

一种基于非完全数据的路径损耗模型选择算法

姜向远 张焕水* 王伟

(山东大学控制科学与工程学院 济南 250061)

摘要: 为尽可能选择反应实际信道衰减特性的路径损耗模型, 该文针对无线传感器网络中的路径损耗提出一种多模型选择算法。首先分析一组路径损耗模型的统计特性, 然后考虑接收信号强度(RSS)的非完全数据, 提出基于期望最大化(EM)的参数估计算法, 最后在准则函数的基础上给出表征模型优劣的权重系数, 建立了一种进行路径损耗模型筛选的量化方法。实验分析表明, 该算法的参数估计较为准确, 且与同类算法相比, 该算法甄别出的最优模型与实验数据有较好的拟合度。

关键词: 无线传感网; 非完全数据; 期望最大化; 准则函数; 路径损耗模型

中图分类号: TP393

文献标识码: A

文章编号: 1009-5896(2012)06-1438-07

DOI: 10.3724/SP.J.1146.2011.01084

A Model Selection Algorithm for Path Loss from Incomplete Data

Jiang Xiang-yuan Zhang Huan-shui Wang Wei

(School of Control Science and Engineering, Shandong University, Ji'nan 250061, China)

Abstract: In order to reflect the actual channel attenuation, a model selection algorithm for path loss in wireless sensor networks is proposed. Firstly, the statistical properties of some path loss models are analyzed, and then Expectation Maximization (EM) algorithm from incomplete data of Received Signal Strength (RSS) is proposed for parameter estimation. Finally, a set of weighted coefficients are given on the basis of criterion function, which could select an appropriate path loss model. Through experiment, the proposed model selection method could estimate parameters effectively, compared with other similar algorithms, this method could pick up a model fitting the experimental data better.

Key words: Wireless sensor network; Incomplete data; Expectation Maximization (EM); Criterion function; Path loss model

1 引言

路径损耗模型是无线传感器网络中进行功率控制、链路分析^[1]乃至跟踪定位^[2]的重要理论基础。因为信号传输环境复杂多变, 单一的路径损耗模型很难满足现实应用的复杂要求, 诸如经典log模型、log双斜率模型^[3]、人体传感器网络模型、双斜率模型^[4]等路径损耗模型陆续被提出。面对如此多的路径损耗模型, 在建模过程中模型选择成为非常重要的一个环节。所谓模型选择, 就是确定能够准确反映数据生成过程的最优模型。基于信息理论的准则函数方法如赤池信息准则(AIC), 贝叶斯信息准则(BIC)^[5]是模型选择中最常用的方法, 我们可以通过比较每个候选模型的准则值并将准则值最小的模型作为真实模型的估计。

目前, 路径损耗模型筛选方面的研究主要集中在模型参数估计^[6]和特定环境建模^[7]。上述研究多是针对某一种或某一类路径损耗模型, 且仅从工程应用的角度分析模型的可靠性, 无法保证模型是最优的, 更没有给出详细的模型选择标准。文献[8]系统分析了2.4 GHz频段下多种路径损耗模型的统计特性, 实验验证各模型在人体周围传输环境的可靠性, 给出基于改进的赤池信息准则(AIC_c)的多模型选择方法。但文献[8]在实验测量及理论分析时没有考虑观测误差对参数估计及模型筛选的影响, 这样势必会影响建模的精度和模型筛选的准确性。

针对上述问题, 本文充分考虑接收信号强度(RSS)测量过程中引入的舍入误差, 借助期望最大化(EM)算法和AIC_c准则函数建立了筛选路径损耗最优模型的量化方法: 使用EM算法求解含有非完全数据的极大似然估计问题, 并在求解过程中, 针对线性模型使用传统的矩阵运算求解待定参数, 针对非线性模型则引入牛顿迭代和矩阵的正交三角(QR)

2011-10-19 收到, 2012-03-04 改回

国家杰出青年科学基金(60825304)和国家 973 重点基础研究基金(2009cb320600)资助课题

*通信作者: 张焕水 hszhang@sdu.edu.cn

分解来估计模型参数,有效改进参数的估计精度;考虑到实际应用环境下 RSS 采样数通常较少,该算法使用适应于小样本空间的 AIC_c 准则函数做模型选择,并引入权重系数判断模型的优劣。

2 路径损耗模型分析

无线传感器网络中常用的路径损耗模型有 4 种,假设第 k 种模型的系数为 $\{\xi_{k0}, \xi_{k1}, \dots\}$,令 d 表示信号传输距离, p 表示发射的基准能量,则各模型中 RSS 均值 $s(d, p)$ 可表示为

(1)经典 log 模型

$$s_1(d, p) = p + \xi_{10} + 10\xi_{11} \lg d \quad (1)$$

该模型适用于无线个人局域网大部分传输环境。在理想的自由传输空间,路径损耗指数 $\xi_{11} = 2$,在实际传输环境, ξ_{11} 取值通常在 2 到 5 之间。

(2)log 双斜率模型

$$s_2(d, p) = p + \begin{cases} \xi_{20} + 10\xi_{21} \lg d, & 1 < d < d_b \\ \xi_{20} + 10(\xi_{22} - \xi_{21}) \lg d_b + 10\xi_{22} \lg d, & d \geq d_b \end{cases} \quad (2)$$

该模型由双射线模型推导得到^[3],适用于无线个人局域网室外传播环境,尤其是视距范围。 d_b 表示信号强度变化规律发生变化的位置,又称拐点; ξ_{21} , ξ_{22} 分别是拐点前后反应信号强度变化规律的指数。通常情况下 $\xi_{21}=2$, ξ_{22} 视传输环境的变化略有波动,取值范围在 2 到 7 之间。

(3)人体传感器网络模型

$$s_3(d, p) = p + \xi_{30} + 10\xi_{31}d \quad (3)$$

该模型已收入 IEEE 802.15.4a 信道模型标准,适用于有衍射的短距离传输环境。

(4)双斜率模型

$$s_4(d, p) = p + 10 \lg(\xi_{40} \exp(\xi_{41}d) + \xi_{42} \exp(\xi_{43}d)) \quad (4)$$

该模型可视为 log 双斜率模型在人体传感器网络应用环境的改进,适用于短距离传输环境,尤其是有柱状障碍物的非视距传输条件。

上述模型中(1)-(3)是线性回归模型,(4)为非线性模型,它们的阴影衰落是均值为 0,方差为 σ^2 的白噪声。令 θ 表示各模型系数及对应阴影衰落的标准差 σ ,作为各模型的一列待定参数,则 RSS 观测的概率密度函数可表示成如下形式:

$$f(r | d, p, \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{r - s(d, p)}{\sigma}\right)^2\right] \quad (5)$$

对应 n 个相互独立的 RSS 观测的条件似然函数可表示为

$$l(\theta) = \prod_{i=1}^n \int f(r^{(i)} | d^{(i)}, p^{(i)}, \theta) dr \quad (6)$$

3 模型选择算法

为了充分利用 RSS 舍入误差的信息,提高模型选择的准确性,本节提出一种基于非完全数据的路径损耗模型选择算法。首先采用 EM 算法估计模型参数并求取似然函数极值,然后依据 AIC_c 准则函数选取最优模型。定义如下一组对应 n 个 RSS 观测的向量:

(1)采用第 k 种模型时接收信号强度 $r_k = (r_k^{(1)}, \dots, r_k^{(n)})$;

(2)采用第 k 种模型时发射信号强度 $p_k = (p_k^{(1)}, \dots, p_k^{(n)})$;

(3)采用第 k 种模型时主从节点间的距离 $d_k = (d_k^{(1)}, \dots, d_k^{(n)})$ 。

3.1 基于非完全数据的参数估计

本小节将采用不同方法分别对线性模型和非线性模型做参数估计。因信号强度检测方法简易,指示器自身精度有限等原因,观测到的 RSS 值 o 往往难以准确反映实际的 RSS 值 r 。假设向量 $o = (o^{(1)}, \dots, o^{(n)})$ 表示 RSS 的观测,则 $o^{(j)} - \lambda/2 \leq r^{(j)} \leq o^{(j)} + \lambda/2$,其观测精度主要由某一常数 λ 决定。

考虑非完全数据,第 k 种模型的 RSS 似然函数可表示为

$$l_k(\theta) = \prod_{i=1}^n \int_{o^{(i)} - \lambda/2}^{o^{(i)} + \lambda/2} f(r_k^{(i)} | d_k^{(i)}, p_k^{(i)}, \theta) dr_k \quad (7)$$

上述非完全数据的似然函数形式上类似常规的似然函数,但很难得到解析解。因各模型完全数据的似然函数存在且数学期望可求,故可采用 EM 算法得到模型参数的局部最优解。为了得到非完全数据似然函数的数学期望,需知道 RSS 非完全数据的概率分布。EM 算法的 E-步先假定未知参数为某一组定值 θ_t ,可得 RSS 的概率密度函数服从正态分布,则 EM 算法可表示为

E-步

$$\psi_k(\theta, \theta_t) = E_{\theta_t} \ln l(\theta) \quad (8)$$

其中 $\psi_k(\theta, \theta_t)$ 表示似然函数的数学期望, $l(\theta)$ 为式(6)所示各模型完全数据的似然函数。

M-步

$$\theta_{t+1} = \operatorname{argmax}_{\theta} \psi_k(\theta, \theta_t) \quad (9)$$

其中 θ_t 表示第 t 步的模型参数,该算法不断重复上述两步,形成一个循环。迭代过程中似然函数值不断增加,从而保证了算法的收敛性,且收敛于似然函数的一个局部极值^[9]。

下面介绍如何通过 EM 算法求取待定参数的估计值,对式(9)取对数,代入式(8)可得

$$\psi_k(\boldsymbol{\theta}, \boldsymbol{\theta}_t) = E_{\boldsymbol{\theta}_t} \sum_{i=1}^n \left[-\frac{1}{2} \ln 2\pi - \ln \sigma - \frac{1}{2} \left(\frac{r_k^{(i)} - s_k^{(i)}}{\sigma} \right)^2 \right] \quad (10)$$

式(10)似然函数的极大值对应残差平方和的最小值, 故参数 $\boldsymbol{\theta}$ 的极大似然估计可转化为求均方误差 SVE_k 的最小值问题。

$$SVE_k = \sum_{i=1}^n E_{\boldsymbol{\theta}_t} (r_k^{(i)} - s_k^{(i)})^2 \quad (11)$$

但不同路径损耗模型的最小二乘估计方法略有差异。

(1)线性模型 模型(1)-模型(3)是标准的线性回归模型, 根据线性回归模型的相关计算性质可得模型参数的最小二乘估计^[9]

$$\hat{\boldsymbol{\xi}}_k = (\mathbf{Z}_k^T \mathbf{Z}_k)^{-1} \mathbf{Z}_k^T \mathbf{Y}_k \quad (12)$$

其中

$$\hat{\boldsymbol{\xi}}_1 = \begin{bmatrix} \hat{\xi}_{10} \\ \hat{\xi}_{11} \end{bmatrix}, \quad \hat{\boldsymbol{\xi}}_2 = \begin{bmatrix} \hat{\xi}_{20} \\ \hat{\xi}_{21} \\ \hat{\xi}_{22} \end{bmatrix}, \quad \hat{\boldsymbol{\xi}}_3 = \begin{bmatrix} \hat{\xi}_{30} \\ \hat{\xi}_{31} \end{bmatrix} \quad (13)$$

$$\mathbf{Z}_1 = \begin{bmatrix} 1 & 10 \lg d_1^{(1)} \\ \vdots & \vdots \\ 1 & 10 \lg d_1^{(n)} \end{bmatrix}, \quad \mathbf{Z}_2 = \begin{bmatrix} 1 & -10 \lg d_b & 10 \lg d_2^{(1)} - 10 \lg d_b \\ 1 & -10 \lg d_b & 10 \lg d_2^{(2)} - 10 \lg d_b \\ \vdots & \vdots & \vdots \\ 1 & -10 \lg d_b & 10 \lg d_2^{(n)} - 10 \lg d_b \end{bmatrix}, \quad \mathbf{Z}_3 = \begin{bmatrix} 1 & 10 d_3^{(1)} \\ \vdots & \vdots \\ 1 & 10 d_3^{(n)} \end{bmatrix}, \quad \mathbf{Y}_k = \begin{bmatrix} E_{\boldsymbol{\theta}_t} r_k^{(1)} - p_k^{(1)} \\ E_{\boldsymbol{\theta}_t} r_k^{(2)} - p_k^{(2)} \\ \vdots \\ E_{\boldsymbol{\theta}_t} r_k^{(n)} - p_k^{(n)} \end{bmatrix} \quad (14)$$

为了求第 k 种模型的系数 $\hat{\boldsymbol{\xi}}_k$, 需先求非完全数据 $r_k^{(i)}$ 的数学期望

$$E_{\boldsymbol{\theta}_t} r_k^{(i)} = \frac{\left(\exp\left(-\frac{1}{2}(a_k^{(i)})^2\right) - \exp\left(-\frac{1}{2}(b_k^{(i)})^2\right) \right) \sigma_t}{\sqrt{2\pi} (\varphi(b_k^{(i)}) - \varphi(a_k^{(i)}))} + s_k^{t(i)}, \quad k = 1, 2, 3 \quad (16)$$

其中 φ 为标准正态分布函数, $s_k^{t(i)}$ 表示第 k 种路径损耗模型中对应待定参数 $\boldsymbol{\theta}_t$ 的 RSS 均值

$$s_k^{t(i)} = s_k(d_k^{(i)}, p_k^{(i)}, \boldsymbol{\theta}_t) \quad (17)$$

$a_k^{(i)}$ 和 $b_k^{(i)}$ 则表示为

$$a_k^{(i)} = \frac{o_k^{(i)} - \lambda/2 - s_k^{t(i)}}{\sigma_t}, \quad b_k^{(i)} = \frac{o_k^{(i)} + \lambda/2 - s_k^{t(i)}}{\sigma_t} \quad (18)$$

(2)非线性模型 式(12)是最小二乘估计的经典表达式, 但考虑计算复杂度及收敛性, 本文在非线性模型中使用 QR 分解和牛顿迭代求解最小二乘估计^[10]。

首先将非线性模型在 $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0)}$ 处做一阶泰勒级数展开

$$s_4(d_4^{(i)}, p_4^{(i)}, \boldsymbol{\theta}) = s_4(d_4^{(i)}, p_4^{(i)}, \boldsymbol{\theta}^{(0)}) + v_{i0}(\xi_{40} - \xi_{40}^{(0)}) + v_{i1}(\xi_{41} - \xi_{41}^{(0)}) + v_{i2}(\xi_{42} - \xi_{42}^{(0)}) + v_{i3}(\xi_{43} - \xi_{43}^{(0)}) \quad (19)$$

其中

$$v_{ij} = \left. \frac{\partial s_4(d_4^{(i)}, p_4^{(i)}, \boldsymbol{\theta})}{\xi_{4j}} \right|_{\boldsymbol{\theta}^{(0)}}, \quad j = 0, 1, 2, 3 \quad (20)$$

对应 n 个观测, 式(19)可表示成向量形式。

$$s_4(\boldsymbol{\xi}_4) = s_4(\boldsymbol{\xi}_4^{(0)}) + \mathbf{V}^{(0)}(\boldsymbol{\xi}_4 - \boldsymbol{\xi}_4^{(0)}) \quad (21)$$

$\mathbf{V}^{(0)}$ 是在 $\boldsymbol{\xi}_4 = \boldsymbol{\xi}_4^{(0)}$ 处求导的 $n \times 4$ 的矩阵, 其组成元素为 $\{v_{ij}\}$ 。由式(21)可得残差

$$\mathbf{w}(\boldsymbol{\xi}_4) = \mathbf{r}_4 - s_4(\boldsymbol{\xi}_4) = \mathbf{w}^{(0)} - \mathbf{V}^{(0)}\boldsymbol{\delta} \quad (22)$$

其中 $\mathbf{w}^{(0)} = \mathbf{r}_4 - s_4(\boldsymbol{\xi}_4^{(0)})$, 高斯增量 $\boldsymbol{\delta} = \boldsymbol{\xi}_4 - \boldsymbol{\xi}_4^{(0)}$ 。

非线性模型系数的牛顿迭代解法可表示为

(i)初始化 $j = 0$, 给定初值 $\boldsymbol{\xi}_4^{(0)}$ 求 $\mathbf{V}^{(0)}$, $\mathbf{w}^{(0)}$, 按步骤(ii)的方法求解 $\boldsymbol{\delta}^{(0)}$ 。

(ii)求解牛顿增量以使系统的均方误差 $|\mathbf{w}(\boldsymbol{\xi}_4)|^2$ 最小, 先对雅可比矩阵 $\mathbf{V}^{(j)}$ 做 QR 分解有 $\mathbf{V}^{(j)} = \mathbf{Q}\mathbf{R} = \mathbf{Q}_1\mathbf{R}_1$, 其中 \mathbf{Q} 是 $n \times n$ 的正交阵, 且有 $\mathbf{Q} = [\mathbf{Q}_1|\mathbf{Q}_2]$, \mathbf{Q}_1 是 \mathbf{Q} 的前 4 列, \mathbf{Q}_2 是 \mathbf{Q} 的后 $n-4$ 列。 \mathbf{R} 是 $n \times 4$ 的矩阵, 其前 4 行可表示成 4×4 的上三角阵 \mathbf{R}_1 , 后 $n-4$ 行各元素均为 0。由 \mathbf{Q}, \mathbf{R} 值可得 $\hat{s}_4(\boldsymbol{\xi}_4^{(j)}) = \mathbf{Q}_1\mathbf{Q}_1^T \mathbf{w}(\boldsymbol{\xi}_4^{(j)})$, 高斯增量 $\boldsymbol{\delta}^{(j)} = \mathbf{R}_1^{-1}\mathbf{Q}_1^T \mathbf{w}(\boldsymbol{\xi}_4^{(j)})$ 。

(iii)更新雅可比矩阵 $\mathbf{V}^{(j+1)}$, 残差 $\mathbf{w}(\boldsymbol{\xi}_4^{(j+1)}) = \mathbf{r}_4 - s_4(\boldsymbol{\xi}_4^{(j+1)})$, 系数 $\boldsymbol{\xi}_4^{(j+1)} = \boldsymbol{\xi}_4^{(j)} + \boldsymbol{\delta}^{(j)}$ 。

(iv)如果 $|\mathbf{s}_4(\boldsymbol{\xi}_4^{(j+1)}) - \hat{s}_4(\boldsymbol{\xi}_4^{(j)})| < \frac{\hat{s}_4(\boldsymbol{\xi}_4^{(j)})}{1000}$, 可得模型系数 $\hat{\boldsymbol{\xi}}_4 = \boldsymbol{\xi}_4^{(j+1)}$; 否则 j 加 1, 转到步骤(ii)。

阴影衰落的标准差 σ 的估计值为 $\hat{\sigma}_k = \sqrt{SVE_k/n}$, 至此, 模型系数 $\boldsymbol{\theta}$ 全部给出, 进而可得似然函数极值。这是下一节进行模型筛选的基础。

3.2 基于权重系数的模型选择

在 3.1 节参数估计的基础上, 本节将通过 AIC_c 准则函数给出一组进行模型选择的度量值。考虑到实际应用中路径损耗模型筛选的样本空间不能太大, 本算法采用适用于小样本空间的 AIC_c 准则函数

作为判断路径损耗模型与测量数据匹配程度的标准^[1]。

$$AIC = -2 \ln l(\theta) + 2K \quad (23)$$

$$AIC_c = AIC + \frac{2K(K+1)}{(N-K-1)} \quad (24)$$

上述准则函数在采样数相对于模型参数个数较少的情况下，即 $N/K < 40$ 时，有良好的统计学特性和工程应用价值。其中 N 为实验采样数， K 为模型参数个数，式(23)中 $\ln l(\theta)$ 由 3.1 节给出。根据极大似然估计的性质可知，信息量 AIC_c 值越小，模型越优。 $2K$ 项可理解为模型的待定参数越多，模型本身越复杂，同样性能下该模型就越差。准则函数是对 Kullback-Leibler 信息丢失的估计，故信息量 AIC_c 值越小，模型与真实数据越匹配。但是在实际应用中，讨论某一个模型的准则函数值是没有意义的，只有通过比较一组模型的准则函数值，才能评判他们的优劣^[8]。

下面给出两个使用准则函数判断模型优劣的性能指标^[8,11]

$$\Delta_i = AIC_{c,i} - \min(AIC_c) \quad (25)$$

$$\omega_i = \frac{\exp(-\Delta_i/2)}{\sum_{m=1}^M \exp(-\Delta_m/2)} \quad (26)$$

$AIC_{c,i}$ 是第 i 种模型的信息量 AIC_c 值， M 表示模型的总个数。很明显，一组模型中最优模型的 Δ_i 值是 0， $\Delta_i < 2$ 表示该模型能很好地匹配实验数据，如果 Δ_i 介于 2 与 7 之间表明该模型能够大体反应现实数据的分布， $\Delta_i > 7$ 则表示该模型没有参考价值。权值 ω_i 是更为直观的模型选择指标，它表示模型 i 在所有模型中最优程度的百分比，值越大则表示该模型越优。此外， ω_i/ω_j 这一度量量化的反映模型 i 相对与模型 j 的优劣程度。

4 实验验证及性能分析

实验系统选用 Crossbow 公司的 MIB520 采集板、Micaz 节点和 TinyOS 嵌入式系统，整套实验装置如图 1 所示。测量时，主节点位置固定，通过 MIB520 与电脑相连，实时采集实验数据。从节点沿水泥路或走廊直线摆放，其距主节点距离不断增大，直到无法接收到信号。

实验在教学楼前水泥路和教学楼内走廊进行，两测量环境均在视距范围内。因为印刷电路板制作、元器件焊接及电池电量的差别，不同 Micaz 节点射频信号的发送和接受存在差异。为了抵消上述影响，取主、从节点 RSS 观测各 20 次求均值，即为节点在该测试地点的 RSS 观测值。为了减少多径衰落的

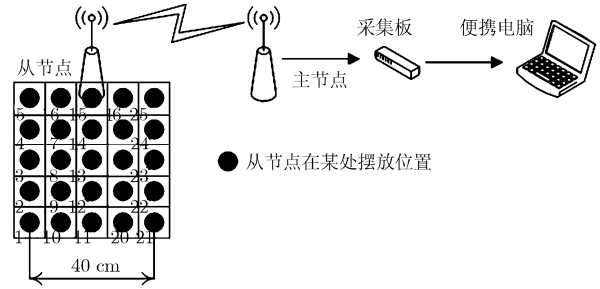


图 1 实验装置

影响，从节点在 40 cm 见方区域内测量 25 组数据，各区域中心点与主节点成一直线，每组数据对应不同发射功率。

4.1 参数估计

log 双斜率模型与 IEEE 802.15.4 的路径损耗模型形式相似，故分别使用水泥路和走廊环境下的观测数据对 log 双斜率模型做参数估计，以验证基于 EM 算法的参数估计的有效性。经多次实验分析，RSS 观测的舍入误差取 $\lambda=1$ 较为合理，故非完全数据 r 的各元素满足 $o^{(j)} - (1/2) \leq r^{(j)} \leq o^{(j)} + (1/2)$ 。随机设定待定参数的初值 $p + \xi_{20} = 3$, $\xi_{21} = 2$, $\xi_{22} = 5$ ，利用式(11)-式(18)进行参数估计，EM 迭代的终止条件为似然函数的增量小于 0.001。设定模型的信号强度变化的拐点在主从节点距离 8 m 处。

图 2 是使用任意两组 RSS 观测数据 a 与 b 时取 30 步 EM 迭代得到的似然函数值变化曲线，从图中可以看出当迭代次数小于 12 时，似然函数的增长比较明显，此时待定参数正在优化，当迭代次数大于 12 后，似然函数几乎没有变化，这表明参数几乎没有进一步优化，即估计值基本上接近真实值。

在水泥路和走廊环境下，表 1 使用基于非完全数据的 EM 迭代算法和基于常规观测数据的极大似然估计(MLE)算法，估计 log 双斜率模型的参数，并与 IEEE 802.15.4 的标准模型参数相对照。不难发现，水泥路环境下 EM 算法得到的模型参数与 IEEE 802.15.4 给出的标准模型参数更为接近，从而

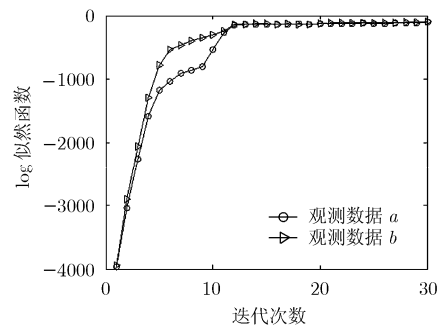


图 2 EM 迭代的似然函数值

表1 log 双斜率模型的参数估计与 IEEE802.15.4 模型参数对照

参数	单位	水泥路		走廊		IEEE 802.15.4
		EM	MLE	EM	MLE	
ξ_{21}	dB/cm	2.143	3.421	1.579	1.632	2.0
ξ_{22}	dB/cm	3.302	2.167	2.415	1.783	3.3
$p+\xi_{20}$	dB	41.179	48.010	48.346	54.348	40.2

可证明本文算法优于 MLE 算法, 参数估计结果更准确。走廊环境下, EM 算法和 MLE 算法估计的参数均与标准模型参数有较大差距, 这是因为走廊内无线信号碰到墙壁发生反射、散射等现象, 而 IEEE 802.15.4 的标准模型是一个适用于室外视距范围的路径损耗模型, 很难反映上述原因造成的多径能量叠加对路径损耗的影响。

4.2 模型选择

本文给出的路径损耗模型选择算法适应于不同环境, 考虑多径能量叠加对路径损耗造成的影响, 水泥路和走廊环境下各模型的参数估计和模型选择结果可能略有差异, 但基本算法是一致的。故后面仅以水泥路环境为例, 讨论模型选择算法的流程, 分析模型选择的性能。

以 EM 算法得到的似然函数值为基础, 表 2 列出了在水泥路环境下使用本文模型选择算法得到的各模型的权重系数, 该结果是综合路径损耗模型与测量数据的拟合程度以及模型简易度两个因素得出的。可见 log 双斜率模型是最优的, 其权值 ω 取值介于 0.77 和 0.78 之间, 经典 log 模型次之, 人体传感器网络模型与双斜率模型则无法与测量数据匹配, 主要原因在于上述两个模型是针对人体无线传感器网络中小尺度非视距范围内的信号传输。

表2 本文算法的模型选择结果

模型	参数个数	待定系数	水泥路	
			Δ	ω
(1)	2	ξ_{10}, ξ_{11}	2.61	0.22
(2)	3	$\xi_{20}, \xi_{21}, \xi_{22}$	0.00	0.78
(3)	2	ξ_{30}, ξ_{31}	21.57	0.00
(4)	4	$\xi_{40}, \xi_{41}, \xi_{42}, \xi_{43}$	11.81	0.00

在确定了最优模型后, 使用 MATLAB 的 Model Browser 工具箱对模型做校正分析。如图 3 所示, 路径损耗与主从节点距离、发射信号强度这两个量密切相关, 主从节点间距离增大, 两节点间能量衰减加大, 且这一衰减速率随着发射能量提升而加大。故在建模过程中需慎重选择发射信号强度及主从节点间距离。

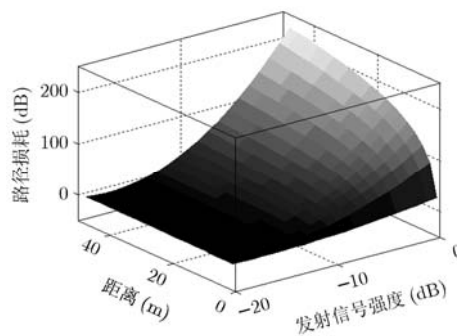


图3 最优模型分析

4.3 算法对比

传统的拟合方法可以较为直观地比较各模型与实验数据的拟合度, 但无法以量化的手段准确判断哪个路径损耗模型更满足应用的要求。传统的 AIC_c , AIC, BIC^[12]等模型选择方法可量化判断模型的优劣, 但都仅考虑观测数据, 而没有考虑观测误差。下面将通过对实验数据的分析, 比对本文模型选择算法与传统方法的性能。

图 4 给出了上述 4 种路径损耗模型与水泥路环境下测量数据的拟合结果, 通过观察可较为直观地看到各路径损耗模型拟合曲线与测量数据的拟合程度, 图中纵坐标表示主从节点间的路径损耗值, 横坐标表示主、从节点间的直线距离。观察图 4 可粗略比较模型的优劣, 比如人体传感器网络模型明显要差, 但其他 3 个模型难以取舍。怎样准确选择适用于水泥路环境的最优模型, 本文给出的路径损耗模型选择算法提供了解决这一问题的有效手段。

表 3 选取水泥路环境下的非完全数据并求各种算法的准则函数值。纵向比较各模型的权重系数 ω 及其他信息量, 可见本文算法给出的最优模型是 log 双斜率模型, 而 AIC_c , AIC, BIC 均选择了经典 log 模型。本文算法的准则函数与 AIC_c 的唯一不同在于计算 $-2\ln l(\theta)$ 项时考虑了非完全数据的影响, 故本文算法选择的最优模型理论上更能反映真实采样数据。

另外, IEEE 802.15.4 的标准路径损耗模型可视为 log 双斜率模型在非视距且无多径能量叠加情况下的一个特例, 通过 4.1 节中表 1 的参数估计结果

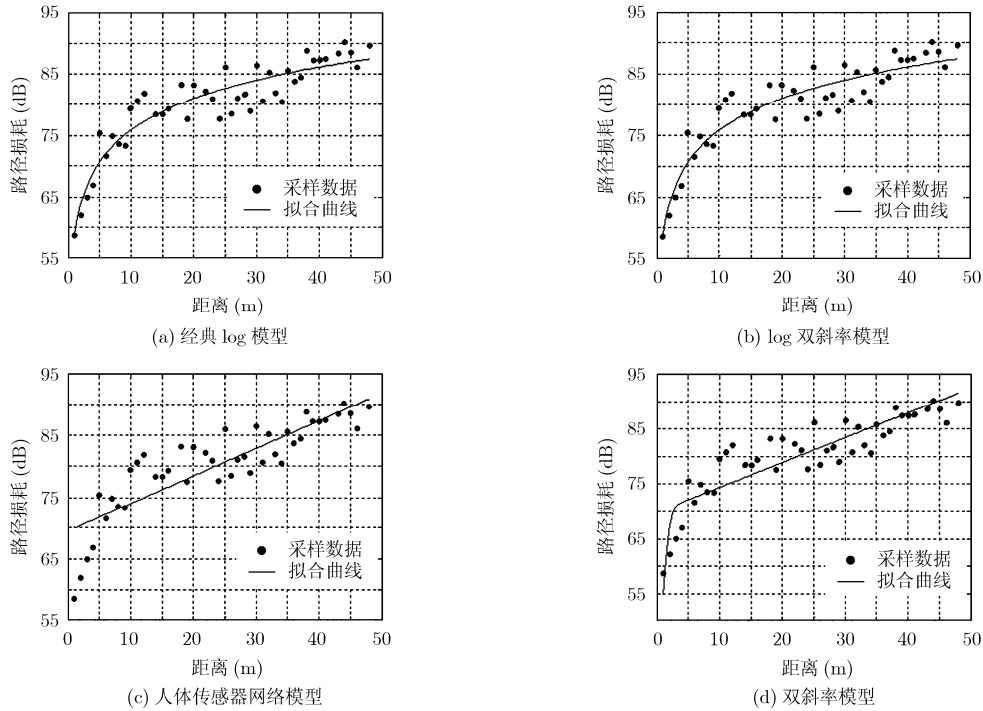


图 4 水泥路环境各模型拟合效果

表 3 文中 4 种模型对应本文权重系数 ω 及 AIC_c , AIC , BIC 等信息量的计算值

模型	ω	AIC_c	AIC	BIC
(1)	0.22	91.15	89.02	90.13
(2)	0.78	96.43	91.90	92.30
(3)	0.00	127.98	110.65	111.14
(4)	0.00	110.70	108.67	109.17

也可发现,在水泥路环境下 log 双斜率模型即为标准模型,性能明显优于经典 log 模型。因此,有理由认为本文算法修正了 AIC_c 准则函数因为观测数据误差造成的误判,选择了最优路径损耗模型。与本文算法相比, AIC , BIC 算法对各模型优劣的区分度不大,且易产生误判。

5 结束语

为了提高路径损耗模型的建模精度及模型选择的准确性,本文提出一种考虑 RSS 舍入误差的模型选择算法。与 AIC_c , AIC , BIC 等模型选择算法相比,本文基于非完全数据的 EM 算法可更为准确地估计模型参数,改善模型筛选的性能。但本文模型选择的准则函数仅对 AIC_c 做了一些延伸,有待对其他准则函数做深入研究,以进一步改进模型选择的性能。该算法同样适用于 RSS 观测存在截断误差或误差服从某一特定概率分布的情况,对路径损耗模型的建模和模型选择有一定指导作用。此外,该算法对其

他研究领域的模型选择算法研究也有积极的借鉴意义。

参考文献

- [1] 刘玉玺, 刘璐, 郑丽娜. 多蜂窝分布式系统下行链路性能分析[J]. 电子与信息学报, 2011, 33(10): 2287-2292.
Liu Y X, Liu J, and Zheng L N. Downlink performance analysis of distributed antenna systems in multi-cell environment[J]. *Journal of Electronics & Information Technology*, 2011, 33(10): 2287-2292.
- [2] 孙晓艳, 李建东, 陈彦辉. 二进制传感器网络加权目标跟踪算法研究[J]. 电子与信息学报, 2010, 32(9): 2053-2057.
Sun X Y, Li J D, and Chen Y H. The study on weighted target tracking algorithm for binary sensor networks[J]. *Journal of Electronics & Information Technology*, 2010, 32(9): 2053-2057.
- [3] Sarkar T K, Zhong J, and Kyung J K. A survey of various propagation models for mobile communication[J]. *IEEE Transactions on Antennas Propagation*, 2003, 45(3): 51-82.
- [4] Benoit L, Bart B, and Ingrid M. A survey on wireless body area networks[J]. *Wireless Networks*, 2011, 17(1): 1-18.
- [5] Hirose K and Kawano S. Bayesian information criterion and selection of the number of factors in factor analysis models[J]. *Journal of Data Science*, 2011, 9(1): 243-259.
- [6] Alam N, Balaie A T, and Dempster A G. Dynamic path loss exponent and distance estimation in a vehicular network using Doppler effect and received signal strength[C].

- Vehicular Technology Conference IEEE 72nd, Ottawa Sept. 6-9, 2010: 1-5.
- [7] Ding H P, Xu Z Y, and Brian M S. A path loss model for non-line-of-sight ultraviolet multiple scattering channels[J]. *EURASIP Journal on Wireless Communications and Networking*, 2010, 10(1): 1-11.
- [8] Fort A and Desset C. Indoor body area channel model for narrowband communications [J]. *IET Microwave Antennas Propagation*, 2007, 1(6): 1197-1203.
- [9] Roos T and Myllymaki P. A statistical modeling approach to location estimation[J]. *IEEE Transactions on Mobile Computing*, 2002, 1(1): 59-69.
- [10] Bates D M and Watts D G. *Nonlinear Regression Analysis and Its Applications*[M]. New York: John Wiley & Sons, 1988: 40-42.
- [11] Burnham K P and Anderson D R. *Model Selection and Multi-model Inference: A Practical Information Theoretic Approach*[M]. New York: Springer Verlag, 2002: 321-323.
- [12] Arlot S and Celisse A. A survey of cross-validation procedures for model selection[J]. *Statistics Surveys*, 2010, 4(10): 40-79.
- 姜向远: 男, 1983年生, 博士生, 研究方向为信号检测与估计、无线测距与定位技术.
- 张焕水: 男, 1963年生, 教授, 博士生导师, 研究方向为无线传感器网络、网络控制系统理论与应用.
- 王伟: 男, 1980年生, 讲师, 研究方向为信道估计和多用户检测、分布式估计.