程中容易受到干扰, 这限制了传统室内定位方法在实际生活中的应用。而基于位置的指纹定位技术具有普适性强的优点, 是当前的研究热点。指纹数据的数量是影响指纹定位精度的重要因素, 但是采集大量指纹数据的成本较大。因此, 如何使用少量指纹数据实现较高定位精度成为指纹定位技术的难点。针对此问题, 提出了一种基于生成对抗网络 (GAN, generative adversarial network) 的高精度室内无线定位方法。首先, 在室内等间隔密集地采集指纹数据, 构造初始指纹数据集, 选取初始指纹数据集中部分指纹数据, 使用 GAN 利用部分指纹数据得到大量指纹数据; 然后, 基于这些生成数据, 使用  $k$  最近邻 (KNN,  $k$ -nearest neighbor) 分类算法模型和随机森林模型进行定位预测。实验结果表明, 该方法能够基于少量指纹数据实现较高的无线定位精度, 定位精度达 15.4 cm。

**关键词:** 指纹定位; 生成对抗网络; 室内; KNN; 随机森林

**中图分类号:** TN961

**文献标识码:** A

**doi:** 10.11959/j.issn.2096-3750.2021.00208

## High-precision indoor wireless positioning method based on generative adversarial network

WANG Fuzhan, ZHU Xiaorong, CHEN Meijuan, ZHU Hongbo

College of Telecommunications & Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

**Abstract:** Because wireless signals are susceptible to interference during the propagation process, the application of traditional indoor positioning methods in real life is limited. Because location-based fingerprint positioning technology has the advantage of strong universality, it has become a current research hotspot. The number of fingerprint data is an important factor affecting the accuracy of fingerprint positioning, but the cost of collecting a large amount of fingerprint data is large. Therefore, how to use a small amount of fingerprint data to achieve higher positioning accuracy is a difficult point of fingerprint positioning technology. Aiming at this problem, a high-precision indoor wireless positioning method based on generative adversarial network was proposed. Firstly, fingerprint data was collected densely at equal intervals indoors, and the initial fingerprint data set was constructed, the part of the fingerprint data was selected in the initial fingerprint data set, and the generative adversarial network was used to obtain a large amount of fingerprint data from part of the fingerprint data. Then, based on these generated data, a KNN ( $k$ -nearest neighbor) model and a random forest model were used for location prediction. Experimental results show that this method can achieve high wireless positioning accuracy based on a small amount of fingerprint data, and the positioning accuracy can reach 15.4 cm.

**Key words:** fingerprint localization, generative adversarial network, indoor, KNN, random forest

### 1 引言

随着无线通信技术的进步, 基于位置信息的服

务已经成为人们关注的热点。《室内外高精度定位导航白皮书》中<sup>[1]</sup>指出, 在当今社会生活中, 人们每天在室内活动的时间约占一天总时间的 80%,

室内环境包括商场、隧道、家庭住宅、公司办公区等，所以在室内环境下对人们进行定位变得越来越重要。

被广泛使用的全球定位系统（GPS, global positioning system）在室内环境下无法较好地工作<sup>[2]</sup>，主要是由于随着人口的增多，建筑群的高度和密集度也日益增加，所以 GPS 信号在传输过程中很容易受到干扰，甚至无法被接收<sup>[3]</sup>。因此，在室内环境下对目标进行精准定位成为当前的研究热点。

文献[4-5]列出了很多常用的室内定位技术，如蓝牙技术、紫蜂技术等。但是各技术受限于自身特性，定位准确度较低，Wi-Fi 的准确度达 1 m 左右，蓝牙技术的准确度为 2.5 m，而且这些定位技术的抗干扰能力较差，很难在阻挡物较多、通信环境较差的情况下使用。本文使用超宽带（UWB, ultra wide band）技术进行定位，UWB 具有抗干扰能力强、数据传输速度快<sup>[6]</sup>、穿透力较强等优点，将 UWB 应用于室内定位的发展前景较好。

由于不规则的室内信号传播等因素，许多传统的室内定位方法在技术上面临着极大的挑战。在使用传统的基于距离的定位技术如信号到达时间差（TDOA, time difference of arrival）、接收信号强度指示（RSSI, received signal strength indication）、到达时间（TOA, time of arrival）和节点的到达角（AOA, angle of arrival）等时，需要确定未知目标节点和相邻参考节点之间的信号传播距离。但是，在阻挡物较多的室内环境下会造成非视距传输和多径信号传输，导致信号损失，造成距离的估计不准确，所以基于位置的指纹定位技术已经逐渐成为人们的研究热点。指纹定位技术需要在现场调查并且创建一个当前场景下的指纹库，此过程既费时又费力。在采集指纹的过程中，如果采集数据量较少，会导致定位精度较低；如果采集数据量过多，则会消耗较多的时间和人力成本，降低采集效率。

目前，国内外相关学者对指纹定位提出了许多技术方案。文献[7]通过详尽的实验结果验证了基于指纹的室内定位的实用性，并认为该方案是一种适用于室内的、可靠且价格低廉的解决方案。文献[8]通过使用费歇耳信息矩阵（Fisher information matrix）说明不规则环境因素对指纹定位的影响以及定位误差与 RSS 分布的理论关系。文献[9]通过支持向量回

归开发了一个连续空间估计器，通过使用连续空间估计器可以覆盖指纹数据库中还未存储的位置，降低了离线阶段的现场测量成本。文献[10]提出了一种基于深度学习的指纹方法，该方案分析了信道状态信息的特点，得到了最优指纹权重。文献[11]表明，基于指纹的定位系统中的关键因素是匹配算法和指纹数据的准确性。文献[12]通过使用移动用户的上、下文感知信息来扩展指纹定位，并利用这些信息建立平面图，减少了所需的信号源数量和无线干扰的影响。文献[13]融合了蓝牙技术在无线局域网（WLAN, wireless local area network）指纹定位中的应用，提出了一种定位算法，通过蓝牙信标数据对 WLAN 指纹数据库进行过滤，从而定位目标；通过初始指纹数据集的碎片，提高了在线阶段的计算性能。文献[14]提出了一种基于指纹协作和辅助节点的定位方法，该方法利用辅助节点和未知目标之间的距离信息进行相似性分析，利用有色噪声的自适应卡尔曼滤波来减小测距误差。

综上所述，很多学者在使用指纹识别方式实现室内定位时，通过降低测量成本或者提高定位精度来进行研究，但是未在指纹数据较少的情况下达到较高的定位精度。若要实现较高的定位精度，则需要足够多的指纹数据，可以通过使用 GAN 解决指纹数据较少的问题。

随着人工智能技术的不断发展，GAN 被应用于图像处理和数据处理方面。文献[15]经过对抗的过程提出了一种用来预测产生式模型的框架，同时训练两个模型：一个可以用来发现数据分布规律的产生式模型  $G$ ；另一个用来区分真实样本和生成数据的判别式模型  $D$ 。但是，在 GAN 实际训练过程中还存在许多问题，如训练较困难、生成样本的多样性不足等。文献[16]根据 GAN 存在的问题，引入了 Wasserstein 距离来替代 JS 散度（Jensen-Shannon divergence）和 KL 散度（kullback-leibler divergence），并将其作为优化目标。与初始 GAN 中的 KL 散度和 JS 散度相比，基于 Wasserstein 距离的 GAN 有较好的平滑特性，可以解决原始 GAN 存在的梯度消失情况。通过引入基本约束条件 1-Lipschitz 函数以便判别器收敛，并采用 weight clipping 方法对判别器的权值进行限制以满足约束条件，但是 weight clipping 方法没有将判别器限制于 1-Lipschitz 函数。文献[17]提出了 WGAN-GP

算法，并且在算法中使用梯度惩罚来执行 1-Lipschitz 约束，该方法在最后的实验中得到了较好的效果。

本文提出了一种基于 GAN 的高精度室内无线定位方法，创新性地将 GAN 应用于室内定位领域。对于部分指纹数据，通过 GAN 可以生成大量的指纹数据，降低了采集指纹的成本。基于这些生成的数据，使用 KNN 模型和随机森林模型进行定位预测，实验结果表明，该方法能够基于少量指纹数据实现较高的无线定位精度，定位精度达 15.4 cm。

## 2 系统模型

### 2.1 定位系统的组成部分

本文考虑的定位系统结构如图 1 所示，系统组成部分包括待定位的盲节点、已知位置的 4 个锚节点和用于计算的服务器与控制终端。系统中的盲节点可以向锚节点广播自身的 TOA 信息，锚节点可以接收盲节点的信息，并将这些信息传输到计算终端，计算终端将这些信息组合成当前位置的指纹信息，用于训练和测试盲节点的位置。

### 2.2 基于指纹的室内定位技术

在使用指纹技术进行定位时，该过程由两个阶段组成：1) 离线阶段：首先，将空间网格化，对空间进行等间隔采样，采集各个虚拟点的特征参数，构造指纹数据集，并使用指纹数据集中的数据训练机器学习模型；2) 在线阶段：当盲节点

运动至空间中某一个位置时，将该位置所采集的指纹数据作为输入，求得训练模型的输出，作为盲位置的虚拟坐标。指纹定位技术流程如图 2 所示。本文使用的指纹定位算法包括 KNN 算法和随机森林算法。

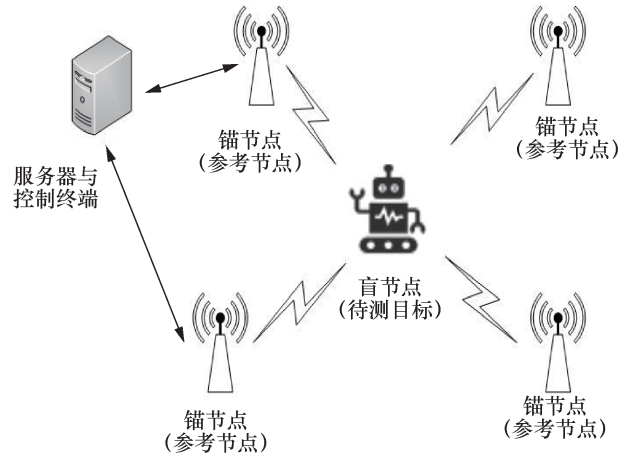


图 1 定位系统结构

## 3 基于 GAN 的指纹数据扩充

GAN 模型经过框架中生成模型和判别模型的相互博弈、学习后，得到较好的输出。生成模型在给定的高斯噪声情况下输出样本，将真实数据中的样本和由生成网络生成的样本输入判别模型中，该模型将对给定样本进行标记。当样本来自生成器时，标记为 0；当样本来自真实数据时，标记为 1。GAN 流程如图 3 所示。

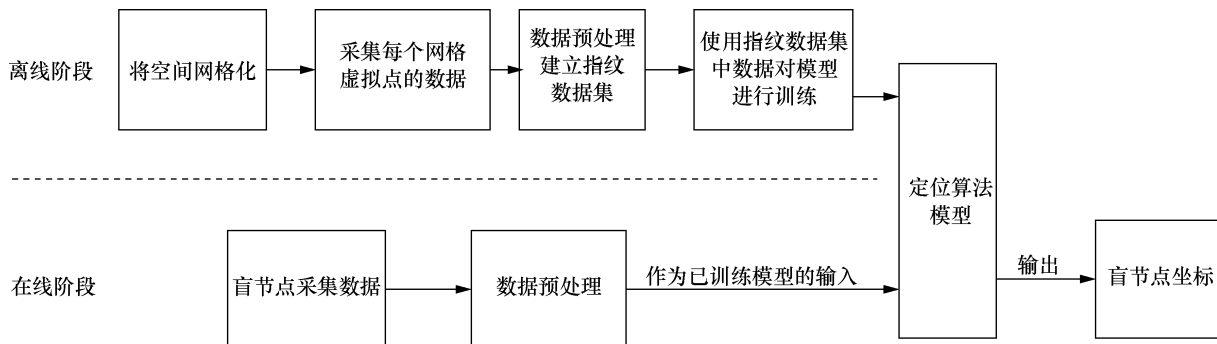


图 2 指纹定位技术流程

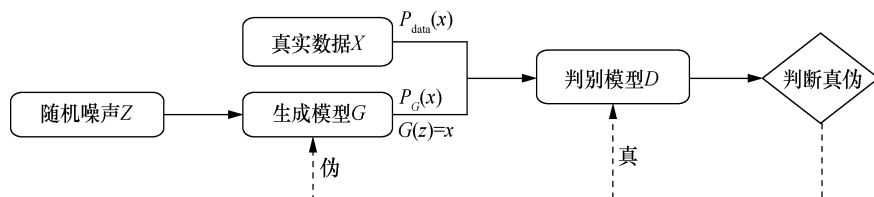


图 3 GAN 流程

一般情况下，把生成模型也称作生成网络，当已知分布规律的数据  $z$  时，生成模型可以把数据  $z$  转换成未知分布规律的数据  $x$ ，数据  $x$  可能与数据  $z$  的分布规律相差较大，把数据  $x$  的分布称作  $P_G(x)$ ，并且  $P_G(x)$  和  $P_{\text{data}}(x)$  之间的散度距越小越好。当散度距足够小时，可以认为数据  $x$  和真实数据的分布较类似，即可以通过当前生成网络生成符合真实数据分布的数据。

对于判别模型来说，最关键的问题是计算散度距。新建一个神经网络，用来衡量  $P_G$  与  $P_{\text{data}}(x)$  之间的散度距，一般称此神经网络为判别网络。

经过多轮迭代后，两个网络间的博弈使两个网络更好地完成任务，式(1)为 GAN 的优化目标函数。

$$\min_G \max_D V(D, G) = E_{x \sim P_{\text{data}}(x)} [\log D(x)] + E_{z \sim P_G(z)} [\log(1 - D(G(z)))] \quad (1)$$

其中， $P_{\text{data}}(x)$  是数据  $x$  上的分布，即指纹的特征向量分布； $P_G(z)$  是输入噪声所服从的分布； $G(z; \theta_g)$  表示一个数据空间的映射，并且  $G$  是一个由参数  $\theta_g$  的多层感知机表示的可微函数； $D(x; \theta_d)$  得到的是一个标量，其中， $D(x)$  代表数据  $x$  来自真实的数据分布而不是  $P_G$  的概率； $E$  表示期望。

在原始 GAN 实际训练中，还存在许多问题，如模型训练困难、生成样本的多样性不足等。文献[16]根据 GAN 存在的问题，使用了 Wasserstein 距离作为优化目标对模型进行训练，保证了梯度在所有位置均可用。通过引入一个基本约束条件 1-Lipschitz 函数以便判别器收敛，并采用 weight clipping 方法对判别器的权值进行限制以满足约束条件，但是 weight clipping 方法没有真正地将判别器限制于 1-Lipschitz 函数。在文献[17]中提出了 WGAN-GP 算法，并且在算法中使用梯度惩罚来执行 1-Lipschitz 约束，在本实验中使用 WGAN-GP 算法对指纹数据进行拟合，优化目标如式(2)所示。

$$L = E_{x_2 \sim P_G} [D(x_2)] - E_{x_1 \sim P_{\text{data}}} [D(x_1)] + \lambda E_{x_3 \sim P_{\text{penalty}}} \left[ \left( \left\| \nabla_{x_3} D_W(x_3) \right\|_2 - 1 \right)^2 \right] \quad (2)$$

在式(2)中， $x_1 \sim P_{\text{data}}$  是真实数据所服从的分布，即本实验中指纹的特征向量分布； $x_2 \sim P_G$  是生成器的数据分布； $x_3 \sim P_{\text{penalty}}$  是从真实数据集和生成数据

集组成的数据集中采样得到的数据分布。本文使用  $\varepsilon \sim \text{uniform}[0,1]$ ，从  $P_{\text{data}}$  中采样得到  $x_1$ ，从  $P_G$  中采样得到  $x_2$ ，在  $x_1$  和  $x_2$  的连线上进行随机插值采样，得到  $x_3 = \varepsilon x_1 + (1 - \varepsilon)x_2$ 。 $\lambda$  为惩罚参数，默认取 10。

## 4 基于机器学习和指纹的室内定位方法

本文通过机器学习算法和指纹相结合的方式实现室内定位，所使用的机器学习模型包括 KNN 模型和随机森林模型。

### 4.1 基于指纹的 KNN 模型

#### 4.1.1 基于指纹的 KNN 算法

KNN 算法的实现较简单，具有较好的普适性。在实验采集的指纹数据训练集中，当输入一个新的指纹特征数据时，从指纹数据训练集中寻找与输入指纹最相近的  $K$  个指纹数据，这  $K$  个指纹数据中最多的类即预测的输入指纹特征的类别。

#### 算法 1 KNN 算法

输入 训练集  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$

$x_i \in X \subseteq R^n$  为指纹数据训练集中的虚拟点指纹特征信息， $y_i \in Y$  为指纹特征信息在实验场景中所对应的虚拟点坐标， $i = 1, 2, 3, \dots, N$ 。

输出 输入指纹  $x$  所属的预测坐标  $y$

通过本实验选择的指纹距离度量从实验中的指纹数据训练集  $D$  中选择与输入指纹  $x$  最相近的  $K$  个虚拟坐标点，将这  $K$  个虚拟坐标点的输入指纹  $x$  的覆盖区域记作  $N_k(x)$ 。

在  $N_k(x)$  中，通过分类决策规则来确定  $x$  的坐标  $y$ 。

$$y = \arg \max_{c_j} \sum_{x_i \in N_k(x)} I(y_i = c_j), \quad (3)$$

$i = 1, 2, \dots, N; j = 1, 2, \dots, K$

其中， $I$  为指示函数，即当  $y_i = c_j$  时， $I$  为 1；否则， $I$  为 0。

#### 4.1.2 指纹距离度量

在本文实验场景中，两个指纹信息之间的距离反映了两个指纹信息的相似度。若两个指纹信息之间的距离小，则两个指纹信息的相似度大；若距离大，则两个指纹信息的相似度小。在 KNN 模型中，常用的距离度量是欧式距离、 $L_p$  距离 ( $L_p$  distance) 或闵氏距离 (Minkowski distance)。

在特征空间  $x$  中,  $x_i, x_j \in x$ ,  $x_i = (x_i^1, x_i^2, \dots, x_i^n)$ ,  
 $x_j = (x_j^1, x_j^2, \dots, x_j^n)$ 。  
 $x_i, x_j$  的  $L_p$  距离为

$$L_p(x_i, x_j) = \left( \sum_{l=1}^n |x_i^l - x_j^l|^p \right)^{\frac{1}{p}}, p \geq 1 \quad (4)$$

$x_i, x_j$  的欧氏距离为

$$L_2(x_i, x_j) = \left( \sum_{l=1}^n |x_i^l - x_j^l|^2 \right)^{\frac{1}{2}} \quad (5)$$

$x_i, x_j$  的曼哈顿距离为

$$L_1(x_i, x_j) = \left( \sum_{l=1}^n |x_i^l - x_j^l| \right) \quad (6)$$

#### 4.1.3 指纹 $K$ 值选择

在本文实验场景中, 指纹  $K$  值的选择会对 KNN 模型的训练结果产生影响。若选择指纹  $K$  值过小, 则在较小的范围内进行预测, 容易发生过拟合。若选择指纹  $K$  值过大, 会使学习的误差增大, 导致预测发生错误。在实际应用中, 一般利用交叉验证法对 KNN 模型的指纹  $K$  值进行选择。

#### 4.1.4 分类决策规则

在本文实验场景中, 基于指纹的 KNN 模型中使用的分类决策规则是多数表决法。该方法通过输入指纹的  $K$  个相近指纹位置中的多数类别来决定输入指纹的坐标。

多数表决法: 若分类的损失函数为 0-1 损失函数, 则分类函数为

$$f: R^n \rightarrow \{c_1, c_2, \dots, c_K\} \quad (7)$$

误分类的概率为

$$P(Y \neq f(X)) = 1 - P(Y = f(X)) \quad (8)$$

在指纹数据库中,  $x \in X$ 。最相近的  $K$  个指纹数据构成的集合为  $N_k(x)$ , 如果  $N_k(x)$  区域的类别为  $c_j$ , 则误分类的概率为

$$\frac{1}{k} \sum_{x_j \in N_k(x)} I(y_i \neq c_j) = 1 - \frac{1}{k} \sum_{x_j \in N_k(x)} I(y_i = c_j) \quad (9)$$

如果要使误分类的概率达到最小, 则让  $\sum_{x_j \in N_k(x)} I(y_i = c_j)$  最大, 多数表决法能够让经验风险达到最小。

## 4.2 基于指纹的随机森林模型

在基于指纹的随机森林模型中, 使用分类回归树作为弱学习器。在使用决策树作为随机森林模型基础时, 对于使用决策树进行特征选择方面作出部分改动。决策树通常会在树节点上  $n$  个指纹特征信息中选出一个最好的指纹特征信息, 通过这个最好的指纹特征信息来划分决策树的子树。在本文使用的基于指纹的随机森林算法中, 随机选择全部指纹特征信息中的部分指纹特征信息, 从所选的指纹特征信息中选出的最好的指纹特征信息来划分决策树的子树。

**输入** 训练集  $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ , 弱分类器迭代次数  $T$ 。其中,  $x_i \in X \subseteq R^n$  为指纹数据训练集中的指纹信息,  $y_i \in Y$  为指纹信息在实验场景中所对应的虚拟点坐标,  $i = 1, 2, 3, \dots, N$ 。

**输出** 最终的强分类器  $f(x)$

1)  $t = 1, 2, 3, \dots, T$ , 对指纹数据训练集进行第  $t$  次采样, 从指纹数据训练集中随机、有放回地选择  $m$  个指纹数据, 并利用所选的  $m$  个指纹数据来对第  $t$  个决策树模型  $G_t(x)$  进行训练。在使用所选的指纹数据训练决策树时, 从全部指纹特征信息中选出部分指纹特征信息。从所选的部分指纹特征信息中, 选出最好的指纹特征信息来划分决策树的子树。

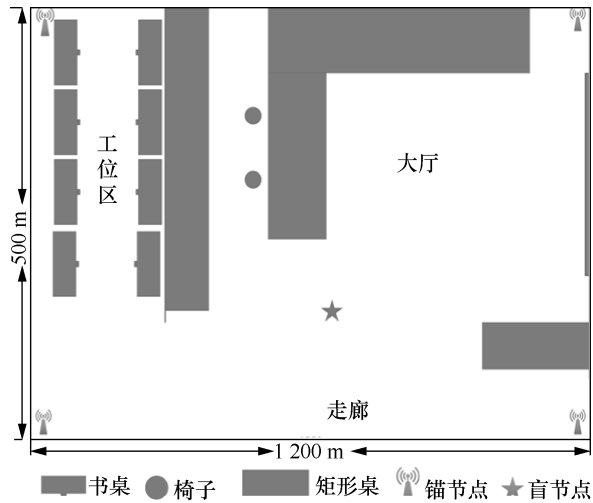
2) 在模型预测中, 将  $T$  个决策树得到的坐标预测结果进行算术平均, 计算出的结果即最终的预测坐标。

使用随机森林弱学习器的数量在很大程度上影响了强学习器的效果, 如果弱学习器的数量过少, 则容易造成欠拟合; 如果弱学习器的数量过多, 则容易造成过拟合。应根据实际场景, 对弱学习器的数量作出调整。

## 5 实验结果分析

### 5.1 实验场景

选择某实验室作为实验环境, 空间规格为 12 m × 9 m, 在该空间大厅内按 20 cm × 10 cm 的虚拟网格进行数据采集。在本文实验中, 对每个虚拟点进行 10 次数据采集, 共采集 4 000 个参考点作为指纹点来建立原始数据库。在本文实验中, 手持终端以 2 次/s 的速度扫描 4 个基站发送的指纹数据, 实验中的 4 个锚节点和手持盲节点终端均为 UWB 模块。空间平面图示意图如图 4 所示。



### 5.2 基于 GAN 的指纹数据扩充实现

在本文实验场景中，所采集的初始指纹数据训练集中包括 40 000 条指纹数据，将初始指纹数据训练集按照 3:1 的比例切割为初始训练集和测试集，测试集的指纹数量为 10 000 条，初始训练集的指纹数量为 30 000 条。在本文实验中，从初始训练集中随机抽取 15 000 条数据，将 15 000 条数据按照空间位置分类后，不同位置分别使用 GAN 模型实现该位置的数据扩充。经过 GAN 模型后，实现了将原始的 15 000 条数据扩充为 30 000 条数据，生成了扩充数据集。本文同时使用初始训练集和扩充数据集作为训练集训练 KNN 模型和随机森林模型，使用已划分的测试集来测试该模型的定位精度。

### 5.3 机器学习模型参数预处理

在测试定位精度前，对使用的 KNN 模型和随机森林模型的参数进行预处理。

#### 5.3.1 基于指纹的 KNN 模型参数选择

在基于指纹的 KNN 模型中，需要考虑的因素包括指纹  $K$  值的选择、指纹数据间的距离度量和指纹的分类决策。

基于指纹的 KNN 模型使用的分类决策为多数表决法，即在  $K$  个邻居节点中，多数节点的决策为最终决策；使用欧氏距离作为指纹距离度量来选择最合适的指纹  $K$  值。在实验空间内，将采集到的初始指纹数据训练集按照 3:1 的比例分为初始训练集和测试集，并且使用交叉验证法对模型进行  $K$  值选择。不同  $K$  值定位精度比较如图 5 所示，给出了不同  $K$  值下 KNN 模型在实验场景中的定位精度。

从图 5 中可以看出，当  $K$  值增大时，基于指

纹的 KNN 模型的定位精度呈下降趋势，但是达到最优值后，随着  $K$  值的增大，定位精度逐渐升高。为了得到较高的定位精度，在实验中设定  $K=10$  时可以达到最优值，在下面的测试中，设置指纹  $K$  值为 10。

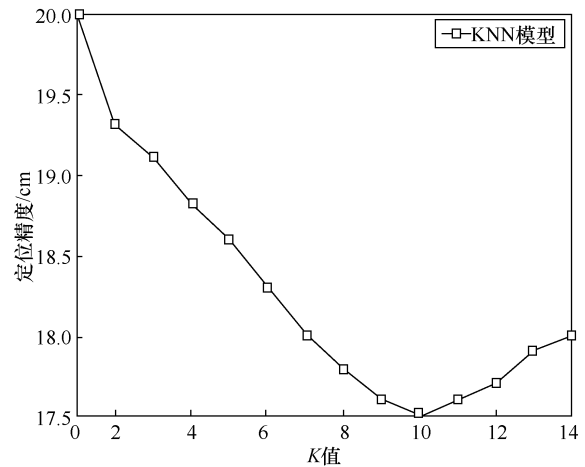


图 5 不同  $K$  值定位精度比较

#### 5.3.2 基于指纹的随机森林模型参数选择

在基于指纹的随机森林模型中，需要考虑的主要因素为决策树的数量。从理论上讲，当使用的决策树数量越多时，基于指纹的随机森林模型定位精度就越高，为了选出实验使用的最优决策树数量，在实验空间内，将采集到的初始指纹数据训练集按照 3:1 的比例切割为初始训练集和测试集，使用不同数量的决策树对模型进行实验分析。不同决策树数量下的定位精度比较如图 6 所示。

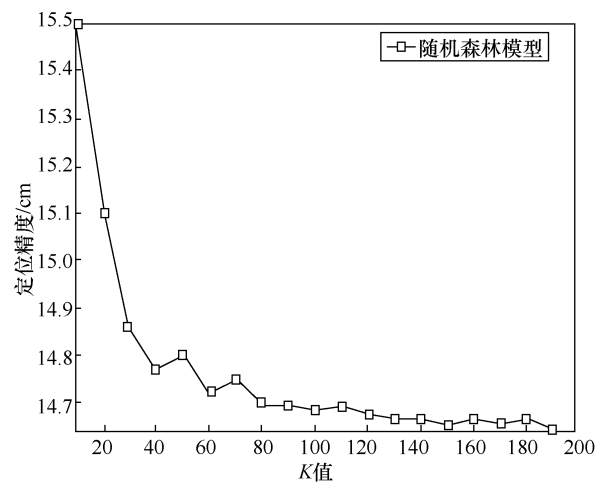


图 6 不同决策树数量下的定位精度比较

从图 6 中可以看出，当决策树数量增多时，随机森林模型的定位精度呈现下降趋势，在决

策树数量达 80 个后，随着决策树数量的增加，定位精度的变化逐渐平缓。为了得到较高的室内定位精度和较快的计算效率，本文实验中将基于指纹的随机森林模型使用的决策树数量设定为 80 个。

### 5.4 基于机器学习模型的指纹定位实现

#### 5.4.1 指纹数据的生成质量

本节使用两种机器学习模型来测试 GAN 模型生成的指纹数据质量，分别使用第 5.2 节中训练 GAN 的 15 000 条数据和 GAN 生成的 15 000 条数据作为机器学习模型的训练集，训练出该训练集下的机器学习模型。使用训练好的机器学习模型对测试集进行位置预测。两种模型在 15 000 条指纹数据下的定位精度如表 1 所示。

表 1 两种模型在 15 000 条指纹数据下的定位精度

机器学习模型	指纹数据来源	定位精度/cm
KNN 模型	初始 15 000 条数据	21.7
KNN 模型	生成 15 000 条数据	20.8
随机森林模型	初始 15 000 条数据	18.3
随机森林模型	生成 15 000 条数据	18.0

由表 1 可知，GAN 模型生成的指纹数据质量较高，定位效果较好。

#### 5.4.2 离线训练阶段

在离线训练阶段，分别使用初始训练集和扩充数据集作为机器学习模型的训练集。从训练集中选取 5 000 条指纹数据分别作为两种机器学习模型训练的数据集，训练出该数据集下的两种机器学习模型。通过递增的方式进行训练，即训练集的指纹数目每次增加 2 000 条，直到所有数据被使用完为止，训练出不同指纹数据量下的机器学习模型。

#### 5.4.3 在线测试阶段

在离线训练阶段的基础上，使用训练好的机器学习模型对测试集进行位置预测，得出定位精度。两种模型在不同指纹数据量下的定位精度如图 7 所示，即使用两种机器学习模型在两种训练集、不同指纹数据量下的定位精度。

由图 7 可以看出，KNN 模型和随机森林模型在不同指纹数据量下的定位精度变化，通过实验发现，GAN 扩充的指纹数据集与真实采样得到的初始训练集在两种机器学习模型下的定位精度非常相

似，可以证明扩充的指纹数据具有较好的效果。两种模型在 30 000 条指纹数据下的定位精度如表 2 所示，可以看出，当指纹数据来源相同时，随机森林模型的定位精度优于 KNN 模型；当基于同一类机器学习模型时，经过 GAN 扩充的数据集在指纹定位时的定位精度与初始数据集的定位精度非常相似。与文献[18]和文献[19]提出的优化方法对比，本文所提方法的定位精度均高于上述文献的方法。

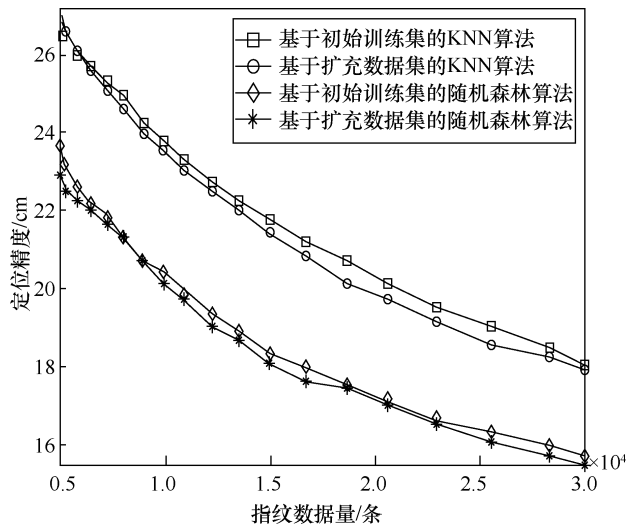


图 7 两种模型在不同指纹数据量下的定位精度

表 2 两种模型在 30 000 条指纹数据下的定位精度

机器学习模型	指纹数据来源	定位精度/cm
KNN 模型	初始训练集	18.1
KNN 模型	扩充数据集	17.9
随机森林模型	初始训练集	15.6
随机森林模型	扩充数据集	15.4

定位误差累积分布如图 8 所示，展示了当指纹数据量为 30 000 条时，上述 4 种情况下的指纹定位误差累积分布。由图 8 可知，基于初始训练集的 KNN 算法的定位精度优于 10 cm 的置信概率为 40%，优于 20 cm 的置信概率为 70%；基于扩充数据集的 KNN 算法的定位精度优于 10 cm 的置信概率为 41%，优于 20 cm 的置信概率为 71%。但是，基于初始训练集的 KNN 算法的指纹定位最大误差为 91 cm，而基于扩充数据集的 KNN 算法的指纹定位最大误差为 81 cm。基于初始训练集的随机森林算法定位精度优于 10 cm 的置信概率为 50%，优于 20 cm 的置信概率为 76%；基于扩充

数据集的随机森林算法定位精度优于 10 cm 的置信概率为 48%，优于 20 cm 的置信概率为 77%。但是，基于初始训练集的随机森林算法的指纹定位最大误差为 82 cm，基于扩充数据集的随机森林算法的指纹定位最大误差为 80 cm。图 8 中的 4 种情况可以证明，基于 GAN 的指纹数据扩充能够实现较好的效果。

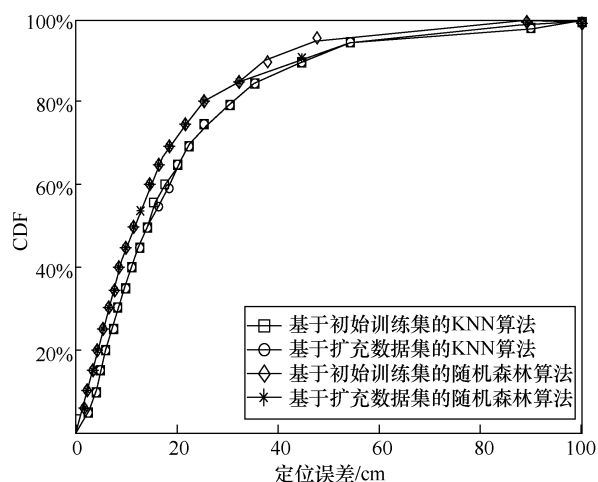


图 8 定位误差累积分布

## 6 结束语

本文针对指纹定位过程中数据量较少的问题，提出了一种基于 GAN 的高精度室内无线定位方法。在使用指纹技术进行室内无线定位时，利用机器学习模型对实验场景中的位置进行预测。在训练机器学习模型时，需要采集实验场景中的指纹特征数据，若选取指纹的数量较少，则定位精度较低；若选取指纹的数量过多，则采集较密集，采集任务烦琐，导致采集效率较低。本文将 GAN 模型应用于指纹室内定位，使用 GAN 实现了将少量指纹数据集扩充为大量指纹数据集，提高了指纹采集效率。实验结果表明，本文提出的方法可以在使用少量指纹数据的情况下，实现与大量指纹数据集相同的定位精度。

## 参考文献:

[1] 导航与位置服务科技专项总体专家组. 室内外高精度定位导航白皮书[R]. 2013.  
General Expert Group of Navigation and Location Service Science and Technology. White paper on indoor and outdoor high precision positioning and navigation[R]. 2013.

[2] HAN G J, JIANG J F, ZHANG C Y, et al. A survey on mobile anchor

node assisted localization in wireless sensor networks[J]. IEEE Communications Surveys and Tutorials, 2016, 18(3): 2220-2243.

[3] ASHKENAZI V, ROBERTS G W. Experimental monitoring of the Humber bridge using GPS[J]. Proceedings of the Institution of Civil Engineers-Civil Engineering, 1997, 120(4): 177-182.

[4] LIU W Y, DING H, HUANG X T, et al. TOA estimation in IR UWB ranging with energy detection receiver using received signal characteristics[J]. IEEE Communications Letters, 2012, 16(5): 738-741.

[5] LIU H, DARABI H, BANERJEE P, et al. Survey of wireless indoor positioning techniques and systems[J]. IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews), 2007, 37(6): 1067-1080.

[6] MEXTORF H, DASCHNER F, KENT M, et al. UWB free-space characterization and shape recognition of dielectric objects using statistical methods[J]. IEEE Transactions on Instrumentation and Measurement, 2011, 60(4): 1389-1396.

[7] TORRES J, BELMONTE O, MONTOLIU R, et al. How feasible is Wi-Fi fingerprint-based indoor positioning for in-home monitoring[C]//2016 12th International Conference on Intelligent Environments (IE). Piscataway: IEEE Press, 2016: 68-75.

[8] ZHOU M, QIU F, TIAN Z S, et al. Positioning error & signal distribution: an analysis towards lower error bound in WLAN fingerprint based indoor localization[C]//2015 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2015: 1-6.

[9] HERNÁNDEZ N, OCAÑA M, ALONSO J, et al. Continuous space estimation: increasing Wi-Fi-based indoor localization resolution without increasing the site-survey effort[J]. Sensors, 2017, 17(12): 147.

[10] WANG X Y, GAO L J, MAO S W, et al. CSI-based fingerprinting for indoor localization: a deep learning approach[J]. IEEE Transactions on Vehicular Technology, 2016: 1.

[11] KAEMARUNGS K, KRISHNAMURTHY P. Analysis of WLAN's received signal strength indication for indoor location fingerprinting[J]. Pervasive and Mobile Computing, 2012, 8(2): 292-316.

[12] D'SOUZA M, SCHOOTS B, ROS M. Indoor position tracking using received signal strength-based fingerprint context aware partitioning[J]. IET Radar, Sonar & Navigation, 2016, 10(8): 1347-1355.

[13] LI Q Y, LI W, SUN W, et al. Fingerprint and assistant nodes based Wi-Fi localization in complex indoor environment[J]. IEEE Access, 2016: 1.

[14] NIU J W, WANG B W, CHENG L, et al. WicLoc: an indoor localization system based on Wi-Fi fingerprints and crowdsourcing[C]//2015 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2015: 2013-2018.

[15] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//International Conference on Neural Information Processing Systems. MIT Press, 2014: 2672-2680.

[16] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein GAN[J]. International Conference on Machine Learning, [S.l.s.n.], 2017.

[17] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved training of Wasserstein GANs[C]//Advances in Neural Information Processing Systems. Massachusetts: MIT Press, 2017: 5767-5777.

[18] 郭妍, 陈晓, 任晓晔. 一种优化随机森林模型的室内定位方法[J]. 激光杂志, 2018, 39(10): 70-74.

GUO Y, CHEN X, REN X Y. An indoor positioning method for optimizing random forest model[J]. Laser Journal, 2018, 39(10): 70-74.

[19] 张萌, 吕艳, 倪益华. 基于密度峰值聚类的随机森林室内定位[J]. 计算机工程与设计, 2018, 39(5): 1490-1496.

ZHANG M, LYU Y, NI Y H. Random forest indoor location based on density peak cluster[J]. Computer Engineering and Design, 2018, 39(5): 1490-1496.

#### [作者简介]



王福展(1995-), 男, 南京邮电大学硕士生, 主要研究方向为无线室内定位。



朱晓荣(1977-), 女, 博士, 南京邮电大学教授、博士生导师, 主要研究方向为5G网络、物联网、无线定位算法等。



陈美娟(1971-), 女, 博士, 南京邮电大学副教授、硕士生导师, 主要研究方向为5G网络、物联网。



朱洪波(1956-), 男, 博士, 南京邮电大学教授、博士生导师, 主要研究方向为泛在无线通信与物联网、宽带移动通信、下一代网络、无线通信与电磁兼容等。