

基于正态分布概率计算和支持向量机计算的 WiFi 定位技术

北京优锐科技有限公司 丁贵金 朱韬 袁万尚

2013 年 7 月 12 日

摘要

基于概率和支持向量机原理的 WiFi 定位技术，不需要依赖专用设备，部署简单使用便捷，对环境无强制依赖，可以在复杂 WiFi 环境下实现移动设备精确定位。该 WiFi 定位技术的核心原理是支持向量机，辅助正态分布的概率计算来优化支持向量机计算过程。

本文针对的设备：是带有 WiFi 功能的移动设备。针对的环境：是分布着大量 WiFi 接入设备的室内环境。实现的主要目标：是通过 WiFi 移动设备，在分布着大量 WiFi 接入设备的室内环境中，实现精确定位。

关键词：WiFi 定位，室内定位，支持向量机。

1 引言

本文的“名词、概念和常识说明”部分，主要解释全文中出现的各种专用名词，涉及到的所有概念，以及阅读本文所需的技术常识。“实

现方法”部分则具体阐述技术实现过程，“讨论”部分主要分析了实现过程中可能出现的各种问题，以及处理问题的方法。“技术创新”部分具体说明该技术的先进性，与同类技术相比下的优势，“应用前景”部分介绍了该技术实际应用的具体形式，以及对采用该技术的行业所产生的积极作用。

2 名词、概念和常识说明

2.1 定位空间 (CR)

定位空间 (calibration room)，指的是提供定位功能的空间。例如采用 WiFi 定位的公司内部空间，公司外部无限大的空间就是不可定位空间。如图1：

2.2 移动终端设备

本文中的“移动终端设备”，是指拥有 WiFi 连接能力的移动设备，如有 WiFi 功能的智能手

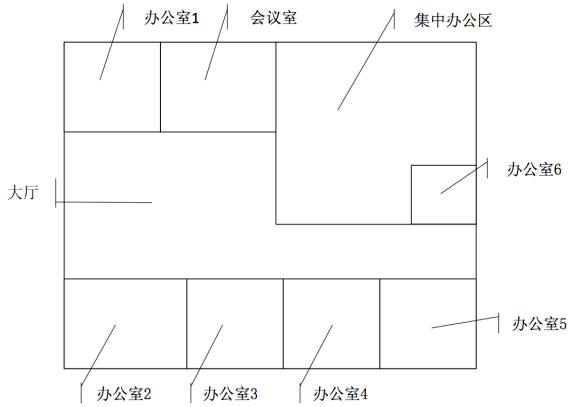


图 1: 定位空间示例图

机、平板电脑和其他便携的移动终端设备。

2.3 WiFi 接入设备 (AP)

提供 WiFi 接入服务的硬件设备，如无线路由器。本文中对 WiFi 接入设备的要求，是必须能够被移动终端设备识别到的 WiFi 接入设备。

2.4 AP 名 (MAC)

本文中所用 AP 名就是 AP 的 MAC 地址，是 WiFi 接入设备的唯一标示。

2.5 AP 信号强度 (RSS)

AP 信号强度，是移动终端设备对 WiFi 接入设备信号强度的识别结果。该结果是一个从 0 到 -100 的值，该值是一个指数值，代表信号的强度指数，并不是真正的物理信号量。不同的移动终端设备，对同一个 WiFi 接入设备，信号强度值的识别结果是不同的。

2.6 AP 信号强度平均值 (mRSS)

同一个移动终端设备，对同一个 WiFi 接入设备，每次信号强度的识别结果也是不同的。一个移动终端设备，对同一个 WiFi 接入设备，进行多次采集，识别到一组信号强度，对该组信号强度的求平均值，就得到了信号强度平均值，计算公式如下：

$$mRSS = 10 \lg \left(\frac{1}{N} \sum_{i=1}^N 10^{\frac{RSS_i}{10}} \right) \quad (1)$$

N 为该组信号强度的个数。

2.7 WiFi 接入设备数据

WiFi 接入设备数据包括，WiFi 接入设备名 MAC 地址，和 WiFi 接入设备识别到 AP 的信号强度。

2.8 待定位点 (LP)

带定位点，是指一个移动终端设备，出现在定位空间内，该移动终端设备自身的位置。

2.9 定位区域 (value)

定位空间是一个完整的封闭空间，要在这个空间内实现 WiFi 定位，就必须将这个完整的空间划分为多个小区域，每个区域都有一个 ID。如图2:

2.10 定位校准点 (CP)

在定位区域中，指定一个位置点，这个点就是定位校准点。一个定位区域内可能会有多

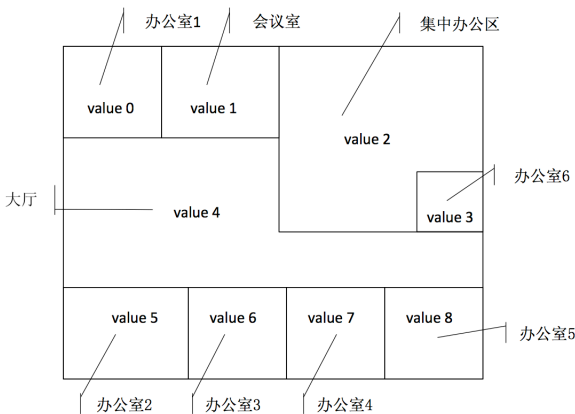


图 2: 定位区域示意图

其中 value 0 中 0 就是该定位区域的 ID。

个定位校准点，在这些点上所识别到的 WiFi 接入设备数据，都输入这个定位区域的定位区域数据。如图3:

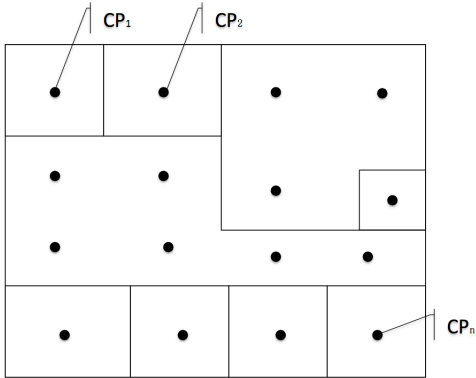


图 3: 定位校准点示意图

2.11 定位区域数据

在定位区域内的定位校准点上，一次采集到的 WiFi 接入设备数据，就是定位区域采集数据。该数据的标示是区域 ID，内容是 WiFi 接

入设备数据。如下表所示:

<i>Value</i>	AP_i	RSS_i
0	MAC_1	RSS_1
0	MAC_2	RSS_2
0	MAC_3	RSS_3
0	MAC_4	RSS_4
0	MAC_5	RSS_5
...

(2)

多次采集定位区域数据后，一个 AP 会对应多个 RSS 值。如下表所示:

<i>Value</i>	AP_i	$RSS_{i,n}$
0	MAC_1	$RSS_{1,1}...RSS_{1,2}...RSS_{1,n}$
0	MAC_2	$RSS_{2,1}...RSS_{2,2}...RSS_{2,n}$
0	MAC_3	$RSS_{3,1}...RSS_{3,2}...RSS_{3,n}$
0	MAC_4	$RSS_{4,1}...RSS_{4,2}...RSS_{4,n}$
0	MAC_5	$RSS_{5,1}...RSS_{5,2}...RSS_{5,n}$
...

(3)

2.12 校准数据

校准数据，是对所有定位区域多次采集到的数据，整理计算后的结果。如下表所示：

$Value(n)$	AP_i	$mRSS_n$	$\sigma_{n,i}$
0	MAC_1	$mRSS_1$	$\sigma_{0,1}$
0	MAC_2	$mRSS_2$	$\sigma_{0,2}$
0	MAC_3	$mRSS_3$	$\sigma_{0,3}$
0	MAC_4	$mRSS_4$	$\sigma_{0,4}$
0	MAC_5	$mRSS_5$	$\sigma_{0,5}$
...
1	MAC_1	$mRSS_1$	$\sigma_{1,1}$
1	MAC_2	$mRSS_2$	$\sigma_{1,2}$
1	MAC_3	$mRSS_3$	$\sigma_{1,3}$
1	MAC_4	$mRSS_4$	$\sigma_{1,4}$
1	MAC_5	$mRSS_5$	$\sigma_{1,5}$
...

(4)

Value 是定位区域，n 代表不同定位区域的 ID 号。AP 是收集到的所有 WiFi 接入设备名，mRSS 是该 WiFi 接入设备被识别到的信号强度平均值，i 代表不同 AP 的序号， $\sigma_{n,i}$ 代表对应一个定位区域中一个 AP 的信号强度标准差。

2.13 待定位数据

移动设备在待定位点，某一时刻采集到的 WiFi 接入设备数据。如下表：

AP_i	RSS_i
MAC_1	RSS_1
MAC_2	RSS_2
MAC_3	RSS_3
MAC_4	RSS_4
MAC_5	RSS_5
...	...

(5)

AP 是识别到的所有 WiFi 接入设备名，RSS 是识别到的所有 WiFi 设备的信号强度。

2.14 概率

2.14.1 正态分布

正态分布是一种常用的概率分布函数，本文阐述的 WiFi 定位技术，在实现过程中采用了这种数学方法。公式如下：

$$Pr_i = \frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{-(RSS_i - mRSS_i)^2}{2\sigma_i^2}} \quad (6)$$

2.14.2 概率值 (Pr)

本文中所讲的概率值，是指一个待定位点，出现在某一定位区域中的概率。

2.15 支持向量机 (SVM)

支持向量机，是一种机器学习原理，是采用数学方法，实现对某种向量值进行特定分类的工具。

2.15.1 样本

本文所说的样本，是指 WiFi 接入设备的数据。

2.15.2 样本空间

多个样本生成的集合，叫做样本空间。特征向量本文所说的特征向量，是指可以代表一个定位区域的样本。是校准数据中，一个定位区域中所有样本的子集。

2.15.3 核函数

核函数，是指支持向量机所采用的某种特征向量训练模式。本文中的核函数采用 RBF 核函数，采用该核函数时，需要提供一个高斯参数 σ 作为训练调优参数，默认值使用 0.75。

2.15.4 惩罚因子 C

惩罚因子，是指在支持向量机计算中，分割两个样本空间时，对不可分样本的包容程度。默认值 1.0。

3 实现方法

3.1 总体流程

如图4

3.2 确定 CR、Value、CP

1. 根据实际需求，确定定位空间整体范围。如图5:

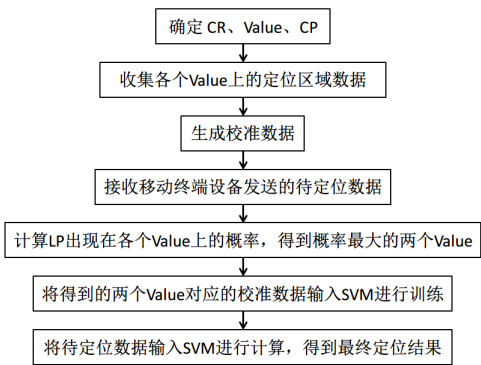


图 4: 总体流程示意图

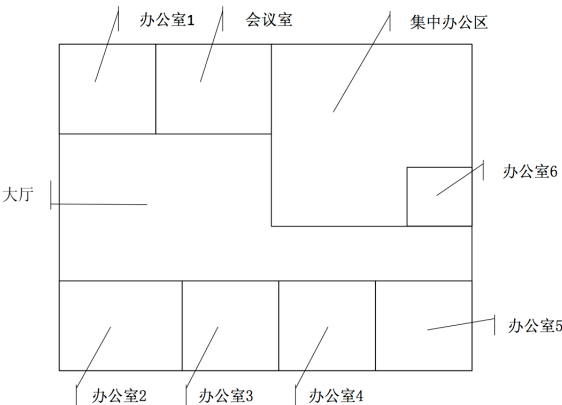


图 5: 确定定位空间

2. 在定位空间内划分出各个定位区域。如图6:
3. 在每个定位区域内确定一个或多个 CP。如图7:

3.3 收集各个 Value 上的定位区域数据

用移动终端设备，在定位空间内，收集所有定位区域中的多组定位区域数据，将这些数据按定位区域 ID，统一保存。

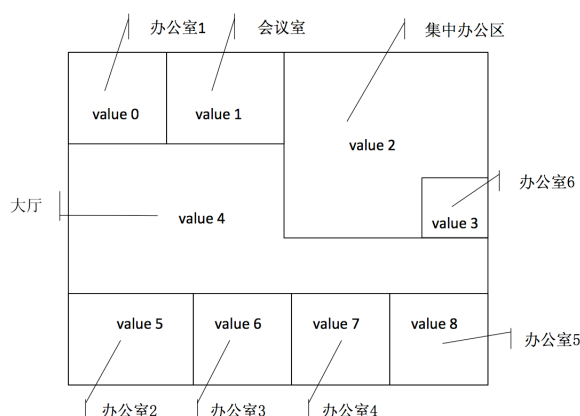


图 6: 确定定位区域

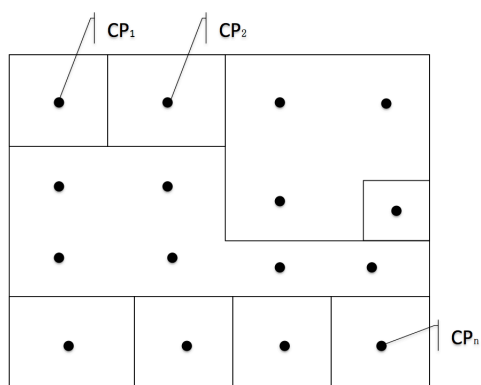


图 7: 确定定位校准点

3.4 生成校准数据

将收集到的区域定位数据，进行处理，计算出其中每个 AP 的信号平均值（mRSS）和信号强度值的标准差。按照各 AP 所对应的区域 ID（CP value）组织计算结果，统一保存生成校准

数据。如下表所示：

$Value(n)$	AP_i	$mRSS_n$	$\sigma_{n,i}$
0	MAC_1	$mRSS_1$	$\sigma_{0,1}$
0	MAC_2	$mRSS_2$	$\sigma_{0,2}$
0	MAC_3	$mRSS_3$	$\sigma_{0,3}$
0	MAC_4	$mRSS_4$	$\sigma_{0,4}$
0	MAC_5	$mRSS_5$	$\sigma_{0,5}$
...
1	MAC_1	$mRSS_1$	$\sigma_{1,1}$
1	MAC_2	$mRSS_2$	$\sigma_{1,2}$
1	MAC_3	$mRSS_3$	$\sigma_{1,3}$
1	MAC_4	$mRSS_4$	$\sigma_{1,4}$
1	MAC_5	$mRSS_5$	$\sigma_{1,5}$
...

(7)

3.5 得到移动终端设备的待定位数据

一个移动终端设备，在定位空间内移动，在不同的位置点上会收集到不同的待定位数据，这些不同的待定位数据都对应不同的待定位点（LP）。将移动终端设备在 LP 上收集到的带定位数据保存。如下表所示：

AP_i	RSS_i
MAC_1	RSS_1
MAC_2	RSS_2
MAC_3	RSS_3
MAC_4	RSS_4
MAC_5	RSS_5
...	...

(8)

3.6 计算 LP 出现在各个 Value 上的概率，得到概率最大的两个 Value

从校准数据中按照不同 Value，依次取出每个 Value 对应的一组校准数据，将 LP 的带定位数据与每组校准数据中共有的 AP，分别进行正态分布概率计算，计算公式如下：

$$Pr_i = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(RSS_i - mRSS_i)^2}{2\sigma_i^2}} \quad (9)$$

再将计算得到的概率值相加，最终得到了该 LP 出现在每个 Value 中的概率。

其中概率值最大的两个 Value，就是该 LP 可能出现的位置范围。

3.7 将得到的两个 Value 对应的校准数据输入 SVM 进行训练

得到了概率最大的两个 Value，说明 LP 出现在这两个 Value 中的可能性最大。将 LP 的待定位数据的 AP 名和这两个 Value 对应的校准数据的 AP 名做交集，得到一组大家都包含的 AP。

从 LP 的待测数据中挑出这个几个共有 AP 的待定位数据，作为 LP 的特征数据。

两个 Value 的校准数据中分别挑出这几个共有的 AP 校准数据，与该 Value 的 ID 一起，作为 SVM 的特征向量。

将特征向量输入 SVM 进行训练。如图8：

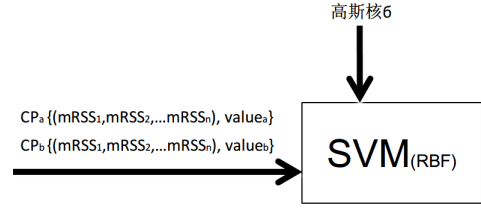


图 8: SVM 训练过程

3.8 将待定位数据输入 SVM 进行计算，得到最终定位结果

将 LP 的特征数据，输入训练完成的 SVM，得到一个 Value 的 ID 值。该 Value 所对应的区域就是 LP 所在的位置。如图9：

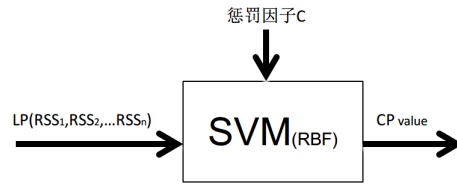


图 9: SVM 计算过程

4 讨论

在整个 WiFi 定位过程中，有时会出现无法定位的状况，具体分析有如下两种情况发生：

1. 待定位数据中的所有 AP，不包含在校准数据中。这种情况说明移动终端设备，出现在了定位空间之外的位置，因此无法实现定位。
2. 从待定位数据和校准数据中，无法找出共有的 AP，或者共有的 AP 少于 4 个。这种情

况下无法生成有效的特征向量, 因此 SVM 无法计算。

3. 在某些情形下, 待定位数据是 SVM 不可分类的。这种情况说明移动终端设备出现在两个定位区域的边界上, SVM 不能准确计算出移动终端设备的位置。为了比较好的解决这个问题, 在使用 SVM 的过程中可以使用惩罚因子作为计算参数, 来提高对不可分类数据的最大识别能力。
4. 对于使用惩罚因子后, 依然无法定位的数据, 采取失败处理。

5 技术创新 (权利保护)

5.1 用边界入口点校准的方法解决 WiFi 定位功能开关和不同类型设备识别的问题

本文所阐述的 WiFi 定位技术, 在实际使用的过程中, 首先会出现两个必须解决的问题:

1. 移动终端设备如何判断自身是否进入了一个可定位的空间, 并开启定位服务?
2. 移动终端设备如何才能让定位系统知道自身的设备类型?

为了解决以上问题, 我们采用边界入口点校准的方法实现。在定位空间的边界上, 设置一个或多个入口校准点, 所有需要 WiFi 定位服务的移动终端设备, 都必须先到这些入口点进

行校准。校准方式有很多种, 如手动开启 WiFi 定位应用, 二维码扫描开启 WiFi 定位应用等。

在开启移动终端设备 WiFi 定位应用的同时, 由于入口校准点是事先约定好的, 因此该校准点的定位区域数据也是相对固定的。根据移动终端设备在该点接收到的区域定位数据, 就可以判断出这个移动终端设备的 WiFi 信号识别能力。不同类型的移动终端设备, 对于同样的 AP, 信号强度识别结果是不同的, 根据这个特点, 便可以基本判断出与这个移动终端设备的最接近的设备类型。

5.2 用分级校准数据解决不同种类型 WiFi 设备识别 AP 信号差异问题

由于不同类型的移动终端设备, 对于同样的 AP, 信号强度识别结果不同, 而一组完整的校准数据是用一台移动终端设备采集的。因此使用一组校准数据, 无法实现多种不同类型的移动终端设备, 进行准确的 WiFi 定位。

为了解决这个问题, 必须准备多组校准数据, 这些数据使用不同类型的移动终端设备进行采集, 根据移动终端设备类型分组保存。

当移动终端设备在边界入口点开启 WiFi 定位应用的同时, 根据设备所采集到的入口点的定位区域数据, 可以判断出移动终端设备的类型, 再根据该移动终端设备类型选择相应的校准数据, 最终实现准确的 WiFi 定位。

5.3 用共有 AP 的方法解决 SVM 训练时 AP 在识别过程中缺失的问题

所有移动终端设备，在每次识别 AP 时都不可能做到绝对准确，因此一台移动终端设备不同时刻识别到的定位区域数据可能会存在不一致的情况。

为了解决以上问题，必须在每次定位时，采用带定位数据与校准数据交集的结果，动态的进行 SVM 训练，使每次 SVM 训练用的特征向量中的 AP，和 LP 的特征数据中的 AP 保证一致。

而每次定位都做 SVM 训练，计算量非常大，并且已知其中绝大部分计算都是无用的。因此在做 SVM 计算之前，我们采用概率的方式筛选出 LP 出现可能最大的两个定位区域。之后之针对这两个定位区域的校准数据做 SVM 计算，这样大大提高了计算效率，同时保证了定位结果的准确。

6 应用前景

WiFi 定位技术可以应用在多种领域，主要为行业提供室内定位基础服务。本文阐述的 WiFi 定位技术，不依赖特定硬件和特殊环境要求，部署应用方便灵活，成本低，可以结合多种具体的业务实现垂直服务。

6.1 室内定位和导航

室内定位和导航，是 WiFi 最主要的应用方式，也是最直接的服务提供模式。

1. 公共室内空间的位置服务，如商场、超市和大规模的综合购物中心。为了方便顾客找到所需商品位置，或快速找到某个区域(例如卫生间、出口、餐饮区，等等)，结合 WiFi 定位和电子地图技术，实现顾客自身位置确定和导航服务功能。
2. 室内区域流量统计服务，如在商场、超市和大规模的综合购物中心，有时业主需要统计一段时间内，某个区域或所有区域的客流信息，或者某个客流最大的热点区域。结合 WiFi 定位和电子地图以及大数据技术，可以实现室内区域的流量统计，顾客使用移动设备上的 WiFi 定位客户端软件，经过的路径会推送至服务器，服务器端根据统计整理，最后分析出某个或所有区域的客流数据。

6.2 地理围栏

地理围栏是一个新兴的移动互联网服务概念，其主要目的是实现互联网应用和移动设备位置信息的结合。

1. 设备位置开启移动应用。当移动设备进入某个特定区域后，便自动开启一些应用软件。这种服务适合应用在商场导购，医院就医指导，机场火车站等公共服务场所。
2. 特定移动应用的使用范围。移动设备上的某些软件，只能在某些特定的环境中才可以使用。这种服务适合公司企业，用来实现现代移动化的企业管理和自动化办公。当

员工进入企业办公区域后，相关工作的信息服务和业务软件才可以工作，保证了企业数据的安全，也将企业管理简单有效的实现。

6.2.1 移动设备防盗

某些非个人使用的移动设备，如餐厅的点餐设备，高级场馆的自助服务设备，都是为了本地服务存在的。为了防止个别人将其私自带出，又要实现该设备的自由使用，就必须采用一种防盗的技术做应用保证。本文阐述的 WiFi 定位技术，可以划出移动设备的安全范围，当设备离开这个安全区域后，便可以通过报警、监控和追踪等方式及时发现并处理。

6.2.2 区域信息安全

信息隔离，是信息安全中的一项重要内容，指信息内容与信息的有效区域之间的对应关系。采用本文阐述的 WiFi 定位技术，可以通过软件安全策略，规定信息的安全范围，当移动设备需要读取某些带有安全级别的数据时，会根据移动设备所处的位置进行操作合法性判断，当发现移动设备所处的位置不合法时，便会组织数据的读取。对于已经读取到的数据，当设备离开该数据的安全区域后，通过软件安全策略谁自动删除该数据，保证数据存在的有效范围与数据的安全范围一致。

6.2.3 地理信息隔离

移动设备的最大特点是可以随身携带，可以随时随地地产生数据，如照相、笔记、下载或应用软件生成数据。利用本文所阐述的 WiFi 定位技术，可以在产生数据的同时，将数据与其产生时的位置做关联。无论在以后的搜索查找，或是统计，位置信息会提供跟多的有效途径，版主使用者对数据的管理。也可以将数据的存储或显示方式，与位置信息关联，是某些数据只在特定位置区域是才可以被操作。

6.3 游戏

电子游戏的场景通常是虚拟的空间，但是随着技术的发展，电子游戏的实现也可以利用现实空间。本文阐述的 WiFi 定位技术，可以实现真实空间在设备上的体现，电子游戏的方式也就从单一的虚拟空间扩展到了真实空间，游戏玩家除了操作设备，还需要亲身行动来完成游戏。这样的组合不仅扩展了电子游戏的功能，甚至可以改变电子游戏的传统运行模式。