



(12)发明专利

(10)授权公告号 CN 103648106 B

(45)授权公告日 2017.03.22

(21)申请号 201310750528.6

H04W 64/00(2009.01)

(22)申请日 2013.12.31

(56)对比文件

CN 103079269 A, 2013.05.01,

US 2011/0164522 A1, 2011.07.07,

邓志安.基于学习算法的WLAN室内定位技术
研究.《哈尔滨工业大学博士学位论文》.2012,

(65)同一申请的已公布的文献号
申请公布号 CN 103648106 A

(43)申请公布日 2014.03.19

审查员 徐意特

(73)专利权人 哈尔滨工业大学

地址 150001 黑龙江省哈尔滨市南岗区西
大直街92号

(72)发明人 谭学治 周才发 马琳 邓仲哲
何晨光 迟永钢 魏守明

(74)专利代理机构 哈尔滨市松花江专利商标事
务所 23109

代理人 杨立超

(51)Int.Cl.

H04W 16/20(2009.01)

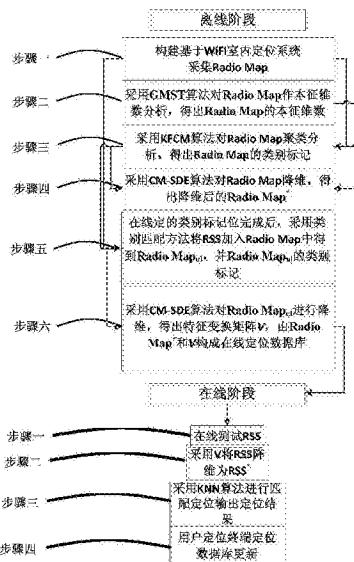
权利要求书3页 说明书13页 附图3页

(54)发明名称

一种基于类别匹配的半监督流形学习的
WiFi室内定位方法

(57)摘要

一种基于类别匹配的半监督流形学习的
WiFi室内定位方法,本发明涉及一种室内定位方法。本发明是要解决现有WiFi室内定位方法中存
在的Radio Map数据库大等问题。一、采集Radio Map;二、对Radio Map作本征维数分析;三、对
Radio Map聚类分析;四、对Radio Map降维;五、将RSS加入Radio Map中得到Radio Map_{ui};六、对
Radio Map_{ui}进行降维,得出特征变换矩阵V,由
Radio Map^{*}和V构成在线定位数据库。一、在线测
试RSS;二、将RSS降维为RSS^{*};三、输出定位结果;
四、数据库更新。本发明应用于网络技术领域。



1. 一种基于类别匹配的半监督流形学习的WiFi室内定位方法，其特征在于基于类别匹配的半监督流形学习的WiFi室内定位方法离线阶段定位过程按以下步骤实现：

一、对待定位的室内区域布置接入点AP，使无线信号覆盖待定位的室内区域，完成无线保真技术WiFi网络构建；

在待定位的室内区域规则选取并记录参考点的相应坐标，测量并依次记录参考点接收到的所有AP的信号强度RSS信号作为位置特征信息，构建网络信号覆盖图Radio Map，并存储Radio Map；

二、采用测地线最小生成树GMST本征维数估计算法对步骤一中构建的Radio Map的本征维数进行分析，得到的本征维数作为基于类别匹配的半监督局部鉴别嵌入CM-SDE算法的输入参数之一，决定Radio Map降维后的维数；

三、采用将核模糊C均值KFCM算法对Radio Map进行聚类分析，实现建立的Radio Map的类别标记，并作为CM-SDE的输入参数之一，并且提供相应的初始聚类中心及类别标记；

四、步骤二中的本征维数与步骤三中的类别标记作为输入参数，采用CM-SDE算法对步骤一中构建的Radio Map降维，得出相应的降维后的网络信号覆盖图RadioMap^{*}，RadioMap^{*}作为在匹配定位数据库用于在线定位阶段；

五、将不同用户在线定位阶段测试得到的未标记RSS，采用类别匹配的方式加入至已有Radio Map中，得到相应的包含未标记信号覆盖图RadioMap_{ul}，通过类别匹配更新的聚类中心作为CM-SDE算法中新的类别输入参数；

六、步骤五中的更新的聚类中心作为输入参数，采用CM-SDE算法对RadioMap_{ul}降维得到特征变换矩阵V'，V'与RadioMap^{*}共同构成在线匹配定位数据库，用于在线阶段定位；其中，所述线阶段定位具体为：

六(一)、在线测试RSS；

六(二)、采用V将RSS降维为降维后的信号强度RSS^{*}；

六(三)、采用K近邻法KNN算法进行匹配定位输出定位结果；

六(四)、用户定位终端定位数据库更新；

即完成了一种基于类别匹配的半监督流形学习的WiFi室内定位方法的离线阶段实现方式。

2. 根据权利要求1所述的一种基于类别匹配的半监督流形学习的WiFi室内定位方法，其特征在于步骤二中采用GMST本征维数估计算法对步骤一中构建的Radio Map的本征维数进行分析，其计算公式为：

$$d_{\text{intrinsic dim}} = \frac{1}{1-a},$$

上式中 $\frac{1}{1-a}$ 中的a表示最小生成树的线性拟合表达式y=ax+b的斜率。

3. 根据权利要求1所述的一种基于类别匹配的半监督流形学习的WiFi室内定位方法，其特征在于步骤四中生成新的降维后的信号覆盖图RadioMap^{*}与步骤六中的V'表达式为：

$$\text{Radio Map}^* = V' \cdot X$$

X是需要降维的Radio Map。

4. 根据权利要求1所述的一种基于类别匹配的半监督流形学习的WiFi室内定位方法，

其特征在于步骤五中类别匹配的实现,其核心流程分两步完成:

第一步,寻找未标记RSS的类别属性,由下式完成类别属性标记:

$$x_i \in G_i \Leftrightarrow \arg \min_{1 \leq j \leq c} D(x_i, v_j)$$

第二步:对RSS进行门限检测:通过计算并判定广义符号值与门限值V_T的关系,从而实现Radio Map及类别标记数据的更新,广义符号值及聚类中心的更新分别由如下两式完成:

$$S_i = \sum_{j=1}^N \text{sgn}(RSS_j - v_j)$$

$$v_i = \frac{1}{|G_i|} \sum_{x_k \in G_i} x_k$$

其中,所述门限值V_T=0.9N;c为聚类中心个数;x_i为第i个数据点;G_i为第i类构成的邻接关系图;D(x_i,v_j)表示计算x_i与v_j之间的欧式距离;|•|表示计算某类中元素的个数;N为室内定位系统中AP的个数;sgn(•)定义为:sgn(x)= $\begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$ 。

5.根据权利要求1所述的一种基于类别匹配的半监督流形学习的WiFi室内定位方法,其特征在于基于类别匹配的半监督流形学习的WiFi室内定位方法在线阶段定位过程通过下述步骤实现:

一、在线测试RSS;

二、将得到的待定位点的RSS采用特征变换矩阵降维变换得到RSS*;

三、采用KNN算法对RSS*与Radio Map*匹配位,对待定位点的具体位置坐标进行预测并进行在线数据的更新,其实现过程为:

(1) 在线阶段,测试点处接收的RSS=[AP₁,AP₂,…,AP_n],与特征变换矩阵V'相乘,从而得出降维后的RSS'=[AP₁,AP₂,…AP_d],其中d表示本征维数;

(2) 采用KNN算法实现RSS'与Radio Map*的匹配,采用与RSS'最近的K个参考点的坐标的平均值作为测试点(x',y'),其表达式为:

$$(x', y') = \frac{1}{K} \sum_{i=1}^K (x_i, y_i)$$

式中,(x',y')为测试点预测的坐标,(x_i,y_i)为第i个近邻点的坐标,K为KNN算法中近邻的数目;

四、用户定位终端定位数据库更新,即完成了基于类别匹配的半监督流形学习的WiFi在线阶段室内定位方法。

6.根据权利要求5所述的一种基于类别匹配的半监督流形学习的WiFi室内定位方法,其特征在于步骤三中对待定位点的具体位置坐标进行预测并进行在线数据的更新由离线数据库和在线数据库实现:

离线数据库的实现方式为:将定位用户本次定位测得的RSS作为未标记数据采用类别匹配方式加入到Radio Map中,并在移动终端上实现对本地在线匹配定位数据库的更新,实现动态的更新本地的数据,从而实现离线数据库定位方式;

在线数据库的实现方式为:用户在线定位完成后,将用户本次在线测得的RSS值上传至在线定位数据库所在的服务器,并在服务器端将在线定位数据库进行更新将将在线定位数

据传回上传RSS数据的定位终端。

一种基于类别匹配的半监督流形学习的WiFi室内定位方法

技术领域

[0001] 本发明涉及一种室内定位方法,具体涉及一种基于类别匹配的半监督流形学习的 WiFi室内定位方法。

背景技术

[0002] 随着无线局域网络在世界范围的飞速发展和移动终端设备的广泛普及,近年来出现了许多室内定位相关的技术和应用。由于多径效应、信号衰减及室内定位环境的复杂性,基于传统的信号传播模型的室内定位方法难以达到高精度的室内定位要求。基于到达时间 (Time of Arrival),到达时间差 (Time Difference of Arrival) 和到达角度 (Angles of Arrival) 等定位方法虽然可以基本满足定位精度需求,然而都需要定位终端有额外的硬件设备支持,具有较大局限性,从而导致基于上述几类定位方法的室内定位系统没有得到普及。

[0003] 目前,基于WLAN位置指纹 (Finger Print) 的WiFi室内定位方法得到了广泛应用。该方法的网络构建方法成本低廉,其使用2.4GHz ISM (Industrial Science Medicine) 公共频段且无需在现有设施之上添加定位测量专用硬件。只需要通过移动终端的无线网卡及相应软件测量接收到的接入点 (Access Point, AP) 的信号强度 (Received Signal Strength, RSS),由此来构建网络信号覆盖图 (Radio Map),进而通过匹配算法来预测移动用户所处位置的坐标,或相对位置。

[0004] 然而通过该方式建立的Radio Map包含有庞大的数据信息,且随着定位区域扩大, Radio Map可能(依据定位匹配方式及算法选择)呈指数形势增长。获得尽可能多的相关数据特征信息对于整个系统来说会提升定位精度,但是处理大量的特征信息增加算法开销,定位算法无法在处理能力有限的移动终端上有效运行,同时某些特征信息可能是对于定位没有作用甚至有负面作用,致使匹配效率降低,从而导致匹配定位算法的实现变得更加复杂,并且定位精度下降。

[0005] 当AP的数目增加及定位的参考点 (Reference Point) 增加时, Radio Map的数据信息增加。此时, Radio Map中代表的AP数目信息表示了数据的维数。因此,当AP数目增加, Radio Map就变成了高维数据。为减轻处理高维数据的负担,降维算法是有效的解决方法之一。高维数据可能包含很多特征,这些特征都在描述同一个事物,这些特征一定程度上是紧密相连的。如当从各个角度对同一个物体同时拍照时,得到的数据就含有重叠的信息。如果能得到这些数据的一些简化的不重叠的表达,将会极大地提高数据处理运行的效率并一定程度上提高准确度。降维算法的目的也正是在于提高高维数据的处理效率。

[0006] 除了可以简化数据使其能够高效处理外,降维方法还可以实现数据可视化。由于很多统计学的和机器学习算法对于最优解的准确性很差,降维的可视化应用可以令用户能够实际看到高维数据的空间结构和算法输出的能力,具有很强的应用价值。

[0007] 目前有很多基于不同目的的降维算法,包括有线性与非线性降维算法。其中PCA (Principal Component Analysis) 和LDA (Linear Discriminant Analysis) 是典型的线性

降维算法。这一类算法对于具有线性结构的高维数据有着良好的处理结果,但对于非线性结构的高维数据没有好的结果。典型的非线性降维算法以流形学习(Manifold Learning)算法。2000年Science杂志上同一期发表了3篇有关于流形学习算法中提出了2种经典的流形学习算法:LLE(Local Linear Embedding)及ISOMAP(Isometric Mapping)。由此,各种基于不同的准则的流形学习算法被提出并有一部分流形学习算法应用于图像处理方面。LDE(Local Discriminant Embedding)算法是流形学习算法中较晚提出的,它是一种的典型的基于特征提取的流形学习算法,而不只基于可视化目标。

[0008] 对于上述的降维算法并不能对在线获得的RSS数据或者新增的RSS来提高降维的精度。由于室内定位环境的变化,在不同时刻,特别是在长时间后,RSS数据之间的相关性就会降低。在已有算法中,有一类算法可以将在不同时间的RSS数据同时加入相应的已有数据中,从而增强不同数据之间的相关性,提高降维精度。这一类算法通常称之为半监督算法。根据半监督算法的特点,提出半监督鉴别嵌入(Semi-supervised Discriminant Embedding,SDE)算法。

[0009] 根据SDE算法的特点,采用类别匹配的方式将在线新得到的RSS加入已有Radio Map,然后进行局部鉴别嵌入降维,从而提出基于类别匹配的半监督局部鉴别嵌入算法(Classification Matching Based Semi-supervised Discriminant Embedding,CM-SDE)。采用CM-SDE算法对Radio Map进行降维,得出降维后的Radio Map,将对降维后的Radio Map室内定位,从而提出基于CM-SDE算法的WiFi室内定位算法。

发明内容

[0010] 本发明是要解决现有WiFi室内定位方法中存在的Radio Map数据库大,以及由于在线定位阶段计算复杂度高而引起的难以应用在线阶段获得RSS数据、难于在移动终端实现以及难于保证定位的实时性要求等问题,而提供了一种基于类别匹配的半监督流形学习的WiFi室内定位方法。

[0011] 基于类别匹配的半监督流形学习的WiFi室内定位方法离线阶段定位过程按以下步骤实现:

[0012] 一、对待定位的室内区域布置AP,使无线信号覆盖待定位的室内区域,完成WiFi网络构建;

[0013] 在待定位的室内区域规则选取并记录参考点的相应坐标,测量并依次记录参考点接收到的所有AP的RSS信号作为位置特征信息,构建Radio Map,并存储Radio Map;

[0014] 二、采用GMST本征维数估计算法对步骤一中构建的Radio Map的本征维数进行分析,得到的本征维数作为CM-SDE算法的输入参数之一,决定Radio Map降维后的维数;

[0015] 三、采用KFCM算法对Radio Map进行聚类分析,实现建立的Radio Map的类别标记,并作为CM-SDE的输入参数之一,并且提供相应的初始聚类中心及类别标记;

[0016] 四、步骤二中的本征维数与步骤三中的类别标记作为输入参数,采用CM-SDE算法对步骤一中构建的Radio Map降维,得出相应的降维后的RadioMap*,RadioMap*作为在匹配定位数据库用于在线定位阶段;

[0017] 五、将不同用户在线定位阶段测试得到的未标记RSS,采用类别匹配的方式加入至已有Radio Map中,得到相应的包含未标记信号覆盖图RadioMap_{u1},通过类别匹配更新的聚

类中心作为CM-SDE算法中新的类别输入参数；

[0018] 六、步骤五中的更新的聚类中心作为输入参数，采用CM-SDE对RadioMap_{ul}降维得到特征变换矩阵V'，V'与RadioMap*共同构成在线匹配定位数据库，用于在线阶段定位；其中，所述线阶段定位具体为：

[0019] 六(一)、在线测试RSS；

[0020] 六(二)、采用V将RSS降维为RSS*；

[0021] 六(三)、采用KNN算法进行匹配定位输出定位结果；

[0022] 六(四)、用户定位终端定位数据库更新；

[0023] 即完成了一种基于类别匹配的半监督流形学习的WiFi室内定位方法的离线阶段实现方式。

[0024] 基于类别匹配的半监督流形学习的WiFi室内定位方法在线阶段定位过程通过下述步骤实现：

[0025] 一、在线测试RSS；

[0026] 二、将得到的待定位点的RSS采用特征变换矩阵降维变换得到RSS*；

[0027] 三、采用KNN算法对RSS*与Radio Map*匹配位，对待定位点的具体位置坐标进行预测并进行在线数据的更新，其实现过程为：

[0028] (1) 在线阶段，测试点处接收的RSS=[AP₁, AP₂, …, AP_n]，与特征变换矩阵V'相乘，从而得出降维后的RSS'=[AP₁, AP₂, …, AP_d]，其中d表示本征维数；

[0029] (2) 采用KNN算法实现RSS'与Radio Map*的匹配，采用与RSS'最近的K个参考点的坐标的平均值作为测试点(x', y')，其表达式为：

$$[0030] (x', y') = \frac{1}{K} \sum_{i=1}^K (x_i, y_i)$$

[0031] 式中，(x', y')为测试点预测的坐标，(x_i, y_i)为第i个近邻点的坐标，K为KNN算法中近邻的数目；

[0032] 四、用户定位终端定位数据库更新，即完成了基于类别匹配的半监督流形学习的WiFi在线阶段室内定位方法。

[0033] 发明效果：

[0034] 针对本发明提出的CM-SDE算法及背景需要求，通过对哈尔滨工业大学科学园2A栋12层走廊构成的室内定位区域进行定位。采用联想V450笔记本电脑并结合NetStumbler软件测试所有参考点处的RSS，构成Radio Map，并采用CM-SDE算法对Radio Map降维并采用KNN算法实现室内定位。在附图4的仿真中，降维后的Radio Map的维数为原始Radio Map维数的三分之一。从附图4的仿真结果来看，采用CM-SDE算法与初始的KNN的算法的定位性能可比拟，但CM-SDE的定位复杂度仅为三分之一，且CM-SDE算法可以有效地应用实时得到的新的RSS来Radio Map的密度，从而有效的提高定位精度。

附图说明

[0035] 图1是本发明中离线数据库定位实施流程图；实线箭头表示步骤之间的数据传输；

[0036] 图2是本发明中在线数据库定位实施流程图；实线箭头表示步骤之间的数据传输；

[0037] 图3是基于WiFi的室内定位网络的构建及实验环境示意图；

[0038] 图4是采样网格图；

[0039] 图5是采用CM-SDE与KNN算法定位性能对比图；其中，表示CM-SDE，表示KNN。

具体实施方式

[0040] 具体实施方式一：本实施方式的基于类别匹配的半监督流形学习的WiFi室内定位方法离线阶段定位过程按以下步骤实现：

[0041] 一、对待定位的室内区域布置AP，使无线信号覆盖待定位的室内区域，完成WiFi网络构建；

[0042] 在待定位的室内区域规则选取并记录参考点的相应坐标，测量并依次记录参考点接收到的所有AP的RSS信号作为位置特征信息，构建Radio Map，并存储Radio Map；

[0043] 二、采用GMST本征维数估计算法对步骤一中构建的Radio Map的本征维数进行分析，得到的本征维数作为CM-SDE算法的输入参数之一，决定Radio Map降维后的维数；

[0044] 三、采用KFCM算法对Radio Map进行聚类分析，实现建立的Radio Map的类别标记，并作为CM-SDE的输入参数之一，并且提供相应的初始聚类中心及类别标记；

[0045] 四、步骤二中的本征维数与步骤三中的类别标记作为输入参数，采用CM-SDE算法对步骤一中构建的Radio Map降维，得出相应的降维后的RadioMap^{*}，RadioMap^{*}作为在匹配定位数据库用于在线定位阶段；

[0046] 五、将不同用户在线定位阶段测试得到的未标记RSS，采用类别匹配的方式加入至已有Radio Map中，得到相应的包含未标记信号覆盖图RadioMap_{ul}，通过类别匹配更新的聚类中心作为CM-SDE算法中新的类别输入参数；

[0047] 六、步骤五中的更新的聚类中心作为输入参数，采用CM-SDE对RadioMap_{ul}降维得到特征变换矩阵V'，V'与RadioMap^{*}共同构成在线匹配定位数据库，用于在线阶段定位；其中，所述线阶段定位具体为：

[0048] 六（一）、在线测试RSS；

[0049] 六（二）、采用V将RSS降维为RSS^{*}；

[0050] 六（三）、采用KNN算法进行匹配定位输出定位结果；

[0051] 六（四）、用户定位终端定位数据库更新；

[0052] 即完成了一种基于类别匹配的半监督流形学习的WiFi室内定位方法的离线阶段实现方式。

[0053] 所述离线阶段与在线阶段均在定位终端完成。

[0054] 具体实施方式二：本实施方式与具体实施方式一不同的是：步骤二中采用GMST本征维数估计算法对步骤一中构建的Radio Map的本征维数进行分析，其计算公式为：

[0055] $d_{\text{intrinsic dim.}} = \frac{1}{1-a}$ ，测地距最小生成树算法

[0056] 上式中 $\frac{1}{1-a}$ 中的a表示最小生成树的线性拟合表达式y=ax+b的斜率。

[0057] 其它步骤及参数与具体实施方式一相同。

[0058] 具体实施方式三：本实施方式与具体实施方式一或二不同的是：步骤四中生成新的降维后的信号覆盖图RadioMap^{*}与步骤六中的V'表达式为：

[0059] Radio Map*= $V' \cdot X$

[0060] X是需要降维的Radio Map。

[0061] 其它步骤及参数与具体实施方式一或二相同。

[0062] 具体实施方式四:本实施方式与具体实施方式一至三之一不同的是:步骤五中类别匹配的实现,其核心流程分两步完成:

[0063] 第一步,寻找未标记RSS的类别属性,由下式完成类别属性标记:

$$[0064] x_i \in G_i \Leftrightarrow \arg \min_{1 \leq j \leq c} D(x_i, v_j)$$

[0065] 第二步:对RSS进行门限检测:通过计算并判定广义符号值与门限值的关系,从而实现Radio Map及类别标记数据的更新,广义符号值及聚类中心的更新分别下两式完成:

$$[0066] S_i = \sum_{j=1}^N \operatorname{sgn}(RSS_j - v_j)$$

$$[0067] v_i = \frac{1}{|G_i|} \sum_{x_k \in G_i} x_k$$

[0068] 其中,所述门限值V_T=0.9N。

[0069] 其它步骤及参数与具体实施方式一至三之一相同。

[0070] 具体实施方式五:基于类别匹配的半监督流形学习的WiFi在线阶段室内定位方法通过下述步骤实现:

[0071] 一、在线测试RSS;

[0072] 二、将得到的待定位点的RSS采用特征变换矩阵降维变换得到RSS*;

[0073] 三、采用KNN算法对RSS*与Radio Map*匹配位,对待定位点的具体位置坐标进行预测并进行在线数据的更新,其实现过程为:

[0074] (1)在线阶段,测试点处接收的RSS=[AP₁, AP₂, ..., AP_n],与特征变换矩阵V'相乘,从而得出降维后的RSS'=[AP₁, AP₂, ..., AP_d],其中d表示本征维数;

[0075] (2)采用KNN算法实现RSS'与Radio Map*的匹配,采用与RSS'最近的K个参考点的坐标的平均值作为测试点(x',y'),其表达式为:

$$[0076] (x', y') = \frac{1}{K} \sum_{i=1}^K (x_i, y_i)$$

[0077] 式中,(x',y')为测试点预测的坐标,(x_i,y_i)为第i个近邻点的坐标,K为KNN算法中近邻的数目;

[0078] 四、用户定位终端定位数据库更新,即完成了基于类别匹配的半监督流形学习的WiFi在线阶段室内定位方法;

[0079] 所述离线阶段在服务器上完成;在线阶段在定位终端上完成。

[0080] 具体实施方式六:本实施方式与具体实施方式五不同的是:步骤三中对待定位点的具体位置坐标进行预测并进行在线数据的更新由权利要求1中所述的离线数据库和在线数据库实现:

[0081] 离线定位数据库的实现方式为:将定位用户本次定位测得的RSS作为未标记数据采用类别匹配方式加入到Radio Map中,并在移动终端上实现对本地在线匹配定位数据库

的更新,实现动态的更新本地的数据,从而实现离线数据库定位方式;

[0082] 在线数据库的实现方式为:用户在线定位完成后,将用户本次在线测得的RSS值上传至在线定位数据库所在的服务器,并在服务器端将在线定位数据库进行更新将将在线定位数据传回上传RSS数据的定位终端。

[0083] 其它步骤及参数与具体实施方式一至五之一相同。

[0084] 仿真实验:

[0085] 一、结合附图3对本仿真实验做出详细说明:图示为哈尔滨工业大学科学园2A栋12层的平面图示意,基于WiFi的室内定位系统就是基于该实验环境下建立。在实验环境中,总共布置27个AP,AP布置的位置为蓝色无线发射信号形状标记所在处。AP离房间地面高度为2米。在离线阶段,在联想V450笔记本上安装NetStumbler软件,在所有参考点的四个不同的方位上连续采样记录AP的100个RSS值,以及AP的相关信息。将所有的采样点的物理坐标及相应的物理坐标及RSS值存储为定位过程所调用的数据,建立Radio Map。在实验环境共有900个参考点,其采样密度为0.5米×0.5米,如附图4所示。Radio Map作为CM-SDE算法的输入参数及本征维数估计算法的输入数据。

[0086] 二、Radio Map的本征维数的获取通过下述步骤实现:

[0087] 本征维数是对于高维数据进行本征空间维数及空间重建所需最小的独立变量的个数。在具体实际计算中,由于高维数据的本征并不明显,通常不是寻求得到确切的本征维数,而是寻求估计本征维数的可信取值。具体的说,给定一个来自高维空间的样本,本征维数估计算法的中心任务和重要内容就是通过这些样本数据来确定这个高维结构的本征维数。

[0088] Radio Map的本征维数的估计是CM-SDE算法的重要输入参数,这关系到降维的结果是否能够代表Radio Map的高维空间的特征,因此准确有效的本征维数的估计至关重要。目前,常用本征维数估计算法分为两类:局部估计与全局估计。采用全局算法估计对Radio Map的本征维数进行估计,并作为CM-SDE算法的输入变量。本实验中采用测地线最小生成树算法(Geodesic Minimum Spanning Tree,GMST)对Radio Map的本征维数进行估计。

[0089] 下面对GMST算法的理论进行分析。

[0090] 测地线最小生成树(GMST)估计是基于测地线最小生成树的长度函数依赖于本征维数d。GMST是指定义在数据集X上的近邻曲线的最小生成树。GMST的长度函数L(X)是在测地线最小生成树中所有边缘对应的欧氏距离之和。

[0091] GMST估计在数据集X上构造一条近邻曲线G,其中,在X内每一个数据点x_i都和它的k个近邻x_j相连接。测地线最小生成树T定义为X上的最小曲线,它具有长度

$$[0092] L(X) = \min_{T \in \Upsilon} \sum_{e \in T} g_e \quad (1)$$

[0093] 其中,Υ是曲线G的所有子树集合,e是树T的一个边缘,g_e是边缘e对应的欧氏距离,其计算公式为式(2)所示。

$$[0094] g_e = ||x_i - x_j||, x_i, x_j \in e \quad (2)$$

[0095] 在GMST估计中,一些子集A ⊂ X由各种大小m组成,并且子集A的GMST的长度L(A)

也需要计算。理论上, $\frac{\log L(A)}{\log m}$ 是线性的, 从而可以由 $y=ax+b$ 这种形式的函数来估计, 通过

最小二乘法可以估算出变量 a 和 b 。可以证明, 由 a 的估算值和 $\hat{d} = \frac{1}{1-a}$ 能够得到本征维数的

估计。由GMST算法给出本征维数 d 的表达式为式(3)所示。本征维数 d 是CM-SDE算法的另一个重要的输入参数。

$$[0096] \quad d = \frac{1}{1-a} \quad (3)$$

[0097] 三、CM-SDE是一种半监督流形学习算法, 在CM-SDE算法实现的过程中需要对所有的参考点进行类标记。考虑到目前的WiFi室内定位环境中, 参考点数目将近1000点, 并没有人为地对所有的参考点进行类标记, 而是采用一定的分类算法对参考点的类别进行标记。

[0098] 聚类的目标是将数据集 $X = \{x_1, x_2, \dots, x_n\}$ 划分为 c 类且各类数据之间互不相关。基本的聚类算法按如下步骤实现:

[0099] (1)生成 c 个聚类中心, 记为 $v_i, i=1, 2, \dots, c$ 。

[0100] (2)将数据集 $X = \{x_1, x_2, \dots, x_n\}$ 的每个元素归类, 采用最近邻(Nearest Neighbor)算法判定元素的归属关系, 其等价表达式为:

$$[0101] \quad x_i \in G_i \Leftrightarrow \arg \min_{1 \leq j \leq c} D(x_i, v_j) \quad (4)$$

[0102] 式(4)中, x_i 为第 i 个数据点, G_i 为第 i 类构成的邻接关系图。 $D(x_i, v_j)$ 表示计算 x_i 与 v_j 之间的欧式距离。

[0103] (3)聚类中心的更新, 对于第 i 类的聚类中心更新为:

$$[0104] \quad v_i = \frac{1}{|G_i|} \sum_{x_k \in G_i} x_k \quad (5)$$

[0105] 式(5)中, $|\cdot|$ 表示计算某类中元素的个数。

[0106] (4)收敛性校验及迭代

[0107] 若满足以下四种情况下的收敛性条件之一, 则迭代停止, 否则重复执行(2)~(3), 直到迭代收敛或者达到最大执行次数。四种收敛性判定条件为:

[0108] 条件一: 聚类中心不变;

[0109] 条件二: 每个聚类的元素不变;

[0110] 条件三: 聚类中心变化收敛于半径 ϵ 内;

[0111] 条件四: 聚类元素变化收敛于半径 ϵ 内。

[0112] 总的来说, 上述收敛性判定条件可以表述为下式:

$$[0113] \quad \max |v_i - v_i'| \leq \epsilon, \epsilon \geq 0 \quad (6)$$

[0114] 其中, v_i' 表示更新后的第 i 类的聚类中心。

[0115] 从上述聚类算法的基本实现方式分析, 用于判定元素的归属关系的算法构成了聚类的核心。不同类型的聚类算法提出不同的归类指标, 一般将该函数称为损耗函数(Loss Function)。基本的聚类方法中采用欧式距离作为损耗函数。本专利采用KFCM算法对Radio Map进行类别分析得出聚类中心和Radio Map的类别标记。KFCM算法的理论分析如下:

[0116] 引入核函数的模糊c均值聚类的目标是将原始的数据集所在空间变换至无穷维的希尔伯特空间(Hilbert Space),再对变换后的空间作相应的聚类分析。通过核函数的变换,将原始数据之间的类别特征进一步变换后更易于表述和区分。基于核函数的模糊c均值聚类算法的目标函数为:

$$[0117] J_{KFCM} = \sum_{i=1}^c \sum_{k=1}^n (u_{ki})^m \| \Phi(x_k) - W_i \|_H^2 \quad (7)$$

[0118] 式(7)中, $\Phi(x_k)$ 、 W_i 分别表示在希尔伯特空间下的数据集及相应的聚类中心。通过推导可以得出KFCM算法的解表述为:

$$[0119] \begin{cases} u_{ki} = \left[\sum_{j=1}^c \left(\frac{\|\Phi(x_k) - W_j\|_H^2}{\|\Phi(x_k) - W_i\|_H^2} \right)^{\frac{1}{m-1}} \right]^{-1} \\ v_i = \frac{\sum_{k=1}^n (u_{ki})^m \Phi(x_k)}{\sum_{k=1}^n (u_{ki})^m} \end{cases} \quad (8)$$

[0120] KFCM的解的关键在于计算希尔伯特空间的损耗函数或者相似度函数。在本文中考虑引入高斯核函数(Gaussian Kernel Function)的FCM(Fuzzy C-Means)算法的理论分析及其实现。高斯核函数如式(9)所示。

$$[0121] K(x, y) = e^{-\lambda \|x-y\|^2} \quad (9)$$

[0122] 希尔伯特空间中,由式 $\|\Phi(x_k) - W_i\|_H^2$ 表述其相应的损耗函数,该式进一步表述为式(10)。

$$[0123] \|\Phi(x_k) - W_i\|_H^2 = \langle \Phi(x_k) - W_i, \Phi(x_k) - W_i \rangle \quad (10)$$

[0124] 式(10)中, $\langle \cdot, \cdot \rangle$ 表示计算相应式的核函数值。而实际上,无穷空间的变换不存在,因此,对式(10)进一步简化为式(11)所示。

$$[0125] D_{ki} = \|\Phi(x_k) - W_i\|_H^2 = \langle \Phi(x_k), \Phi(x_k) \rangle - 2 \langle \Phi(x_k), W_i \rangle + \langle W_i, W_i \rangle \quad (11)$$

[0126] 式(11)的全展开式为式(12)所示。

$$[0127] \begin{aligned} D_{ki} &= \langle \Phi(x_k), \Phi(x_k) \rangle - \frac{2}{\sum_{k=1}^n (u_{ki})^m} \sum_{j=1}^n \langle \Phi(x_k), \Phi(x_j) \rangle \\ &\quad + \frac{1}{\left(\sum_{k=1}^n (u_{ki})^m \right)^2} \sum_{j=1}^n \sum_{l=1}^n \langle \Phi(x_l), \Phi(x_j) \rangle \end{aligned} \quad (12)$$

[0128] 式(12)中, $\langle \Phi(x_k), \Phi(x_j) \rangle$ 由高斯核函数计算,即:

$$[0129] \langle \Phi(x_k), \Phi(x_j) \rangle = K(x_j, x_k) = e^{-\lambda \|x_j - x_k\|^2} \quad (13)$$

[0130] 在算法实现中,不是随机生成聚类中心,而是从数据集 $X = \{x_1, x_2, \dots, x_n\}$ 中随机选择 c 个元素作为聚类中心,构成集合 $Y = \{y_1, \dots, y_c\}$ 。因此,初始化的损耗函数值计算如式所示。

$$[0131] D_k^o = \|\Phi(x_k) - \Phi(y_i)\| = K(x_k, x_k) - 2K(x_k, y_i) + K(y_i, y_i) \quad (14)$$

[0132] 四、运用CM-SDE算法实现对Radio Map进行降维并获取特征权值矩阵过程通过下述步骤实现:

[0133] CM-SDE算法是基于标记数据与未标记数据的类间散度及类内散度最大化的一种流形学习算法。在对CM-SDE算法进行理论分析之前对CM-SDE算法给定的输入数据做如下说明:输入高维数据点 $\{x_i\}_{i=1}^m \in R^n$,数据点 x_i 的类标记为 $y_i \in \{1, 2, \dots, P\}$,其中P表示将高维数据划分为P个子流形,即将输入的高维数据分成P类,记P类的聚类中心为 $V = \{v_1, v_2, \dots, v_P\}$ 。将输入的高维数据表示成矩阵的形式: $X = [x_1, x_2, \dots, x_m] \in R^{n \times m}$ 。从矩阵表示的形式来看,矩阵中的列代表一个高维数据点。

[0134] 对于包含未标记数据的RadioMap u ,其中所有的未标记数据 $X_u = [x_{u1}, x_{u2}, \dots, x_{uk}] \in R^{n \times k}$ 进行类别匹配,同时已标记数据记为 $X_l = [x_{l1}, x_{l2}, \dots, x_{lc}] \in R^{n \times c}$ 。对于 X_u 中的所有数据进行有序类别匹配。有序的含义是当分配某一个未标记数据后,会影响相应类的聚类中心,因此会对下一未标记数据的类判别会有影响。本专利中主要的考虑信号的采集的时间顺序。假定 X_u 是按时间顺序排列。采用式(15)计算类的归属 x_{u1} ,并采用式(5)更新相应的聚类中心。然后依次将所有的未标记数据进行类别匹配

$$[0135] x_{ui} \in G_j \Leftrightarrow \arg \min_{1 \leq j \leq p} D(x_{ui}, v_j) \quad (15)$$

[0136] CM-SDE算法的目标函数为:

$$[0137] \begin{cases} \text{minimize} & \frac{X^T S_w X}{X^T S_t X} \\ \text{maximize} & \frac{X^T S_b X}{X^T S_t X} \end{cases} \quad (16)$$

[0138] 式(16)中 S_w 、 S_b 、 S_t 分别表示类内散度、类间散度及总散度可以由式(17)计算:

$$[0139] \begin{cases} S_w = \sum_{i=1}^P \left(\sum_{x_k \in G_i} (x_k - m^{(i)}) (x_k - m^{(i)})^T \right) \\ S_b = \sum_{i=1}^P (m - m^{(i)}) (m - m^{(i)})^T \\ S_t = S_w + S_b \end{cases} \quad (17)$$

[0140] 式(17)中, $m^{(i)} = \sum_{k=1}^{l_i} x_k / l_i$ 为第*i*类的均值, l_i 为第*i*类的采样点的数目;

$m = \sum_{k=1}^N x_k / N$ 为全体采样点的均值, N 为采样点的数目。

[0141] 对于式(16)所示的目标函数同样可以表示局部鉴别嵌入流形学习算法的目标函数形式,其表达式为:

$$[0142] \quad \begin{cases} \text{Maximize } J(V) = \sum_{i,j} \|V^T x_i - V^T x_j\|^2 w_{ij}' \\ \text{subject to } \sum_{i,j} \|V^T x_i - V^T x_j\|^2 w_{ij} = 1 \end{cases} \quad (18)$$

[0143] 式(18)中, w_{ij} 表示同类数据间的权重分配, w_{ij}' 表示不同类数据间的权重分配, 分别表示为 $W_{N \times N}$ 和 $W_{N \times N}'$ 。权重计算过程由两步完成。第一步:构造邻域图。根据高维数据点的类标记信息及其近邻关系构造无方向图 G 及 G' 。其中近邻关系是采用KNN算法给出的准则, 即选择数据点最近的 K 个点作为其邻居, G 表示当 x_i 与 x_j 的类标记信息 $y_i = y_j$ 时且 x_i, x_j 互为 K 近邻关系; G' 表示当 x_i 与 x_j 的类标记信息 $y_i \neq y_j$ 时且 x_i, x_j 互为 K 近邻关系。第二步:计算权值矩阵。根据第一步构造的邻接图采用类高斯函数进行权值矩阵的计算。其表达式(19)、(20)为所示。公式中 w_{ij} 表示近邻点 x_i 与 x_j 之间的权值, $\|x_i - x_j\|^2$ 为近邻点 x_i 与 x_j 之间的距离, 采用矩阵方式计算距离, t 为权值归一化参数, U, L 分别表示未标记和已标记的采样点的数目。根据分析可以知道, $W_{N \times N}$ 和 $W_{N \times N}'$ 可以由三部分构成, 分别是:已标记数据与已标记数据之间的权重、已标记数据与未标记数据之间的权重及未标记数据与未标记数据之间的权重, 分别表示为: $W_{L \times L} = \{w_{ij}^ll\}$ 、 $W_{L \times U} = \{w_{ij}^{lu}\}$ 、 $W_{U \times U} = \{w_{ij}^{uu}\}$ 。

$$[0144] \quad W_{N \times N} = \begin{cases} w_{ij}^l = \begin{cases} \exp(-\|x_i - x_j\|^2/t); & \text{if } x_i, x_j \in G; x_i \in X_L, x_j \in X_L \\ 0; & \text{otherwise} \end{cases} \\ w_{ij}^{lu} = \begin{cases} \exp(-\|x_i - x_j\|^2/t); & \text{if } x_i, x_j \in G; x_i \in X_L, x_j \in X_u \\ 0; & \text{otherwise} \end{cases} \\ w_{ij}^{uu} = \begin{cases} \exp(-\|x_i - x_j\|^2/t); & \text{if } x_i, x_j \in G; x_i \in X_u, x_j \in X_u \\ 0; & \text{otherwise} \end{cases} \end{cases} \quad (19)$$

$$[0145] \quad W'_{N \times N} = \begin{cases} w_{ij}^{ll} = \begin{cases} \exp(-\|x_i - x_j\|^2/t); & \text{if } x_i, x_j \in G'; x_i \in X_L, x_j \in X_L \\ 0; & \text{otherwise} \end{cases} \\ w_{ij}^{lu} = \begin{cases} \exp(-\|x_i - x_j\|^2/t); & \text{if } x_i, x_j \in G'; x_i \in X_L, x_j \in X_u \\ 0; & \text{otherwise} \end{cases} \\ w_{ij}^{uu} = \begin{cases} \exp(-\|x_i - x_j\|^2/t); & \text{if } x_i, x_j \in G'; x_i \in X_u, x_j \in X_u \\ 0; & \text{otherwise} \end{cases} \end{cases} \quad (20)$$

[0146] 由上述计算公式及矩阵的性质可以得: $W_{L \times L}^T = W_{L \times L}$ 、 $W_{U \times U}^T = W_{U \times U}$ 、 $W_{L \times U}^T = W_{L \times U} = W_{U \times L}^T = W_{U \times L}$ 。由此可以推导出 $W_{N \times N}$ 和 $W_{N \times N}'$ 表示为分块矩阵的形式, 如式(21)所示。

$$[0147] \quad \begin{cases} W_{N \times N} = \begin{bmatrix} W_{L \times L} & W_{L \times U} \\ W_{L \times U}^T & W_{U \times U} \end{bmatrix} \\ W'_{N \times N} = \begin{bmatrix} W'_{L \times L} & W'_{L \times U} \\ W'_{L \times U}^T & W'_{U \times U} \end{bmatrix} \end{cases} \quad (21)$$

[0148] 根据矩阵的计算式: $\|A\|^2 = \sum_{i,j} a_{ij}^2$, 计算式表示为矩阵A的矩阵的计算方法, 计算式给出的方法与矩阵的迹的计算式一致, 即: $\|A\|^2 = \text{tr}(AA^T)$ 。由此式(18)可以表示为矩阵的迹的计算方式:

$$[0149] \quad J(V) = \sum_{i,j} \left\{ \text{tr} \left[(V^T x_i - V^T x_j)(V^T x_i - V^T x_j)^T w_y' \right] \right\} \quad (22)$$

[0150] 式(22)可以简化为:

$$[0151] \quad J(V) = \sum_{i,j} \left\{ \text{tr} \left[V^T (x_i - x_j)(x_i - x_j)^T V \right] w_y' \right\} \quad (23)$$

[0152] 由矩阵迹的计算的标量性质及权值元素均为实数, 可以将式(23)简化为:

$$[0153] \quad J(V) = \text{tr} \left\{ V^T \sum_{i,j} \left[(x_i - x_j) w_y' (x_i^T - x_j^T) \right] V \right\} \quad (24)$$

[0154] 根据简单的数学关系, 可以将式(24)简化为:

$$[0155] \quad J(V) = 2 \text{tr} \{ V^T [X(D' - W'_{N \times N}) X^T] V \} \quad (25)$$

[0156] 式(25)中: X为输入数据, λ 和v为特征值与特征向量, W和W'分别为G及G'对应的权值矩阵, D及D'均为对角阵, 其对角元素可以由式(26)表示。

$$[0157] \quad \begin{cases} d_{ii} = \sum_j w_y' \\ d'_{ii} = \sum_j w_y' \end{cases} \quad (26)$$

[0158] 根据式(25)的推导方式, 同理可以将式(18)中的约束条件写成如式(25)相似的形式, 由此, 可以将(18)表示为如下形式:

$$[0159] \quad \begin{cases} \text{Maximize } J(V) = 2 \text{tr} \left\{ V^T \left[X \left(D' - W'_{N \times N} \right) X^T \right] V \right\} \\ \text{subject to } 2 \text{tr} \left[X \left(D - W_{N \times N} \right) X^T \right] = 1 \end{cases} \quad (27)$$

[0160] 对式(27)应用拉格朗日(Lagrange)乘数法, 可以得出式(28)所示:

$$[0161] \quad X(D' - W'_{N \times N}) X^T v = \lambda X(D - W_{N \times N}) X^T v \quad (28)$$

[0162] 对式(28)进行广义特征值分解, 得出其特征值分解的特征值及特征向量, 表示为: $\lambda = [\lambda_1, \lambda_2, \dots, \lambda_n]^T$, 其对应的特征向量为: $v = [v_1, v_2, \dots, v_n]^T$ 。取前d个最大的特征值对应的特征向量构成变换矩阵 $V = [v_1, v_2, \dots, v_d]$ 。由CM-SDE算法的输出数据变换方法可以得出, 降维后数据为:

$$[0163] \quad z_i = V^T x_i \quad (29)$$

[0164] 式(29)中, z_i 表示输入高维数据点 x_i 变换后的低维输出数据。从本专利中发明内容

给出的离线阶段的实施步骤为：第一步先采用CM-SDE算法对所有参考点Radio Map进行降维处理，得到相应的参考点的降维后的Radio Map，即作为在线阶段的匹配定位数据库(RadioMap*)。第二步对添加未标记数据的Radio Map，即RadioMap_{u1}进行降维处理，得到特征变换矩阵V'。由此可以建立离线阶段所需要数据库：RadioMap*和V'。

[0165] 五、由不同的用户在线阶段定位阶段获得的RSS是未标记类别属性的，其加入Radio Map，并构成Radio Map_{u1}的过程称为类别。其实现方法如下所述：

[0166] 将不同用户在线阶段定位阶段测试得到的未标记RSS，采用类别匹配的方式加入至已有Radio Map中，得到相应的包含未标记信号覆盖图RadioMap_{u1}；通过类别匹配方法增加Radio Map的数据量，进而提高Radio Map的密度，为CM-SDE算法提供新的降维数据，同时可以更新聚类中心，为CM-SDE算法提供新的类别数据。类别匹配方法分为两步，其实现过程如下所述：

[0167] 第一步，寻找未标记RSS的类别属性。记一组未标记RSS为RSS_i，与步骤三中的聚类中心进行匹配，由式(4)完成RSS_i的类别标记。

[0168] 第二步：对RSS进行门限检测。对于聚类中心v_i表示为v_i=(v_{i1}, v_{i2}, ..., v_{iN})，N为室内定位系统中AP的个数。RSS_i表示为RSS_i=(RSS_{i1}, RSS_{i2}, ..., RSS_{iN})。计算下式所定义的广义符号值：

$$[0169] S_i = \sum_{j=1}^N \text{sgn}(RSS_{ij} - v_{ij}) \quad (30)$$

[0170] 其中，sgn(•)定义为：

$$[0171] \text{sgn}(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (31)$$

[0172] 当S_i大于设定门限值时，则将RSS_i加入Radio Map中，并更新聚类中心，否则舍弃RSS_i，不加入Radio Map中。在本专利中，门限值V_T=0.9N。聚类中心的更新由式(5)完成：

[0173] 六、基于CM-SDE算法的WiFi室内定位方法的离线数据库实现方式：

[0174] 离线数据库方式由三部分构成。第一，所有参考点的Radio Map的建立，并采用CM-SDE算法得到RadioMap*。第二，再随机采样U点未标记数据并添加入原有的Radio Map中，并用CM-SDE算法得到V'，并将形成的定位数据库下载(存储)到定位的移动终端。第三，在线定位实现及Radio Map更新。第三部分的具体实现如下所述：

[0175] 在线阶段，测试点处接收的RSS=[AP₁, AP₂, ..., AP_n]，n表示室内定位系统布置的AP的数目。将RSS与特征变换矩阵V'相乘，从而得出降维后的RSS'=[AP₁, AP₂, ..., AP_d]，其中d表示本征维数。再采用KNN算法实现RSS'与RadioMap*的匹配。采用与RSS'最近的K个参考点的坐标的平均值作为测试点(x', y')，其表达式为：

$$[0176] (x', y') = \frac{1}{K} \sum_{i=1}^K (x_i, y_i) \quad (32)$$

[0177] 将定位用户本次定位测得的RSS作为未标记数据采用类别匹配方式加入到Radio Map中，并在移动终端上实现对本地在线匹配定位数据库的更新，实现动态的更新本地的数据，从而实现离线数据库定位方式，附图1所示于类别匹配的半监督流形学习的WiFi室内定位方法在用户定位终端实现；

[0178] 七、基于CM-SDE算法的WiFi室内定位方法的在线数据库实现方式：

[0179] 在线数据库方式由四部分构成。第一，所有参考点的Radio Map的建立，并采用CM-SDE算法得到RadioMap*。第二，再随机采样U点未标记数据并添加入原有的Radio Map中，并用CM-SDE算法得到V'，并将形成的定位数据库下载(存储)到定位的移动终端。第三，在线定位实现及Radio Map更新。第三部分的具体实现如下所述：

[0180] 在线阶段，测试点处接收的RSS=[AP₁, AP₂, …, AP_n]，n表示室内定位系统布置的AP的数目。将RSS与特征变换矩阵V'相乘，从而得出降维后的RSS'=[AP₁, AP₂, …, AP_d]，其中d表示本征维数。再采用KNN算法实现RSS'与RadioMap*的匹配。采用与RSS'最近的K个参考点的坐标的平均值作为测试点(x', y')，其表达式为：

$$[0181] (x', y') = \frac{1}{K} \sum_{i=1}^K (x_i, y_i) \quad (33)$$

[0182] 第四部分：用户在线定位完成后，将用户本次在线测得的RSS值上传至在线定位数据库所在的服务器，并在服务器端将在线定位数据库进行更新将将在线定位数据传回上传RSS数据的定位终端，即附图2所示的离线阶段在线定位数据库所在服务器上完成，而在线阶段在定位终端完成。

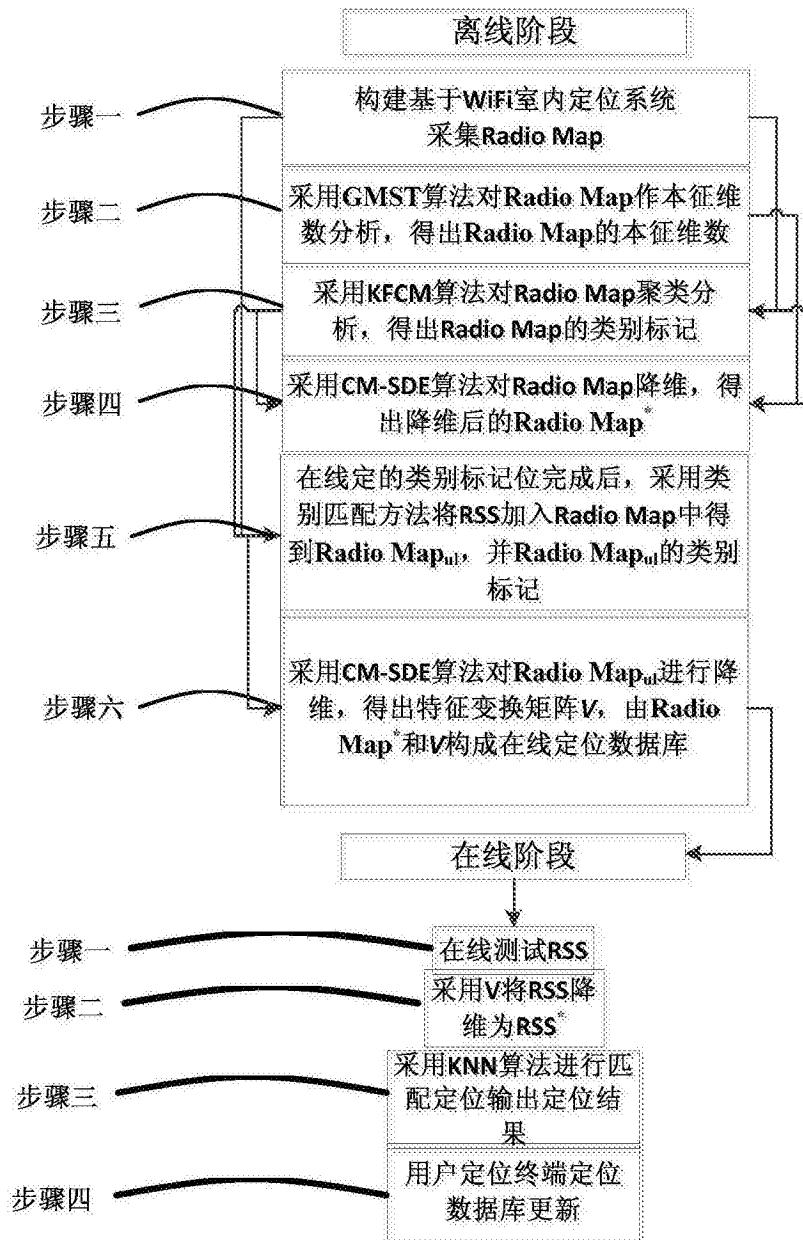


图1

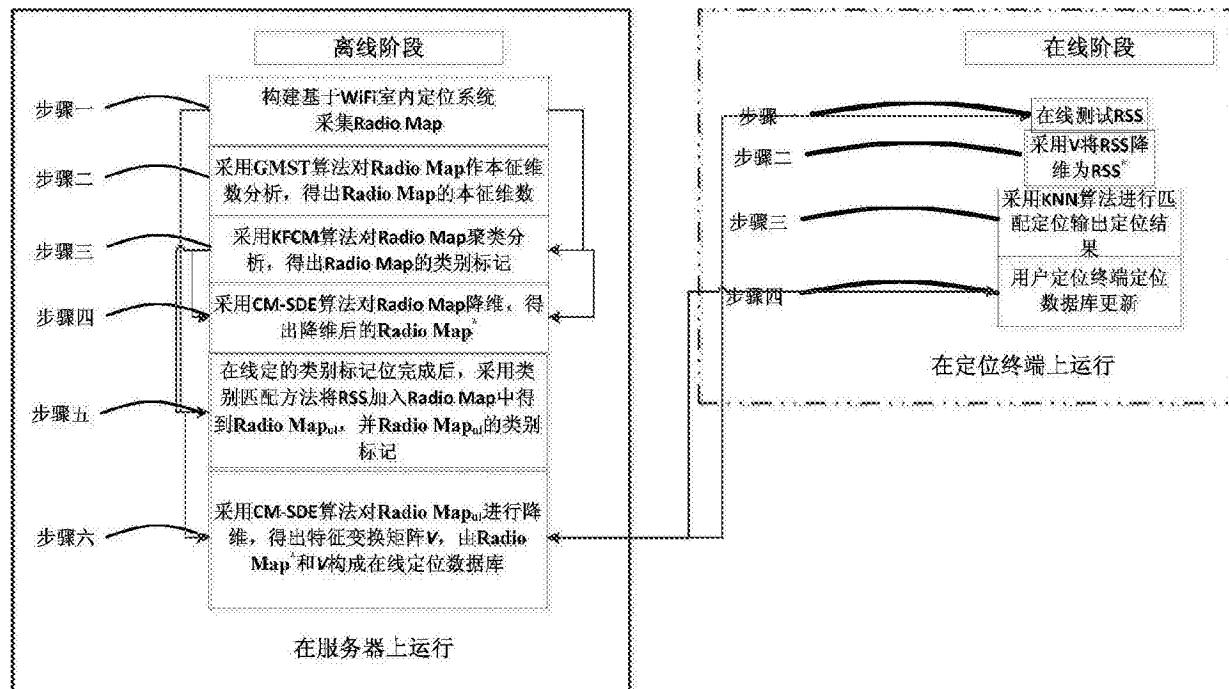
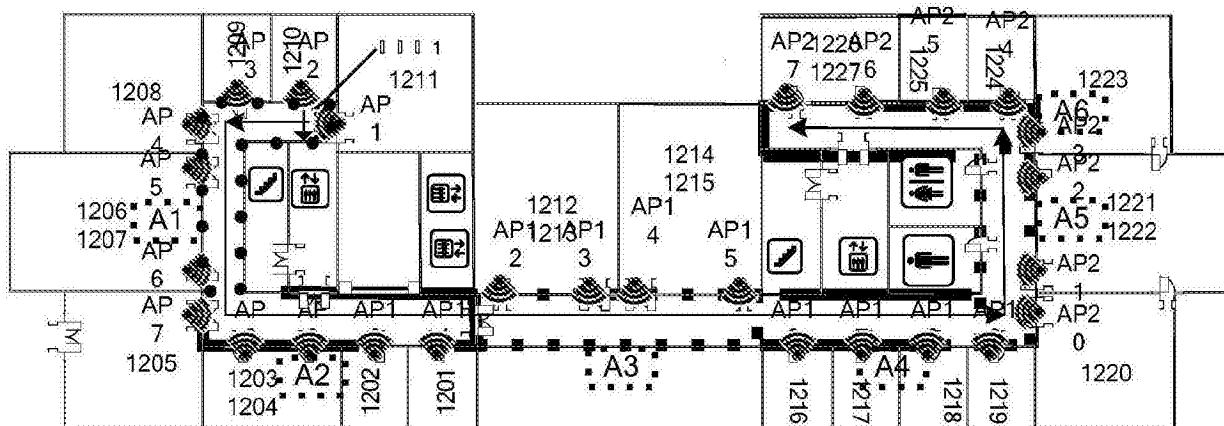


图2



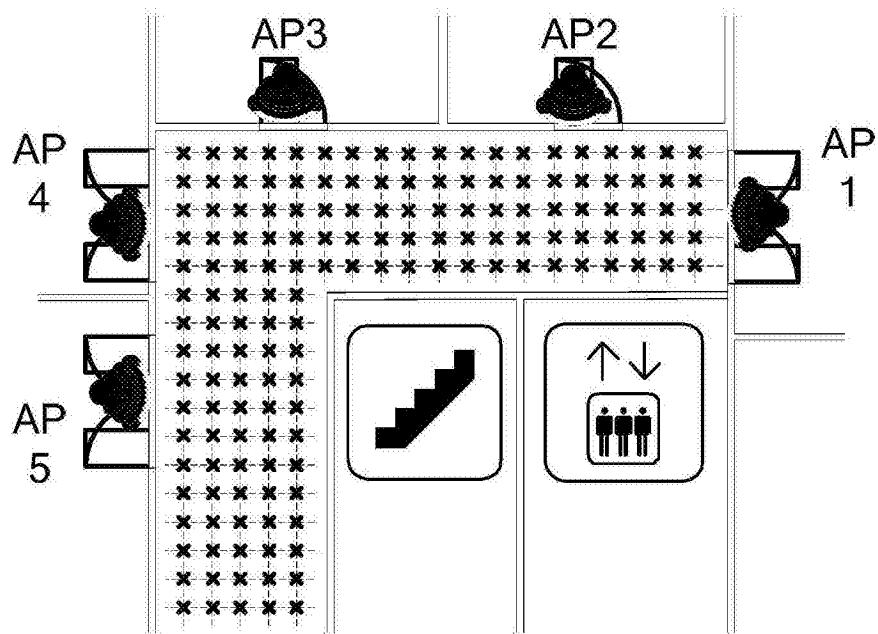


图4

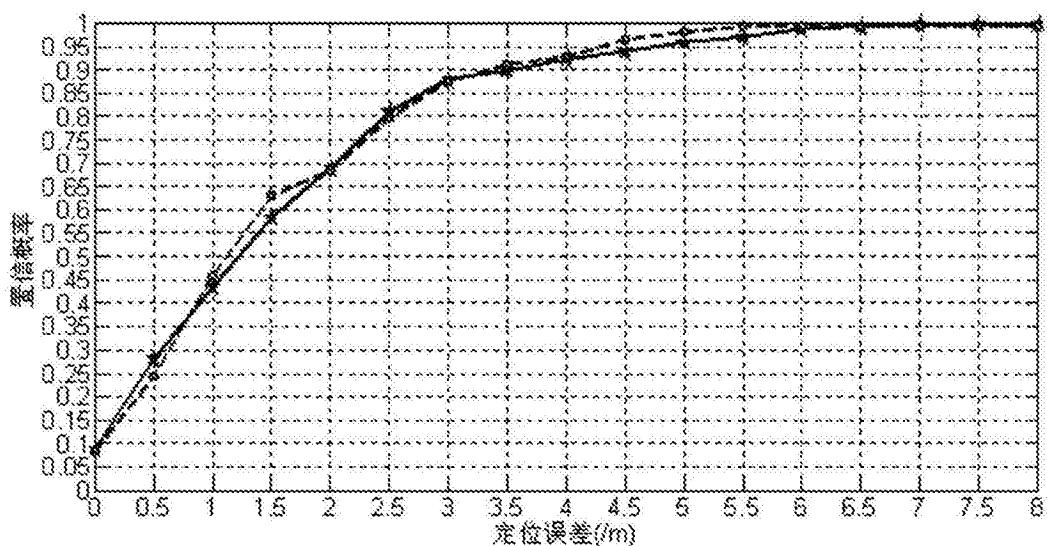


图5