

基于无线传感器网络拓扑结构的物联网定位模型

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 南京 210016)

摘要 无线传感器网络是物联网(Internet of Things)的重要组成部分,利用其实现物联网中目标的定位技术已成为研究热点之一.由于受环境、障碍物、网络攻击和硬件错误等诸多因素的影响,传感器节点所采集的数据易产生较大误差,形成错误数据,从而对定位造成严重影响.尽管已发展出了众多定位算法和模型,但针对错误数据实现定位的研究还较罕见,尤其在國內,几乎是空白.文中针对上述问题,旨在利用网络(几何)拓扑结构信息,提出一种用局部信息刻画全局分布密度信息的新颖物联网定位模型:鲁棒的局部保持的典型相关分析定位模型 LE-RLPCCA.与现有同类典型方法在真实环境中的实验结果相比,LE-RLPCCA 具有更高的定位鲁棒性和稳定性.

关键词 物联网;无线传感器网络;定位;拓扑结构;错误数据;鲁棒性

中图法分类号 TP393 **DOI号**: 10.3724/SP.J.1016.2010.01548

Wireless Sensor Network-Based Topology Structures for the Internet of Things Localization

GU Jing-Jing CHEN Song-Can ZHUANG Yi

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

Abstract Wireless Sensor Network-based localization technique, as one fundamental component of the Internet of Things, has attracted more and more attentions. However, some of the collected data easily result in faulty data with large biases due to the network attacks, large obstacles, network environments and hardware errors, which brings a big challenge for localization. The localization research in such a scenario is still relatively few. The authors try to develop a novel localization algorithm to solve this problem, by characterizing the density information of data distribution with sufficiently exploiting topology structures in WSN, and name it as Location Estimation- Robust Locality Preserving Canonical Correlation Analysis (LE-RLPCCA). Experimental results in realistic environments show that LE-RLPCCA achieves higher robust and stability for faulty data than other publicly available advanced localization algorithms.

Keywords Internet of Things; wireless sensor network; localization; topology structures; faulty data; robust

1 引言

物联网(The Internet of Things)是把具有标

识、感知和智能处理能力的物体,借助通信技术互连而成的网络.这些无需人工干预即可实现协同和互动,目的在于为人们提供智能服务^[1-2].由于物联网可广泛应用于各种自然灾害的监测、医疗卫生、电力

收稿日期:2010-04-25;最终修改稿收到日期:2010-08-17. 本课题得到航天创新基金(2009-042-004)、国家“八六三”高技术研究发展计划项目基金(2008AA706103)、国家自然科学基金(60773061)和江苏省自然科学基金(BK2008381)资助. 顾晶晶,女,1983年生,博士研究生,主要研究方向为无线传感器网络定位研究、模式识别. E-mail: gujingjing@nuaa.edu.cn. 陈松灿,男,1962年生,教授,博士生导师,主要研究领域为神经计算、模式识别、数据挖掘. 庄毅,女,1956年生,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.

系统、智能交通、智能小区管理、军事、航空和航天等领域,因此美国、日本等国大力投资着手开展对其相关技术的研究,并建立或正在建立国家级感知中心。目前我国也正在建立自己的“传感信息中心”。无疑,物联网正在进入迅速发展的时期。

随着物联网的发展,传感器技术已开始得到广泛关注和应用,通过它构成的无线传感器网络(Wireless Sensor Network, WSN)可连接物理世界和数字世界。目前国际上已有研究工作将其应用于环境监测和保护以及时发现和定位事故源、航空和航天的落点控制、军事目标的定位与跟踪等方面。在各种应用中,位置信息对物联网的监测活动至关重要,虽然可以通过全球定位系统 GPS(Global Position System)实现定位,但其适应于无遮挡的室外环境,且用户节点通常能耗高、体积大,成本较高,还需要固定的基础设施^[3]。因此在 GPS 应用受限的场景下,或是在人类难以胜任或无法到达的复杂环境中,采用体积小、能量消耗低、价格低廉的传感器网络可很好地解决目标发现及定位等问题。因而,作为物联网应用的基础和关键技术,传感器节点的定位问题已得到越来越多的关注。

目前,典型的 WSN 定位技术有基于接收信号强度指示(Received Signal Strength Indicator, RSSI)^[4]、基于到达时间(Time of Arrival, TOA)^[5]、基于到达时间差(Time Difference on Arrival, TDOA)^[6]和基于到达角度(Angle of Arrival, AOA)^[7]等方法,其中,因 RSSI 定位技术无需额外硬件设备支持,且符合低功率、低成本等要求,故得到了更广泛的应用。

然而基于 RSSI 的定位结果常不稳定^[8],故应用性受到一定限制。为了解决这一问题,研究者们已提出了众多改进策略和算法^[9-11]。近年来,借助机器学习对定位机制进行建模和算法设计已成为一个研究热点之一^[9-12]。该类方法的实质是将传感器定位过程视为一个机器学习问题,通过深入挖掘可用传感数据所隐含的信息(如网络拓扑结构、数据的非线性表示等),学习并建立一个从信号空间到物理空间的映射,而后运用该映射估计出未知节点的空间(位置)坐标,从而实现定位。其中,在最近的工作^[12]中,作者发展出了一个联合充分利用信号空间和物理空间局部拓扑结构信息的无线传感器网络定位模型:基于局部保持的典型相关分析的定位(Location Estimation-Locality Preserving Canonical Correlation Analysis, LE-LPCCA)。与目前同类方法相比,其定位精度和稳定性显著提高。

然而,在大规模物联网中,由于受硬件错误、网络攻击、能量不足、恶劣天气等实际环境因素的影响,所采集的数据常含较大误差,从而严重影响定位结果,在本文中,我们称这种数据为错误数据,这不同于含有一些误差和噪声的普通数据。上述介绍的所有定位算法或模型虽然可作用于含有一定误差的非理想数据,然而当错误数据存在时,它们的效果不尽如人意。虽然可以通过某种手段预先检测出它们,并加以剔除,但因 WSN 对传感器节点能量、存储量等都有严格的限制和要求,并且简单的检测算法也无法保证能全部消除这些错误数据,因此无需网络检测工具、且又具有较好鲁棒性的错误容忍技术^[13]可发挥更有效的作用。即由于物联网自身的脆弱性,采用鲁棒模型和算法,以最大限度地降低错误数据对网络的影响程度。本文就是针对该问题,在我们先前提出的 LE-LPCCA 基础上,通过对信号空间和物理空间局部信息的刻画进行改造,提出了具有较好鲁棒性的 LE-Robust LPCCA (LE-RLPCCA) 定位模型,使其在包含局部结构信息的基础上包含相对全局的拓扑结构信息,并利用这些结构信息最大程度地减小错误数据在整个建模过程中的影响。

本文在第 2 节介绍基于 WSN 的物联网定位机制结构;第 3 节简要回顾 LE-LPCCA 定位模型;第 4 节详细介绍 LE-RLPCCA 算法的设计思想和定位模型;实验结果及分析在第 5 节中给出;最后,第 6 节进行全文总结。

2 基于传感器网络的物联网定位机制

物联网的层次结构自底向上可分为 4 层:传感器网络层、接入网络层、中间件层和应用层^[1]。传感器网络层处于最底层,它将各种设备上的传感器联结起来形成一个信息采集与控制的网络。由图 1 可见,定位机制涉及到物联网的 4 个层次:首先,在最低层由传感器采集各类监控信号,经过 WSN 的汇聚节点或基站传输到上一层网络空间;其次,在网络层对数据进行融合(预处理),通过网桥、网关、路由等网络设备接入核心网络;接着,在中间件层,传输来的数据被存储在相应服务器中,并由定位模型或算法进行定位;最后,定位结果一方面在应用层呈现给管理员,另一方面根据应用需求,反馈给物联网底层,实现对应的控制。本文的重点在中间件层,研究如何根据已知节点的信号强度和物理坐标,建立定位模型。

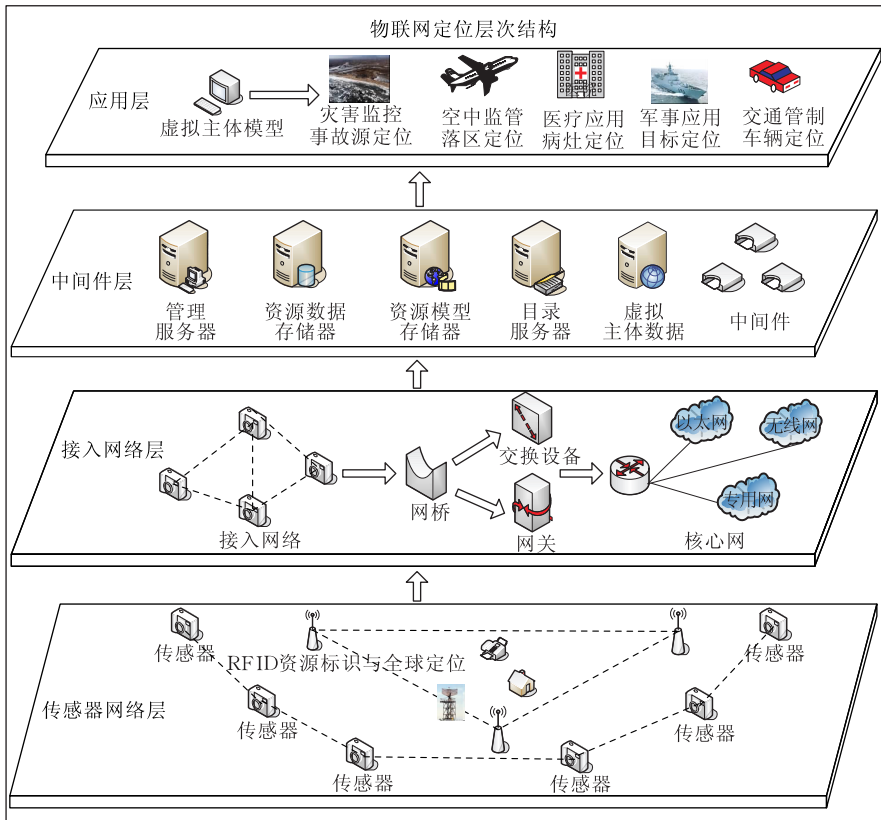


图 1 物联网定位层次结构

在基于 WSN 的物联网定位机制中,利用信号空间和物理空间的两个相配对的数据集,来建立两个空间之间的映射是一个关键步骤.在 WSN 中,不难得出结论:在节点功率相同、传送模式相似的前提下,若两个传感器节点接收到的信号强度值相似,那么它们在网络中的实际物理位置也相邻.这一特性说明信号数据集和物理位置数据集之间有着紧密的相关性.然而由于采集的数据易含有噪声,可能会导致这一相关性无法明显反应在所采集到的原始数据中^[11].

针对该问题,典型相关分析(Canonical Correlation Analysis, CCA)^[14]正是一种建立两个数据集间映射且使它们之间相关性最大的建模方法.然而 CCA 只能挖掘数据间的线性关系,这通常较难适用于实际环境.文献[11]沿此思想,利用非线性核 CCA(Kernel CCA, KCCA)^[15]挖掘数据间的非线性关系,由此实现了所需的非线性映射,提出了 LE-KCCA^[11]传感器定位算法.但 KCCA 模型采用了一个统一的全局非线性映射,而没有顾及网络的局部结构特点.

另一方面,拓扑结构复杂且多变的物联网同样使 KCCA 中核参数的选择成为难点,而选择的好坏

对 KCCA 模型的有效构建至关重要.因此,我们利用 LPCCA 方法^[12]来挖掘信号空间和物理空间的局部拓扑结构信息,从而建立了 LE-LPCCA 传感器定位模型,实现了节点定位精度的显著提高.然而,LE-LPCCA 模型并不针对本文所关注的错误数据,为此,我们发展出了具有鲁棒性的 LE-RLPCCA 定位模型.下一部分简要介绍了基于 CCA 框架的 LPCCA 模型是如何建立信号空间到物理空间的映射.

3 LE-LPCCA 的模型建立

在基于 RSSI 的 WSN 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)^[4].在训练阶段,通过对已知节点的信号强度和物理坐标来得出信号空间和物理空间的映射,建立定位模型;在定位阶段,运用得出的映射对未知节点进行位置估计.在此部分,我们主要回顾 LE-LPCCA 模型.

在 WSN 中,已知传感器所采集到的信号空间的和物理空间数据可以分别表示成两组数据集,记 $\mathbf{X}=[x_1, x_2, \dots, x_n]^{\rho \times n}$ 为 n 个已知节点所接收到的

信号强度,其中每个信号向量 $\mathbf{x}_i (i=1,2,\dots,n)$ 的维数为 p , p 为 Access Point (AP 节点) 个数, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]^{q \times n}$ 为相应节点的物理坐标, 通常在实际空间中, 坐标为二维或三维, 因此 $q=2$ 或 3 . 构建定位模型的首要任务就是建立两个数据集之间的映射.

CCA^[14] 是用来构建两组数据间映射的经典方法, 其目标是分别为 \mathbf{X} 和 \mathbf{Y} 寻找两组基向量 $\mathbf{w}_x \in R^p$ 和 $\mathbf{w}_y \in R^q$, 使得变换后的数据 $\mathbf{w}_x^T (\mathbf{x}_i - \bar{\mathbf{x}})$ 和 $\mathbf{w}_y^T (\mathbf{y}_i - \bar{\mathbf{y}})$ 之间的相关性达到最大. 其中, $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$ 和 $\bar{\mathbf{y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i$ 分别是 \mathbf{X} 和 \mathbf{Y} 的样本均值. 通过相关性定义^[14] 以及若干步数学推导, CCA 可表述为式(1)优化问题的解^[16]:

$$\begin{aligned} \max_{\mathbf{w}_x, \mathbf{w}_y} \quad & \mathbf{w}_x^T \sum_{i=1}^n \sum_{j=1}^n (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{y}_i - \mathbf{y}_j)^T \mathbf{w}_y \\ \text{s. t.} \quad & \mathbf{w}_x^T \sum_{i=1}^n \sum_{j=1}^n (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{w}_x = 1 \\ & \mathbf{w}_y^T \sum_{i=1}^n \sum_{j=1}^n (\mathbf{y}_i - \mathbf{y}_j)(\mathbf{y}_i - \mathbf{y}_j)^T \mathbf{w}_y = 1 \end{aligned} \quad (1)$$

通过解优化问题(1)^[14-15], 我们可以得出 \mathbf{w}_x 和 \mathbf{w}_y , 进而对数据做形如 $\mathbf{w}_x^T (\mathbf{x}_i - \bar{\mathbf{x}})$ 和 $\mathbf{w}_y^T (\mathbf{y}_i - \bar{\mathbf{y}})$ 的变换, 变换后的两数据集之间相关性最大. 然而, 若基于 CCA 建立 WSN 中信号空间到物理空间的映射并构建定位模型, 只能挖掘两组数据之间的线性相关现象, 而且没有利用网络的局部结构信息. 为了弥补这个问题, LE-LPCCA 算法^[12] 在构建映射时, 将网络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, 因此通过局部线性的方法达到解决非线性问题的目的. LE-LPCCA 模型首先给出了 WSN 中近邻节点的定义^[12].

定义 1. 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 x_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^[16] 划分: 若 $\mathbf{x}_j (\mathbf{y}_j)$ 是 $\mathbf{x}_i (\mathbf{y}_i)$ 的 k -近邻样本, 则称 $\mathbf{x}_j (\mathbf{y}_j)$ 是 $\mathbf{x}_i (\mathbf{y}_i)$ 的局部近邻.

根据上述定义, WSN 的拓扑结构在信号空间和物理空间的相似度矩阵为 $\mathbf{S}_x = \{S_{ij}^x\}_{i,j=1}^n$ 和 $\mathbf{S}_y = \{S_{ij}^y\}_{i,j=1}^n$, 其中矩阵元素^[12]:

$$S_{ij}^x = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / t_x), & j \in ne(i) / i \in ne(j) \\ 0, & \text{其它} \end{cases} \quad (2a)$$

$$S_{ij}^y = \begin{cases} \exp(-\|\mathbf{y}_i - \mathbf{y}_j\|^2 / t_y), & j \in ne(i) / i \in ne(j) \\ 0, & \text{其它} \end{cases} \quad (2b)$$

参数 t_x 取为信号空间的平均距离 $\sum_{i=1}^n \sum_{j=1}^n \|\mathbf{s}_i - \mathbf{s}_j\|^2 / n(n-1)$, 参数 t_y 也作类似处理. 因此, 可以看出 S_{ij}^x (或 S_{ij}^y) 越大意味着 \mathbf{x}_i 与 \mathbf{x}_j (或 \mathbf{y}_i 与 \mathbf{y}_j) 距离越近, 若 \mathbf{x}_i 与 \mathbf{x}_j (或 \mathbf{y}_i 与 \mathbf{y}_j) 互不在邻域范围内, 则相似度为零. 且 S_{ij}^x (或 S_{ij}^y) 依赖于传感器节点的布局, 因此 S_{ij}^x (或 S_{ij}^y) 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.

有了上述相似度矩阵的定义, 网络局部邻域内的典型相关性可定义为

$$\mathbf{w}_x^T \cdot \sum_{j=1}^n S_{ij}^x (\mathbf{x}_i - \mathbf{x}_j) S_{ij}^y (\mathbf{y}_i - \mathbf{y}_j)^T \cdot \mathbf{w}_y,$$

因此, 一个全局网络的非线性问题可分解为 n 个局部的(准)线性子问题. 反过来, 这些子问题的组合可作为原问题的近似. 因此, 考虑了数据的局部分布特点以后, LPCCA 可描述为如下优化问题^[16,12]:

$$\begin{aligned} \max_{\mathbf{w}_x, \mathbf{w}_y} \quad & \mathbf{w}_x^T \sum_{i=1}^n \sum_{j=1}^n S_{ij}^x (\mathbf{x}_i - \mathbf{x}_j) S_{ij}^y (\mathbf{y}_i - \mathbf{y}_j)^T \mathbf{w}_y \\ \text{s. t.} \quad & \mathbf{w}_x^T \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{x2} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{w}_x = 1 \\ & \mathbf{w}_y^T \sum_{i=1}^n \sum_{j=1}^n S_{ij}^{y2} (\mathbf{y}_i - \mathbf{y}_j)(\mathbf{y}_i - \mathbf{y}_j)^T \mathbf{w}_y = 1 \end{aligned} \quad (3)$$

求解(3)的过程类似 CCA, 详细请参考文献[16]. 注意这里的数据不需要加入样本均值 $\bar{\mathbf{x}}$ 和 $\bar{\mathbf{y}}$, 这一点和 CCA 有显著区别. 至此, LE-LPCCA 模型对 WSN 中的信号空间到物理空间的映射已建立完成.

4 LE-RLPCCA 定位模型的提出

4.1 错误数据的几何表现

在复杂或不可信网络环境中, 受网络攻击、硬件错误、环境障碍物等因素影响, 数据在传输或定位过程中易发生失真或错误, 与普通网络中易产生的简单误差不同, 这种错误数据会严重地影响定位结果. 本文正是侧重对此网络环境下的定位方法开展研究.

这种具有严重偏差的错误数据, 亦可称为“野值”, 其值与所采集到的正常数据有很大偏差, 从而严重影响定位结果^[17]. 换言之, 这一“野值”特性使错误数据在几何空间分布上远离其它正常数据点, 从而具有较低分布密度, 因此 LE-RLPCCA 的动

机则是在同一数据集中,使分布密度越低的点对定位的影响越低,以实现较好的鲁棒性。

4.2 RLPCCA 模型描述

通过对 LE-LPCCA 的刻画,我们有了对相似度矩阵 $\mathbf{S}_X = \{S_{ij}^X\}_{i,j=1}^n$ 和 $\mathbf{S}_Y = \{S_{ij}^Y\}_{i,j=1}^n$ 的定义. 从而, 可得 $D_i^X = \sum_{j=1}^n S_{ij}^X$ 和 $D_i^Y = \sum_{j=1}^n S_{ij}^Y$, 其中 D_i^X 刻画了节点 i 在信号空间中所处的区域密集程度. 若 D_i^X (D_i^Y) 越大, 表示节点 i 所处的区域密度越高. 进一步, 我们可刻画出基于节点拓扑结构的密度:

$$M_i^X = \frac{D_i^X}{\sum_{j=1}^n D_j^X}; M_i^Y = \frac{D_i^Y}{\sum_{j=1}^n D_j^Y} \quad (4)$$

显然, 当 M_i^X (M_i^Y) 越大, 则表示节点 i 在信号 (物理) 空间的密度越大. 由此可得出信号空间的密度矩阵 $\mathbf{M}_X = \text{diag}[M_1^X, M_2^X, \dots, M_n^X]$ 和物理空间的密度矩阵 $\mathbf{M}_Y = \text{diag}[M_1^Y, M_2^Y, \dots, M_n^Y]$. 将 M_i^X 与 M_i^Y 代入式(3), 分别代替 S_{ij}^X 和 S_{ij}^Y , 可得到如下优化问题:

$$\begin{aligned} \max_{\mathbf{w}_x, \mathbf{w}_y} & \mathbf{w}_x^T \sum_{i=1}^n \sum_{j=1}^n M_i^X (\mathbf{x}_i - \mathbf{x}_j) M_i^Y (\mathbf{y}_i - \mathbf{y}_j)^T \mathbf{w}_y \\ \text{s. t.} & \mathbf{w}_x^T \sum_{i=1}^n \sum_{j=1}^n M_i^{X^2} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{w}_x = 1 \\ & \mathbf{w}_y^T \sum_{i=1}^n \sum_{j=1}^n M_i^{Y^2} (\mathbf{y}_i - \mathbf{y}_j) (\mathbf{y}_i - \mathbf{y}_j)^T \mathbf{w}_y = 1 \end{aligned} \quad (5)$$

通过对式(5)的进一步展开与合并, 有

$$\begin{aligned} \max_{\mathbf{w}_x, \mathbf{w}_y} & \mathbf{w}_x^T \mathbf{X} \mathbf{M}_{XY} \mathbf{Y}^T \mathbf{w}_y \\ \text{s. t.} & \mathbf{w}_x^T \mathbf{X} \mathbf{M}_{XX} \mathbf{X}^T \mathbf{w}_x = 1 \\ & \mathbf{w}_y^T \mathbf{Y} \mathbf{M}_{YY} \mathbf{Y}^T \mathbf{w}_y = 1 \end{aligned} \quad (6)$$

这里 $\mathbf{M}_{XY} = \mathbf{D} \mathbf{M}_{XY} - \mathbf{M}_X \circ \mathbf{M}_Y$, $\mathbf{M}_{XX} = \mathbf{D} \mathbf{M}_{XX} - \mathbf{M}_X \circ \mathbf{M}_X$, $\mathbf{M}_{YY} = \mathbf{D} \mathbf{M}_{YY} - \mathbf{M}_Y \circ \mathbf{M}_Y$, 符号 \circ 表示算子 (对 $\mathbf{A}, \mathbf{B} \in \mathcal{R}^{n \times n}$, $(\mathbf{A} \circ \mathbf{B})_{ij} = A_{ij} B_{ij}$, A_{ij} 表示 \mathbf{A} 的第 ij 元素); $\mathbf{D} \mathbf{M}_{XY}$ ($\mathbf{D} \mathbf{M}_{XX}$, $\mathbf{D} \mathbf{M}_{YY}$) 是大小为 $n \times n$ 的对角矩阵, 其第 i 个对角元素等于 $\mathbf{M}_X \circ \mathbf{M}_Y$ ($\mathbf{M}_X \circ \mathbf{M}_X$, $\mathbf{M}_Y \circ \mathbf{M}_Y$) 的第 i 行 (因其对称性, 或第 i 列) 元素之和.

利用 Lagrange 乘子法, 求解优化问题(6), 容易得到如下 RLPCCA 的广义特征值方程:

$$\begin{pmatrix} 0 & \mathbf{X} \mathbf{M}_{XY} \mathbf{Y}^T \\ \mathbf{Y} \mathbf{M}_{YX} \mathbf{X}^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{w}_x \\ \mathbf{w}_y \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{X} \mathbf{M}_{XX} \mathbf{X}^T & \\ & \mathbf{Y} \mathbf{M}_{YY} \mathbf{Y}^T \end{pmatrix} \begin{pmatrix} \mathbf{w}_x \\ \mathbf{w}_y \end{pmatrix} \quad (7)$$

求解(7), 得到基向量组 $(\mathbf{w}_x, \mathbf{w}_y)$ 后, 即可对原有数据进行形如 $\mathbf{w}_x^T \mathbf{x}$ 和 $\mathbf{w}_y^T \mathbf{y}$ 的变换.

文献[15]给出了利用奇异值分解(SVD)技术求解CCA的方法. 本文同样利用SVD求解RLPCCA.

令 $\mathbf{H} = (\mathbf{X} \mathbf{M}_{XX} \mathbf{X}^T)^{-\frac{1}{2}} (\mathbf{X} \mathbf{M}_{XY} \mathbf{Y}^T) (\mathbf{Y} \mathbf{M}_{YY} \mathbf{Y}^T)^{-\frac{1}{2}}$, $\mathbf{u} = (\mathbf{X} \mathbf{M}_{XX} \mathbf{X}^T)^{\frac{1}{2}} \mathbf{w}_x$, $\mathbf{v} = (\mathbf{Y} \mathbf{M}_{YY} \mathbf{Y}^T)^{\frac{1}{2}} \mathbf{w}_y$, 则式(7)可整理为

$$\begin{cases} \mathbf{H} \mathbf{H}^T \mathbf{u} = \lambda^2 \mathbf{u} \\ \mathbf{H}^T \mathbf{H} \mathbf{v} = \lambda^2 \mathbf{v} \end{cases} \quad (8)$$

式(8)实际上对应于矩阵 \mathbf{H} 的SVD分解, 记 $\mathbf{H} = \mathbf{U} \mathbf{D} \mathbf{V}^T = \sum_{i=1}^d \lambda_i \mathbf{u}_i \mathbf{v}_i^T$, 为矩阵 \mathbf{H} 的SVD分解, 其中对角矩阵 \mathbf{D} 的第 i 个元素恰好等于 λ_i , \mathbf{u}_i 和 \mathbf{v}_i 分别是矩阵 \mathbf{U} 和 \mathbf{V} 的第 i 列, 对应奇异值 λ_i , 则有

$$\begin{cases} \mathbf{w}_{xi} = (\mathbf{X} \mathbf{M}_{XX} \mathbf{X}^T)^{-\frac{1}{2}} \mathbf{u}_i \\ \mathbf{w}_{yi} = (\mathbf{Y} \mathbf{M}_{YY} \mathbf{Y}^T)^{-\frac{1}{2}} \mathbf{v}_i \end{cases} \quad (9)$$

利用式(9)即可得RLPCCA问题的第 i ($i=1, 2, \dots, d$) 对基向量, 这里 $d \leq \min(p, q)$. 利用SVD分解求解RLPCCA具有计算稳定的特点, 且 λ_i 和 \mathbf{w}_{xi} 和 \mathbf{w}_{yi} , $i=1, \dots, d$, 可一次性计算得到. 现设 $\mathbf{W}_X = [\mathbf{w}_{x1}, \dots, \mathbf{w}_{xd}]$ 和 $\mathbf{W}_Y = [\mathbf{w}_{y1}, \dots, \mathbf{w}_{yd}]$.

4.3 LE-RLPCCA 定位算法

已知信号强度矩阵 \mathbf{X} 和对应的物理坐标矩阵 \mathbf{Y} . 定位模型的目标就是通过未知节点 g 所接收到的信号向量 $\mathbf{x}_g = (x_{g1}, x_{g2}, \dots, x_{gm})^T$, 估计出 g 的位置坐标 \mathbf{y}_g . 在4.2节, 我们已详细讨论了训练阶段中利用RLPCCA建立信号空间到物理空间的映射的方法; 接下来的工作就是如何在定位阶段中快速定位出未知节点: 由于RLPCCA算法的特点是使数据变换后的信号空间和物理空间的相关性最大, 且仍保持原始空间局部拓扑结构信息. 因此, 坐标邻近的点在变换后信号空间中依旧相邻. 根据这一特性, 我们可对未知节点所接收到的信号强度值做数据变换后, 找出与之 K 个距离最近的已知节点, 因而这 K 个节点的物理坐标 $[\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K]^{q \times K}$ 必定也在 \mathbf{y}_g 附近. 最终我们可以通过质心方法估算出 g 点的物理坐标:

$$\mathbf{y}_g = \frac{\mathbf{y}_1 + \mathbf{y}_2 + \dots + \mathbf{y}_K}{K} \quad (10)$$

现在总结LE-RLPCCA算法的具体构建步骤, 其训练阶段和定位阶段分别描述如下.

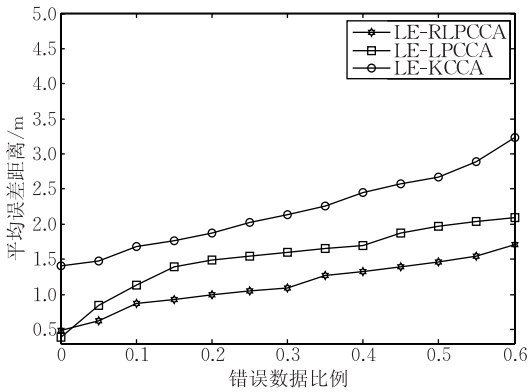
训练阶段.

1. 通过式(2), 计算相似度矩阵 $\mathbf{S}_X = \{S_{ij}^X\}_{i,j=1}^n$ 和 $\mathbf{S}_Y = \{S_{ij}^Y\}_{i,j=1}^n$.
2. 通过式(4), 得出密度矩阵 \mathbf{M}_i^X 与 \mathbf{M}_i^Y .
3. 解最优化问题(7)~(9), 学习物理空间和信号空间的关系, 求出映射矩阵 \mathbf{W}_X 和 \mathbf{W}_Y .
4. 对已知数据 $\{\mathbf{X}, \mathbf{Y}\}$ 做线性变换: $\{P^X, P^Y\} = \{\mathbf{W}_X^T \times \mathbf{X}, \mathbf{W}_Y^T \times \mathbf{Y}\}$. 即对原始空间数据做相应映射, 从而使 P^X, P^Y 相关

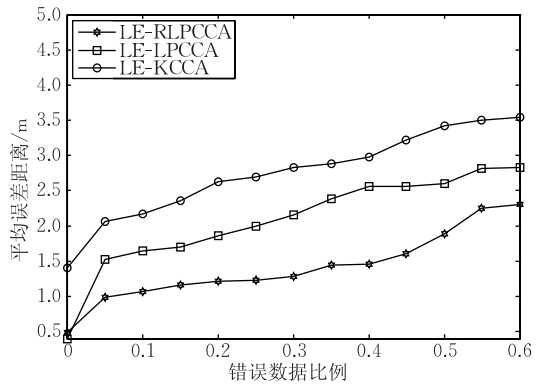
定位效果及鲁棒性,把训练数据分成 3 种情形:(1)信号强度数据有误;(2)已知节点的物理坐标有误;(3)两组数据集都有错误.其中,对第(1)和(2)两种情形,随机选取训练数据的 0%~50%作为错误数据,并设定错误值分别为正常值的 3 倍、6 倍、10 倍;对第(3)种情形,假定错误数据占整个训练数据的 30%,并设定当错误信号值分别为正常值的 3 倍、6 倍、10 倍时,错误物理坐标的错误值为正常值的 1~10 倍.对于这 3 种情形,定位平均误差如图 3~图 5 所示.

由图 3~图 5 可见:(1)LE-RLPCCA 定位模型

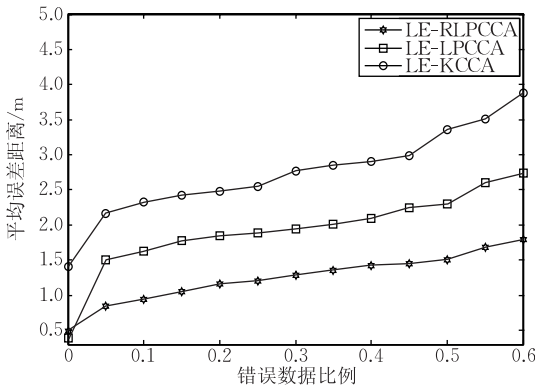
拥有最低的平均定位误差,随着错误数据比例的增加及错误数据值增大,LE-RLPCCA 的定位优势越发明显,且相对稳定;(2)LE-LPCCA 在无错误数据时(如图 1 及图 2 中错误数据比例为 0%时),平均定位误差较小,然而随着错误数据逐渐增加(图 3、图 4),或者错误数据值增大时(图 5),通过 LE-LPCCA 模型计算出的平均误差增加幅度较大,但结果仍好于 LE-KCCA 模型.(3)LE-KCCA 模型在无错误数据时,已显示出较差的定位结果,而存在错误数据时,平均定位误差进一步大幅度增加;同时还能发现,随着错误数据逐渐增加(图 3、图 4),或者



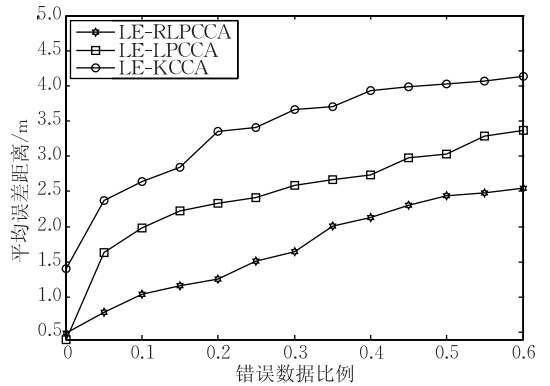
(a) 错误信号强度3倍于正常值



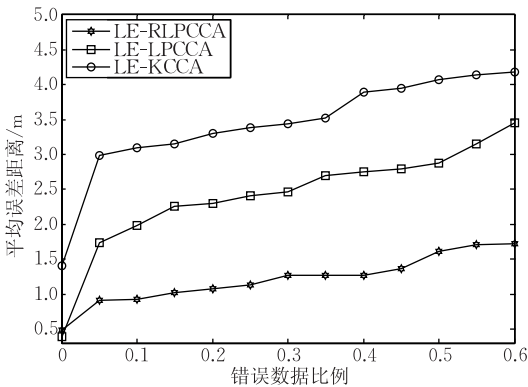
(a) 错误物理坐标3倍于正常值



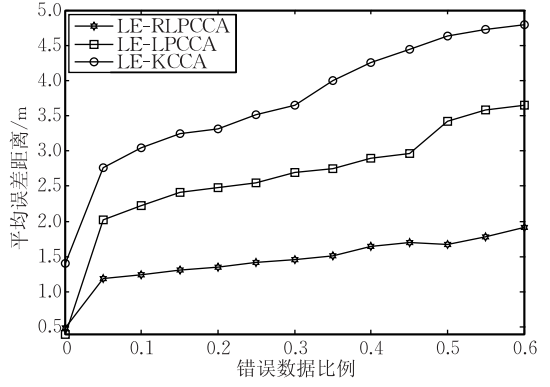
(b) 错误信号强度6倍于正常值



(b) 错误物理坐标6倍于正常值



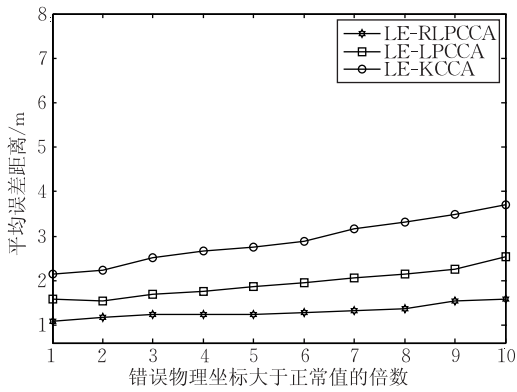
(c) 错误信号强度10倍于正常值



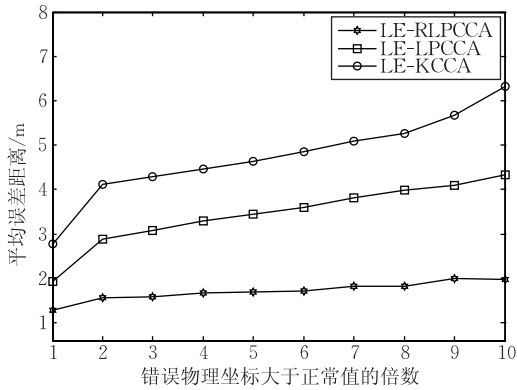
(c) 错误物理坐标10倍于正常值

图 3 当错误数据比例逐渐增大时,不同的错误信号强度所产生的定位平均误差

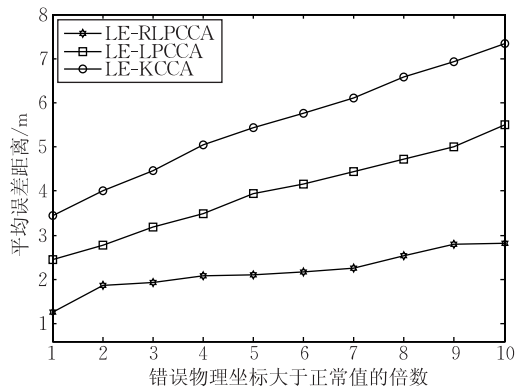
图 4 当错误数据比例逐渐增大时,不同的错误地址坐标所产生的定位平均误差



(a) 错误信号强度3倍于正常值



(b) 错误信号强度6倍于正常值



(c) 错误信号强度10倍于正常值

图5 当错误物理坐标值成倍增大时,不同的错误信号强度所产生的定位平均误差

错误数据值逐渐增大时(图5), LE-KCCA 与 LE-LPCCA 间的差距进一步加大,这说明虽然两种定位模型都没有针对错误数据,但因 LE-LPCCA 模型基于了网络的拓扑结构信息,所以错误数据主要作用于相应的局部领域空间,而在执行 LE-KCCA 时,错误数据会影响整个网络空间,因此显示出更差的鲁棒性。

6 结 论

本文提出了一种 LE-RLPCCA 物联网定位模

型,通过充分利用 WSN 中信号空间和物理空间的拓扑结构和密度信息,可有效应用于数据采集易产生较大错误的物联网环境,以克服现有算法或模型定位精度不高、鲁棒性差等问题。

与同类研究相比,LE-RLPCCA 模型在存在错误数据的场景中,具有定位精确度高且性能相对稳定的优点。下一步的工作拟把 LE-RLPCCA 细化到分布式计算领域中,以更方便适用于大规模物联网的定位估计。

参 考 文 献

- [1] Commission of the European Communities. Internet of Things—An Action Plan for Europe. 1st Edition. Brussels: COM (2009) 278. 2009; 1-12
- [2] Kranenburg R V. The Internet of Things: A critique of ambient technology and the all-seeing network of RFID. Network Notebooks 02. Amsterdam: Institute of Network Cultures Press, 2007; 10-28
- [3] Sun Li-Min, Li Jian-Zhong, Chen Yu, Zhu Hong-Song. Wireless Sensor Network. 1st Edition. Beijing: Tsinghua University Press, 2005; 136(in Chinese)
(孙利民, 李建中, 陈渝, 朱红松. 无线传感器网络. 第1版. 北京: 清华大学出版社, 2005; 136)
- [4] Bahl P, Padmanabhan V N. Radar: An in-building RF-based user location and tracking system//Proceedings of the IEEE INFOCOM. Tel Aviv, Israel, 2000; 775-784
- [5] Harter A, Hopper A, Steggles P, Ward A, Webster P. The anatomy of a context-aware application//Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking. In Mobile Computing and Networking. Seattle, USA, 1999; 59-68
- [6] Girod L, Estrin D. Robust range estimation using acoustic and multimodal sensing//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 01). Maui, Hawaii, USA, 2001; 1312-1320
- [7] Niculescu D, Nath B. Ad hoc positioning system (APS) using AoA//Proceedings of the IEEE INFOCOM. San Francisco, USA, 2003; 1734-1743
- [8] Wang Fu-Bao, Shi Long, Ren Feng-Yuan. Self-localization systems and algorithms for wireless sensor networks. Journal of Software, 2005, 16(5): 857-868(in Chinese)
(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)
- [9] Wang C Q, Chen J M, Sun Y X. Sensor network localization using kernel spectral regression. Wireless Communications and Mobile Computing, 2010, 10(8): 1045-1054
- [10] Weiss A J, Picard J S. Network localization with biased range measurements. IEEE Transactions on Wireless Communications, 2008, 7(1): 298-304

- [11] Pan J J, Kwok J T. Multidimensional vector regression for accurate and low-cost location estimation in pervasive computing. *IEEE Transactions on Knowledge and Data Engineering*, 2006, 18(9): 1181-1193
- [12] Gu Jing-Jing, Chen Song-Can, Zhuang Yi. Location estimation in wireless sensor network using locality preserving canonical correlation analysis. *Journal of Software*, in Press. Available at http://parnec.nuaa.edu.cn/publication.htm#Technical_Report (in Chinese)
(顾晶晶, 陈松灿, 庄毅, 基于局部保持的典型相关分析的无线传感器网络定位算法, 软件学报, 09 年录用. 相应的技术报告 2009.C01 可从 http://parnec.nuaa.edu.cn/publication.htm#Technical_Report 下载)
- [13] Han X F, Cao X, Lloyd E L, Shen C C. Fault-tolerant relay node placement in heterogeneous wireless sensor networks. *IEEE Transactions on Mobile Computing*, 2010, 9(5): 643-656
- [14] Hardoon D R, Szedmak S, Shawe-Taylor J. Canonical correlation analysis; an overview with application to learning methods. *Neural Computation*, 2004, 16(12): 2639-2664
- [15] Melzer T, Reiter M, Bischof H. Appearance models based on kernel canonical correlation analysis. *Pattern Recognition*, 2003, 36(9): 1961-1971
- [16] Sun T K, Chen S C. Locality preserving CCA with applications to data visualization and pose estimation. *Image and Vision Computing*, 2007, 25(5): 531-543
- [17] Heinz C, Seeger B. Statistical modeling of sensor data and its application to outlier detection. University of Stuttgart, Germany; Technical Report; LSIR-. REPORT-2006-006, 2006



GU Jing-Jing, born in 1983, Ph. D. candidate. Her research interests include wireless sensor network, pattern recognition.

CHEN Song-Can, born in 1962, professor, Ph. D. supervisor. His research interests include neural computation, pattern recognition, data mining.

ZHUANG Yi, born in 1956, professor, Ph. D. supervisor. Her research interests include information safety, distributed computing, computer network.

Background

This paper focuses on the wireless sensor network (WSN)-based localization technique in The Internet of Things. The localization of WSN has been one of the foundational functions and obtained satisfactory accuracies in some environments. However, some of the collected data easily result in faulty data with large biases due to the network attacks, large obstacles, network environments and hardware errors, which brings a big challenge for localization. The localization research in such a scenario is still relatively few. Herein, the authors try to develop a novel localization algorithm to solve this problem, by characterizing the density information of data distribution with sufficiently exploiting topology structures in WSN, and name it as Location Estimation- Robust Locality Preserving Canonical Correlation Analysis (LE-RLPCCA). Experimental results in realistic environments show that LE-RLPCCA achieves higher robust and stability for faulty data than other publicly available advanced

localization algorithms.

This research was supported by National Natural Science Foundation of China (NSFC) (60773061), Aeronautics Innovation Fund (2009-042-004) and the National High Technology Research and Development Program (863 Program) (2008AA706103). The NSFC project has designed LPCCA model which built a solid fundament for studying LE-RLPCCA model. The latter two projects were applying localization models and algorithms of wireless sensor network (WSN) in the Internet of Things. Results of this study partially resolved the structural design of localization systems in WSN-based Internet of Things, and some issues of the target localization model and localization algorithm in Internet of Things containing faulty data. The team has published several research articles and scientific reports in WSN localization, as described in reference, and developed two software & hardware achievements.