

面向 WLAN 室内定位的 T 检验样本容量优化方法

周 牧, 卫亚聪, 田增山, 李玲霞
(重庆邮电大学移动通信重点实验室, 重庆 400065)

摘 要: WLAN(Wireless Local Area Networks)室内定位已受到人们广泛的关注,而在离线指纹采集阶段常常容易造成位置指纹 RSS 数据采集的盲目性和不可靠性,并忽略所需采集 RSS(Received Signal Strength)样本容量与定位性能的关系.为了解决这一问题,本文提出一种面向 WLAN 室内定位的 T 检验样本容量优化方法.该方法在离线阶段利用 OC(Operating Characteristics)函数优化指纹数据库允许的最小 RSS 样本容量,而在在线阶段利用 T 检验方法对目标终端进行粗定位,并进而提出基于 T 检验的 KNN(K-Nearest Neighbour)算法以完成对目标终端的精定位.此方法用有限的样本容量获得较稳定的定位性能分析结果,显著地减少了大量的人力和时间开销.

关键词: 室内定位; 样本容量; T 检验; OC 函数; K 近邻

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2018)06-1351-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.06.011

T-test Based Sample Capacity Optimization for WLAN Indoor Localization

ZHOU Mu, WEI Ya-cong, TIAN Zeng-shan, LI Ling-xia

(Chongqing Key Lab of Mobile Communication Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: WLAN indoor localization has caught significantly wide attention. In offline phase, the location fingerprint RSS data acquisition often results in blindness and unreliability, and ignores the relations between the required RSS sample capacity and localization performance. To solve this problem, a new T-test based sample capacity optimization approach for WLAN indoor localization is proposed. In offline phase, the Operating Characteristics (OC) function is used to optimize the allowable minimum RSS sample capacity for the fingerprint database construction. In online phase, we perform the rough localization by using the T-test approach, and then propose the T-test based KNN algorithm for the fine localization of target terminal. This method uses a limited sample capacity to obtain a more stable localization performance analysis results, significantly reducing the amount of manpower and time overhead.

Key words: indoor localization; sample capacity; T-test; OC function; K-nearest neighbor

1 引言

在过去十年,位置服务(LBS, Location Based Service)需求的不断增加推动了定位技术的飞速发展^[1].虽然全球定位系统(GPS, Global Positioning System)可在室外提供米级的定位精度,但在室内环境下由于存在非视距(NLOS, Non-line-of-sight)传播的影响,GPS难以实

现高精度定位.因此,出现了诸如基于超声波(UW, Ultrasonic Wave)、红外(IR, Infrared Ray)、超宽带(UWB, Ultra-WideBand)、无线射频识别(RFID, Radio Frequency Identification)和无线局域网(WLAN, Wireless Local Area Network)等技术的室内定位系统.从基础设施成本和环境适应性两方面考虑,基于位置指纹的WLAN室内定位系统得到了广泛的研究和应用.

收稿日期:2017-01-06;修回日期:2017-06-05;责任编辑:梅志强

基金项目:国家自然科学基金资助项目(No. 61301126, No. 61471077);长江学者和创新团队发展计划基金资助项目(No. IRT1299);重庆市科委重点实验室专项经费基金资助项目;重庆市基础与前沿研究计划基金资助项目(重点)(No. cstc2015jcyjBX0065);重庆市高校优秀成果转化基金资助项目(No. KJZH17117);重庆市科委重点实验室专项经费,重庆邮电大学青年科学研究项目(No. A2013-31)

基于位置指纹的 WLAN 室内定位系统在离线指纹采集阶段需要耗费大量的人力和时间开销,同时,为了实现高精度定位,通常要求在每个参考点处采集大量的 RSS 样本.于是,为了降低指纹采集开销,基于传播模型的 WLAN 室内定位系统应运而生,其利用对数距离损耗模型^[2]、MK 模型、RADAR 模型或高斯过程隐变量模型 (GPLVM, Gaussian Process Latent Variable Models)^[3]等传播模型构建位置指纹数据库,但此类方法存在的主要问题是需要对室内环境中的信号分布特性进行精确估计.与此同时,基于信号插值的 WLAN 室内定位系统^[1,4]也可在一定程度上降低位置指纹数据库的构建开销,然而,上述方法均未考虑所需采集 RSS 样本容量的优化问题,从而容易造成位置指纹数据采集的盲目性和不可靠性.

在 WLAN 室内定位系统中,常采用欧式距离、余弦相似度^[5]和索罗森距离^[6]等来刻画不同 RSS 样本间的相似度.然而,上述相似度计算方法均未考虑 RSS 样本方差及样本容量对样本间相似度的影响,从而容易造成相似度计算的不稳定性.为了解决这一问题,本文提出了一种面向 WLAN 室内定位的 T 检验样本容量优化方法.该方法在离线阶段利用 OC (Operating Characteristics) 函数确定每个参考点处允许的最小 RSS (Received Signal Strength) 样本容量,从根本上解决位置指纹数据采集的盲目性和不可靠性;而在在线阶段利用 T 检验方法对终端进行粗定位,并进而提出基于 T 检验的 KNN 算法以完成对终端的精定位.实验仿真表明,T 检验样本容量优化方法,在减少大量的时间开销下,仍然保持定位性能的稳定性.

2 RSS 样本容量优化

2.1 RSS 样本特性

由大数中心极限定理可知,当重复试验的次数充分多时,随机事件 A 的概率分布 $f(A)$ 趋于正态分布.基于此,本系统在目标区域内均匀标记若干参考点,且在每个参考点处采集 m 个 RSS 样本值 $\text{RSS}_i = \{ \text{rss}_{i_1}, \text{rss}_{i_2}, \dots, \text{rss}_{i_m} \}$ ($1 \leq i \leq n$),其中, $\text{rss}_{i_j} = (\text{rss}_{i_j^1}, \text{rss}_{i_j^2}, \dots, \text{rss}_{i_j^r})$ ($1 \leq j \leq m$), $\text{rss}_{i_j^l}$ ($1 \leq l \leq r$) 为第 i 条 RSS 样本序列内第 j 个样本矢量中来自第 l 个 WLAN 接入点的样本值, m 为 RSS 样本序列长度 (即样本矢量个数), r 为 AP (Access Point) 个数.假设第 i 条 RSS 样本序列内来自第 l 个 WLAN 接入点的样本值 $\text{rss}_{i_1^l}, \text{rss}_{i_2^l}, \dots, \text{rss}_{i_m^l}$ 相互独立且对应的数学期望和方差分别为 $E(\text{rss}_{i_j^l}) = \mu_1$ 和 $D(\text{rss}_{i_j^l}) = \sigma^2$,则该样本值求和得到的新变量所对应的标准化变量 Y_m 为:

$$Y_m = \frac{\sum_{j=1}^m \text{rss}_{i_j^l} - E(\sum_{j=1}^m \text{rss}_{i_j^l})}{\sqrt{D(\sum_{j=1}^m \text{rss}_{i_j^l})}} = \frac{\sum_{j=1}^m \text{rss}_{i_j^l} - m\mu_1}{\sqrt{m\sigma^2}} \quad (1)$$

其中, Y_m 的分布函数满足:

$$\lim_{m \rightarrow \infty} F_m(x) = \lim_{m \rightarrow \infty} P \left\{ \frac{\sum_{j=1}^m \text{rss}_{i_j^l} - m\mu_1}{\sqrt{m\sigma^2}} \leq x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \varphi(x) \quad (2)$$

由上式可知,当 m 值充分大时, Y_m 趋于正态分布.于是,第 i 条 RSS 样本序列内来自第 l 个 WLAN 接入点的样本值的算术平均值 $\bar{X} = \frac{\sum_{j=1}^m \text{rss}_{i_j^l}}{m}$ 将近似服从均值为 μ_1 且方差为 σ^2/m 的正态分布.

2.2 最小样本容量

假设在每个参考点处 RSS 的理想、总体和样本均值分别为 μ, μ_1 和 \bar{X} ,则根据 3σ 准则,令 RSS 均值误差接受范围为 $|\mu_1 - \mu| \leq \sigma$.于是,利用假设检验理论中的 OC 函数, RSS 样本容量的优化问题可描述为:

构建如下假设检验问题:

$$H_0: |\mu_1 - \mu| \leq \delta \quad H_1: |\mu_1 - \mu| > \delta \quad (3)$$

为了解决上述问题,令 OC 函数为 $\beta(\mu_1) = P_{\mu_1}$ (接受 H_0),其计算过程如下:

$$\begin{aligned} \beta(\mu_1) &= P_{\mu_1} (\text{接受 } H_0) \\ &= P_{\mu_1} \left\{ -z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{m}} < z_{\alpha/2} \right\} \\ &= \Phi(Z_{\alpha/2} - \lambda) + \Phi(Z_{\alpha/2} + \lambda) - 1 \end{aligned} \quad (4)$$

其中, $\lambda = \sqrt{m}(\mu_1 - \mu)/\sigma$.

若真实情况为 $|\mu_1 - \mu| \leq \delta$,但经过假设检验后得到 $|\mu_1 - \mu| > \delta$,则说明发生“弃真”错误,即 $P\{\text{当 } H_0 \text{ 为真时,拒绝 } H_0\}$.为了尽可能避免犯此类错误,通常要求控制犯此类错误的概率不超过给定的阈值 α ,其中, α 的取值范围一般介于 0 到 0.1 之间^[7].

在 OC 函数中,若要求假设检验满足 $H_1: |\mu_1 - \mu| > \delta$ 的错误概率不超过给定阈值 β_1 (即 $\beta(\mu_1) \leq \beta_1$),从而可以得到 $\beta_1 = \Phi(Z_{\alpha/2} - \sqrt{m}\delta/\sigma) + \Phi(Z_{\alpha/2} + \sqrt{m}\delta/\sigma) - 1$.若令 $\beta_1 = 0.01$ ^[8],则根据式(2),可得:当 $x > 3.5$ 时, $\Phi(x) \approx 1$.通常在实际情况下 m 值较大,于是有 $\Phi(Z_{\alpha/2} + \sqrt{m}\delta/\sigma) \approx 1$ (即 $\beta(\mu_1) \approx \Phi(Z_{\alpha/2} - \lambda)$).此外,由于 $\beta(\mu_1)$ 为关于 λ 的单调递减函数,可得:当 $\lambda = \sqrt{m}(\mu_1 - \mu)/\sigma \geq \sqrt{m}\delta/\sigma$ (即接受 H_1) 时,有 $\beta(\mu_1) \leq \beta_1 \approx \Phi(Z_{\alpha/2} - \sqrt{m}\delta/\sigma)$.综上所述,当 $Z_{\alpha/2} - \sqrt{m}\delta/\sigma \leq -Z_{\beta_1}$ (即 $\sqrt{m} \geq (Z_{\alpha/2} + Z_{\beta_1})\sigma/\delta$) 时,假设检验满足 $H_1: |\mu_1 - \mu| > \delta$ 的错误概率不超过 β_1 .基于此,可计算得到 $m \geq 21.6$,同时考虑 WLAN 室内环境的复杂性,本系统选取的最小 RSS

样本容量为 30.

3 T 检验定位

本系统采用统计学理论中的 T 检验方法判断任意两个 RSS 样本序列是否来自同一正态总体,以分辨不同物理位置处采集 RSS 样本序列间的差异性,进而实现对目标终端的定位.对于来自 AP_i 的两条样本容量分别为 m_1 和 m_2 的 RSS 样本序列,假设其分别来自总体均值为 μ_1 和 μ_2 且总体方差为 σ_1^2 和 σ_2^2 的 RSS 样本集合,可分别计算其样本均值 $\bar{X} = \sum_{j=1}^{m_1} \text{rss}_{i_j}^l / m_1$ 和 $\bar{Y} = \sum_{j=1}^{m_2} \text{rss}_{i_j}^l / m_2$ 及样本方差 $S_1^2 = \sum_{j=1}^{m_1} (\text{rss}_{i_j}^l - \bar{X})^2 / (m_1 - 1)$ 和 $S_2^2 = \sum_{j=1}^{m_2} (\text{rss}_{i_j}^l - \bar{Y})^2 / (m_2 - 1)$.

3.1 构建双边假设检验模型

构建如下双边假设检验模型:

$$H_0: \mu_1 = \mu_2 \quad H_1: \mu_1 \neq \mu_2 \quad (5)$$

3.2 确定检验统计量

选择 t 统计量为检验统计量,可表示为:

$$t = \frac{(\bar{X} - \bar{Y})}{S_w \sqrt{\frac{1}{m_1} + \frac{1}{m_2}}} \quad (6)$$

3.3 确定拒绝域

在双边假设检验中,若真实情况下两条 RSS 样本序列来自同一正态总体,但双边假设检验的结果为其来自不同正态总体,则说明发生“弃真”错误.于是,由检验统计量的定义,可得拒绝域为:

$$|t| = \left| \frac{(\bar{X} - \bar{Y})}{S_w \sqrt{\frac{1}{m_1} + \frac{1}{m_2}}} \right| \geq k > 0 \quad (7)$$

其中, k 为 $|t|$ 过大时拒绝 H_0 的门限.从而有:

$$P\{\text{当 } H_0 \text{ 为真时, 拒绝 } H_0\} \\ = P_{|\mu_1 - \mu_2| = 0} \left\{ \left| \frac{(\bar{X} - \bar{Y})}{S_w \sqrt{\frac{1}{m_1} + \frac{1}{m_2}}} \right| \geq k \right\} \quad (8)$$

由此可得拒绝域为:

$$|t| = \left| \frac{(\bar{X} - \bar{Y})}{S_w \sqrt{\frac{1}{m_1} + \frac{1}{m_2}}} \right| \geq t_{\alpha/2}(m_1 + m_2 - 2) \quad (9)$$

其中, $t_{\alpha/2}(m_1 + m_2 - 2)$ 为双边假设检验中 t 分布的 $\alpha/2$ 分位点.

3.4 匹配定位

由式(13)可知,当 $|t|$ 在拒绝域时,说明假设检验在显著性水平 α 下拒绝 H_0 ,反之,则说明假设检验在显著性水平 α 下接受 H_0 .

根据上述过程,可对来自每个 WLAN 接入点的 RSS

样本序列进行粗定位,得到每个测试点所对应的匹配参考点集合 $\{G_1, G_2, \dots, G_r\}$,其中, G_r 为在测试点处来自 AP_r 的 RSS 样本序列所对应的匹配参考点集合.然后,通过计算测试点与每个匹配参考点处 RSS 样本间的欧式距离,结合 KNN^[2] 算法实现对目标终端的精定位(即基于 T 检验的 KNN 定位算法).

4 实验分析

在 $57\text{m} \times 25\text{m}$ 的目标区域内,布置 6 个 D-Link2310 AP 接入点(即 AP1、2、3、4、5 和 6),均匀标记 88 个参考点和 25 个测试点,如图 1 所示.利用三星 S7568 手机及自主研发的 WLAN 信号采集软件采集 RSS 样本.

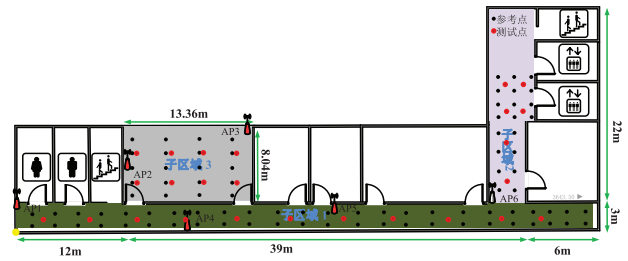


图1 实验环境

4.1 参数 α 讨论

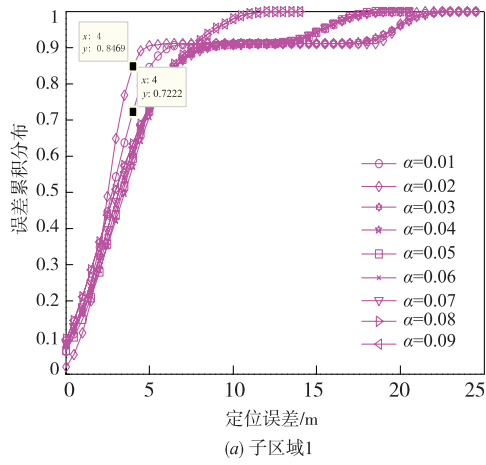
图 2 给出了在不同子区域内定位误差累积分布与 α 值的关系.由图可知,在子区域 1 内,当 $\alpha = 0.02$ 时,定位精度最高,而在子区域 2 和 3 内, α 值的变化对定位精度的影响不大.该结果的主要原因是:子区域 1 内参考点位置较为分散,若 α 值较大,则犯“弃真”错误的概率也较大,反之,则 T 检验定位中匹配的邻近参考点数较少,定位精度下降.综上考虑,系统选择 $\alpha = 0.02$.

4.2 样本容量讨论

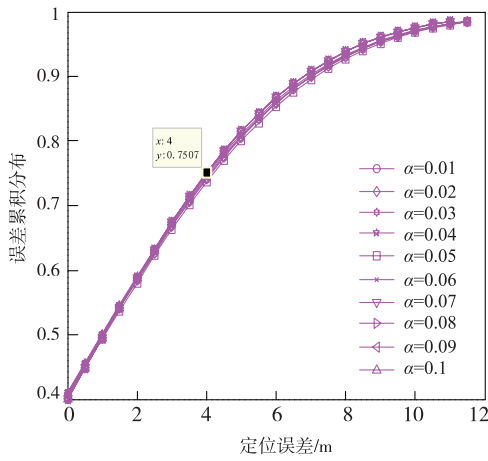
图 3 比较了在不同子区域内信号热度分布图皮尔逊相似度与 RSS 样本容量的关系.由图可知,在子区域 1 内,不同 RSS 样本容量所对应的信号热度分布图皮尔逊相似度差别不大,而在子区域 2 和 3 内,不同 RSS 样本容量所对应信号热度分布图的差异性较大.例如,对于子区域 2,来自 AP1 的 RSS 样本容量为 10 和 100 所对应信号热度分布图之间的皮尔逊相似度仅为 83.05%,而 RSS 样本容量为 30 和 100 所对应信号热度分布图之间的皮尔逊相似度增大至 97.39%.由此可见,RSS 样本容量在 30 和 100 时具有相似的样本特性.综上所述,系统允许的最小样本容量为 30.

4.3 T 检验性能讨论

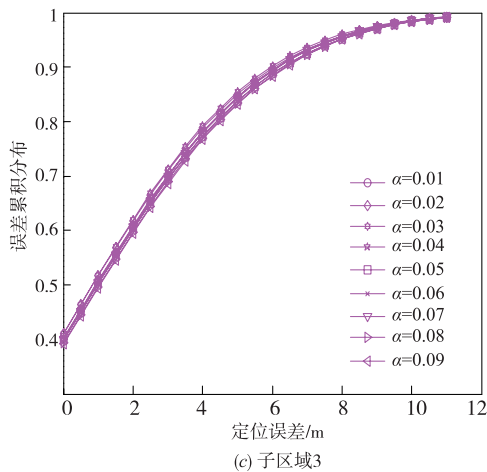
图 4 给出了来自不同 WLAN 接入点的 RSS 样本所对应的 T 检验性能(即 25 个测试点所对应的参考点匹配结果).以来自 AP1 的 RSS 样本为例,子区域 1、2 和 3 内测试点匹配到其所在子区域内参考点的概率(即正确区域匹配概率)分别为 69.23%、57.3% 和 85.7%.



(a) 子区域1



(b) 子区域2

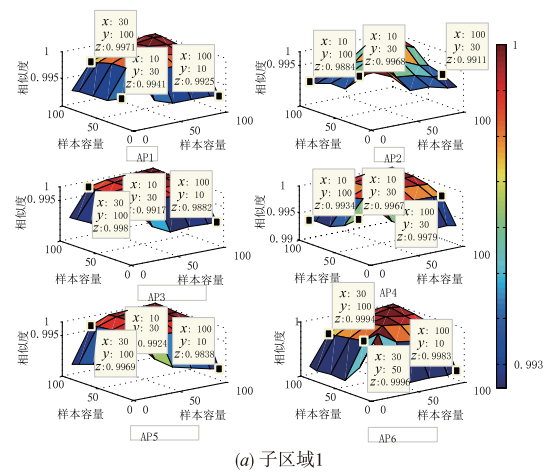


(c) 子区域3

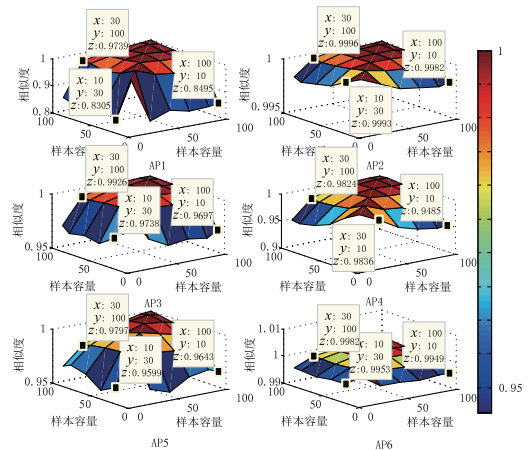
图2 不同子区域内定位误差与 α 值的关系

4.4 定位性能讨论

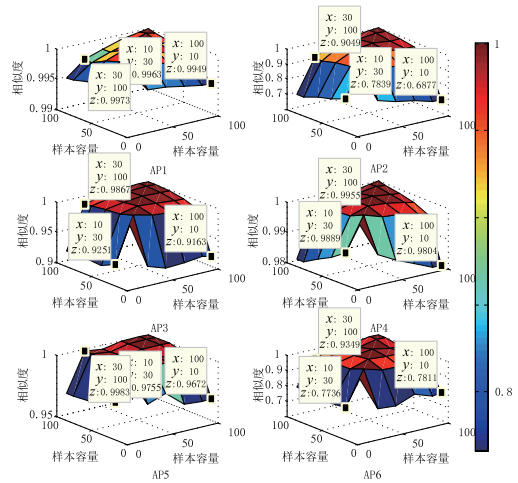
图5给出了本文所提T检验定位算法与传统KNN^[2]、WKNN^[2]和贝叶斯定位算法的性能比较.以样本容量30为例,本文算法的定位误差在3.5米内的置信概率为79.91%,高于KNN、WKNN和贝叶斯算法的71.78%、76.25%和55.33%.由图5(b)和5(c)可知,当RSS样本容量为30和100时,不同定位算法之间的



(a) 子区域1



(b) 子区域2



(c) 子区域3

图3 不同子区域内信号热度分布图相似性与样本容量的关系

性能差异不大,且均优于RSS样本容量为10时的情况.通过图5可见,在指纹数据库构建过程中,通过T检验样本容量的优化使得每个参考点采集数据时间开销减少了70%,显著降低了工作量,但还是可以保证良好的定位性能.

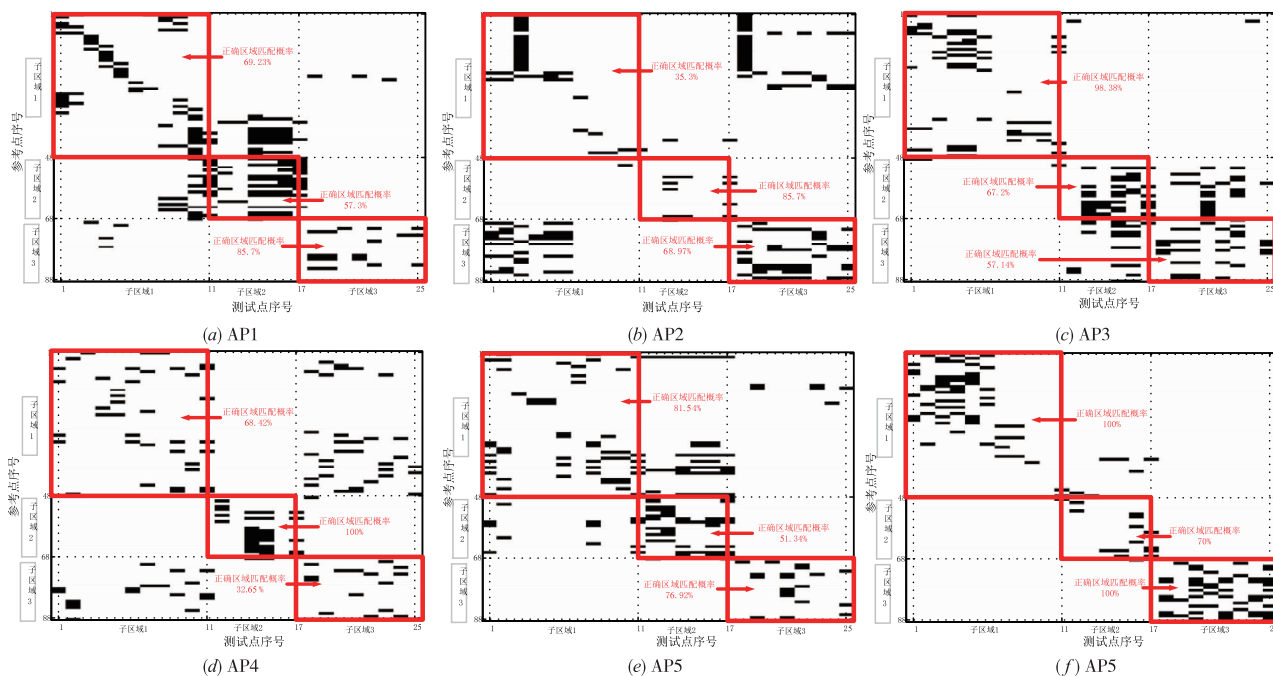


图4 来自不同WLAN接入点的RSS样本所对应的T检验性能

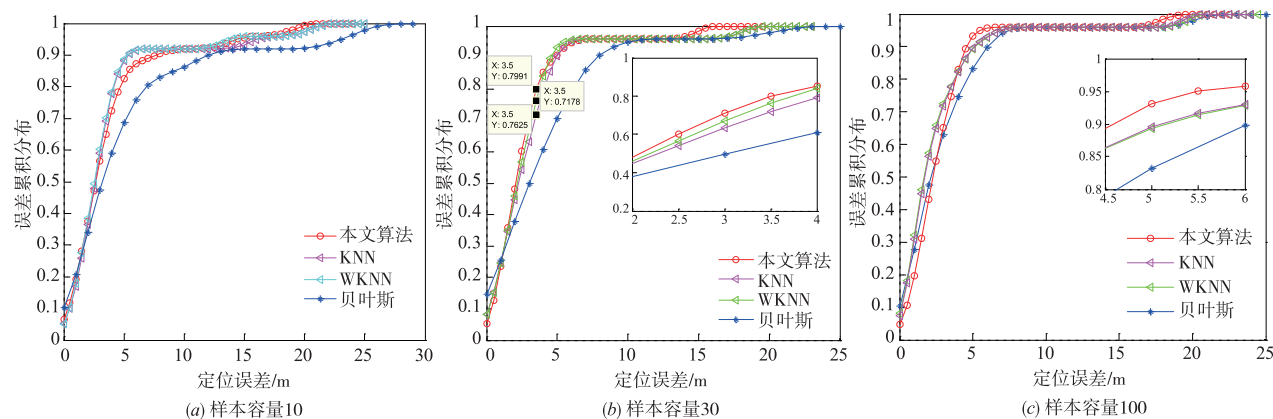


图5 不同定位算法的定位性能比较

5 结论

针对现有 WLAN 室内定位系统构建位置指纹数据库需要大量人力和时间开销的问题,本文提出面向 WLAN 室内定位的 T 检验样本容量优化方法.该方法最主要是通过考虑 RSS 样本序列的方差和均值与样本容量的推导关系,可以推导出用于指纹数据库构建的最小 RSS 样本容量,从而对目标区域内每个参考点处需要采集的 RSS 样本容量进行优化.然后,利用 T 检验方法,完成目标位置与参考点之间的匹配,即粗定位过程.最后,提出基于 T 检验的 KNN 算法以实现对目标终端的精定位.实验结果表明,本文算法显著降低了指纹数据库构建的时间开销,并在提高系统的工作效率下,保证定位性能的稳定性.

参考文献

[1] SHIN H, CHON Y, KIM Y, et al. MRI: Model-based radio interpolation for indoor war-walking[J]. IEEE Transactions on Mobile Computing, 2015, 14(6): 1231 - 1244.

[2] JUNG S, LEE C, HAN D. Wi-Fi fingerprint-based approaches following Log-distance path loss model for indoor positioning[A]. 2011 IEEE MTT-S International Microwave Workshop Series on Intelligent Radio for Future Personal Terminals[C]. United States: IEEE Computer Society, 2011. 1 - 2

[3] 刘剑, 龚志恒, 吴成东, 等. 一种基于改进高斯过程隐变量模型的多角度人脸识别算法[J]. 电子与信息学报, 2013, 35(9): 2033 - 2039.

- LIU J, GONG Z H, WU C D, et al. A multi-angle face recognition algorithm based on modified gaussian process latent variable model[J]. Journal of Electronics and Information Technology, 2013, 35(9): 2033 – 2039. (in Chinese)
- [4] GU Z, CHEN Z Q, ZHANG Y X, et al. Reducing fingerprint collection for indoor localization[J]. Computer Communications, 2015, 83: 56 – 63.
- [5] HAN S, ZHAO C, MENG W X. Cosine similarity based fingerprinting algorithm in WLAN indoor positioning against device diversity[A]. IEEE International Conference on Communications, ICC 2015[C]. London: Institute of Electrical and Electronics Engineers Inc, 2015. 2710 – 2714.
- [6] TORRES-SOSPEDRA J, MONTOLIU R, MARTINEZ-USO A, et al. UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems[A]. 5th International Conference on Indoor Positioning and Indoor Navigation, IPIN 2014[C]. Korea: Institute of Electrical and Electronics Engineers Inc, 2014. 261 – 270.
- [7] KARIM T, REMI C, FLORENT R. Fully automatic detection of anomalies on wheels surface using an adaptive accurate model and hypothesis testing theory[A]. 24th European Signal Processing Conference, EUSIPCO 2016[C]. Hungary: European Signal Processing Conference, 2016. 508 – 512.
- [8] 甘伦知. 假设检验中控制第二类错误的探讨[J]. 统计与决策, 2011, 346(22): 35 – 37.
- GAN L Z. On the control of errors of the second kind in hypothesis test[J]. Statistics and Decision, 2011, 346(22): 35 – 37. (in Chinese)

作者简介



周 牧 男, 1984 年生于重庆. 现为重庆邮电大学教授. 主要研究方向为无线定位技术、机器学习与人工智能、凸优化理论.
E-mail: zhoumu@cqupt.edu.cn



卫亚聪 女, 1993 年生于山西临汾. 现为重庆邮电大学硕士生. 主要研究方向为无线定位技术、数值计算.
E-mail: 2540462563@qq.com



田增山 男, 1968 年生于河南固始. 现为重庆邮电大学教授、博士生导师. 主要研究方向为蜂窝网无线定位系统、个人通信、GPS 精密定位和姿态测量、数据压缩和数据融合.
E-mail: tianzs@cqupt.edu.cn



李玲霞 女, 1976 年生于湖北. 现为重庆邮电大学高级工程师, 硕士生导师. 主要研究方向为未来移动通信理论与技术、宽度无线接入技术.
E-mail: lilx@cqupt.edu.cn