

基于 RSS 典型性判定的室内定位算法研究

吴佳英^{1,2}, 徐蔚鸿^{1,2}, 陈顺明², 李平²

(1. 南京理工大学 计算机与工程学院, 江苏 南京 210094; 2. 长沙理工大学 计算机与通信工程学院, 湖南 长沙 410114)

摘 要: 在基于 RSS 指纹集的定位算法中, 相似样本集的质量, 是影响定位精度的一个关键性因素; 而待定位点的 RSS 向量, 则是影响相似样本点质量的一个重要元素。通过对 D-RSS 分布规律分析, 提出了 RSS 典型性的概念, 并且提出了基于 RSS 典型性判定的室内定位算法。该算法根据 RSS 的典型性特征与有效的相似样本点之间的关系, 提出了 RSS 典型性的判别方法以及与典型性相关的动态 K 值。通过实验证明, 该算法不仅能完整地找出有效的相似样本点, 排除非实质性相似点的干扰, 而且在不同的定位场景中具有较强的适应性, 同时具有较高的定位精度。

关键词: 室内定位; 典型性判定; 相似点集; KNN

中图分类号: TP391

文献标识码: A

文章编号: 1000-436X(2014)Z2-0140-07

Indoor positioning algorithm research based on the typicality judgment of RSS

WU Jia-ying^{1,2}, XU Wei-hong^{1,2}, CHEN Shun-ming², LI Ping²

(1. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China;

2. School of Computer and Telecommunications, Changsha University of Science and Technology, Changsha 410114, China)

Abstract: In the process of indoor location based on RSS fingerprint, the quality of the obtained similar point set is a key factor for a successful position. And the locating point's RSS is an important reason which affects the quality of the similar point set. By analyzing the distribution of D-RSS, the concept of RSS's typicality was proposed firstly, and an indoor localization algorithm based on typicality judgment of RSS was also presented. According to the principle that the RSS values and the effective similar sample points, a typicality discrimination method for RSS values and a self-adapting K value were presented. Confirmed by the experiments, the algorithm not only can find the effective similarity sample points completely, but also can eliminate the non-substantive similarities points, and then can adapt to the different scenes, then have the higher positioning accuracy.

Key words: indoor positioning; typicality judgment; similar point set; KNN

1 引言

近年来, 随着 WLAN 技术的迅速发展, 人们对于无线终端位置的需求, 室内定位技术越来越受到人们的关注。一般地, 室内定位算法主要包括两大类: 基于距离和基于指纹^[1]。基于距离的方法^[2], 即获得待定位点与 AP(access point)之间的距离, 并利用三角定位法估计位置。而基于指纹的算法, 根据指纹信息的不同分为多种^[3], 如 TOA (time of arrival)、TDOA (time difference of arrival)、AOA

(angle of arrival)和 RSS (received signal strength)。但是由于 TOA、TDOA 和 AOA 对室内环境因素(如多径效应、反射和非视距传播等)非常敏感, 并且需要特殊设备进行测量。而 RSS 却可以非常方便地获取, 并不需要特别的硬件支撑。因此, 基于 RSS 指纹的室内定位方法已经成为目前最流行的室内定位算法之一^[4]。

基于 RSS 指纹的定位算法包括 2 个阶段: 训练阶段, 建立样本指纹信息库; 定位阶段, 估计待定位点位置^[5]。定位阶段为了更精确获得待定位点位置, 会

收稿日期: 2014-10-19

基金项目: 湖南省教育厅重点基金资助项目 (14A004)

Foundation Item: Project Supported by the Research Foundation of Hunan Province (14A004)

使用不同的匹配算法，其中包括概率性算法^[6]、确定性算法^[7]以及神经网络算法^[8]。其中应用最为广泛的是确定性算法中的 KNN(K-nearest neighbor)算法^[9]。

KNN 算法的主要过程为：计算待定位点 RSS 向量与训练集中的样本点的 RSS 向量之间的欧式距离，选取距离最小的 K 个，作为待定位点的相似样本点，构成相似样本集，而相似样本集的中心坐标即为待定位点的估计位置^[10]。

在基于 KNN 的这一类算法中，相似样本点集的获取最为关键。但是，在实际定位过程中，相似样本集的质量一直不高，即有效相似样本点的遗漏以及相似样本集中混有与待定位点仅具有逻辑相似关系的非实质性相似样本点。从与相似点集关系最密切的 RSS 向量入手，通过分析样本点之间的 D-RSS 分布规律，提出了 RSS 典型性的基本概念，并通过逐一考查 RSS 向量在各个维度上的 RSS 值的典型性与否，排除不具备典型性特征的 RSS 值，再利用与典型性 RSS 值相关的动态 K 值，尽可能地找出最有效的相似样本点，从而提高定位计算的精确度。

2 已有工作

对于如何获取到高质量相似样本集，目前的方法主要包括两大类。第一类为加权算法，即将获取到的 RSS 向量与样本集中的各个样本点的 RSS 向量进行某些差异计算（如欧几里德距离），将其中差异小的样本点置为高权值，因而在定位计算中能区别对待每个样本点以减小误差。如 K 近邻加权法，它将所测 RSS 向量与 k 个相似样本点 RSS 向量之间的欧式距离作为加权系数，然后再进行坐标计算，而不是取相似样本点中心点作为估计坐标，因此，能提高定位的精度；Yu Jiankun 等^[11]提出的将待定位点与所有样本点之间的模糊数学关系作为加权系数，再取其中权值较大的进行定位计算，这一方法能避免在获取相似点集时造成的有效相似样本点的遗漏，从而能更为精确地进行定位计算。第二类为聚类算法，则是利用具有某些相似属性的元素通过一些聚类算法能聚在一块的原理，即通过聚类的方法获取与待定位点 RSS 向量属性相似的样本点集，从而确定待定位点坐标。Sun Yongliang 等^[12]提出的 KNN-FCM 算法则更加的精细，首先，通过 KNN 算法得到 k 个相似样本点，然后利用 FCM 算法将这 k 个样本点均作为聚类中心，再分别进行

聚类，得到的不同类之间进行交集运算，从而获取到精确的相似样本点集。

上述这些方法，都能在一定程度上提高相似样本集的质量，但是，在过程中所有的 RSS 值都参与到了定位计算中，并没有考虑到这些 RSS 值能否都适合纳入到定位计算中，即是否具备典型性的特征，因为非典型性的 RSS 值在样本集中很可能没有能与之对应的相似样本点的存在，强行加入计算，相似点集中则会混入一些非实质性相似样本点，从而给定位带来非常大的误差。针对此点，提出的基于 RSS 典型性判定的室内定位算法(TJR, typical judgment of RSS algorithm)则是在进行定位计算之前，首先对 RSS 向量在各个维度上的 RSS 值是否具备典型性进行考查，以便排除那些不具备典型性特征的 RSS，同时提出与具备典型性特征的维度中所含样本点数量相关的动态 K 值，最后利用优化后的 RSS 向量以及最优的 K 值，获取到待定位的坐标。

3 算法思想

众所周知，RSS 值极易受到室内复杂环境的影响而产生不规则上下波动。由于这些波动，当次的 RSS 测量值很可能不能代表待定位点的区域位置特征，因而不适合进行定位计算。然而，通过一些滤波手段来提高 RSS 质量的方法过于复杂，设备需求过于苛刻，所以使用起来并不划算。其实，只需要对测量到的 RSS 值是否具备典型性进行判定，即可知道此 RSS 值是否能够代表待定位点的坐标位置特征。如果 RSS 向量在某一个维度上的值具备典型性特征，则说明此 RSS 值能纳入定位计算，而与之对应的则是计算过程中能够更容易获取到高质量的相似样本集。

3.1 算法分析

给定训练样本集为 $A=(F, D)$ ，其中 F 、 D 分别表示样本集 A 中的 RSS 矩阵和距离矩阵，分别表

$$\text{示为 } F = \begin{pmatrix} s_{11} & \cdots & s_{1m} \\ \vdots & \ddots & \vdots \\ s_{n1} & \cdots & s_{nm} \end{pmatrix}, \quad D = \begin{pmatrix} d_{11} & \cdots & d_{1m} \\ \vdots & \ddots & \vdots \\ d_{n1} & \cdots & d_{nm} \end{pmatrix}, \quad \text{其}$$

中， m 、 n 分别表示 AP 个数及样本点个数。待定位点的 RSS 向量则表示为 $f_0 = (s_1, s_2, s_3, \dots, s_m)$ 。

研究中发现任意维度上，RSS 值与距离值分布都呈现“纺锤型”，图 1 所示为在其中一个维度上的距离值与 RSS 值分布情况。

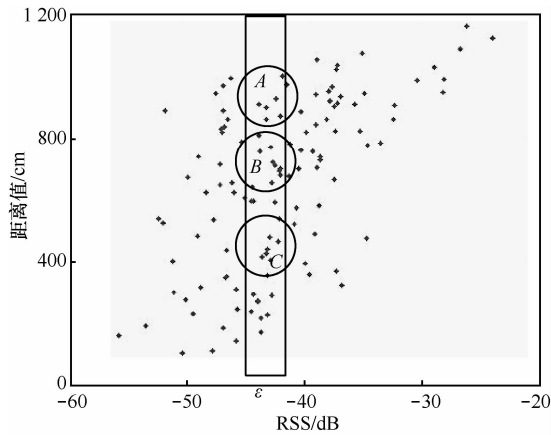


图 1 D-RSS 分布

由图 1 可以看出 RSS 值与距离值之间存在“多对多”现象，正因为如此，在相似点集中才会出现只具有逻辑相似关系的非实质相似样本点。要排除这些非实质性相似点，首先要找出待定位点与相似点集之间的联系。通过分析发现：在看似杂乱无章的 D-RSS 分布中，实际上待定位点与有效相似样本点集之间在各个维度上存在着一些必然的联系，这些联系即所提出的典型性特征，不具备典型性的 RSS 值在其维度上很难有效地匹配到与之相对应的相似样本点，也就是说，没有与之相对应的邻居节点，这样的 RSS 值在计算中必须舍弃。同样，根据此点，只需对 RSS 值的 D-RSS 分布情况进行分析即可判定该 RSS 值是否具备典型性特征。

3.2 相关定义

定义 1 (相似区间) 给定 f_0 以及在第 i 维度上的 RSS 值 $s_i (i=1, 2, \dots, m)$ ，若存在元素 $s_{ij} \in F_i^T$ ，且满足 $s_{ij} \in [s_i - \varepsilon_i, s_i + \varepsilon_i]$ ，则称由这样的元素构成的区间为相似区间，记为 a_i ，而 ε_i 为该区间的相似半径。如图 1 中黑色框内所示。

定义 2 (相似簇) 给定维度 i ，若存在样本点 $(p_1, p_2, \dots, p_\xi)_i \in a_i$ ，且任意样本点之间的距离为： $\|p_\tau - p_\nu\| \leq \Delta \varepsilon$ ，则称由这些聚集较为紧密的样本点构成的区域为相似簇，记为 c_{ij} 。如图 1 中 A、B、C 所示区域。

定义 3 (相似区间平均密度) 在维度 i 中，将所有符合 $(p_1, p_2, \dots, p_\kappa)_i \in a_i$ 的样本点的个数记为 $Num(a_i)$ ，且 a_i 所占面积记为 $S(a_i)$ ，则称在相似区间 a_i 单位面积上样本点个数为相似区间平均密度，记为 $\rho(a_i) = Num(a_i)/S(a_i)$ 。

定义 4 (簇密度) 在维度 i 中，将所有符合 $(p_1,$

$p_2, \dots, p_\xi) \in c_{ij}$ 的样本点个数记为 $Num(c_{ij})$ ，且 c_{ij} 所占面积记为 $S(c_{ij})$ ，则称在相似簇区域内单位面积上样本点的个数为簇密度，记为 $\rho(c_{ij}) = Num(c_{ij})/S(c_{ij})$ 。

定义 5 (典型性和典型维度) 给定 i 维度的 RSS 值 $s_i (i=1, 2, \dots, m)$ ，在相似区间 a_i 中，将所有的相似簇的簇密度值 $(\rho(c_{i1}), \rho(c_{i2}), \dots, \rho(c_{i\eta}))$ 与平均密度值 $\rho(a_i)$ 进行比较，如果其中存在某一个簇密度值 $\rho(c_i) > \rho(a_i)$ ，则称此 s_i 具备典型性，而此时维度 i 称为典型维度。

定义 6 (维度 K 值) 获取典型维度中簇密度大于平均密度的所有相似簇中所包含样本点数量 $(Num(c_{i1}), Num(c_{i2}), \dots, Num(c_{ip}))$ 。其中最小值则为此维度的维度 K 值： $K_i = \min(Num(c_{ij}))$ 。

3.3 算法描述

在本算法中，为了判定 RSS 是否具备典型性，在各个维度中首先获取 RSS 值相似区间内的所有相似簇的簇密度，再比较簇密度与相似区间平均密度，然后确定该 RSS 值是否具有典型性特征。如果此 RSS 值具备典型性，则将此维度设置为典型维度，接着获取所有的维度 K 值，再经过比较、筛选得到最终用于计算的 K 值。由于在每次定位计算过程中的相似簇样本点的数量都是变化的，因此， K 也是一个动态变化的值，而且它的变化与获取到的 RSS 向量相关。因此，该 K 值能耗更适应当前的测量 RSS 向量，从而能寻找到更为精确的相似样本点。

4 算法流程

TJR 算法的具体流程如图 2 所示。

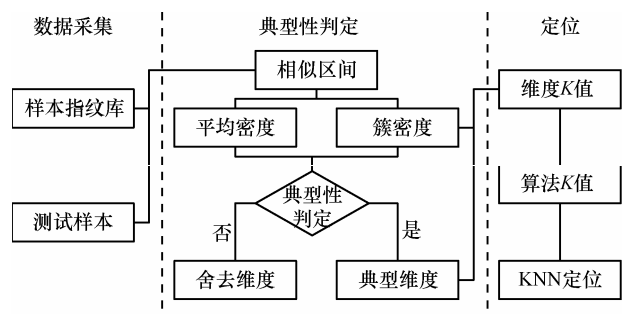


图 2 算法流程

假设测试样本 RSS 向量为 $f_0 = (s_1, s_2, s_3, \dots, s_m)$ ，以其中第 i 维度为例 ($i=1, 2, 3, \dots, m$)，算法的主要过程如下。

- 1) 根据该维度的 RSS 值 s_i ，得到其相似区间 a_i 。
- 2) 利用基于密度的聚类算法——DBSCAN 算法^[13]，得到相似区间内所有的相似簇 $(c_1, c_2, \dots, c_\psi)_i$ 。
- 3) 计算相关密度
 - ① 相似区间平均密度根据定义 3 有： $\rho(a_i) = Num(a_i)/S(a_i)$ ，其中 $S(a_i) = 2\varepsilon\Delta d$ ，且 $\Delta d = d_h - d_l$ ， d_h 为相似区间内坐标位置最高的样本点的纵坐标， d_l 为相似区间内位置最低样本点的纵坐标。
 - ② 簇密度根据定义 4 有： $\rho(c_i) = Num(c_{i,j})/S(c_{i,j})$ ，其中， $S(c_{i,j}) = 2\varepsilon\Delta d_c$ ， $\Delta d_c = d_{c_h} - d_{c_l}$ ， d_{c_h} 为相似簇内坐标位置最高样本点的纵坐标，而 d_{c_l} 为相似簇内最低位置样本点的纵坐标。
- 4) 根据定义 5 可以很容易地获取到所有的典型维度，如果不是典型维度的则应当舍去该维度。
- 5) 确定 K 值，根据定义 6 可以获取到所有典型维度的维度 K 值 $(K_1, K_2, \dots, K_\gamma)$ ，确定最终 K 值为： $K_0 = \min(K_\gamma)$ 。

5 实验仿真

5.1 实验场景

实验设置在综合教学楼的阶梯教室当中，教室长 15 m，宽 10 m，在其中设置 10 个 AP，具体场

景布置如图 3 所示。

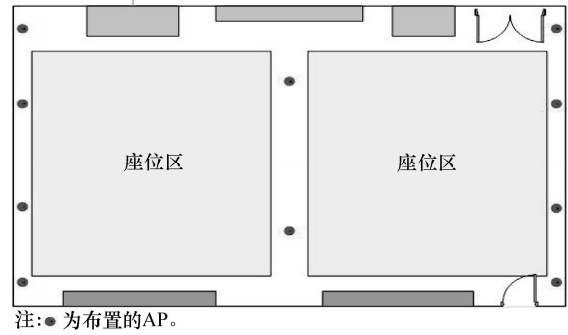


图 3 实验室布局

5.2 维度 D-RSS 分布

根据训练样本集以及获取的待定位样本点的 RSS 向量，得到各个维度的 D-RSS 分布图，图 4 是其中部分维度的 D-RSS 分布。

由图 4 可以看出，在各个维度上 D-RSS 都呈现多对多的“纺锤型”分布，这是由于室内复杂的环境造成的，同时，由于这种分布的存在使室内定位变得非常困难。但是，看似无规则的分布，其中也存在着一些必然联系，TJR 算法就是利用待定位点 RSS 值在相似区间内与邻居样本点之间的分布关系进行相似判定的。

5.3 DBSCAN 参数设置

利用 DBSCAN 算法对相似区间内的样本点分

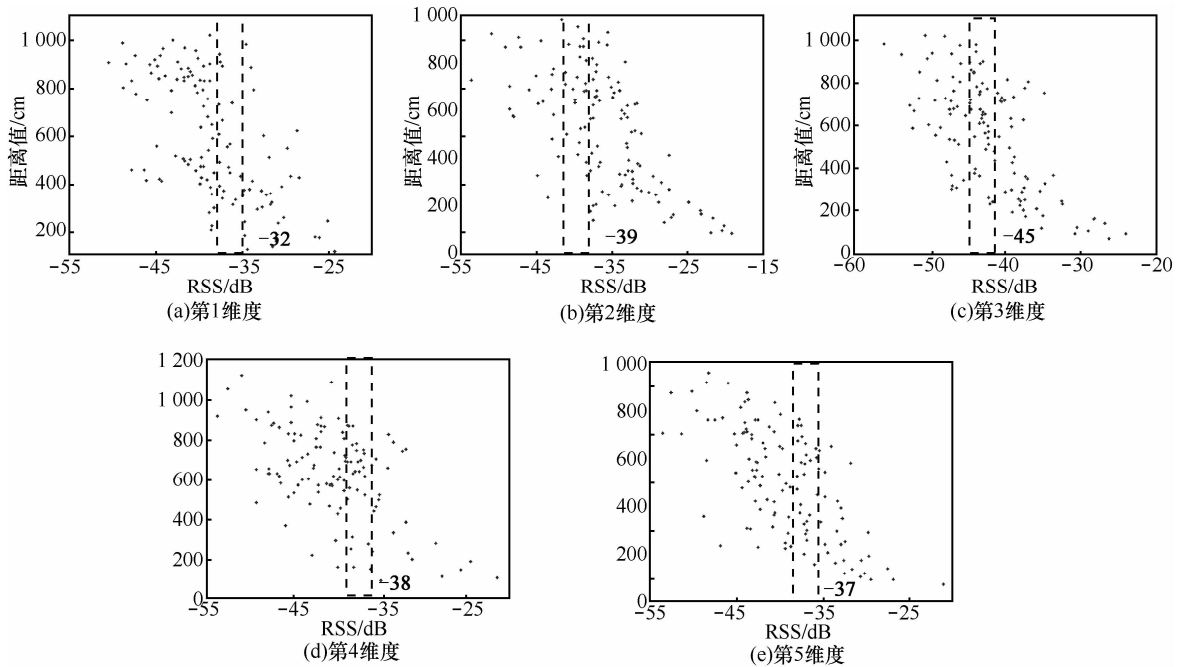


图 4 部分维度的 D-RSS 分布

注：其中黑色框区域为每一个维度的相似区间，黑色加深数字为待定位点的 RSS 向量在此维度的测量值。

布进行分析。经过多次实验，最终将 DBSCAN 算法的相关参数设置如下： E 邻域设置为 $\epsilon/2$ ，为相似区间半径的一半，取此值能够避免在相似区间样本点较为稀疏时不能形成有效的聚类，MinPts 设置为 3，根据相似区间内样本点的平均集中程度，将此值设为 3 最合适，而 D 域则是相似区间内的所有样本点。

通过聚类分析，此次测量到的待定位点 RSS 向量，在第 1、6、8 维度内所有相似簇的密度均小于区间平均密度，因此，在这些维度上 RSS 值不具有典型性特征，同时，图 4 中也可以看出，在第 1 维度 D-RSS 分布图中，在黑色框区域内，样本点较为稀疏，所形成的相似簇的簇密度明显要小于该区间的区间密度，这而侧面证明了，利用 DBSCAN 算法得到相似簇，并进行典型性判定是正确的。

对可聚样本点进行研究中发现如下性质。

性质 若存在高维 $k(k \leq m)$ 可聚样本点 c ，且存在相对应的低维可聚样本点 $(c_1, c_2, \dots, c_\theta)$ ，则 $(c_1, c_2, \dots, c_\theta)$ 均分布在以 c 为圆心， $2\epsilon_d$ 为半径的圆内。

证明 由假设得知，样本点 $(c_1, c_2, \dots, c_\theta)$ 在 $\alpha(\alpha \geq 2)$ 个维度上可聚，取其中任意 2 个可聚维度 i, j 。在 i 维度中，样本点 $(c_1, c_2, \dots, c_\theta)$ 与 c 均分布在以与其相对应的相似簇的几何中心 o_i 为圆心， ϵ_i 为半径的圆内，而在 j 维度中，样本点 $(c_1, c_2, \dots, c_\theta)$ 与 c 则分布在以 o_j 为圆心， ϵ_j 为半径的圆内。因此， c 与 2 个圆内的所有样本点之间的距离都不会超过 $\max(2\epsilon_i, 2\epsilon_j)$ ，同理可知，在其他可聚维度上， c 与样本点 $(c_1, c_2, \dots, c_\theta)$ 之间的距离都不会超过 $2\epsilon_d$ 。因此，命题得证。

5.4 动态 K 值

在本算法中， K 值能根据每一次测量数值的变化进行动态的变，不同的场景中均会出现相适应的变化。由于此次测量值中第 3 维度中样本点的分布较为理想，因此，将它作为例子来描述维度 K 值获取过程。通过 DBSCAN 聚类后，在第三维度相似区间内会形成若干相似簇，如图 5 所示。

从图 5 中可以看出，在相似区间内有 6 个相似簇，可以将其依次编号为： A 、 B 、 C 、 D 、 E 、 F ，同时，计算每一个相似簇的簇密度，如表 1 所示。

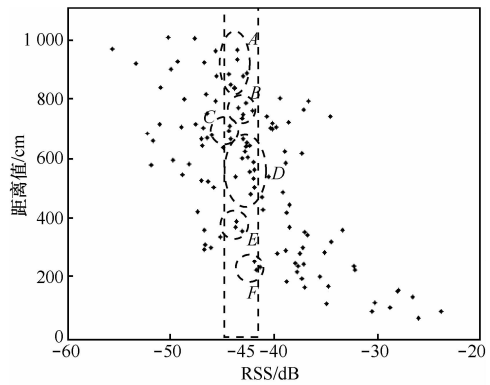


图 5 第 3 维度相似簇分布

表 1 第三维度相似簇密度值

编号	密度值	样本点个数
A	2.1	7
B	2.3	5
C	1.1	3
D	2.8	14
E	0.7	3
F	0.9	3
相似区间	1.6	35

对这些密度值进行描点，如图 6 所示。

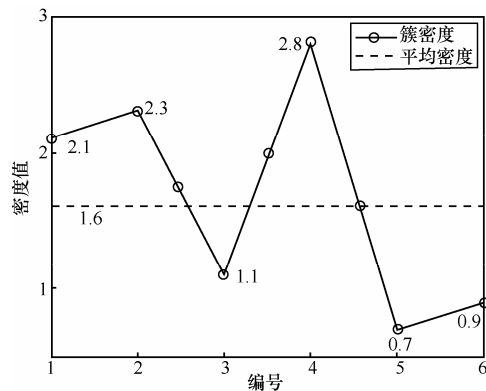


图 6 簇密度分布

从图中可以看出，有 3 个密度值是大于平均密度的，而这 3 个相似簇所含样本点个数分别为 7、5、14，最小值为 5，因此可以得到第三维度 K 值： $K_3=5$ 。使用同样的方法，得到其他典型维度的维度 K 值，如表 2 所示。

表 2 典型维度 K 值

维度	K 值
2	6
3	5
4	9
5	7
7	6
9	5
10	7

从表 2 中可以看出，最小 K 值为第三，九维度的维度 K 值，因此，将最终 K 值设置为 5。

5.5 相似点集

通过不同的匹配算法，得到待定位点的不同的相似点集，其具体分布如图 7 所示。

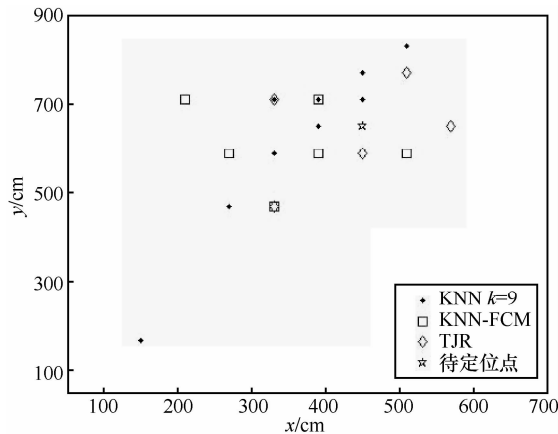


图 7 相似点集分布

由图 7 可以看出，传统的 KNN 算法由于是对整个样本空间进行的无差别选择计算，因此，获取到的相似样本点的误差较大，不仅仅遗漏了其中一些较为关键的相似样本点，而且还存在一些非实质性相似样本点。KNN-FCM 算法则有较大的提升，基本能选择到与待定位点位置相关的相似样本点，但是也具备相似点集中心不明确的缺点。而典型性算法所获得的相似样本点则更加精确，样本点基本上都分布在以待定位点为圆心的圆内，更加有利于定位计算。

5.6 误差分析

通过对最终获取到的待定位点的估计坐标与待定位点的实际位置的比较，表 3 是几种定位算法的平均误差比较。

表 3 几种算法的误差比较

算法	KNN	WKNN	KNN-FCM	TJR
	$K=6$	$K=6$		
MEE/m	1.69	1.27	0.97	0.71

从表 3 中可以看出，传统的 KNN 算法误差还是比较大的，而进行加权计算后的 WKNN 算法，在精度则有较大提升；而 KNN-FCM 聚类算法则是在 KNN 基础上，再聚类合并的一种算法，因此定位精度也相对较高；典型性算法在进行计算之前先剔除了不具备典型性的 RSS 维度值，

因而将计算带来的误差减到最小，因此，精度有着更进一步的提升，定位误差几乎仅是传统 KNN 算法的一半。

图 8 则是几种算法的积累误差数据比较。

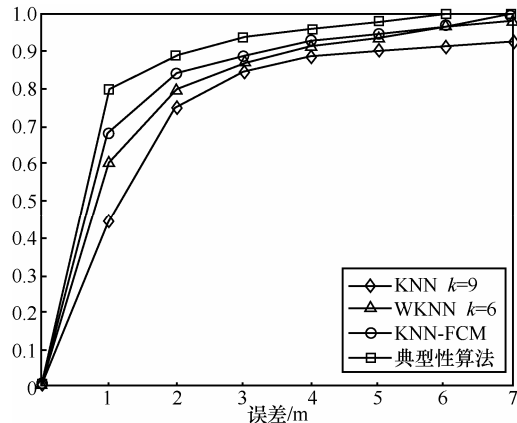


图 8 几种算法的积累误差比较

从图 8 中可以看出，典型性算法误差变化范围在 0.5 m 到 1.8 m 之间，变化较为稳定；而其他几种算法误差范围在 0.8 m 至 2.5 m 之间。因此，典型性算法在定位精度这一性能上有着较大优势。

6 结束语

测量到的待定位点 RSS 是否能适用于本次定位计算，可以通过对该 RSS 值进行典型性判定，适用于计算的 RSS 值能带来精确的定位计算，从而获取到高质量的相似点集。提出的判定 RSS 典型性的室内定位算法通过实验证实是成功的，它不仅能够剔除不适用于计算的 RSS 值，从而有效地排除非实质性相似样本点的干扰，而且能获取到基本完整的有效相似样本点，使更为精确地确定待定位点坐标，完成定位。因此，基于 RSS 典型性判定的室内定位算法有着较大的应用前景。

参考文献：

- [1] HATAMI A. Application of Channel Modeling for Indoor Localization Using TOA and RSS[D]. Worcester Polytechnic Institute, 2006.
- [2] SPIRITO M A. On the accuracy of cellular mobile station location estimation[J]. IEEE Tran Veh Tech, 2001,50(3):674-685.
- [3] MILES S B, SARMA S E, WILLIAMS J R. RFID Technology and Applications[M]. Cambridge University Press, 2008.
- [4] LI B, WANG Y, LEE H K, et al. Method for yielding a database of location fingerprints in WLAN[J]. IEEE Proc Commun, 2005, 152(5): 580-586.

- [5] ASSAD M A, HEIDARI M, PAHLAVAN K. Effects of channel modeling on performance evaluation of Wi-Fi RFID localization using a laboratory testbed[A]. IEEE Global Telecommunications Conference[C]. 2007. 366-370.
- [6] YOUSSEF M A, AGRAWALA A, SHANKAR A U. WLAN location determination via clustering and probability distributions[A]. IEEE International Conference on Pervasive Computing and Communications[C]. 2003. 143-150.
- [7] LIN T N, LIN P C. Performance comparison of indoor positioning techniques based on location fingerprinting in wireless networks[J]. International Conference on Wireless Networks, 2005, 2(6): 1569-1574.
- [8] BATTITI R, NHAT T L, VILLANI A. Location-aware Computing: A Neural Network Model for Determining Location in Wireless LANs[R]. Trento, Italy University of Trento, 2002.
- [9] DUDA R O, HART P E, *et al.* Pattern Classification[M]. Second Edition, John Wiley, 2000.
- [10] MA J, LI X, TAO X, *et al.* Cluster filtered KNN: a WLAN based indoor positioning scheme[A]. International Symposium on A World of Wireless, Mobile and Multimedia Networks[C]. 2008.1-8.
- [11] YU J K, LIU J Y. A KNN indoor positioning algorithm that weighted by the membership of fuzzy set[A]. 2013 IEEE and Internet of Things, IEEE International Conference on and IEEE Cyber[C]. 2013. 1899-1903.
- [12] SUN Y L, XUY B, MA L. KNN-FCM hybrid algorithm for indoor location in WLAN[A]. Power Electronics and Intelligent Transportation System(PEITS), 2009 2nd International Conference[C]. 2009. 251-254.
- [13] WANG X F, HUANG D S. A novel density-based clustering framework by using level set method[J]. Knowledge and Data Engineering, 2009, 21:1515-1531.

作者简介:



吴佳英 (1977-), 女, 湖南益阳人, 副教授, 主要研究方向为人工智能、模式识别、无线网络。



徐蔚鸿 (1963-), 男, 湖南湘潭人, 教授、博士生导师, 主要研究方向为模式识别、人工智能与图像处理。



陈顺明 (1988-), 男, 湖南郴州人, 长沙理工大学硕士生, 主要研究方向为无线网络、室内定位。



李平 (1972-), 男, 湖南新化人, 博士, 长沙理工大学教授, 主要研究方向为物联网、信息安全、数据挖掘。