

基于半监督仿射传播聚类 and KLDA 的室内定位算法^{*}

金纯^{1,2}, 邱灿^{1†}, 王腾¹, 刘谦¹

(1. 重庆邮电大学移动通信重点实验室, 重庆 400065; 2. 重庆金瓯科技发展有限公司, 重庆 400041)

摘要: 室内定位中位置指纹库采集的密集程度往往与定位精度密切相关, 针对离线阶段时指纹库稀疏的情况下定位精度低的问题, 提出了一种基于半监督仿射传播聚类 and KLDA 的室内定位算法。该算法结合了在线阶段采集无位置标签的 RSSI 数据, 通过建立局部邻域图将无位置标签的 RSSI 信息反映到离线指纹数据库的结构中, 并使用 KLDA 方法抽取位置指纹库中最大的特征信息, 有效利用了无位置标签的 RSSI 信息从而提高定位精度。实验结果表明, 该算法结合在线阶段 RSSI 数据后定位精度得到了明显的提高, 而且在仅保留离线指纹数据库三分之二的情况下, 几乎能够取得与传统 KNN 算法使用全指纹库时相同的定位精度, 相当于减少了离线阶段采集指纹库的工作开销。

关键词: 聚类; 线性判别分析; 位置指纹; 接收信号强度指示

中图分类号: TP393; TP301.6

文献标志码: A

文章编号: 1001-3695(2018)09-2659-04

doi:10.3969/j.issn.1001-3695.2018.09.022

Indoor localization algorithm based on semi-supervised affinity propagation clustering and KLDA

Jin Chun^{1,2}, Qiu Can^{1†}, Wang Teng¹, Liu Qian¹

(1. Key Laboratory of Wireless Transmission, Chongqing University of Posts & Telecommunications, Chongqing 400065, China; 2. Chongqing Jinou Science & Technology Development Co., Ltd, Chongqing 400041, China)

Abstract: Positioning accuracy is often closely related with the intensity of location fingerprint for indoor localization. When location fingerprint collected during offline stage is sparse, the accuracy is very low. Aiming at solving the problem above, this paper proposed an indoor localization algorithm based on semi-supervised affinity propagation clustering and KLDA. The algorithm took advantages of unlabeled data and reflected information of it into the structure of offline fingerprint through establishing the local neighborhood graph. It used KLDA method to extract the greatest feature of the location fingerprint and effectively used unlabeled data, thus to improve positioning accuracy. The experimental results show that the algorithm combining with unlabeled data collected online can improve the localization accuracy obviously. In addition, when keeping only two-thirds of offline fingerprint database, it also can achieve almost the same positioning accuracy as the traditional KNN algorithm using complete database, which is equivalent to reduce working overhead of collecting fingerprint in the offline phase.

Key words: clustering; linear discriminant analysis; location fingerprint; RSSI

0 引言

随着无线通信迅速的发展, 人们已经逐渐习惯使用 GPS 导航来到达目的地, GPS 在移动终端的普及性给人类带来了很大的便利, 但是在室内环境中由于多径效应的原因 GPS 不能够提供较高的定位精度, 这远不能满足人们对于室内位置服务的需求。无论是商场导购、地铁等公共场所定位, 还是紧急情况下的伤员救援, 这些场景中都需要精确的位置服务。随着人们对室内位置服务的需求不断增长, 越来越多的学者投入到室内定位技术的研究中来。

目前, 室内定位中最为普遍的定位方法为位置指纹定位法。位置指纹定位方法可分为概率性和确定性两类方法。典型的确定性算法有 NN^[1] 和 KNN^[2] 算法, 即通过计算到每个指纹数据的欧氏距离找到最近的一个 RP 或几个 RP 来确定待测目标位置。文献[3]中提出了一种基于贝叶斯模型的定位法, 引用贝叶斯模型统计各信号强度的高斯分布, 通过最大似然法估计定位位置, 相比于传统的 KNN 算法定位精度得到了很大的提高。K-means^[4]、APC (affinity propagation clustering, 仿射传播聚类)^[5]、C-means^[6] 等聚类算法将离线阶段的指纹库

进行分类处理, 大大减少了对指纹库全局搜索的匹配时间。在文献[7]中提出了一种基于 PCA 变换的聚类算法, 该算法利用 PCA 变换对离线阶段指纹数据库进行去噪和降维处理, 有效提高了聚类的准确率并减少了数据存储开销。文献[8]提出了一种基于 LDA (linear discriminant analysis, 线性判别分析) 的算法, 该算法利用线性判别嵌入式将聚类时的目标优化问题转换为低维空间特征求解问题, 把高维的离线数据样本投影到最佳鉴别矢量空间, 提取了离线数据指纹库的特征并压缩了样本空间维数, 在投影到的样本子空间中保证了最大类间距离和最小类内距离。

在室内定位过程中, 位置指纹数据库的密集程度会直接影响定位的精度, 通常采集间隔越小, 对在线定位阶段越有利。但是采集密集指纹库需要花费大量的时间、精力。针对该问题本文提出了一种基于仿射传播聚类 and KLDA (kernel linear discriminant analysis, 核函数线性判别分析) 的定位算法。在该算法中, 利用部分未知位置坐标的 RSSI 数据与离线位置指纹库, 通过 KLDA 算法来提取数据库中的特征信息, 并将样本投影到低维子空间中并保留最有用的特征信息。该算法结合了未知位置的 RSSI 信息对数据库进行特征处理, 通过建立局部

收稿日期: 2017-04-19; 修回日期: 2017-05-31 基金项目: 重庆市重点产业共性关键技术创新专项项目 (cstc2015zdcy-ztx40008)

作者简介: 金纯, 男, 重庆人, 教授, 博士, 主要研究方向为蓝牙及物联网; 邱灿 (1994-), 男 (通信作者), 硕士研究生, 主要研究方向为无线定位 (377899484@qq.com); 王腾 (1993-), 男, 硕士研究生, 主要研究方向为无线通信; 刘谦 (1993-), 男, 硕士研究生, 主要研究方向为无线通信。

邻域图将未知位置的 RSSI 信息反映到离线指纹数据的结构中,有效利用无位置标签的 RSSI 数据,从而提高定位精度。值得注意的是,未知位置的 RSSI 数据是很容易采集的,与传统的算法相比该算法可以节省数据库采集的工作开销,同时保证较高的定位精度。

1 基于半监督 APC 和 KLDA 算法

位置指纹定位技术一般分为离线和在线两个阶段。离线阶段时搜集某个场景中已知位置无线信号的位置指纹信息,再利用采集的信息预先建立该场景的位置指纹地图。在线实时阶段根据移动终端接收到的 RSSI 信息与离线数据库匹配得到定位坐标。文中提出了一种基于半监督仿射传播聚类和 KLDA 的室内定位算法。首先在 APC 聚类过程中考虑了 AP 集合相似度对两两指纹点相似度的贡献,对指纹点的相似度函数作出修正;然后在该算法中考虑了无位置的 RSSI 数据,有效利用了无位置标签的 RSSI 信息,从而提高了定位精度。

1.1 基于 AP 集合相似度的 APC 算法

仿射传播聚类是由 Frey 等人提出的一种聚类算法,该算法将数据集中的每一个样本点视为候选的聚类中心,为每个样本点建立亲和度和吸引度,然后在循环迭代过程中,每个样本数据竞争最终的聚类中心。仿射传播聚类相比于 K-means 聚类算法等,不需要初始化聚类中心和聚类中心数量,可以有效避免聚类中心选择问题导致聚类陷入局部极值的情况。因此在本文中采用该聚类算法。定义离线阶段两样本 RSSI 矢量之间的相似度为

$$\text{sim}(i, j) = -\|x_i - x_j\|^2 \quad (1)$$

聚类时位置相近的 RSSI 数据更有可能在一个类之中,但是由于多径效应相邻位置的 RSSI 数据可能有较大的波动,式(1)没有充分考虑指纹数据的位置信息,会造成聚类时相邻位置不在同一类之中。针对此问题考虑各样本邻近的 AP 集合对两样本相似度作出了修正。假设有 a 个 AP 离某离线指纹的位置在 $d_{\text{threshold}}$ 以内,那么就认为这 a 个 AP 属于样本点的邻近 AP 集合 $\{ap_1, ap_2, \dots, ap_a\}$ 。可以定义两样本点之间邻近 AP 集合相似度为

$$\text{sim}_{AP} = \frac{A \cap B}{A \cup B} \quad 0 \leq \text{sim}_{AP} \leq 1 \quad (2)$$

两样本的 AP 集合相似度越大,表示这两样本点在较多相同 AP 的邻近范围之内。考虑 AP 集合相似度来修正仅以 RSSI 欧氏距离来评判样本点的相关性函数,可以在一定程度上约束位置相隔较远的指纹数据因 RSSI 的波动性而聚于同类之中。定义修正后的样本相似度函数为

$$s(i, j) = -(1 - p \times \text{sim}_{AP}) \times \|x_i - x_j\|^2 \quad 0 < p < 1 \quad (3)$$

其中: p 为考虑 AP 相似的集合对样本指纹相似度影响的权重, p 在不断的实验中取得合适的值。根据式(3)求出离线指纹样本的相似矩阵 S 。在聚类之前,应该对 $s(i, i)$ 设置一个合理的初值,通常在实验中取矩阵 S 的中值。在仿射传播聚类过程中,根据两两样本点的吸引度消息和归属感消息来传递信息并通过不断迭代和更新最终确定聚类中心。其中吸引度消息定义为 $r(i, j)$,该式表示除 j 以外的样本点作用下 j 作为聚类中心对样本点 i 的吸引度; $a(i, j)$ 为归属感消息,表示除 j 以外的样本点作用下 i 认为样本点 j 为聚类中心的归属感。吸引度和归属感在聚类过程中的更新式如下:

$$r(i, j) = s(i, j) - \max\{a(i, j') + s(i, j')\}_{j' \neq j} \quad (4)$$

$$a(i, j) = \min\{0, r(j, j) + \sum_{i' \neq i, j} \max\{0, r(i', j)\}\} \quad (5)$$

通过以上两种消息在样本点之间传递实现中心和归属点的划分。经过不断地迭代完成聚类。

1.2 结合无位置标签 RSSI 数据的 KLDA 算法

1.2.1 LDA 算法

线性判别式分析^[9],也叫做 Fisher 线性判别(Fisher linear discriminant, FLD),是模式识别的经典算法,它是在 1996 年由 Belhumeur 引入模式识别和人工智能领域的。其基本思想是将高维的模式样本投影到最佳鉴别矢量空间,以达到抽取分类信息和压缩特征空间维数的效果,并且能够保证投影后模式样本在新的子空间有最大的类间距离和最小的类内距离,即模式在该空间中有最佳的可分离性。

假定训练样本位置指纹集 $X = \{x_i^T\}_{i=1}^n$,该样本集的分类为 $C_i \in \{1, 2, \dots, c\}$,在这里 c 表示所划分子类的个数,在聚类过程中,应该尽量达到类间距离最大化和类内距离最小化,即保持类间离散,类内聚合。在嵌入投影算子矩阵之后,上述优化问题可以描述为^[10]

$$\text{minimize} \quad \frac{V^T S_W V}{V^T S_T V} \quad (6)$$

$$\text{maximize} \quad \frac{V^T S_B V}{V^T S_T V} \quad (7)$$

其中: S_W 、 S_B 分别表示类内散度矩阵和类间散度矩阵; V 为投影矩阵,散度矩阵的表达式为

$$S_W = \sum_{k=1}^c \left[\sum_{x_i \in k} (x_i - a^{(k)}) (x_i - a^{(k)})^T \right] \quad (8)$$

$$S_B = \sum_{k=1}^c l_k (a^{(k)} - a) (a^{(k)} - a)^T \quad (9)$$

其中: $a^{(k)} = \frac{1}{l_k} \sum_{i=1}^{l_k} x_i$ 表示第 k 类中的指纹平均矢量; l_k 表示该类中的样本数量; $a = \frac{1}{n} \sum_{i=1}^n x_i$ 表示总样本集的平均矢量。总散度矩阵为 $S_T = S_W + S_B$ 。

利用 LDA 算法解决上述聚类时的目标函数,主要可以分为以下几步:

a) 构建邻接矩阵。

构建无向图 G 和 G' 。 G 表示聚类中同类别样本数据的矢量图矩阵, G' 表示聚类中不同类别样本数据的矢量图矩阵。无向图 G 和 G' 分别在同类和异类中根据 K 近邻准则而求得。若两位置指纹属于一对近邻组合,则认为它们之间有连边。

b) 计算邻接权重。

根据构建的无向图求出邻接矩阵 W ,求法如下:

$$w_{ij} = \begin{cases} 1 & x_i \in N_K(x_j) \text{ or } x_j \in N_K(x_i) \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

其中: $N_K(\cdot)$ 表示邻近指纹点集合; W 是一个实对称矩阵。

c) 得出投影算子。

投影算子 V 的求取是 LDA 算法中的关键,得到的矩阵 V 即可将高维模式矩阵进行降维特征处理,投影到最佳鉴别矢量空间。式(8)和(9)中的优化问题可等价于

$$\begin{cases} \text{maximize} & J(V) = \sum_{i,j} \|V^T x_i - V^T x_j\|^2 w_{ij} \\ \text{s. t.} & \sum_{i,j} \|V^T x_i - V^T x_j\|^2 w_{ij} = 1 \end{cases} \quad (11)$$

因此聚类过程中的无约束优化问题转换成了约束问题,该式表示最大化聚类时的类间距离和最小化类内距离。将式(11)进一步转换为矩阵迹的形式:

$$\begin{cases} \text{maximize} & J(V) = 2\text{tr}\{V^T X(D' - W)X^T V\} \\ \text{s. t.} & 2\text{tr}\{V^T X(D - W)X^T V\} = 1 \end{cases} \quad (12)$$

利用拉格朗日乘法对式(12)求解,可得

$$X(D' - W)X^T V = \lambda X(D - W)X^T V \quad (13)$$

其中: D 和 D' 为对角线矩阵,其中的元素分别为 $d_{ii} = \sum_j w_{ij}$ 、 $d'_{ii} = \sum_j w'_{ij}$ 。对式(13)求广义特征值并选取 d 个较大的特征值,然后分别求出 d 个特征值对应的特征向量即可得到投影矩

阵 V 。因此 LDA 对指纹集合 X 降维处理得到 $Z^* = V^T X$ 。

1.2.2 结合在线 RSSI 数据的 KLDA 算法

当数据样本量较小时,会导致高维度数据经降维之后变成奇异矩阵。这样本文在对目标函数最优化时无法完成对于高维数据特征的提取。因此提出在离线阶段指纹数据样本量少的情况下,考虑在线阶段大量的 RSSI 指纹信息来最优化目标函数,便可以对上述问题有效求解。由于在线阶段未知位置的 RSSI 数据可以容易获取,所以利用这些数据和离线阶段位置指纹的局部关系,通过判别分析抽取有利的特征信息,从而达到利用少量的离线位置指纹数据实现较高定位精度的目的。此外,线性学习的模式分类能力是有限的,不能够充分解决非线性复杂的分类问题^[11]。而 RSSI 波动和距离是非线性相关的,LDA 不能充分解决该分类问题。核函数因其较强的非线性学习能力在 SVM^[12]中得到普遍应用,因此在本文中结合核函数来解决该问题。在下文中将介绍结合在线未知位置 RSSI 数据的 KLDA 算法。

假设离线阶段已知位置坐标的 RSSI 集合为 X^L ,则矢量 $x_i^L = [x_{i1}^L, x_{i2}^L, \dots, x_{iM}^L]^T$ 表示第 i 个参考位置接收到 M 个 AP 的 RSSI 指纹。 X^U 表示在线阶段随机点采集到的 RSSI 矢量集合,这些随机点的坐标是未知的。 $x_j^U = [x_{j1}^U, x_{j2}^U, \dots, x_{jM}^U]^T$ 表示未知坐标位置点的 RSSI 矢量集合。显然,在线阶段只需采集 RSSI 数据而不需要与位置关联,这比离线阶段收集位置指纹的工作量要小得多。

在线指纹集合和离线指纹集合经核函数映射到高维空间 $\varphi: R^M \rightarrow o$ 。映射函数为 $k(x_1, x_2) = \varphi(x_1)^T \varphi(x_2)$ 。那么映射后的高维空间的投影算子可以通过下式求得

$$v = \sum_i \alpha_i \varphi(x_i) \quad (14)$$

考虑未知位置 RSSI 数据和离线阶段的位置指纹数据的局部关系,通过 KLDA 来提取有用的特征信息。由于结合了在线阶段未知位置采集的 RSSI 样本,式(10)中对样本邻接矩阵求法变为

$$W = \begin{cases} w_{ij}^{LL} = \begin{cases} 1 & x_i^L \in N_K(x_j^L) \text{ 或 } x_j^L \in N_K(x_i^L) \\ 0 & \text{otherwise} \end{cases} \\ w_{ij}^{UU} = \begin{cases} 1 & x_i^U \in N_K(x_j^U) \text{ 或 } x_j^U \in N_K(x_i^U) \\ 0 & \text{otherwise} \end{cases} \end{cases} \quad (15)$$

$$\tilde{W} = \begin{cases} \tilde{w}_{ij}^{LL} = \begin{cases} 1 & x_i^L \in N_K(x_j^L) \text{ 或 } x_j^L \in N_K(x_i^L) \\ 0 & \text{otherwise} \end{cases} \\ \tilde{w}_{ij}^{UU} = \begin{cases} 1 & x_i^U \in N_K(x_j^U) \text{ 或 } x_j^U \in N_K(x_i^U) \\ 0 & \text{otherwise} \end{cases} \end{cases} \quad (16)$$

其中:上标“~”表示对于不同类别。结合未知位置的 RSSI 集合 X^U 可以转换为

$$\begin{cases} \text{maximize } J(V) = \sum_{i,j} \|V^T x_i - V^T x_j\|^2 \tilde{w}_{ij}^{LL} + \sum_{i,j} \|V^T x_i - V^T x_j\|^2 \tilde{w}_{ij}^{UU} \\ \text{s.t. } \sum_{i,j} \|V^T x_i - V^T x_j\|^2 \omega_{ij}^{LL} + \sum_{i,j} \|V^T x_i - V^T x_j\|^2 \omega_{ij}^{UU} + \\ \sum_{i,j} \|V^T x_i - V^T x_j\|^2 \omega_{ij}^{LU} = 1 \end{cases} \quad (17)$$

将式(14)代入式(17)中,并转换成求矩阵迹的形式。那么 KLDA 算法最终得到广义特征值求解问题为

$$K(\tilde{L}^{LL} + \tilde{L}^{UU} + \tilde{L}^{LU})K\alpha = \lambda K(L^{LL} + L^{UU} + L^{LU})K\alpha \quad (18)$$

其中: K 表示核矩阵, $[K]_{ij} = k(x_i, x_j)$ 。对式(18)求广义特征值并将求出的特征值降序排列,然后选出 d 个较大的特征值 $\{\lambda_1, \lambda_2, \dots, \lambda_d\}$ 和其对应的特征向量 $\{\alpha_1, \alpha_2, \dots, \alpha_d\}$ 组成投影之后的空间 o 中的投影算子 α 。

1.2.3 定位算法步骤

与传统位置指纹定位算法相同,本算法主要可以分为离线和在线两个阶段。离线阶段首先通过上文中所提的仿射传播聚类算法找到各位置指纹数据对应的类标签,然后未知位置

RSSI 数据根据与聚类中心的欧氏距离找到各自所在的类,再利用 KLDA 算法求出投影算子对指纹数据非线性投影后的高维空间进行降维处理 $Z^* = \alpha^T \varphi(X)$ 。在线阶段待定位目标接收到的 RSSI 数据映射到高维空间然后通过投影矩阵 α 降维处理与离线阶段匹配,利用 KNN 算法求解待测目标坐标。本文算法主要可以分为以下几个步骤:

a)将离线阶段的位置指纹数据样本进行聚类,然后标记每个数据样本点所在的分类。

b)将随机位置采集的未知位置的 RSSI 集合 X^U 与已知坐标的 RSSI 数据进行比对,将每个样本点放入欧氏距离小的类之中,然后标记未知坐标样本所在的分类,根据 KNN 准则计算出邻接矩阵和拉普拉斯矩阵。

c)将约束问题转换为广义特征求解问题,根据式(18)求解广义特征值和对应的特征向量。然后将非零特征值从大到小降序排列。

d)设定所取特征值阈值,选取 d 个特征值,得到映射后高维空间对应的降维投影矩阵 α 。

e)对待测位置进行实时定位时,测试点所接收到的 RSSI 数据 x_{test} 经非线性映射到 $\varphi(x_{\text{test}})$,然后通过投影矩阵 α 降维处理。那么与离线阶段指纹库 X^L 进行匹配时,利用式(19)找到 X^L 映射的高维空间中与 $\varphi(x_{\text{test}})$ 相关性最大的向量:

$$\text{minimize } k(\alpha^T x_i, \alpha^T x_{\text{test}}) \quad (19)$$

其中: $k(\alpha^T x_i, \alpha^T x_{\text{test}})$ 越小,表示与待测试位置的 RSSI 数据越相似。利用此式找出 k 个最相似位置指纹点,然后利用 KNN 算法计算待测位置坐标:

$$x_{\text{test}}(x, y) = \frac{1}{k} \sum_{i=1}^k (x_{ix}, x_{iy}) \quad (20)$$

2 实验仿真分析

2.1 实验测试场景

为了验证文中算法的有效性,选取了逸夫楼 5 楼的一部分作为测试环境(面积为 $16.2 \text{ m} \times 24.5 \text{ m}$)。实验场景布置如图 1 所示。选取的实验场景中都贴有地板,每块地板长宽都是 0.6 m ,在离线阶段过程中采集指纹时每隔 1.2 m 采集数据。在实验环境中一共布置了 10 个 AP,在部署 AP 时应该尽量按照均匀布置的规则,这样可以使得所有 RP 尽量有邻近的 AP。每个 RP 采样 30 次来自各个 AP 的 RSSI,对其均值进行处理,构建完整的指纹数据库。

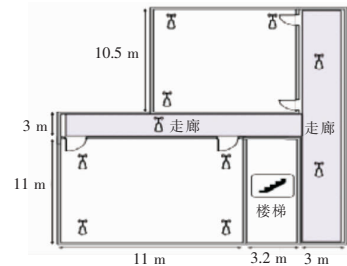


图 1 实验场景分布示意图

2.2 参数 p 对于 AP 集合仿射传播聚类结合 KLDA 算法的影响

在室内环境中,多径效应会引起 RP 接收到的 RSSI 产生波动,仅利用 RSSI 信息进行聚类时相邻位置可能不在同一类,这将会造成 KLDA 算法在构建矢量图时认为相邻位置没有连边,导致投影算子不准确影响定位精度。基于 AP 集合仿射传播聚类算法考虑了 AP 集合相似度对 RSSI 相似函数作出修改,即认为 AP 集合相似度对指纹点的相关程度有一定的贡献,指纹点邻近的 AP 集合相似度越大,越有可能在同类之中。

该方法一定程度上可以约束相邻位置聚到不同类之中,更有利于后面对于样本空间的鉴别分析。

为了验证 AP 集合仿射传播聚类算法有利于最终定位精度的提高,对于传统的仿射传播聚类结合 KLDA 和 AP 集合仿射传播聚类结合 KLDA 算法进行了实验对比,即对不同的参数 p 取值进行实验分析,其中当 p 为 0 时表示不考虑 AP 集合对聚类时的约束作用。实验对比如图 2 所示。从图 2 中可以看出,当 p 取值为 0.1 和 0.2 时,定位精度有所提高, p 为 0.2 时定位性能最优,定位误差 2 m 以内的概率较传统的 APC 结合 KLDA 算法提高了 6.2%。这说明在 AP 数量较多的情况下,基于 AP 集合的 APC 算法较传统的 APC 算法聚类效果好,因而定位精度得到了提高。这是因为基于 AP 集合的 APC 算法还考虑了 AP 集合相似度对指纹点相似度的贡献,AP 集合相似度越高则表示位置邻近的可能性越大,从而可以尽可能把位置相近的指纹点聚于同类之中,避免因 RSSI 波动相邻位置不在同一类。从图 2 中还可看出,当 p 取值过大时定位精度有所下降,并且当 p 取值越大时,定位效果越差,当 $p=0.4$ 时,定位精度在 2 m 范围以内的概率比不考虑 p 的情况下要低 13.6%,因此在选取 p 时应该从 0~1 逐渐从小到大大选取直至定位精度不再提升,对于下文中的实验分析都取 p 值为 0.2。

2.3 有无结合未知位置 RSSI 数据的仿真分析

本节比较了本文算法与传统的 KNN 和 NN 算法的定位精度,为了验证结合在线阶段随机采集未知位置的 RSSI 数据有利于提升定位性能,在实验过程中利用相同的离线位置指纹情况下,对比了有无结合 UP (unlabeled point, 无位置标签点) 的 RSSI 数据定位精度。其中这些未知位置的 RSSI 数据是随机位置采集的。实验对比仿真如图 3 所示。

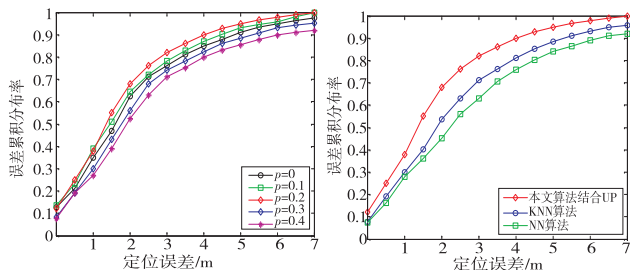


图2 不同的参数 p 对本文算法的影响

图3 本文算法结合UP后与传统算法的对比

从图 3 中可以看出,本文所提的算法定位精度有了明显的提高,本文算法误差在 2 m 以内的概率较传统的 KNN、NN 算法分别提高了 15.2%、23.0%。这是因为文中所提算法结合了未知位置的 RSSI 信息,通过其与离线位置指纹数据的局部邻域分析更有利于判别分析;此外该算法将高维的数据库进行降维处理并去除了不利于定位的 RSSI 数据,减少了离线指纹库的存储开销。

2.4 稀疏离线位置指纹库情况下的实验分析

为了进一步验证本文算法的有效性,在采集较少离线阶段位置指纹的情况下对本文算法作出了实验分析。在该实验中本文算法对预先采集的指纹库只保留三分之二,抽取三分之一的数据出来,并结合部分未知位置的 RSSI 数据进行定位。相较于传统算法使用全指纹库作出了定位精度的对比实验。对比结果如图 4 所示。

从图 4 中可以看出,文中提出的 APC 结合 KLDA 算法在使用比较稀疏的离线位置指纹数据库的情况下也能取得较好的定位精度。基于 APC 和 KLDA 算法定位误差在 2 m 以内的概率为 60.2%,传统的 KNN、NN 算法定位误差在 2 m 以内的概率分别为 63.4%、47.0%。实验过程中本文算法在原始采集的指纹库中去除了三分之一的指纹数据,而 KNN、NN 算法

使用的指纹库数据保持不变。本文算法定位误差在 2 m 以内的概率比 KNN 算法小 3.2%,这相当于本文算法采集更少的离线指纹数据能保证取得与传统算法几乎相同的定位精度,减少了离线采集的工作开销。另外从图 4 中还可得到使用 KLDA 算法比 LDA 算法的定位精度也略有提高,定位误差在 2 m 以内的概率提高了 5.6%,这是因为核函数具有较强的非线性学习能力,更适合于解决指纹库分类问题。

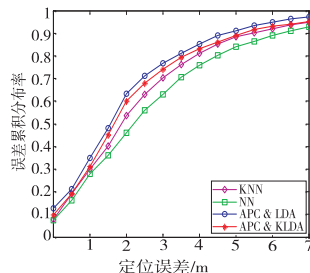


图4 本文算法使用稀疏指纹库与传统算法使用全指纹库情况下的实验比较

3 结束语

在文中提出了一种基于半监督仿射传播聚类 and KLDA 的室内定位算法。首先在 APC 聚类过程中考虑了 AP 集合相似度对两两指纹点相似度的贡献,在实验过程中验证了当 p 取值为 0.2 时,最有利于文中算法的定位。其次在该算法中考虑了无位置的 RSSI 数据,通过 KLDA 算法将原始的指纹库投影到子空间,抽取了最有用的 RSSI 特征信息并在新的子空间中保证了最佳可分离性。通过实验分析验证了本文算法的有效性,在稀疏指纹库的情况下依然能取得较高的定位精度,定位误差在 3 m 以下的概率达到了 74.2%,仅比传统的 KNN 算法小 2.6%,但是却大大减小了离线采集阶段的工作量。

参考文献:

- [1] Xu Xiaolong, Tang Yu, Wang Xinheng. Variance-based fingerprint distance adjustment algorithm for indoor localization [J]. *Journal of Systems Engineering and Electronics*, 2015, 26(6): 1191-1201.
- [2] Kim J, Ji M, Jeon J I. KNN based positioning performance estimation for fingerprint localization [C]//Proc of the 8th International Conference on Ubiquitous and Future Networks. Piscataway, NJ: IEEE Press, 2016: 1-6.
- [3] 张志涛,殷业,杜军辉. 一种基于契合度模型的室内定位方法[J]. *计算机工程*, 2016, 42(3): 295-300.
- [4] Razavi A, Valkama M, Lohan E S. K-means fingerprint clustering for low-complexity floor estimation in indoor mobile localization [C]//Proc of IEEE Globecom Workshops. Piscataway, NJ: IEEE Press, 2015: 1-7.
- [5] Peng Jianye, Li Taoshen, Ge Zhihui, et al. An indoor positioning algorithm based on geometry and RSS clustering [C]//Proc of World Automation Congress. Piscataway, NJ: IEEE Press, 2016: 1-6.
- [6] Sheshasaayee A, Sridevi D. Fuzzy C-means algorithm with gravitational search algorithm in spatial data mining [C]//Proc of International Conference on Inventive Computation Technologies. 2016: 1-5.
- [7] 杨明极,刘恺悌,邵丹. 用于 WLAN 室内定位的 PCA 聚类算法[J]. *电信科学*, 2016, 32(7): 21-26.
- [8] Xia Ying, Ma Lin, Zhang Zhongzhao, et al. Semi-supervised positioning algorithm in indoor WLAN environment [C]//Proc of the 81st IEEE Vehicular Technology Conference. Piscataway, NJ: IEEE Press, 2015: 1-5.
- [9] 李文斌,陈巖璵,张娟,等. 使用 Fisher 线性判别方法的提取分类器[J]. *计算机工程与应用*, 2010, 46(14): 132-134.
- [10] Shevlyakov G L, Oja H. Robust estimation of scatter and correlation matrices [M]//Robust Correlation: Theory and Applications. Hoboken: Wiley, 2016.
- [11] Chen H T, Chang Huangwei, Liu T L. Local discriminant embedding and its variants [C]//Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington DC: IEEE Computer Society, 2005: 846-853.
- [12] 韦燕华,周彦,王冬丽. 基于 LS-SVM 的位置指纹室内定位[J]. *计算机工程与应用*, 2016, 52(9): 122-125.