基于图谱分解的无线定位算法

林 权[1,2] 赵 方[3] 罗海勇[4] 康一梅[2]

摘 要 基于有监督学习的射频指纹定位方法是室内高精度无线定位技术的一个研究热点. 针对有监督学习方法存在训练数据集采集代价较高的问题, 本文提出了一种基于半监督学习的室内无线定位算法. 该算法采用基于 Laplacian 矩阵谱分解的方法获取训练数据在特征向量空间上的表示, 然后通过有标记数据在特征向量空间上的标记对齐, 实现对未标记数据的标记. 实验结果表明, 仅需少量的有标记数据 (20 % 左右), 便能以较高的精度 (80 % 左右) 实现对未标记数据的标记, 从而有效降低了训练开销.

关键词 室内无线定位, 半监督学习, Laplacian 矩阵, 谱分解

DOI 10.3724/SP.J.1004.2011.00316

A Wireless Localization Algorithm Based on Spectral Decomposition of the Graph Laplacian

LIN Quan[1,2] ZHAO Fang[3] LUO Hai-Yong[4] KANG Yi-Mei[2]

Abstract Fingerprint localization based on supervised learning is a hot spot for high-accuracy indoor wireless localization. In order to reduce the training cost of supervised learning method, this paper presents a novel localization algorithm based on semi-supervised learning, which applies spectral decomposition of Laplacian matrix to labeling the unlabeled data through aligning the labeled data in the eigenvectors space. The experimental results show that this algorithm can label the unlabeled data with a high accuracy (about 80 %) using only a small amount of labeled data (about 20 %), which effectively reduces the data collection cost.

Key words Indoor wireless localization, semi-supervised learning, Laplacian matrix, spectral decomposition

随着移动计算设备的迅速发展和逐渐普及, 室内环境下的各种基于位置服务需求 (Location-based service, LBS) 日益迫切. 由于卫星定位系统, 如美国的全球定位系统 (Global positioning system, GPS) 和我国的北斗卫星定位系统等, 在室内环境或拥有高大建筑物的密集城区, 卫星信号受到建筑物的阻隔, 难以有效定位. 目前室内定位一般采用红外线、超声波、射频等传感信号, 其中基于红外线、超声波的定位技术精度相对较高, 不过由于需要使用专门的硬件设施, 且信号需要视距传输, 因此定位范围相对受限, 难以大规模部署. 近年来, 随着无线网络的大规模部署, 基于无线局域网 (Wireless LAN, WLAN) 的室内定位技术受到广泛关注. 其主要原因是基于 WLAN 射频信号定位技术不仅传输距离远、无视距要求, 而且无需增加额外硬件, 添加定位软件简单易行, 与其它使用特定设备的定位技术相比, 成本优势十分明显. 目前基于 WLAN 射频信号强度定位技术已成为室内定位技术的研究热点.

基于射频信号强度定位方法一般采用模式分类的思想, 把定位问题转换为射频信号强度与位置的模式识别问题. 这类方法主要分为离线训练和在线定位两个阶段, 在离线训练阶段, 首先将实际应用场景按一定间距划分为规则的网格, 然后在每个网格上采集一定数量的射频信号强度信息, 构建出 "射频地图 (Radio map)" 定位模型; 在在线定位阶段, 根据待定位目标实时观测到的射频信号强度信息, 通过定位模型进行位置计算, 从而实现对待定位目标的定位. 这类方法将信号强度向量看作是对应位置的信号模式 (Pattern) 在信号强度空间中的特征 (Feature), 而实际的位置则可以看作是该模式的真实标记 (Label), 将信号强度向量及其对应的位置坐标组成训练数据集, 训练得到的分类器 (Classifier) 或者回归函数 (Regression function) 就能作为一个

收稿日期 2009-12-31 录用日期 2010-12-02
Manuscript received December 31, 2009; accepted December 2, 2010
国家自然科学基金 (60873244, 61003307, 60973110), 国家科技重大专项 (2010ZX03006-002-03), 北京自然科学基金 (4102059) 资助
Supported by National Natural Science Foundation of China (60873244, 61003307, 60973110), Key National Science and Technology Specific Projects (2010ZX03006-002-03), and Beijing Natural Science Foundation (4102059)
1. 中航工业综合技术研究所 北京 100028 2. 北京航空航天大学 北京 100191 3. 北京邮电大学 北京 100876 4. 中国科学院计算技术研究所 北京 100190
1. Avic Aero-Polytechnology Establishmen, Beijing 100028 2. Beijing University of Aeronautics and Astronautics, Beijing 100191 3. Beijing University of Posts and Telecommunications, Beijing 100876 4. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190

定位模型. 总之, 这类定位方法通过学习射频信号强度与位置间的内在映射规律, 基于推理实现目标定位.

如果采集了足够的有标记数据, 便可以采用有监督学习 (Supervised learning) 的算法来建立定位模型. 而常见的有监督学习定位算法分为确定性定位模型和概率性定位模型两大类.

确定性定位模型主要采用确定性的推断方法来估计目标位置, 如 k-最近邻 (k-nearest neighbor, k-NN) 算法、决策树 (Decision tree) 算法等. Bahl 等[1−2] 采用了 k-最近邻算法来估算目标的位置, 该方法将实时采集到的信号强度向量与训练数据集中的所有信号强度向量进行匹配, 选出最相似信号的强度向量所处的位置作为当前目标的估计位置. Chen 等[3] 则采用了决策树的方法进行目标定位, 该方法首先对目标环境进行聚类, 然后在每一类中根据接入点 (Access point, AP) 信息熵 (Entropy) 的大小建立决策树, 在在线定位阶段, 首先确定目标所处可能的类, 然后运用该类中的决策树进行目标的细粒度定位.

概率性定位模型主要是构建应用环境中信号强度在不同位置下的条件概率分布. Ladd 等[4] 采用了概率推断方法进行定位, 首先使用从环境中采集的信号强度向量计算每个位置上的信号强度概率分布, 然后加入对目标运动轨迹的空间约束条件, 限制目标的剧烈位置跳变, 提高了定位精度. Roos 等[5] 把最近邻方法与朴素贝叶斯方法 (Naive Bayes) 和直方图分布概率方法进行了对比, 实验结果表明基于概率的定位方法优于基于最近邻的定位方法. Hightower 等[6] 将基于粒子滤波 (Particle filters) 的概率方法应用于定位, 不仅可获得与确定性定位模型近似的定位精度, 并且能够有效融合多种传感信息. Krumm 等[7] 将隐马尔科夫模型 (Hidden Markov model, HMM) 应用于目标运动轨迹建模, 采用状态转移矩阵表示位置间的转移关系, 然后利用 Viterbi 算法[8] 推断出最大可能的运动轨迹序列.

总之, 基于射频信号强度定位方法是一种基于模式分类的方法, 只需要知道每个位置所对应的信号强度特征, 而不需要知道 AP 位置的先验知识. 但是已有的这类定位算法一般需要在离线训练阶段采集大量带位置标记的数据作为训练数据, 需要较大的人力开销. 为此, 本文提出了一种基于图半监督学习的定位算法, 该方法在离线训练阶段仅需采集少量有位置标记数据和一定数量的无位置标记数据, 通过利用所有训练数据之间的关系进行半监督学习, 有效降低了离线训练阶段数据采集代价.

本文第 1 节将简要介绍图的 Laplacian 矩阵及其谱性质, 通过对原始数据的 Laplacian 矩阵进行谱分解, 获得原始数据中潜在的簇类信息; 在这种簇类信息的辅助下, 第 2 节介绍了利用有标记数据的标记信息标记无标记数据的半监督学习方法; 第 3 节将这种半监督学习算法应用于无线定位; 本文第 4 节评测了该算法的性能.

1 图的 Laplacian 矩阵及谱分解

构建由所有训练数据组成的图 $G(V, E)$, 其中节点集 V 包括所有训练数据样本 (l 个有标记数据和 u 个无标记数据), 节点之间的边集 E 为 $(l+u)\times(l+u)$ 的非负值矩阵 W, 其中 w_{ij} 表示第 i 个和第 j 个训练数据之间的关系, 可以采用 k-NN 或者高斯函数的方式进行赋值. 对角矩阵 $D=\mathrm{diag}\{d_{11}, d_{22}, \cdots, d_{ii}, \cdots, d_{nn}\}$, 其中, $d_{ii}=\sum_j w_{ij}$ 表示与节点 i 相关的所有边权值 w_{ij} 之和 (采用 k-NN 构图方式时, D 的对角元素表示了节点的度), 则图的 Laplacian 矩阵可定义如下[9]:

定义 1. 图的 Laplacian 矩阵为 $L=D-W$; 或 $\hat{L}=D^{-1/2}LD^{-1/2}=I-D^{-1/2}WD^{-1/2}$ (归一化).

将 Laplacian 矩阵进行特征值分解后, L 的特征值和特征向量分别为 λ_i、$\boldsymbol{\phi}_i$, 并且特征值以非降次序排列, 则有:

$$L=\sum_{i=1}^{l+u}\lambda_i\boldsymbol{\phi}_i{\boldsymbol{\phi}_i}^{\mathrm{T}} \tag{1}$$

则 Laplacian 矩阵的谱定义如下[9]:

定义 2. Laplacian 矩阵的谱为矩阵 L 的特征值集合 $\{\lambda_i\}$.

Laplacian 矩阵的谱具有许多有意义的性质[10], 下面介绍与本文相关的两个性质.

性质 1. Laplacian 矩阵谱值的大小体现了特征向量的平滑程度.

证明. 对于任一个 $n\times 1$ 向量 $\boldsymbol{f}$, $\boldsymbol{f}$ 的平滑性可以根据向量 $\boldsymbol{f}$ 中两个元素的值以及对应的权值来衡量, 即如果向量 $\boldsymbol{f}$ 是平滑的, 那么列向量 $\boldsymbol{f}$ 中连接权值 w_{ij} 较大的两个元素的值应该是相似的. 向量 $\boldsymbol{f}$ 的平滑性可用下式表示:

$$S(\boldsymbol{f})=\frac{1}{2}\sum_{i,j}w_{ij}(f_i-f_j)^2=\boldsymbol{f}^{\mathrm{T}}L\boldsymbol{f} \tag{2}$$

其中, L 为向量 $\boldsymbol{f}$ 中元素的 Laplacian 矩阵.

因此, 向量 $\boldsymbol{f}$ 的平滑性可以通过 $\boldsymbol{f}^{\mathrm{T}}L\boldsymbol{f}$ 来衡量, 进一步地, 可以利用 ${\boldsymbol{\phi}_i}^{\mathrm{T}}L\boldsymbol{\phi}_i$ 来衡量特征向量 $\boldsymbol{\phi}_i$ 的平滑性, 根据式 (1) 以及特征向量之间的正交性, 有:

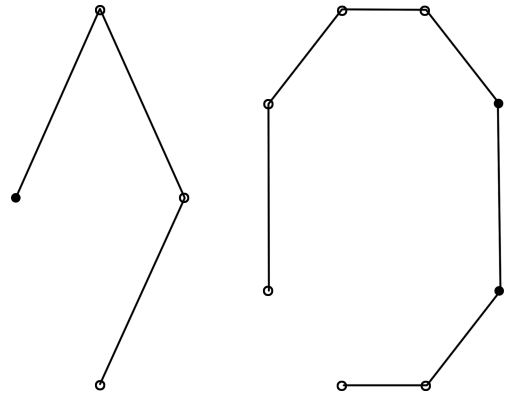
$${\boldsymbol{\phi}_i}^{\mathrm{T}}L\boldsymbol{\phi}_i={\boldsymbol{\phi}_i}^{\mathrm{T}}\left(\sum_{j=1}^{l+u}\lambda_j\boldsymbol{\phi}_j{\boldsymbol{\phi}_j}^{\mathrm{T}}\right)\boldsymbol{\phi}_i=\lambda_i \tag{3}$$

所以, Laplacian 矩阵谱值的大小体现了特征向量的平滑程度. □

性质 2. 如果图 G 具有 n 个连通分量, 那么 Laplacian 矩阵具有 n 个 0 特征值.

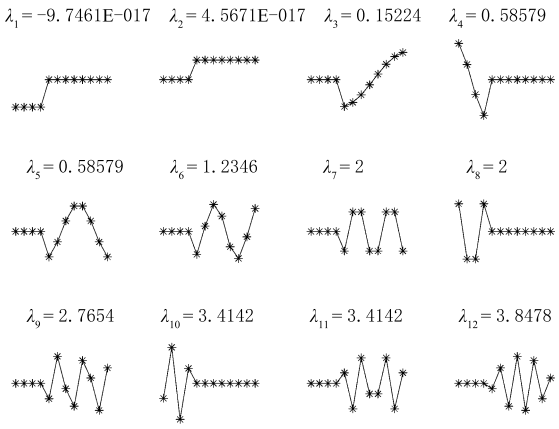
图 1 直观地显示了 Laplacian 矩阵的谱的这两个性质. 其中图 1 (a) 为一个具有两个连通分量的链图 (边权值为 1), 图 1 (b) 为其谱及对应的特征向量.

从图 1 可以看出, 一个具有两个连通分量、边权值为 1 的图, 其谱具有 2 个值为 0 的特征值, 其对应的特征向量是平滑的, 而且随着特征值的增大, 对应的特征向量也变得更加振荡. 图 1 (a) 中实心节点表示有标记数据, 空心节点表示未标记数据.



(a) 具有两个连通分量的链图

(a) Chain graph with doubly connected components



(b) 谱及对应的特征向量

(b) Spectrum and its corresponding eigenvectors

图 1 图、谱及其特征向量

Fig. 1 Graph, spectrum and its corresponding eigenvectors

图 2 显示了图 1 (a) 中所有节点在 2 个特征向量 (特征向量总数的 20 % 左右) 所张成的正交空间上的邻接关系, 其中两个节点之间的虚线表示了这两个节点在该正交空间上是邻接的. 从图 2 可以发现, 图的 Laplacian 矩阵的特征向量隐含了原始图中节点的簇信息, 利用该簇信息, 通过有标记数据的标记, 可实现对无标记数据的标记.

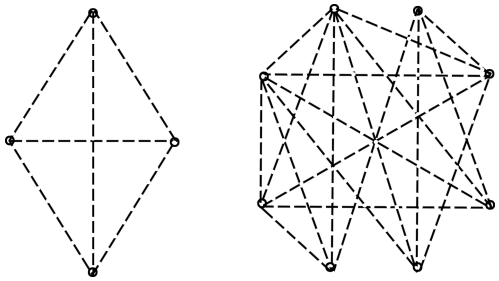


图 2 节点在特征向量空间上的邻接关系

Fig. 2 Adjacent relationship between nodes in the eigenvector space

2 基于图的半监督学习

如第 1 节所述, Laplacian 矩阵的特征向量体现了原始图节点的簇信息, 利用这些簇信息, 根据 "聚类假设[11]" — 在相同聚类 (Cluster) 中的样本示例有较大的可能拥有相同的标记, 可以实现有标记数据的标记在整个数据图中的传播.

由于 Laplacian 矩阵的所有特征值满足 $0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_{l+u}$[9], 因此, 矩阵 L 是半正定的 (Positive semi-definite), 其特征向量形成 Hilbert 空间 $L^2(M)$ 的一组正交基[12]. 那么, 任何函数 $f(x) \in L^2(M)$ 均可表示为正交基的线性组合:

$$f(x) = \sum_{i=1}^{l+u} \alpha_i \phi_i(x) \tag{4}$$

由于数据集中有标记数据有限, 因此选择 p ($p \le l$) 个特征向量形成 p 维正交空间, 定义标记函数为 f, 那么所有数据的标记可以近似表示为

$$f_j \approx \sum_{i=1}^{p} \alpha_i \phi_{ji}, \;\; j = 1, \cdots, l+u, \;\; p \le l \tag{5}$$

其中, ϕ_{ji} 表示第 i 个特征向量的第 j 个元素的值.

在有标记数据上定义误差函数为

$$E(\boldsymbol{\alpha}) = \sum_{j=1}^{l} \left(f_j - \sum_{i=1}^{p} \alpha_i \phi_{ji} \right)^2 \tag{6}$$

通过最小化误差函数, 可获得最优的拟合参数 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_p)^{\mathrm{T}}$, 即

$$\bar{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \sum_{j=1}^{l} \left(f_j - \sum_{i=1}^{p} \alpha_i \phi_{ji} \right)^2 \tag{7}$$

其最小二乘解为

$$\bar{\boldsymbol{\alpha}} = (E_{\mathrm{lab}}^{\mathrm{T}} E_{\mathrm{lab}})^{-1} E_{\mathrm{lab}}^{\mathrm{T}} \boldsymbol{f}_{\mathrm{lab}} \tag{8}$$

其中, $\boldsymbol{f}_{\text{lab}} = (f_1, f_2, \cdots, f_l)^{\text{T}}$ 为有标记训练数据的标记, 而

$$E_{\text{lab}} = \begin{bmatrix} \phi_{11} & \phi_{12} & \cdots & \phi_{1p} \\ \phi_{21} & \phi_{22} & \cdots & \phi_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{l1} & \phi_{l2} & \cdots & \phi_{lp} \end{bmatrix} \tag{9}$$

为有标记数据投影到 p 个特征向量所生成的正交空间上的 p 维表示. 使用拟合参数 $\boldsymbol{\alpha}$, 便可以实现对未标记数据的标记, 如式 (10) 所示:

$$f_j = \sum_{i=1}^{p} \alpha_i \phi_{ji}, \ \ j = l+1, \cdots, l+u \tag{10}$$

3　基于半监督学习的定位算法

本节将基于图的半监督学习理论应用于室内无线定位, 以减少训练开销.

为描述方便, 假设在离线训练阶段所采集的数据为 $\boldsymbol{X} = \{X^{\text{labeled}}, X^{\text{unlabeled}}\}$, 其中 $X^{\text{labeled}} = \{(RSS_1, Location_1), \cdots, (RSS_l, Location_l)\}$ 为所采集的 l 个带位置标记训练数据; $X^{\text{unlabeled}} = \{RSS_{l+1}, \cdots, RSS_{l+u}\}$ 为采集的 u 个无位置标记数据. 而 $RSS_i = \{RSS_i^1, \cdots, RSS_i^{ap}\}$ 和 $Location_i$ 分别表示在位置 i 上所收集到无线射频信号强度向量 (ap 为环境中 AP 的数量) 和位置信息. 定位算法的主要框架如下:

步骤 1. 根据带标记数据对未标记数据进行标记, 获得一个全带标记的训练数据集:

1) 基于 k-NN 与高斯函数结合构建邻接图 $G(V, E)$. 根据所有训练数据集, 包括有标记数据和未标记数据, 采用 k 最近邻与高斯函数结合的方式构建邻接图 $G(V, E)$. 即节点 i 与节点 j 之间存在连接 (权值 $w_{ij} > 0$), 当且仅当节点 j 是节点 i 的 k 近邻. 然后根据节点 i 与节点 j 之间在信号强度空间上的距离, 采用高斯核函数的方式进行赋权:

$$w_{ij} = \frac{\exp\{-\dfrac{dist_{ij}^2}{2\sigma^2}\}}{\sum\limits_{l=1}^{k} \exp\{-\dfrac{dist_{il}^2}{2\sigma^2}\}}$$

2) 计算邻接图的 Laplacian 矩阵. 在邻接图 $G(V, E)$ 的基础上, 根据定义 1 计算 Laplacian 矩阵 $L = D - W$, 其中 D 为节点的权值度对角矩阵, W 为权值矩阵;

3) 特征分解. 在对 Laplacian 矩阵进行特征值分解, 根据定义 2 获得矩阵 L 的谱以及对应的特征向量 $\{\lambda_i, \boldsymbol{\phi}_i\}$, 将所有训练数据映射到由 p 个特征向量所张成的正交空间上;

4) 构建分类器. 基于最小化误差函数 (式 (6)) 准则, 构建射频指纹数据在特征向量空间上的分类器, 并使用式 (8) 计算得到分类器的参数 $\bar{\boldsymbol{\alpha}}$;

5) 标记未标记数据. 根据分类器参数 $\bar{\boldsymbol{\alpha}}$, 采用式 (10) 对未标记数据进行标记, 获得一个完整的全标记训练数据集.

步骤 2. 根据完整的全标记训练数据集, 采用有监督学习的定位方法, 如最大后验 (Maximum a posteriori, MAP)[13], 对目标进行定位.

4　实验

本文的实验数据采用香港科技大学所提供的实际数据集[14], 该数据集提供了室内环境不同时间段收集的射频信号强度向量. 本文所采用的是数据集中早上 11 点时间段的数据, 进行实验的网格数为 38 格, 如图 3 所示, 其中每一网格的大小为 1.5 m × 1.5 m, 在每一网格上有 60 个无线射频信号强度向量. 本文采用留一法 (Leave-one-out) 对数据集进行划分, 即将数据集均等地划分为 q 份, 每次仅选择其中 1 份作为有标记数据集, 其余 $q-1$ 份作为未标记数据, 然后使用本文介绍的算法, 对未标记数据进行标记, 并循环 q 次对结果进行统计.

定位标记精度定义为

$$Accu = \frac{\text{标记准确的数量}}{\text{无标记数据的数量}} \times 100\,\% \tag{11}$$

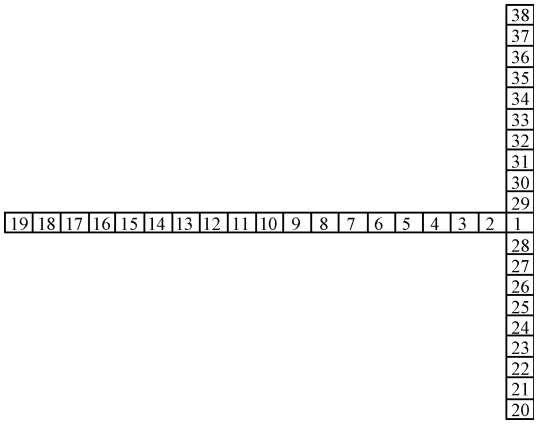


图 3　实验环境布局

Fig. 3　The layout of experimental environment

4.1　不同数目特征向量

本节评测式 (5) 中特征向量个数 p 对定位标记精度的影响. 实验中将每一网格上的 60 个数据集划分为 6 份, 然后迭代地选择其中 1 份作为有标记数据, 剩下的作为未标记数据, 实验结果如图 4 所示.

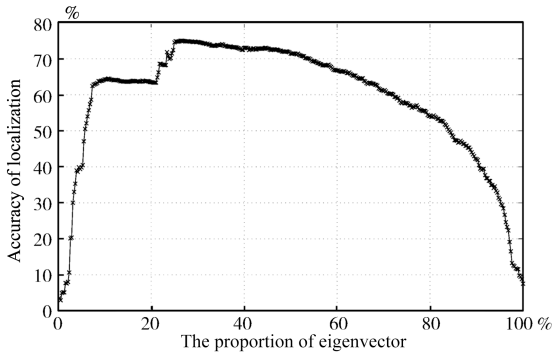


图 4 17 % 比例有标记数据下的标记精度

Fig. 4 The accuracy of localization under 17 % labeled data

从图中可以看出, 随着所用特征向量个数的增加, 即 p 的增大, 定位标记精度呈现出先迅速提高然后缓慢降低的趋势. 这是因为, 当 p 较小时, 采用的特征向量个数相对较少, 产生了拟合不足的现象, 导致式 (10) 的泛化性能较弱, 因此定位标记精度较低. 而当 p 增大到一定程度时, 如有标记数据数量达到 50 % 时, 定位标记精度开始下降, 其原因是过多的特征向量导致式 (7) 过度拟合, 定位精度反而较低.

图 5 显示了当有标记数据占 25 % 时, 定位标记精度与所采用特征向量个数 p 之间的关系. 除了与图 4 有类似的结论外, 对比图 4 和图 5 可以看出, 当采用的特征向量个数占所有特征向量个数的 20 % ∼ 25 % 左右时, 定位标记精度达到最高值.

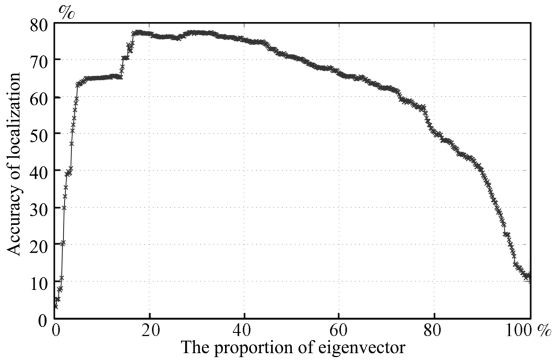


图 5 25 % 比例有标记数据下的标记精度

Fig. 5 The accuracy of localization under 25 % labeled data

4.2 不同比例有标记数据

本节进一步评测使用不同比例的有标记数据对剩余训练数据进行标记的性能. 随着有标记数据比例的增多, 有标记数据所提供的标记信息越多, 定位标记精度呈现出递增的趋势, 如图 6 所示. 当有标记数据比例增长到一定程度时 (如 15 %), 定位标记的精度的递增趋势变得缓慢, 这是因为, 当式 (7) 中有标记数据的数量 l 达到一定的程度时已经能够获得较好的拟合系数, 进一步增多有标记数据的数量并不能得到同等的回报.

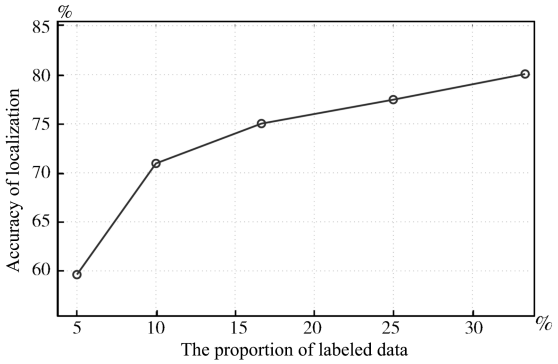


图 6 不同比例有标记数据下的标记精度

Fig. 6 The accuracy of localization under different proportion of labeled data

4.3 不同构图方式

本节将比较使用不同的构图方式对精度的影响. 本文主要采用 k-NN 与高斯函数结合的方式, 每个训练数据样本点选择 k 个最近 (根据式 (12) 计算距离) 的训练数据样本点作为邻居, 采用高斯函数进行权值计算, 最后对权值进行归一化, 如式 (13) 所示.

$$dist_{ij} = \|RSS_i - RSS_j\| = \sqrt{\sum_{l=1}^{ap}(RSS_i^l - RSS_j^l)^2} \tag{12}$$

$$w_{ij} = \frac{\exp\left\{-\dfrac{dist_{ij}^2}{2\sigma^2}\right\}}{\sum\limits_{l=1}^{k}\exp\left\{-\dfrac{dist_{il}^2}{2\sigma^2}\right\}} \tag{13}$$

选择不同的最近邻个数 k, 对于无标记数据的标记精度的影响情况如图 7 所示.

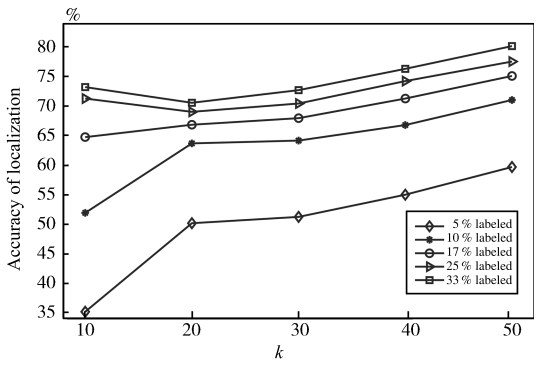


图 7 不同构图方式对标记精度的影响

Fig. 7 The accuracy of localization under different values of k

从图 7 中可以看出, 随着最近邻个数 k 的增大, 定位标记精度大体上呈现出缓慢上升的趋势, 这是因为, 在数据分布的流型上, 更多的局部信息能够更好地表征出流型结构的局部信息. 在实际应用中, 可根据训练数据集的大小以及实验结果来确定最近邻的个数 k.

5 结论

针对现有基于射频指纹定位方法存在带标记训练数据集采集代价较高等问题, 本文基于半监督学习理论, 把基于图 Laplacian 矩阵谱分解方法应用于室内无线定位. 该算法通过构建少量带标记训练数据与大量无标记训练数据之间的标记传播关系, 对无标记训练数据进行标记, 有效降低了训练数据集的采集代价. 实验结果表明, 仅需少量的有标记数据 (20 % 左右), 便能以较高的精度 (接近 80 %) 实现对未标记数据的定位标记.

我们的下一步研究工作是将不同构图方式中邻居数量 k 的选择看作一个超参数选择的机器学习问题, 探索能够在不同训练数据集下自适应调整参数的方法, 提高室内无线定位算法的鲁棒性.

References

1 Bahl P, Padmanabhan V N, Balachandran A. Enhancements to the Radar User Location and Tracking System, Technical Report MSR-TR-2000-12, Microsoft Research, USA, 2000

2 Bahl P, Padmanabhan V N. RADAR: an in-building RF-based user location and tracking system. In: Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies. Tel Aviv, Israel: IEEE, 2000. 775−784

3 Chen Y Q, Yang Q, Yin J, Chai X Y. Power-efficient access-point selection for indoor location estimation. *IEEE Transactions on Knowledge and Data Engineering*, 2006, **18**(7): 877−888

4 Ladd A M, Bekris K E, Rudys A, Marceau G, Kavraki L E, Wallach D S. Robotics-based location sensing using wireless ethernet. In: Proceedings of the 8th Annual International Conference on Mobile Computing and Networking. New York, USA: ACM, 2002. 227−238

5 Roos T, Myllymaki P, Tirri H, Misikangas P, Sievanen J. A probabilistic approach to WLAN user location estimation. *International Journal of Wireless Information Networks*, 2002, **9**(3): 155−164

6 Hightower J, Borriello G. Particle filters for location estimation in ubiquitous computing: a case study. In: Proceedings of the 6th International Conference on Ubiquitous Computing. Nottingham, UK: Springer. 2004. 88−106

7 Krumm J, Horvitz E. LOCADIO: inferring motion and location from WiFi signal strengths. In: Proceedings of the 1st Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services. Washington D. C., USA: IEEE, 2004. 4−13

8 Rabiner L R. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989, **77**(2): 257−286

9 Chung F R K. *Spectral Graph Theory*. Providence: American Mathematical Society, 1997

10 Zhu X J. Semi-Supervised Learning with Graphs [Ph. D. dissertation], Carnegie Mellon University, USA, 2005

11 Zhou Zhi-Hua, Wang Jue. *Machine Learning and Its Application*. Beijing: Tsinghua University Press, 2007. 259−275 (周志华, 王珏. 机器学习极其应用. 北京: 清华大学出版社, 2007. 259−275)

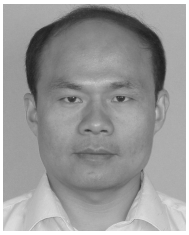
12 Rosenberg S. *The Laplacian on a Riemannian Manifold: an Introduction to Analysis on Manifolds*. England: Cambridge University Press, 1997

13 Youssef M, Agrawala A. On the Optimality of WLAN Location Determination Systems, Technical Reports UMIACS-TR-2003-29 and CS-TR-4459, University of Maryland, USA, 2003

14 Data set of WiFi received signal strength values in an indoor environment [Online], available: http://www.cs.ust.hk/~qyang/ICDMDMC07/data/tr_data.zip, December 10, 2010

林 权 中航工业综合技术研究所助理工程师. 主要研究方向为软件可靠性和无线定位技术. 本文通信作者.
E-mail: linquanisaac@yahoo.cn
(**LIN Quan** Assistant engineer at Avic Aero-Polytechnology Establishmen. His research interest covers software reliability and wireless localization. Corresponding author of this paper.)

赵 方 北京邮电大学副教授. 主要研究方向为无线传感器网络和分布式计算.
E-mail: zhaofanglhy@yahoo.com.cn
(**ZHAO Fang** Associate professor at Beijing University of Posts and Telecommunications. Her research interest covers wireless sensor networks and distributed computing.)



罗海勇 中国科学院计算技术研究所高级工程师. 主要研究方向为普适计算, 无线定位技术和嵌入式系统.
E-mail: yhluo@ict.ac.cn
(**LUO Hai-Yong** Associate professor at the Institute of Computing Technology, Chinese Academy of Sciences. His research interest covers pervasive computing, wireless localization, and embedded systems.)

康一梅 北京航空航天大学教授. 主要研究方向为无线传感器网络, 图像目标识别与跟踪和嵌入式软件工程.
E-mail: kangyimei@buaa.edu.cn
(**KANG Yi-Mei** Professor at Beijing University of Aeronautics and Astronautics. Her research interest covers wireless sensor network, image recognition, and embedded software engineering.)