

用局部保持典型相关分析定位无线传感器网络节点^{*}

顾晶晶, 陈松灿⁺, 庄毅

(南京航空航天大学 计算机科学与工程系, 江苏 南京 210016)

Localization in Wireless Sensor Network Using Locality Preserving Canonical Correlation Analysis

GU Jing-Jing, CHEN Song-Can⁺, ZHUANG Yi

(Department of Computer Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

+ Corresponding author: E-mail: s.chen@nuaa.edu.cn

Gu JJ, Chen SC, Zhuang Y. Localization in wireless sensor network using locality preserving canonical correlation analysis. Journal of Software, 2010,21(11):2883-2891. <http://www.jos.org.cn/1000-9825/3686.htm>

Abstract: In this paper, the deployment of wireless network is analyzed in accordance with the manifold distribution. Next, the previous Locality Preserving Canonical Correlation Analysis (LPCCA) algorithm is used to build a mapping from the signal-vector space to the physical location space and develop a location algorithm, Location Estimation-Locality Preserving Canonical Correlation Analysis (LE-LPCCA, for short), which sufficiently takes the local structure of the network topology into account. Finally, experimental results on benchmark show that this method achieves a higher accuracy and stability than other location algorithms.

Key words: wireless sensor network; localization; canonical correlation analysis; manifold

摘要: 在分析了无线传感器网络的布局基本符合流形(manifold)特点的基础上,借助所提出的局部保持的典型相关分析(locality preserving canonical correlation analysis,简称LPCCA)进行建模,以建立从信号强度空间到现实物理空间的映射;进而提出了一种能够体现网络拓扑结构局部信息的无线传感器网络定位算法——LE-LPCCA(location estimation-locality preserving canonical correlation analysis).与现有同类算法在benchmark上的实验结果相比,LE-LPCCA具有更高的定位精度和稳定性。

关键词: 无线传感器网络;定位;典型相关分析;流形

中图法分类号: TP393 **文献标识码:** A

无线传感器网络具有非常广泛的应用前景,目前国内外已有将其应用于环境监测和保护、医疗护理、航空航天军事等领域的研究工作.其中,无线传感器网络的定位技术是一个研究难点,也是应用研究中的关键技术.传感器节点的自身定位是无线传感器网络应用的基础,常见的定位技术是RSSI(received signal strength indicator)技术^[1].其主要原理是,先通过信号在传播中的衰减,根据信道模型来估计节点之间的距离;再通过距离

^{*} Supported by the National Natural Science Foundation of China under Grant No.60773061 (国家自然科学基金); the National High-Tech Research and Development Plan of China under Grant No.2006AA706103 (国家高技术研究发展计划(863)); the Aeronautical Science Foundation of China under Grant No.05F2037 (航空基金)

Received 2008-09-01; Accepted 2009-07-08

信息进行节点位置的计算.RSSI 技术符合传感器网络定位低功率和低成本的要求,已有很多研究工作都基于此技术,如文献[2-5]等.还有一些比较常见的定位技术,如 TOA(time of arrival)^[6]、TDOA(time difference on arrival)^[7]、AOA(angle of arrival)^[8]、DV-Hop^[9]、凸规划^[10]等.这些技术既有优势,也存在着诸如定位精度不高、网络部署方式受限、硬件要求过高、无法实现低功耗和低成本等问题^[11].本文侧重于 RSSI 的定位技术.

节点的定位过程可以看作是模式的识别过程,其中,节点的信号强度(向量)作为输入模式,对应的节点地址作为(离散或连续的)标号.定位过程基于统计模型,提取出能够表征传感器节点坐标位置的特征量,建立模型,以此确定定位算法.再根据该算法对未知传感器节点(测试样本)进行位置估计,即计算其空间(位置)坐标.其中已有的典型算法包括文献[2-5].文献[2]在定位过程中结合了核技巧,并依照分类方法计算出未知节点可能存在的矩形区域,然后将该矩形的质心作为未知节点的位置作为定位结果;文献[3]结合了聚类分析和决策树分类技术确定出合理的已知节点集合,并弃用无法提供合适定位信息的多余先验信息,从而降低了定位的复杂度;文献[4]将定位过程描述成凸规划问题,进而借助半定规划(semidefinite programming,简称 SDP)优化该问题,以获得最优的定位结果;文献[5]则通过建立贝叶斯层次模型来解决室内定位问题.这些算法在一定场合内取得了较为有效的定位结果,但由于没有充分利用网络的局部拓扑结构以及其与信号空间的相关性,因而定位算法的定位精度仍有待提高.

本文的出发点正是尝试利用无线传感器网络的拓扑结构与信号空间的自身特性来进一步提高定位算法的精度.目前有关这方面的研究工作相对较少,在国内尤其如此.

在节点功率相同、传送模式相似的前提下,由于传感器网络中任意两节点若相邻,则接收到同一信标节点所发射的信号强度也相似,所以物理空间和信号空间存在一一映射关系^[12].定位就是构建出这一映射.文献[12]中提出的 LE-KCCA 算法通过核典型相关分析(kernel canonical correlation analysis,简称 KCCA)^[13]实现所要的映射.然而,KCCA 的计算采用了一个统一的非线性映射而没有顾及网络具有的局部分布特点,即流形结构.另外,拓扑结构复杂且多变的无线传感器网络同样使 KCCA 难以充分体现网络的结构信息.因此,出于对上述两点不足的考虑,本文以我们之前提出的能够同时有效处理两组分别具有流形结构数据的局部保持型典型相关分析(locality preserving canonical correlation analysis,简称 LPCCA)^[14]作为定位算法的建模工具,尝试利用信号空间和物理空间的局部信息,提出一种基于流形学习的无线传感器网络定位新算法 LE-LPCCA(location estimation-locality preserving canonical correlation analysis).LE-LPCCA 算法着眼于每个节点邻域内的典型相关,通过将全局非线性问题分解成多个局部线性问题,以获得整体的相关性.该算法的主要特点就是,通过引入网络分布的全局和局部信息,达到显著提高节点定位精度和稳定性的双重效果.

本文第 1 节分析 LE-LPCCA 算法的动机.第 2 节介绍典型相关性分析(canonical correlation analysis,简称 CCA)^[15]和基于局部的 LPCCA 方法.第 3 节详细介绍 LE-LPCCA 算法的设计思想.实验结果及分析在第 4 节中给出.最后,第 5 节进行全文总结.

1 基于流形学习的 LE-LPCCA 算法

1.1 基于拓扑学的流形学习

流形(manifold)是拓扑学中的概念,一般可以认为是局部具有欧氏性质的空间^[16].而欧氏空间是流形最简单的实例.

流形学习方法提供了一种新颖的根据数据内在分布来分析数据的研究途径,旨在发现数据分布的内在规律性.流形学习加强了数据在观测空间的全局与局部性质,加强了几何上的直观解释.

近年来,流形学习已成为机器学习领域中的一个研究焦点,已有一些研究成果,如线性局部嵌入(locally linear embedding,简称 LLE)算法^[17]、等度规映射(isomap 算法)^[18]、保局投影(locality preserving projections,简称 LPP)算法^[19]、Laplacianface 方法^[20]和局部追踪嵌入(locality pursuit embedding,简称 LPE)算法^[21].这些基于流形局部信息方法的共同点在于,通过保持样本的局部特征,可揭示出隐藏于原始空间整体的低维流形结构.但上述诸方法仅限于单数据集,而无线传感器网络的定位涉及两个分别来自于物理空间和信号空间的数据集,且

它们高度相关.因此,需要寻求更合适的建模方法来实现两个空间之间的映射.

1.2 LE-LPCCA算法的动机

本节通过分析无线传感器网络的拓扑结构基本吻合流形特点来阐明用 LPCCA 进行定位的动机.以无线传感器网络的信号空间到物理空间的坐标映射视作一双射为基础,文献[12]通过 KCCA 算法建立了这两个空间之间的非线性映射,揭示两组不同空间数据中隐含存在的非线性相关现象.因此,通过一个节点接收到来自信标节点的信号强度向量,能够反推出其在物理空间的相对拓扑结构,进而计算出该节点的位置,实现无线传感器网络节点的定位.然而,KCCA 计算着眼于统一的非线性映射而没有考虑节点的局部分布特性.对于拓扑结构复杂多变的无线传感器网络而言,KCCA 算法难以充分体现网络的结构信息.另外,由于不同传感器网络之间的拓扑结构差异较大,接收到的信号模式也可能很不相同,试图通过调整 KCCA 中的核函数或核参数使之实现拓扑结构的依赖同样不太可行,故而该算法的推广性和灵活性受到一定的限制.

流形在数学中可用于描述几何形体,欧氏空间是流形最简单的实例.由于传感器节点分布于地球表面或由若干个局部欧氏空间组成的物理空间中,所以无线传感器网络的布局符合流形的特点.无线传感器节点的布置可用于对流形信息的采样,布置越密,采样效果越好.对于其对应的信号空间,由于每个传感器节点的信号发射范围理论上是一个球面,位于球面内的节点才能接收到该节点发射的信号,这同样与局部欧式空间的分布相吻合.所以,无线传感器网络的信号空间也符合流形特点.

本文考虑通过利用与信号空间的自身特性来提高定位算法的精度,目前有关利用无线传感器网络拓扑结构与流形的关系进行定位的算法研究工作相对较少.文献[22]中虽然提出了用流形的方法进行无线传感器网络定位,但其出发点并未考虑信号空间的局部信息,只是单纯假设采集的数据符合流形,此算法的应用面不宽^[22].文献[23]中运用了半监督学习的流形方法,但其动机是将所采集的数据应用于半监督学习,并未分析传感器网络结构本身与流形的关系.本文弥补了这一缺陷,并提出了利用 LPCCA^[14]实现流形结构和定位.事实上,原有 LPCCA 是作为一种降维工具应用于机器学习领域中的目标姿态估计.而本文力图进一步拓展其应用领域,通过结合具体的定位问题将其作为回归而非降维来适应新的应用,从而发展出了一种能够体现网络局部拓扑结构信息的无线传感器网络的新颖定位算法 LE-LPCCA.

图 1 显示了传感器网络节点的近邻关系图,图中体现了无线传感器网络拓扑结构中的某些局部特征,其中数据来自真实的实验环境,由香港科技大学计算机科学系提供.局部近邻采用 k -近邻定义法, k 取值为 4.图 1 中,圆圈范围内的 4 个节点能够接收到 8 个信标节点发射的信号,它们的物理空间分布聚集在一起,图示可见对应的接收信号强度相近.比如该数据矩阵中的第 1 行,表示 4 个节点接收到一号信标节点发射的信号,信号强度均在 300 左右.由于接收到的信号强度实际为负数,单位:db,这里,为了方便研究和实验,我们已把数据全部处理为正数,数值越大,则表示信号越强.

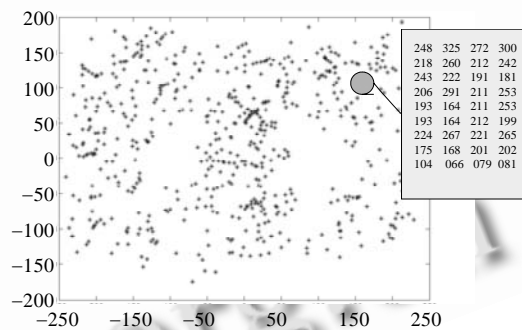


Fig.1 Neighbor relationship of sensor nodes

图 1 传感器节点的近邻关系图

2 相关知识回顾

为完整起见,本节对相关模型作简单的回顾.

信号空间向量和物理空间向量可以分别表示两组数据集,在机器学习领域中,典型相关性分析^[15]是用来构建两组数据间映射的经典方法.

2.1 CCA

设有一批成对的观察样本集 X 和 Y , 记 $X=[x_1, \dots, x_n] \in R^{p \times n}$ 和 $Y=[y_1, \dots, y_n] \in R^{q \times n}$, 且样本已经中心化. CCA 算法的目标是, 分别为 X 和 Y 寻找两组基向量 $w_x \in R^p$ 和 $w_y \in R^q$, 使随机变量 $w_x^T x$ 和 $w_y^T y$ 之间的相关性达到最大. 换言之, 就是使原始数据 X 和 Y 作一定的变换后, 达到相关性最大. 具体可表述为求式(1)相关系数的最大值问题^[15]:

$$\rho = \frac{E[w_x^T x y^T w_y]}{\sqrt{E[w_x^T x x^T w_x] E[w_y^T y y^T w_y]}} = \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x \cdot w_y^T C_{yy} w_y}} \quad (1)$$

这里, $E[\cdot]$ 表示数学期望, $C_{xx}=XX^T \in R^{p \times p}$ 和 $C_{yy}=YY^T \in R^{q \times q}$ 表示集合内协方差(within-set covariance)矩阵, $C_{xy}=XY^T \in R^{p \times q}$ 表示集合间协方差(between-set covariance)矩阵, 且有 $C_{yx} = C_{xy}^T \in R^{q \times p}$.

注意到公式(1)定义的相关函数 ρ 关于 w_x 和 w_y 尺度无关, 则 CCA 可表述为式(2)优化问题的解^[15]:

$$\left. \begin{aligned} & \max_{w_x, w_y} w_x^T C_{xy} w_y \\ & \text{s.t. } w_x^T C_{xx} w_x = 1, w_y^T C_{yy} w_y = 1 \end{aligned} \right\} \quad (2)$$

文献[15]给出了一种把求解 CCA 方程(2)转换成求取特征方程(3)的方法:

$$C_{xy} C_{yy}^{-1} C_{yx} w_x = \lambda^2 C_{xx} w_x \quad (3)$$

解基向量 w_x 和特征值 λ , 则 $w_y = \frac{1}{\lambda} C_{yy}^{-1} C_{yx} w_x$. 通常, 我们可以获得一组特征向量(即基向量)对 $(w_{x_i}, w_{y_i}) (i=1, \dots, d)$, 并可把它们组合成两个映射矩阵: $W_x = [w_{x_1}, \dots, w_{x_d}]$ 和 $W_y = [w_{y_1}, \dots, w_{y_d}]$. 可利用任意一对 w_{x_i} 和 w_{y_i} 进行形如 $w_{x_i}^T x$ 和 $w_{y_i}^T y$ 的特征抽取, 这里, 抽取的特征 $w_{x_i}^T x$ 和 $w_{y_i}^T y$ 可称其为典型变量(canonical variate).

CCA 只能抽取两组数据之间的线性相关关系, 但不能解决无线传感器网络中存在的非线性相关现象. 核方法为这个问题提供了一个有效的解决途径, 即通过把数据映射到高维空间来增加传统的线性学习器的计算能力. 尽管其核化的非线性版 KCCA^[13]能揭示两组数据间隐含存在的非线性相关现象, 然而这种映射仍然是全局的, 没有考虑数据的局部分布情况.

2.2 LPCCA 算法介绍

LPCCA 算法^[14]将样本的局部结构信息引入到 CCA 中, 将原来的全局非线性问题变成若干局部线性问题, 计算每个小邻域内的典型相关问题; 然后对这些子问题求解, 通过优化计算得到一组样本投影基向量. 因此, 通过局部线性的方法达到解决非线性问题的目的. 该方法在保持各样本集的局部结构信息的同时, 还可得到两组样本之间的最大化相关信息.

LPCCA 算法定义了相似度矩阵 $C_X = \{C_{ij}^X\}_{i,j=1}^n$ 和 $C_Y = \{C_{ij}^Y\}_{i,j=1}^n$, 其包含了数据分布的局部信息, 即 C_{ij}^X (或 C_{ij}^Y) 越大, 意味着 x_i 与 x_j (或 y_i 与 y_j) 的距离越近. 所以, LPCCA 的优化问题表述如式(4)所示^[9]:

$$\left. \begin{aligned} & \max_{w_x, w_y} w_x^T X C_{XY} Y^T w_y \\ & \text{s.t. } w_x^T X C_{XX} X^T w_x = 1 \\ & \quad w_y^T Y C_{YY} Y^T w_y = 1 \end{aligned} \right\} \quad (4)$$

这里, $X=[x_1, \dots, x_n], Y=[y_1, \dots, y_n], C_{XY} = D_{XY} - C_X \circ C_Y, C_{XX} = D_{XX} - C_X \circ C_X, C_{YY} = D_{YY} - C_Y \circ C_Y$. 符号 \circ 表示算子; $D_{XX}(D_{YY}, D_{XY})$ 是大小为 $n \times n$ 的对角矩阵, 其第 i 个对角元素等于矩阵 $C_X \circ C_X (C_Y \circ C_Y, C_X \circ C_Y)$ 第 i 行(因其对称性, 或第 i 列)元素之和.

利用 Lagrange 乘子法, 求解优化方程(4), 容易得到如下广义特征值方程:

$$\begin{pmatrix} X C_{XY} Y^T \\ Y C_{YX} X^T \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} = \lambda \begin{pmatrix} X C_{XX} X^T \\ Y C_{YY} Y^T \end{pmatrix} \begin{pmatrix} w_x \\ w_y \end{pmatrix} \quad (5)$$

得到基向量组 (w_x, w_y) 后, 即可对原有数据进行形如 $w_x^T x$ 和 $w_y^T y$ 的线性变换. 注意, LPCCA 不需要对数据进行中心化, 这点与 CCA 和 KCCA 不同. 在整个计算过程中, LPCCA 保持局部化因子 C_{ij}^X, C_{ij}^Y 不变, 而尽量使得在原

空间中近邻的数据点在变换后仍然保持近邻^[14].

3 LE-LPCCA 算法

在基于 RSSI 的无线传感器网络定位机制中,节点的定位通常分为训练阶段(offline training phase)和定位阶段(online localization phase)^[12].在训练阶段,通过对已知节点的数据训练来学习定位算法;在定位阶段,运用学习得出的算法对未知节点进行位置计算.

记 $S=[s_1, s_2, \dots, s_n] \in R^{m \times n}$ 为 n 个无线传感器已知节点所接收到的信号强度,其中,每个信号向量 $s_i(i=1, \dots, n)$ 的维数为 m ,表示有 m 个信标节点在广播消息.若第 $j(j=1, \dots, m)$ 个信标节点位于 i 点的信号接收范围外,则 $s_{ji}=0$. $L=[l_1, l_2, \dots, l_n] \in R^{2 \times n}$ 为相对应节点的物理坐标.算法设计的目标就是,通过未知节点 g 所接收到的信号向量 $s_g=(s_{g1}, s_{g2}, \dots, s_{gm})'$ 来计算出 g 点的位置坐标 y_g .

定义. 在信号空间中, $ne(i)$ 表示与 i 节点所接收到的信号强度相似的节点标号集合,即 s_i 的局部近邻样本下标集;在物理空间中, $ne(i)$ 表示与 i 节点位置近邻的节点下标集.其中,局部近邻由 k -近邻定义法划分:若 $s_j(l_j)$ 是 $s_i(l_i)$ 的 k -近邻样本,则称 $s_j(l_j)$ 为 $s_i(l_i)$ 的局部近邻^[14].

根据上述定义,无线传感器网络的拓扑结构在信号空间和物理空间的相似度矩阵为 $C_S = \{C_{ij}^S\}_{i,j=1}^n$ 和 $C_L = \{C_{ij}^L\}_{i,j=1}^n$,其中,矩阵元素^[14]:

$$C_{ij}^S = \begin{cases} \exp(-\|s_i - s_j\|^2 / t_s), & \text{若 } j \in ne(i) \text{ 或 } i \in ne(j) \\ 0, & \text{其他情形} \end{cases}, C_{ij}^L = \begin{cases} \exp(-\|l_i - l_j\|^2 / t_l), & \text{若 } j \in ne(i) \text{ 或 } i \in ne(j) \\ 0, & \text{其他情形} \end{cases} \quad (6)$$

参数 t_s 取为信号空间的平均距离 $\sum_{i=1}^n \sum_{j=1}^n \|s_i - s_j\|^2 / n(n-1)$,参数 t_l 也作类似处理.因此,相似度矩阵包含了网络拓扑结构的局部信息,即 C_{ij}^S (或 C_{ij}^L) 越大,意味着 s_i 与 s_j (或 l_i 与 l_j) 距离越近;若 s_i 与 s_j (或 l_i 与 l_j) 互不在邻域范围内,则相似度为 0.可以看出, C_{ij}^S (或 C_{ij}^L) 依赖于传感器节点的布局.因此, C_{ij}^S (或 C_{ij}^L) 可随网络拓扑结构的变化而变化,从而显示出高度的灵活性.

根据以上描述,可得如下优化问题:

$$\left. \begin{aligned} \max_{w_S, w_L} & w_S^T S C_{SL} L^T w_L \\ \text{s.t.} & w_S^T S C_{SS} S^T w_S = 1 \\ & w_L^T L C_{LL} L^T w_L = 1 \end{aligned} \right\} \quad (7)$$

这里, $C_{SL} = D_{SL} - C_S \circ C_L$, $C_{SS} = D_{SS} - C_S \circ C_S$, $C_{LL} = D_{LL} - C_L \circ C_L$.利用 Lagrange 乘子法和消元法解上述优化方程组,可以得到映射矩阵 W_S 和 W_L .

接下来,我们给出 LE-LPCCA 算法的具体步骤,其中,Step 1~Step 3 在训练阶段完成,Step 4~Step 6 在定位阶段完成.

LE-LPCCA 算法描述

Input: 已知节点所接收到的信号强度矩阵 $S=(s_1, s_2, \dots, s_n)$; 已知节点的物理坐标矩阵 $L=(l_1, l_2, \dots, l_n)$; 未知节点 g 所接收到的信号向量 $s_g=(s_{g1}, s_{g2}, \dots, s_{gm})'$.

Output: g 点的位置坐标 y_g .

Step 1. 通过式(6),计算相似度矩阵 C_S 和 C_L .

Step 2. 通过计算式(7),对搜集的两空间的数据进行局部典型相关性分析,学习物理空间和信号空间的关系,求出映射矩阵 W_S 和 W_L 以及广义特征值 λ .

Step 3. 对已知数据 $\{S, L\}$ 作线性变换: $\{P^S, P^L\} = \{W_S^T \times S, W_L^T \times L\}$. 即对原始空间数据作相应映射,从而使 P^S 与 P^L 相关性达到最大,并且数据由原始维度变换为二维.

Step 4. 对于未知节点 g 所接收到的信号向量 $s_g=(s_{g1}, s_{g2}, \dots, s_{gm})'$,用 W_S 来对其进行线性变换,得 $P^g = W_S^T s_g$.

Step 5. 计算 g 点与所有已知节点在线性变换后,带有权重的欧式距离^[12]:

$$d_i = \sum_{j=1}^2 \lambda_j (P_{ji}^s - P_j^g)^2, i=1, \dots, n \quad (8)$$

其中, d_i 表示第 i 个已知节点与 g 点的距离; λ_{ij} 为第 i 个特征值第 j 个分量, 表示第 j 大的相关性; P_{ji}^s 表示第 i 个已知节点作线性变换后第 j 个分量; P_j^g 表示 g 点作线性变换后的第 j 个分量.

Step 6. 计算 g 点 K 近邻节点: K 个 $d_i (i=1, \dots, n)$ 的最小值及其相对应的 K 个节点 (即 i 的值). 由 LPCCA 的数据变换后仍然保持近邻的特性可知, 信号空间中的距离越近, 相应的物理空间的距离也越小. 所以, 找出这 K 个节点的物理坐标 $\{y_1, y_2, \dots, y_K\}$, 令其质心为 g 节点的物理坐标:

$$y_g = \frac{y_1 + y_2 + \dots + y_K}{K} \quad (9)$$

至此, y_g 则为 g 点的位置计算结果.

4 实验

4.1 实验描述

实验数据来自香港科技大学计算机科学系. 实验布置在该系大楼内, 由 Cross-bow MICA2 构建而成. 传感器节点部署在有 4 个走廊的室内, 分布有 8 个信标节点 (包括非同一平面). 所以, 每个传感器节点接收到的信号强度向量为 8 维, 即同时接收到 8 个信标节点发送来的信号. 实验设备为拥有 2.4GHz 频率带宽的 IEEE 802.11b 无线网络. 所有数据的采集都是通过 IBM 的笔记本电脑链接一个外部的无线 USB 网络适配器采集得到^[12].

在无线传感器网络实际运用中, 通常是把网络分割成若干网格, 只要节点定位至相应网格内, 则定位结果就为正确, 网格大小的划分根据实际网络用途确定. 本实验把整个网络平面分割成若干个网格, 每个网格大小为 $1.5\text{m} \times 1.5\text{m}$ ^[12]. 每个网格我们搜集 13~14 个样本点, 总共有 2 999 个样本. 其中: 一部分用来作训练样本, 学习定位算法; 剩下一部分作测试样本, 计算定位准确率.

本次实验定位误差由 δ 表示, 实验精度分别由下列 3 项指标表示: $\delta \leq 1.5\text{m}$ (即节点分布在同一个网格中)、 $\delta \leq 3.0\text{m}$ (即节点可分布在相邻网格中) 和定位计算的平均误差 $\bar{\delta}$. 为了进行对比实验, 我们还运行了 LE-KCCA, CCA 和基于流形正则化方法的定位算法 LeMan^[23]. LE-CCA 是为了实验比较而专门设计的算法, 是通过 CCA 构建出信号空间和物理空间之间的全局线性映射. LE-KCCA 与 LeMan 选取的都是高斯核, 且核参数选为 0.05^[23]. 实验环境是在 MATLAB 7.0 中进行. 数据样本在实验前进行了归一化处理. 为了保证结果的准确性, 每组实验均运行了 20 次以上.

4.2 实验结果分析

如图 2 所示为当训练样本 (即已知节点) 取 10%~70% 时, 4 种算法的定位准确度.

由于传感器网络经常应用于大规模分布部署, 已知节点的部署需要按照当地物理情况而定. 所以, 本次实验训练样本的选取完全是随机选取, 但每次运用于 4 种算法中的同比例数据是相同的.

由图 2 可看出, LE-LPCCA 算法的定位结果显示有较大优势. 对 $\delta \leq 1.5\text{m}$ (如图 2(a) 所示), 定位准确率保持在 85% 左右; 对 $\delta \leq 3.0\text{m}$ (如图 2(b) 所示), 定位准确率基本保持在 100% 左右; 定位计算的平均误差 $\bar{\delta}$ 也最小, 保持在 1m 以下 (如图 2(c) 所示). 由图 2 还可以看出, 随着训练样本数的增大, 定位结果更加稳定.

LeMan 算法在 $\delta \leq 1.5\text{m}$ 时 (如图 2(a) 所示) 的最好定位准确率有 78.92% 左右, 平均误差 (如图 2(c) 所示) 都在 1m 以上. 该算法虽然也基于流形方法, 但其仅考虑了信号空间的流形结构, 且不如 LPCCA 更能增强两数据集的相关性. 此外, 该算法可能更适用于采集样本较密集的空间, 文献[23]中的节点密度约是本文的 2.34 倍.

对于 LE-KCCA 算法, 定位结果的准确率不是特别稳定. 只有当训练样本在 50% 左右, 定位结果较好. 当 $\delta \leq 1.5\text{m}$ 时的准确率为 86.02%; 当 $\delta \leq 3.0\text{m}$ 时的准确率接近 100%; 平均误差 0.873m. 但由于训练样本的分布为随机选取, 并不能保证完全均匀分布 (这与实际应用保持一致), 且核参数和核函数选得不一定合适. 所以, 进行整体非线性的 LE-KCCA 算法运算, 结果相当不稳定. 我们观察了随机得到的原始训练数据发现, 当训练样本分布均匀

且数量较多时,LE-KCCA 算法的定位结果较好,反之则不同.

LE-CCA 算法的精确度较低,误差为 1.5m 之内定位准确率只能保持在 70%左右;平均误差也在 1.2m 左右.这是由于其本身仅是两组空间的全局线性映射,对于复杂的非线性空间,LE-CCA 算法不再适用;但另一方面,其简单性也使其较为稳定.

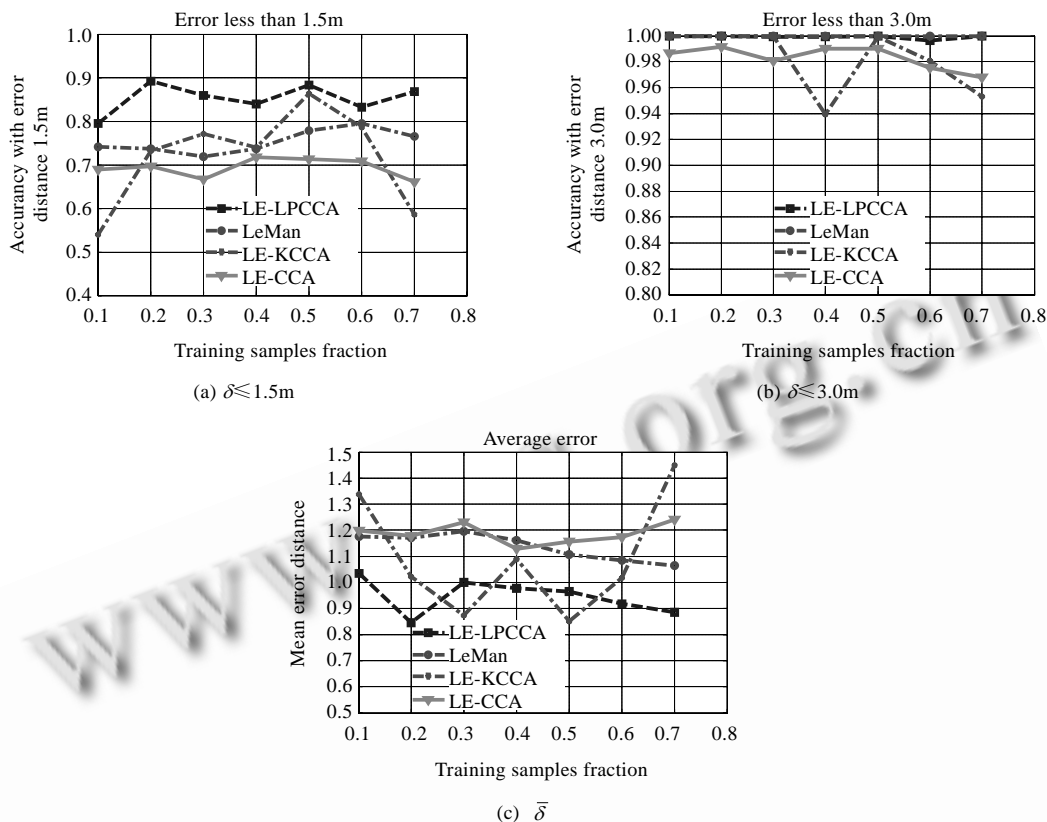


Fig.2 Localization accuracy of four algorithms vs. the number of training samples

图 2 当训练样本比例逐渐增大时,4 种算法运行结果比较

4.3 参数对定位的影响

在 LE-LPCCA 算法中,主要的参数选择是计算 $ne(i)$ 时所用到的 k -近邻法中对 k 的赋值.假设传感器节点分布均匀,图 3 表示了训练样本和测试样本各为 50% 的情况下,当 k 取不同值时定位精度的变化.

由于 k 取 0 和 1 时,网络的拓扑结构无法很好表示,所以我们从 $k=2$ 时开始取值.当 k 取 2 并逐渐增大时,由图 3(b)可见,定位的平均误差逐渐减小.近邻节点越多,定位越准确.但当 k 取到 8 及以上时,变化已相当小,平均误差在 0.75m 左右,结果趋向于稳定. $\delta \leq 3.0m$ 的准确率一直为 100%,而 $\delta \leq 1.5m$ 的准确率也保持在 85% 以上.

通过不同的参数选择实验,以无线传感器网络的局部信息作为考察重点的 LE-LPCCA 算法,对不同的 k 具有良好的稳定性,说明其对 k 不是十分敏感.

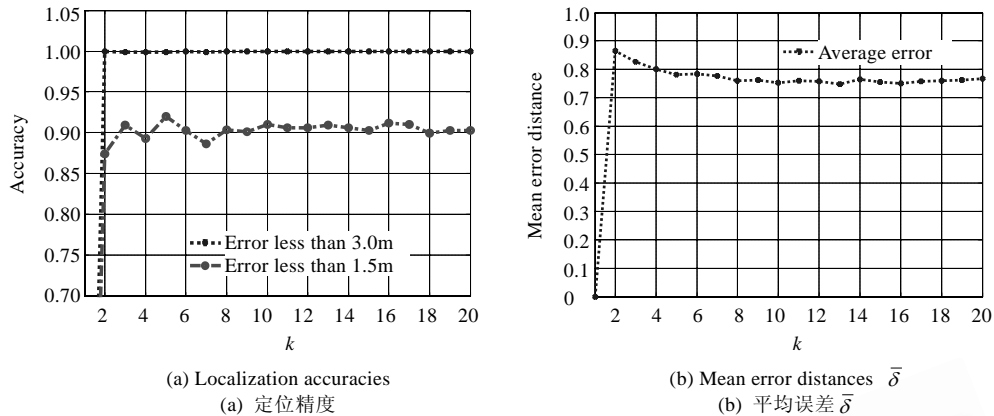


Fig.3 Localization accuracies (on test set) and mean error distances vs. k
图3 (测试集上的)定位精度与平均误差随 k 的变化

5 结论

本文提出的 LE-LPCCA 定位算法通过利用无线传感器网络结构的局部信息,运用流形方法建立了从信号强度空间到物理空间的映射,可有效处理无线传感器网络中拓扑结构复杂的问题。

与同类算法相比,LE-LPCCA 算法有定位精确度高且性能较稳定的优点,但由于其先对局部求解再求总和,所以消耗时间相对略高.进一步的工作是在此基础上进行改进以提高效率。

致谢 实验数据来自香港科技大学计算机科学系,在此特别表示感谢。

References:

- [1] Bahl P, Padmanabhan VN. RADAR: An in-building RF-based user location and tracking system. In: Proc. of the IEEE INFOCOM 2000, Vol.2. Tel Aviv: IEEE Computer and Communications Societies, 2000. 775-784.
- [2] Nauyen XL, Jordan MI, Sinopoli B. A kernel-based learning approach to ad hoc sensor network localization. ACM Trans. on Sensor Networks, 2005,1(1):134-152. [doi: 10.1145/1077391.1077397]
- [3] Chen YQ, Yang Q, Yin J, Chai XY. Power-Efficient access-point selection for indoor location estimation. IEEE Trans. on Knowledge And Data Engineering, 2006,18(7):877-888. [doi: 10.1109/TKDE.2006.112]
- [4] Biswas P, Lian TC, Wang TC, Ye YY. Semidefinite programming based algorithms for sensor network localization. ACM Trans. on Sensor Networks, 2006,2(2):188-220. [doi: 10.1145/1149283.1149286]
- [5] Madigan D, Ju WH, Krishnan P, Krishnakumar AS, Zorych I. Location estimation in wireless networks: A Bayesian approach. Statistica Sinica, 2006,16(2):495-522.
- [6] Harter A, Hopper A, Steggle P, Wark A, Webster P. The anatomy of a context-aware application. Wireless Networks, 2002,8(2/3): 187-197.
- [7] Girod L, Estrin D. Robust range estimation using acoustic and multimodal sensing. In: Proc. of the IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS 2001), Vol.3. Maui: IEEE Robotics and Automation Society, 2001. 1312-1320.
- [8] Niculescu D, Nath B. Ad hoc positioning system (APS) using AoA. In: Proc. of the IEEE INFOCOM, IEEE Computer and Communications Societies. San Francisco: IEEE Press, 2003. 1734-1743.
- [9] Wang FB, Shi L, Ren FY. Self-Localization systems and algorithms for wireless sensor networks. Journal of Software, 2005,16(5): 857-868 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/857.htm> [doi: 10.1360/jos160857]
- [10] Doherty L, Pister KSJ, Ghaoui LE. Convex position estimation in wireless sensor networks. In: Proc. of the IEEE INFOCOM, IEEE Computer and Communications Societies. Anchorage: IEEE Computer and Communications Societies, 2001. 1655-1663.

- [11] Sun LM, Li JZ, Chen Y, Zhu HS. Wireless Sensor Network. Beijing: Tsinghua University Press, 2005. 141–146 (in Chinese).
- [12] Pan JF, Kwok JT, Yang Q, Chen YQ. Multidimensional vector regression for accurate and low-cost location estimation in pervasive computing. *IEEE Trans. on Knowledge and Data Engineering*, 2006,18(9):1181–1193. [doi: 10.1109/TKDE.2006.145]
- [13] Melzer T, Reiter M, Bischof H. Appearance models based on kernel canonical correlation analysis. *Pattern Recognition*, 2003, 36(9):1961–1971. [doi: 10.1016/S0031-3203(03)00058-X]
- [14] Sun TK, Chen SC. Locality preserving CCA with applications to data visualization and pose estimation. *Image and Vision Computing*, 2007,25(5):531–543. [doi I: 10.1016/j.imavis.2006.04.014]
- [15] Hardoon DR, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 2004,16(12):2639–2664. [doi: 10.1162/0899766042321814]
- [16] Bishop E. Differentiable manifolds in complex Euclidean space. *Duke Mathematical Journal*, 1965,32(1):1–21. [doi: 10.1215/S0012-7094-65-03201-1]
- [17] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000,290(5500):2323–2326. [doi: 10.1126/science.290.5500.2323]
- [18] Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000, 290(5500):2319–2323. [doi: 10.1126/science.290.5500.2319]
- [19] He XF, Niyogi P. Locality preserving projections (LPP). In: *Proc. of the NIPS, Advances in Neural Information Processing Systems*. Vancouver: MIT Press, 2004. 96–103.
- [20] He XF, Yan SC, Hu YX, Niyogi P, Zhang HJ. Face recognition using Laplacianfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2005,27(3):328–340. [doi: 10.1109/TPAMI.2005.55]
- [21] Min W, Lu K, He XF. Locality pursuit embedding. *Pattern Recognition*, 2004,37(4):781–788. [doi: 10.1016/j.patcog.2003.09.005]
- [22] Patwari N, Hero III AO. Manifold learning algorithms for localization in wireless sensor networks. In: *Proc. of the IEEE Int'l Conf. on Acoustic Speech and Signal Processing*. Montreal: IEEE Press, 2004. 857–860.
- [23] Pan JF, Yang Q, Chang H, Yeung DY. A manifold regularization approach to calibration reduction for sensor-network based tracking. In: *Proc. of the American Association for Artificial Intelligence (AAAI 2006)*. Boston: AAAI Press, 2006. 988–993.

附中文参考文献:

- [9] 王福豹,史龙,任丰原.无线传感器网络中的自身定位系统和算法.软件学报,2005,16(5):857–868. <http://www.jos.org.cn/1000-9825/16/857.htm> [doi: 10.1360/jos160857]
- [11] 孙利民,李建中,陈渝,朱红松.无线传感器网络.北京:清华大学出版社,2005.141–146.



顾晶晶(1983—),女,江苏南京人,博士生,主要研究领域为模式识别,无线传感器网络定位.



庄毅(1956—),女,教授,博士生导师,主要研究领域为信息安全,分布计算,计算机网络.



陈松灿(1962—),男,教授,博士生导师,主要研究领域为神经计算,模式识别,图像处理与分析,核方法及其应用,数据挖掘.