

基于随机森林的链路质量预测

刘琳岚¹, 高声荣¹, 舒坚²

(1. 南昌航空大学信息工程学院, 江西 南昌 330063; 2. 南昌航空大学软件学院, 江西 南昌 330063)

摘要: 链路质量预测对无线传感器网络的上层协议设计至关重要, 通过链路质量预测方法选择高质量的链路通信, 可以提高数据传输的可靠性和网络通信的效率。基于无监督聚类的高斯混合模型划分链路质量等级, 采用零相位分量分析白化法去除样本间的相关性, 计算信噪比、链路质量指示及接收信号强度指示的均值和方差, 并将其结果作为链路质量参数; 基于随机森林分类算法构建链路质量评估模型, 采用随机森林回归算法构建链路质量预测模型, 预测下一时刻的链路质量等级。在不同的实验场景下, 与指数加权移动平均、三角度量、支持向量回归机和线性回归预测模型相比, 所提预测模型具有更高的预测准确率。

关键词: 无线传感器网络; 链路质量预测; 随机森林; 链路质量等级

中图分类号: TP393

文献标识码: A

doi:10.11959/j.issn.1000-436x.2019025

Link quality prediction based on random forest

LIU Linlan¹, GAO Shengrong¹, SHU Jian²

1. School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China

2. School of Software, Nanchang Hangkong University, Nanchang 330063, China

Abstract: Link quality prediction is vital to the upper layer protocol design of wireless sensor networks. Selecting high quality links with the help of link quality prediction mechanisms can improve data transmission reliability and network communication efficiency. The Gaussian mixture model algorithm based on unsupervised clustering was employed to divide the link quality level. Zero-phase component analysis (ZCA) whitening was applied to remove the correlation between samples. The mean and variance of signal to noise ratio, link quality indicator, and received signal strength indicator were taken as the estimation parameters of link quality, and a link quality estimation model was constructed by using a random forest classification algorithm. The random forest regression algorithm was used to build a link quality prediction model, which predicted the link quality level at the next moment. In different scenarios, comparing with exponentially weighted moving average, triangle metric, support vector regression and linear regression prediction models, the proposed prediction model has higher prediction accuracy.

Key words: wireless sensor network, link quality prediction, random forest, link quality level

收稿日期: 2018-07-30; 修回日期: 2019-03-12

通信作者: 舒坚, shujian@nchu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61762065, No.61363015); 江西省自然科学基金资助项目 (No.20171BAB202009, No.20171BBH80022); 江西省教育厅科学技术重点基金资助项目 (No.GJJ150702); 江西省研究生创新专项资金资助项目 (No.YC2017024)

Foundation Items: The National Natural Science Foundation of China (No.61762065, No.61363015), The Natural Science Foundation of Jiangxi Province (No.20171BAB202009, No.20171BBH80022), The Key Research Foundation of Education Bureau of Jiangxi Province (No.GJJ150702), The Innovation Foundation for Postgraduate Student of Jiangxi Province (No.YC2017024)

1 引言

无线传感器网络 (WSN, wireless sensor network) 是一种由各种具有感知能力、计算能力和通信能力的廉价微型传感器节点通过无线通信的方式以自组织形式构成的网络^[1]。由于被监测区域的复杂性和不确定性^[2], 如果数据分组在低质量链路上进行传输, 当节点之间进行通信时就容易丢失数据, 从而引发消息重传, 虽有重传机制保证数据分组的完整性, 但这易导致传输效率的下降和能耗的增加。通过链路质量预测选择高质量链路进行通信, 不仅可以保障数据的可靠传输, 还可以降低节点的能耗, 延长网络寿命。因此, 降低节点能耗、提高能源利用率的关键在于提高节点间通信效率。

在通信过程中, 准确的链路质量预测是保证上层协议 (数据链路层的 MAC 协议、网络层的路由协议、网络管理层的拓扑控制等) 性能的基础, 对路由协议的设计者而言至关重要^[3]。因此, 建立一个良好的 WSN 链路质量预测机制, 进而动态地调整路由协议, 保证数据正确传输显得十分必要^[4]。本文研究链路质量预测方法, 为上层路由选择提供参考。

2 相关研究

无线传感器网络由于节点移动、多径衰落、噪声和干扰等因素影响导致无线链路动态变化, 实时、准确的链路质量预测面临巨大挑战, 吸引了国内外学者的广泛关注和深入研究。目前, 主要有基于链路特性、基于概率估计理论和基于智能学习的方法。

2.1 基于链路特性的方法

此类方法主要是采用硬件参数接收信号强度指示 (RSSI, received signal strength indicator)、链路质量指示 (LQI, link quality indicator)、信噪比 (SNR, signal to noise ratio) 对链路质量进行预测。文献[5]使用一种帧计数器 meter, 能够在低功率无线传感器网络中实现精确和实时的链路评估, 当存在信道竞争和共存干扰时, 使用分布式轻量级方法统计损坏的数据帧数, 并将其应用到 4-Bit 评估器中, 实验结果表明, 该方法能够提高链路质量评估的准确性和实时性。文献[6]为了解决输电检测系统中对上层通信网络的链路质量问题, 通过分析 WSN 的网络特征, 同时依据跳数、网络环境, 在路由建立阶段选择最优下一跳节点。文献[7]提出了一种简单、

准确、低成本的链路质量估计技术, 适用于资源受限的 WSN 场景, 并通过卡尔曼滤波和模糊逻辑优化了低成本下 RSSI 和 LQI 对链路质量的影响, 最后在锅炉厂的实验场景中, 以 2 种不同的速率实现流式传输, 实验结果表明, 该方法以较小时延为代价实现了无差错传输。文献[8]提出了一种综合性的链路质量度量方法——三角度量 (Triangle), 利用几何学的方法, 结合分组接收率 (PRR, packet reception rate)、LQI 和 SNR 信息来衡量链路质量, 能通过较少的链路探测分组快速且可靠地评估链路质量。

2.2 基于概率估计理论的方法

此类方法主要是通过发送大量的探测分组, 计算接收端的分组接收率。文献[9]基于常用的对数正态路径损耗模型, 将由 SNR 表征的无线链路质量分解为具有不同特性的 2 个部分: 时变非线性部分和非平稳随机部分。通过对这 2 个部分的分别处理, 提出了一种新的链路质量估计方法——WNN-LQE, 获得链路质量的置信区间。文献[10]将指数加权移动平均 (EWMA, exponentially weighted moving average) 数学模型应用到链路质量评估中, 解决数据分组传输速率的不稳定问题, 实验结果表明, 通过在 EWMA 模型中建立相关参数, 可以调节评估结果的灵敏性和稳定性, 以达到特定环境下的要求。文献[11]中的 4-Bit 采用期望传输次数 (ETX, expected transmission count) 作为指标的评估器, 使用 4-Bit 刻画物理层、链路层和网络层的状况, 选择 RSSI、LQI 指标, 同时考虑了链路的非对称性, 实验结果表明, 该评估方法可以在大幅度减少能量消耗的同时提高准确性。然而, 重传数据分组数并未将信道质量与平稳性考虑在内。

2.3 基于智能学习的方法

此类方法主要采用机器学习^[12]和模式匹配等智能学习方法进行建模。文献[13]采用一种基于支持向量机的多分类链路质量评估机制, 选择 RSSI 和 LQI 作为评估参数, 根据 PRR 将链路质量分为 5 个等级, 该模型可以使用较少数量的探测数据分组准确地估计出当前的链路质量。文献[14]使用一种基于模式匹配的方法来预测链路质量变化, 将链路的 SNR 值存储到邻居节点, 以获得 SNR 的时间序列, 当需要对链路的未来状态进行预测时, 节点使用互相关函数查找过去与当前时刻相匹配的 SNR, 但文中只使用了 SNR 来度量链路质量, 并不

能反映出链路的真实情况，且实验只与线性预测模型进行对比，也不能很好地反映出模型的性能。

上述研究为解决链路质量预测问题提供了相关的经验和思路。其中，基于链路特性的方法主要是使用物理层参数预测链路质量，其所需的参数获取简单，可以直接从节点中读取，但节点本身存在校准误差；基于概率估计理论的方法需要发送大量的探测分组，对链路的长期度量较为准确，但对链路的短期变化不敏感；基于智能学习的方法通过数据驱动能够挖掘出数据之间潜在的特征，是未来研究的一个趋势。

本文采用无监督聚类高斯混合模型（GMM, Gaussian mixed model）划分链路质量等级，作为链路质量的评价指标；采用零相位分量分析（ZCA, zero-phase component analysis）白化方法对数据进行预处理，去除样本间的相关性；应用随机森林分类（RFC, random forest classifier）算法评估链路质量等级；通过链路质量的时序信息，应用随机森林回归（RFR, random forest regression）算法预测下一时刻的链路质量等级。

3 链路质量等级的划分

根据不同距离下链路 PRR 分布的不同，将链路的通信范围划分为 3 个不同的区域，分别为连通区、过渡区和不连通区。传统方法中评判链路质量的优劣是通过计算 PRR 值，而不同的文献中划分的等级数不同，划分等级的标准也不同。本文应用 GMM 无监督聚类方法，将 PRR 值划分为 5 个等级，等级划分的标准由 GMM 学习得到。

GMM 根据高斯概率密度函数来判断样本是否属于该高斯模型，每个高斯模型都对应着一种类别，通过样本的输入计算出概率密度，与设定好的阈值相比较，从而确定样本所属的类别。由于 GMM 具有多个模型，划分更为精细，适用于多类别的划分。样本集 $D = \{PRR_1, PRR_2, \dots, PRR_m\}$ 的高斯混合分布和概率密度函数分别为

$$p(PRR_i) = \sum_{j=1}^k \alpha_j N_j(PRR_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \quad (1)$$

$$N_j(PRR_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) = \frac{1}{(2\pi)^{\frac{m}{2}} |\boldsymbol{\Sigma}_j|^{\frac{1}{2}}} e^{-\frac{1}{2}(\text{PRR}_i - \boldsymbol{\mu}_j)^T \boldsymbol{\Sigma}_j^{-1}(\text{PRR}_i - \boldsymbol{\mu}_j)} \quad (2)$$

其中，式(1)为高斯混合分布，该分布由 k 个混合高斯模型组成， $\boldsymbol{\Sigma}_j$ 为协方差矩阵， $\boldsymbol{\mu}_i$ 为第 i 个高斯模

型均值向量， α_j 为混合系数，且 $\sum_{j=1}^k \alpha_j = 1$ ；式(2)为第 j 个高斯模型的概率密度函数。

令 $\theta_j = (\alpha_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ ，GMM 共有 k 个高斯模型，通过样本集 D 来估计 GMM 的所有参数 $\boldsymbol{\Theta} = (\theta_1, \theta_2, \dots, \theta_k)^T$ ，对于给定的样本集 D 采用极大似然估计，即最大化似然函数

$$LL(D) = \sum_{i=1}^m \ln\left(\sum_{j=1}^k \alpha_j p(x_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)\right) \quad (3)$$

似然函数的求解采用最大期望（EM, expectation maximum）算法进行迭代，具体的步骤如下。

步骤 1 初始化高斯混合分布的模型参数 $\{(\alpha_j, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) | 1 \leq j \leq k\}$ 。

步骤 2 E 步：依据当前模型参数，计算 PRR_{*i*} 各混合成分的后验概率，即

$$\gamma_{ji} = E(\alpha_j | PRR_i; \boldsymbol{\Theta}) = \frac{\alpha_j N_j(PRR_i; \boldsymbol{\Theta})}{\sum_{l=1}^k \alpha_l N_l(PRR_i; \boldsymbol{\Theta})} \quad (4)$$

步骤 3 M 步：计算新一轮迭代的模型参数 $\boldsymbol{\mu}'_j$ 、 $\boldsymbol{\Sigma}'_j$ 、 α'_j ，采用式(5)~式(7)分别计算新均值向量、新协方差矩阵和新混合系数为

$$\boldsymbol{\mu}'_j = \frac{\sum_{i=1}^m \gamma_{ij} PRR_i}{\sum_{i=1}^m \gamma_{ij}} \quad (5)$$

$$\boldsymbol{\Sigma}'_j = \frac{\sum_{i=1}^m \gamma_{ij} (PRR_i - \boldsymbol{\mu}'_j)(PRR_i - \boldsymbol{\mu}'_j)^T}{\sum_{i=1}^m \gamma_{ij}} \quad (6)$$

$$\alpha'_j = \frac{\sum_{i=1}^m \gamma_{ij}}{m} \quad (7)$$

步骤 4 重复步骤 2 和步骤 3，直到每一个高斯分布的均值和方差收敛，再将样本划入相应的等级中。

4 链路质量评估

采用 ZCA 白化方法去除冗余样本，采用 Bootstrap 重抽样方法生成不同的数据集；对数据集进行训练，生成多棵链路质量决策树，组合构成 RFC 模型，评估链路质量。

4.1 样本数据预处理

传感器节点收集到的物理层数据（RSSI、LOI、

SNR) 和链路层数据 (PRR) 之间存在着相关性。文献[15]指出, PRR 较高的链路中, 其 RSSI 也较高的概率为 95%, RSSI 与 PRR 是相关的; 文献[16]指出, LQI 与 PRR 具有较强的相关性, 由 LQI 可以估计 PRR 值。由于样本数据各特征之间存在着相关性, 若直接将样本数据用于训练会存在冗余现象, 因此本节采用白化的方法降低样本数据之间的相关性, 经白化处理后的数据特征之间相关性降低, 且所有特征具有相同的方差。

常用的白化处理方法包括主成分分析 (PCA, principal component analysis) 白化和 ZCA 白化。PCA 白化是将数据进行 PCA 处理, 使处理后数据的协方差矩阵为单位矩阵; ZCA 白化是在 PCA 白化的基础上进行一个选择操作, 通过将 PCA 处理后的旋转矩阵左乘特征矩阵, 将数据旋转回去, 保留数据的所有特征, 使 ZCA 白化后的数据更接近原始数据, ZCA 白化的具体步骤如下。

步骤 1 计算数据集的协方差矩阵 Σ 为

$$\Sigma = \frac{1}{m} \sum_{i=1}^m (\mathbf{x}^{(i)})(\mathbf{x}^{(i)})^T \quad (8)$$

其中, m 为数据集样本的个数, $\mathbf{x}^{(i)}$ 为物理层参数特征组成的向量。

步骤 2 求出数据集的协方差矩阵 Σ 后, 计算出协方差矩阵 Σ 的特征向量, 按列排放组成矩阵 U , 即

$$U = \begin{bmatrix} \vdots & \vdots & \vdots \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \quad (9)$$

其中, \mathbf{u}_1 是最大的特征值对应的特征向量, n 是特征的个数, 特征值越大, 代表着该特征包含的信息越多。

步骤 3 将向量 u 组成新基, 得到数据集旋转后的结果 \mathbf{x}_{rot} 为

$$\mathbf{x}_{\text{rot}} = U^T \mathbf{x} = \begin{bmatrix} \mathbf{u}_1^T \mathbf{x} \\ \mathbf{u}_2^T \mathbf{x} \\ \vdots \\ \mathbf{u}_n^T \mathbf{x} \end{bmatrix} \quad (10)$$

其中, $\mathbf{u}_n^T \mathbf{x}$ 就是样本点 x 在特征向量 \mathbf{u}_n 上的投影幅值, 此时的 \mathbf{x}_{rot} 已经消除了输入特征 $\mathbf{x}^{(i)}$ 之间的相关性。

步骤 4 对数据集进行 PCA 白化, 如式(11)所示。经 PCA 白化后, 数据所有特征具有相同的方差, 其中, λ_i 是 \mathbf{x}_{rot} 协方差矩阵中对角元素的值。

$$\mathbf{x}_{\text{PCA},i} = \frac{\mathbf{x}_{\text{rot},i}}{\sqrt{\lambda_i}} \quad (11)$$

步骤 5 将 PCA 处理后的旋转矩阵左乘特征矩阵, 得到 ZCA 白化, 如式(12)所示。ZCA 白化保留了数据的所有特征, 使 ZCA 白化后的数据更加接近原始数据。

$$\mathbf{x}_{\text{ZCA},i} = U \mathbf{x}_{\text{PCA},i} \quad (12)$$

4.2 构建评估模型

上述获得的样本空间中, 应用 ZCA 白化去除了样本之间的相关性, 并用 GMM 无监督聚类算法得到了链路质量等级。由于链路质量是由多个因素共同决定的, 而链路质量等级能够很好地反映出链路的好坏, 通过链路质量等级的确定可以实现链路质量的评估。因此, 链路质量评估是根据多个因素确定链路质量等级的, 其实质就是一个多分类问题。本文利用 RFC 算法在分类问题上的优势, 且不容易出现过拟合的特点, 构建链路质量评估模型。首先通过 Bootstrap 重抽样方法构建不同的数据集, 对每一个数据集分别训练生成链路质量决策树, 然后对决策树的结果进行组合投票产生最优分类, 最终得到链路质量等级。模型将链路物理层的参数组成的向量作为输入, 将评估出来的链路质量等级值作为输出, 输入向量 Input_i 的组成为

$$\text{Input}_i = [\overline{\text{PHY}_i}, \sigma^2(\text{PHY}_i)], \text{PHY}_i \subset (\text{RSSI}, \text{LQI}, \text{SNR}) \quad (13)$$

其中, $\overline{\text{PHY}_i}$ 是物理层参数的均值, $\sigma^2(\text{PHY}_i)$ 是物理层参数的方差。原始训练集 T_k 由 2 类数据构成: 一类为通过 GMM 无监督聚类算法得到的链路质量等级值; 另一类为影响链路质量的 6 种物理层参数, 作为评估模型的输入。利用 Bootstrap 重抽样方法从 T_k 中随机选取训练样本, 经过 w 次抽样后可以构造出不同的训练子集 $T_{k1}, T_{k2}, \dots, T_{kw}$, 并用训练子集分别构建出决策树, 构建过程采用 C4.5 算法, 即在决策树节点划分属性时, 用信息增益比来选择特征。

在生成决策树的过程中, 由于采用 Bootstrap 重抽样方法从原始数据中抽取训练集, 有 63% 的数据会被重复抽取, 而 37% 的数据从未抽取, 未抽取的数据叫作袋外数据 (OOB, out of bag)。用 OOB 来检验决策树的分类效果, 得到的误差为袋外数据误差 (OOBE, out of bag error) [17]。袋外数据误差用于计算决策树的平均误分率, 可以得到 OOB 误

差估计。将 w 棵决策树的误差估计值取平均，可以得到随机森林的泛化误差估计值，OOBE 越小，随机森林算法的分类性能越好。

RFC 通过构造训练集之间的差异增加分类模型间的差异性，从而提高组合模型的泛化能力。通过 w 轮训练，得到 w 棵决策树分类模型序列 $\{h_1(X), h_2(X), \dots, h_w(X)\}$ ，最后组合成一个多分类模型系统。基于 RFC 的评估模型，最终的输出结果采用投票方式产生。

$$H(x) = \arg \max_Y \sum_{i=1}^w I(h_i(x) = Y) \quad (14)$$

其中， $H(x)$ 表示组合分类模型； $h_i(x)$ 表示第 i 棵决策树分类模型； Y 表示输出变量，即链路质量等级值； $I(\cdot)$ 表示示性函数。

5 链路质量预测

采用 RFR 算法对各个实际场景构建模型，建模过程如图 1 所示。

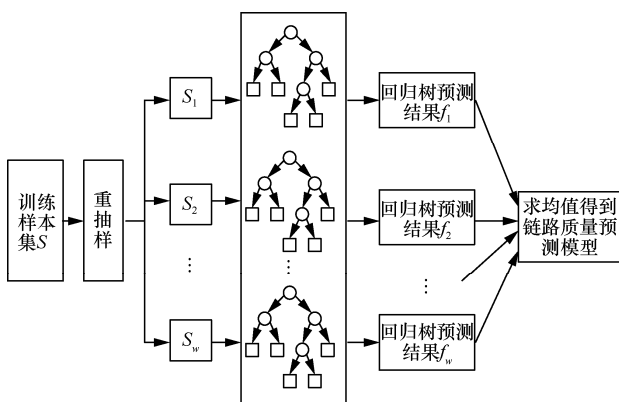


图 1 基于随机森林回归算法的链路质量预测过程

利用 Bootstrap 方法从训练样本集 $S_k (k=1, \dots, n)$ 中进行 w 次抽样，组成 w 个样本子集，每个子集中的元素都不完全相同，对每个样本子集分别构建回归树，并在测试集中对模型进行预测，预测结果需要 w 棵回归树共同决定。

5.1 样本集的构建

RFC 评估模型的输出为链路质量等级值 $lv = \{lv_1, lv_2, \dots, lv_n\}$ ，其中， lv_n 为第 n 个时刻评估模型输出的等级值。利用滑动窗口将 lv 变换成样本集 $S = \{(x_k, y_k)\}$ ，其中， $x_k = (lv_k, \dots, lv_{k+n-2})$ 为预测模型的输入， $y_k = lv_{k+n-1}$ 为标签值， k 为样本集的序号， n 为滑动窗口的大小，窗口的滑动步长为 1。

5.2 回归树生成

回归树的生成就是递归地构建二叉树的过程，对回归树采用平方误差最小化原则，利用分类与回归树 (CART, classification and regression tree) 中的回归树算法生成一棵回归树，共生成 w 棵回归树，从而形成“森林”。为了保证回归树的多样性，降低树之间的相似性，从而保证随机森林模型不容易产生过拟合问题。在回归树的构建过程中，从 M 个影响链路质量等级的特征中随机选取 m 个作为随机特征变量，用来构建回归树， m 在模型中是一个超参数，对最终的结果有一定的影响。随机森林中回归树数量 w 也是模型的一个超参数，超参数之间的组合会对最终的预测效果产生重要的影响。在不同的实验场景中，最优组合也是不同的，根据经验值所构建的随机森林并不适合每个实验场景。因此，本文采用网格搜索算法对超参数进行优化，网格搜索算法将会遍历所有变量可能的取值，计算出每种可能性对应的目标值，从而选出变量最优的组合。回归树生成算法如算法 1 所示。

算法 1 回归树生成算法

输入 训练数据集 D

输出 回归树 $f(x)$

步骤 1 选择最优切分变量 j 与切分点 s ，求解

$$\min_{j,s} \left[\min_{c_1} \sum_{x_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(j,s)} (y_i - c_2)^2 \right] \quad (15)$$

步骤 2 用选定的 (j, s) 划分区域并确定相应的输出值为

$$R_1(j, s) = \{x | x^{(j)} \leq s\}, R_2(j, s) = \{x | x^{(j)} > s\} \quad (16)$$

$$\hat{c}_m = \frac{1}{N_m} \sum_{x_i \in R_m(j,s)} y_i, x \in R_m, m=1,2 \quad (17)$$

步骤 3 继续对 2 个子区域调用步骤 1 和步骤 2，直到子区域不能继续划分，生成叶子节点。

步骤 4 将输入空间划分为 M 个区域 R_1, R_2, \dots, R_M ，生成回归树为

$$f(x) = \sum_{m=1}^M \hat{c}_m I, x \in R_m \quad (18)$$

5.3 预测阶段

当 w 棵回归树模型构建完成后，采用测试集数据对链路质量进行预测。将测试集中的 6 个物理层参数作为链路质量预测模型的输入，分别为 $\{\overline{RSSI}_i, \overline{LQI}_i, \overline{SNR}_i, \sigma^2(\overline{RSSI}_i), \sigma^2(\overline{LQI}_i), \sigma^2(\overline{SNR}_i)\}$ ，得到各回归树模型预测的等级序列。基于 RFR 算法

的预测模型最终输出的链路质量等级值是各回归树模型的均值。

$$Y_k = \frac{1}{w} \sum_{i=1}^w f_i(x_k), k = 1, 2, \dots, n \quad (19)$$

其中, Y_k 为链路质量组合预测模型, $f_i(x_k)$ 为第 i 棵回归树预测模型。

5.4 模型评价

本文采用五折交叉验证法将样本数据集划分为 5 个大小相似的子集, 每个子集在划分的过程中保持数据分布的一致性。在训练过程中, 将 4 个子集的并集作为训练集, 剩下的作为测试集, 进行 5 次训练, 对 5 次预测的结果取均值作为最终的预测结果。采用均方误差 (MSE, mean square error) 评价预测模型的优劣, 即

$$MSE = \frac{1}{N} \sum_{k=1}^N (y_k - lv_k)^2 \quad (20)$$

其中, lv_k 为模型的预测值, y_k 为测试集中的真实值。MSE 越小, 模型的预测值与真实值的差异越小, 模型的性能越好。

6 实验设计与分析

在数据收集过程中, 选用美国 Crossbow 公司的 Telos B 节点, 并采用如图 2 所示的无线传感器网络链路质量测试平台对收集到的数据进行分析, 数据的预处理和模型的构建通过使用服务器上的 Python 平台实现。

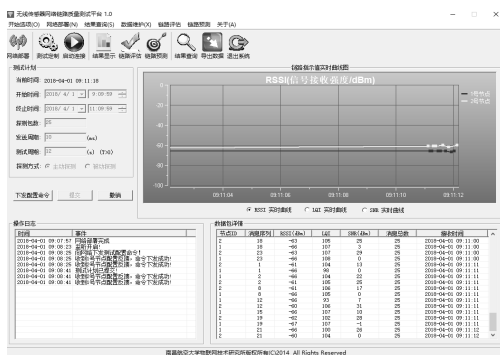


图 2 无线传感器网络链路质量测试平台

6.1 实验参数设置与数据收集

实验场景的设置主要从 WSN 的常用场景出发, 考虑了在真实的环境中可能会遇到的干扰, 例如无线电波、障碍物、邻近信道等。因此本文设定了 4 个实验场景, 有校园广场、校园室内走廊、校

园小树林和校园停车场, 具体的实验场景分别如图 3~图 6 所示, 在每个场景都部署了一个小型的星型 WSN 网络。其中, 校园广场、校园小树林和校园停车场部署了 9 个节点, 一个为 Sink 节点, 8 个为感知节点; 校园室内走廊部署了 5 个节点, 一个为 Sink 节点, 4 个为感知节点; 节点间距均为 10 m。校园广场、校园小树林和校园停车场按照如图 7 所示的星型网络部署实验节点, 校园室内走廊按照如图 8 所示的位置部署实验节点, 实验参数设置如表 1 所示。

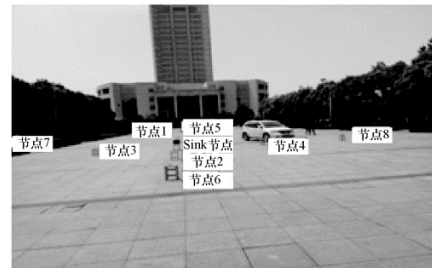


图 3 校园广场实验场景

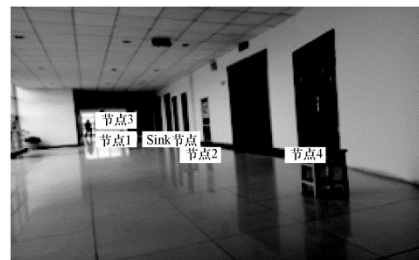


图 4 校园室内走廊实验场景

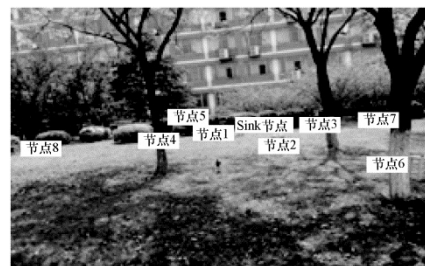


图 5 校园小树林实验场景

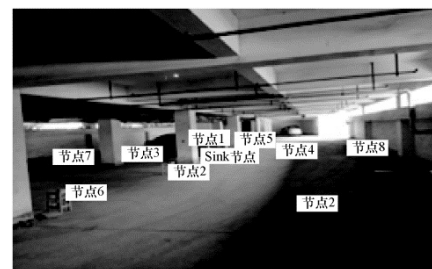


图 6 校园停车场实验场景

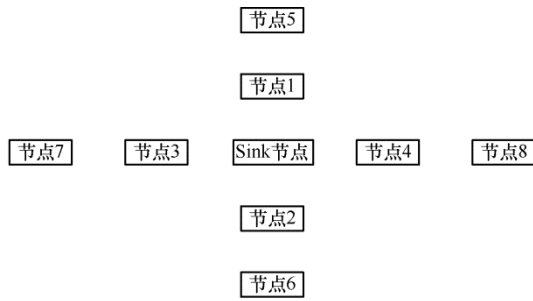


图 7 星型网络部署

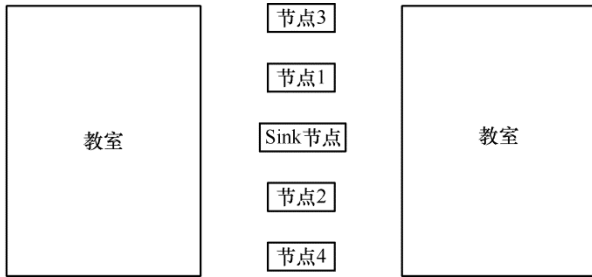


图 8 室内走廊部署

表 1 实验参数设置

参数	参数值
发送功率/dBm	31
信道	26
探测分组数量/个	30
分组发送速率/(个·秒 ⁻¹)	5
发送周期/ms	200
测试周期/s	10

6.2 实验结果分析

为了保证数据的多样性与可靠性，本文通过连续几天的测量获得不同场景下的 PRR，其部分结果分别如图 9~图 12 所示。

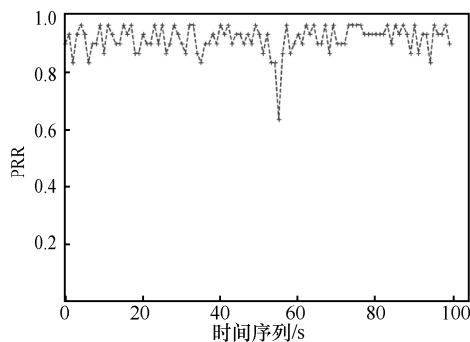


图 9 校园广场节点 1 的 PRR

校园广场场景中比较空旷且干扰源较少，在相同的距离下链路整体表现近似，因此选用节点 1 描述该场景的链路质量。由图 9 可知，校园广场实验

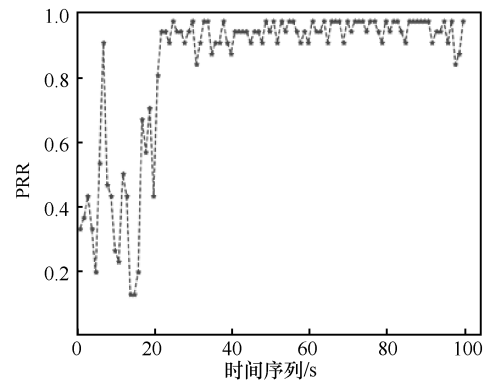


图 10 校园室内走廊节点 2 的 PRR

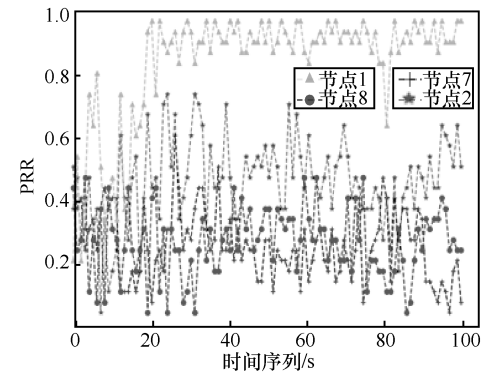


图 11 校园小树林 4 个节点的 PRR

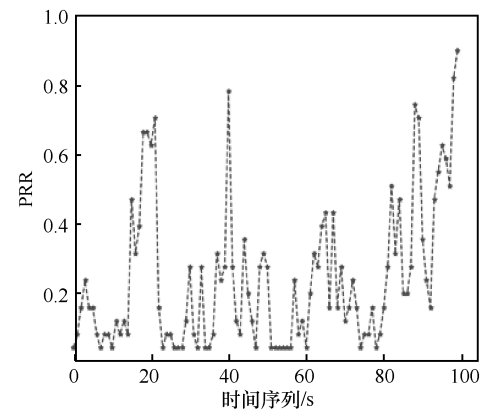


图 12 校园停车场节点 6 的 PRR

场景的 PRR 大部分都在 0.8 以上，链路质量处于较高水平，且稳定性较好，这是由于广场比较空旷且干扰源较少。选用节点 2 描述校园室内走廊场景的链路质量，由图 10 可知，在 0~25 s 链路呈现较高的波动性和突发性，之后链路趋于平稳，这是因为该时段有人员的走动对链路产生了干扰。由图 11 可知，链路质量整体波动较大，节点 7 和节点 8 的 PRR 在 0.1~0.5 之间波动，链路质量处于较差水平，这是因为该场景中的主要干扰源是树木，且分布无规律，容易产生干扰，节点 1 和节点 2 距离 Sink

节点的距离相同，但处于不同的方向，其 PRR 也呈现出不同的波动情况，这是由于不同方向上树木的遮挡情况不同。选用节点 6 描述校园停车场场景的链路质量，由图 12 可知，链路的 PRR 在 0~0.9 之间波动，链路有着较大的波动性和突发性，这是因为该场景中有建筑物和车辆的遮挡，且时常有车辆进出停车场，车辆上移动设备所产生的无线信号会对链路产生干扰。

为了进一步验证 RFR 模型的预测性能，在 4 种实验场景下，将 RFR 模型与 EWMA 模型、Triangle 模型、支持向量回归机 (SVR, support vector regression) 模型和线性回归 (LR, linear regression) 模型进行对比，其中，EWMA 和 Triangle 为传统学习方法，LR 和 SVR 为常见的机器学习方法。各个场景下的预测值与真实值的对比分别如图 13~图 16 所示，从图 13、图 14 和图 16 可知，链路大部分时间都是处于 1、2 等级链路，链路质量较好，5 个模型都能预测出链路的情况，但是当链路发生突变时，RFR 模型能够预测出链路的突变情况，模型的准确性较好；从图 15 可知，此时的链路主要是在 4、5 等级波动，链路质量较差，但 RFR 模型仍能捕捉到链路的变化，模型的准确性好，并能够适应环境的变化。

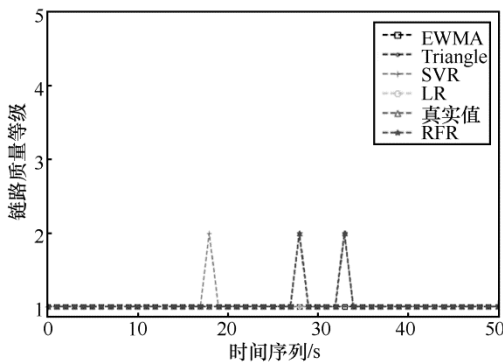


图 13 校园广场的实验预测结果

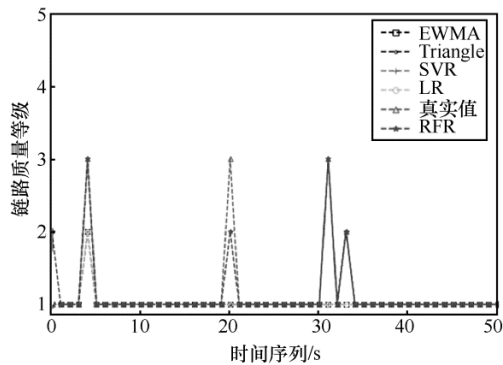


图 14 校园室内走廊的实验预测结果

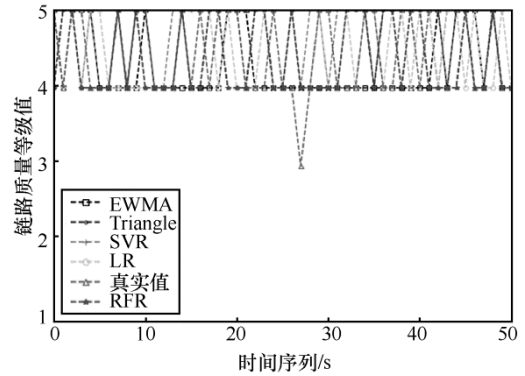


图 15 校园小树林的实验预测结果

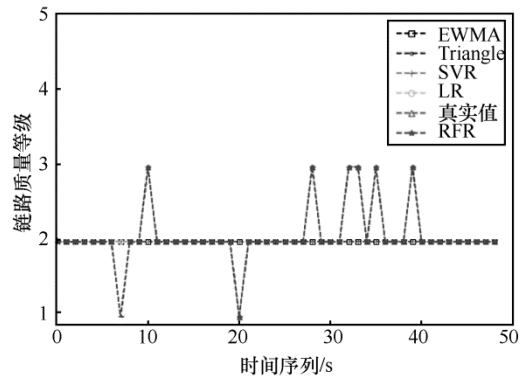


图 16 校园停车场的实验预测结果

为进一步精确对比预测模型的预测效果，计算了预测模型在不同实验场景下的 MSE，其结果如表 2 所示。

实验场景	EWMA	Triangle	SVR	LR	RFR
校园广场	0.193 9	0.155 8	0.089 3	0.104 9	0.019 6
校园室内走廊	0.160 1	0.125 7	0.117 0	0.099 2	0.027 0
校园小树林	0.925 3	0.656 2	0.281 1	0.428 5	0.255 2
校园停车场	0.502 6	0.404 7	0.358 0	0.375 0	0.143 8

由表 2 可知，在 4 种实验场景中，EWMA 模型的 MSE 最大，有着较大的预测误差，Triangle 模型和 EWMA 模型的预测效果均不如机器学习模型。校园停车场、校园室内走廊和校园广场实验场景中，RFR 模型的预测性能均明显优于 EWMA 模型、Triangle 模型、SVR 模型和 LR 模型，其中，校园室内走廊和校园广场是较为稳定的实验场景，Triangle 模型、SVR 模型和 LR 模型均具有较高的预测准确性，但是在校园停车场实验场景中，由于车辆的运动和车载无线设备的干

扰, 链路出现了一定程度的波动, 且呈现出突发性, 导致 SVR 模型和 LR 模型的预测效果均不理想, 但 RFR 模型却能及时地预测出链路的波动情况, 有着更好的抗干扰能力。和其他场景相比, 校园广场和校园室内走廊场景中 RFR 模型的 MSE 均较小, 取得了较好的预测效果, 尤其是在校园广场场景中, RFR 模型的 MSE 最小, 其预测效果优于校园室内走廊场景。这是因为在校园室内走廊中, 会有人员来回走动, 对链路造成干扰, 导致链路不稳定。在校园小树林场景中, 由于受到树木的直接遮挡和多径效应的影响, 链路质量处于较差的水平, 相比其他 3 个实验场景, 各预测模型的预测误差均有增加, 但 RFR 模型的 MSE 最小, 有着更好的预测效果。

在不同的时间段设计了 4 种实验场景, 通过收集不同干扰源下的链路质量测试模型性能。实验结果表明, RFR 预测模型在 4 种场景中表现良好, 均有较高的预测准确性, 体现出模型拥有良好的泛化能力。

在校园广场实验场景中验证模型的预测时间, 5 个模型的平均预测时间如表 3 所示。

表 3 5 个模型的平均预测时间

模型	平均预测时间/s
EWMA	0.002
Triangle	0.004
SVR	0.015
LR	0.005
RFR	0.023

由表 3 可知, EWMA 模型的预测时间最小, 其为传统模型, 计算量最少。RFR 模型通过构造多棵回归树, 保证了模型的预测准确性, 但并未大量增加平均预测时间, 和 SVR 模型处于一个量级上, 就链路质量预测问题而言, 增加的时间处于可接受范围。

7 结束语

本文采用 GMM 划分链路质量等级, 用于评价链路质量; 运用 ZCA 白化去除样本之间的冗余性, 经白化处理后的数据之间相关性较低, 且所有特征具有相同的方差; 基于 RFC 构建链路质量评估模型, 使用 Bootstrap 重抽样方法生成不同的数据集,

训练后生成多棵链路质量决策树, 组合构成 RFC 模型; 通过链路质量的时序信息, 应用 RFR 构建预测模型, 预测下一时刻的链路质量等级, 采用 MSE 评价预测模型的优劣。4 种真实实验场景中的预测结果表明, 与 EWMA 模型、Triangle 模型、SVR 模型和 LR 模型相比, RFR 模型具有较高的预测准确性和较良好的泛化性能。

参考文献:

- [1] 蒋婵, 李陶深, 梁俊斌. 移动低占空比无线传感器网络中低时延的数据持续性提高算法[J]. 通信学报, 2018, 39(3): 53-62.
JIANG C, LI T S, LIANG J B. Low-latency algorithm for improving data persistence in mobile low-duty-cycle wireless sensor network[J]. Journal on Communications, 2018, 39(3): 53-62.
- [2] 张策, 张霞, 李鸥, 等. 不可靠链路下基于压缩感知的 WSN 数据收集算法[J]. 通信学报, 2016, 37(9): 131-141.
ZHANG C, ZHANG X, LI O, et al. Compressive sensing based data gathering algorithm over unreliable links in WSN[J]. Journal on Communications, 2016, 37(9): 131-141.
- [3] NOUHA B, LUCA M, HABIB Y, et al. Radio link quality estimation in wireless sensor networks: a survey[J]. ACM Transactions on Sensor Networks, 2012, 8(4): 688-722.
- [4] TANG Q, WEN H, ZHAO J, et al. Load balanced and efficient data collection protocol for wireless sensor networks[J]. International Journal of High Performance Computing & Networking, 2017, 10(6): 463-473.
- [5] LIU D, CAO Z, ZHANG Y, et al. Achieving accurate and real-time link estimation for low power wireless sensor networks[J]. IEEE/ACM Transactions on Networking, 2017, 25(4): 2096-2109.
- [6] LU J, ZHU Y P, XU Z Q, et al. A reliable wireless sensor network routing method for power transmission line monitoring[J]. Power System Technology, 2017, 41(2): 644-650.
- [7] JAYASRI T, HEMALATHA M. Link quality estimation for adaptive data streaming in WSNs[J]. Wireless Personal Communications, 2017, 94(3): 1543-1562.
- [8] BOANO C A, ZÚÑIGA M A, VOIGT T, et al. The triangle metric: fast link quality estimation for mobile wireless sensor networks[C]//International Conference on Computer Communication and Networks. 2010: 1-7.
- [9] SUN W, LU W, LI Q, et al. WNN-LQE: wavelet-neural-network-based link quality estimation for smart grid WSNs[J]. IEEE Access, 2017, 5: 12788-12797.
- [10] MI X, ZHAO H, ZHU J. Research on EWMA based link quality evaluation algorithm for WSN[C]//Cross Strait Quad-Regional Radio Science and Wireless Technology Conference (CSQRWC). 2011: 757-759.
- [11] FONSECA R, GNAWALI O, JAMIESON K, et al. Four bit wireless link estimation[C]//ACM HotNets. 2007: 1-6.

- [12] GOMES R D, QUEIROZ D V, LIMAFILHO A C, et al. Real-time link quality estimation for industrial wireless sensor networks using dedicated nodes[J]. *Ad Hoc Networks*, 2017, 59: 116-133.
- [13] SHU J, LIU S, LIU L, et al. Research on link quality estimation mechanism for wireless sensor networks based on support vector machine[J]. *Chinese Journal of Electronics*, 2017, 26(2): 377-384.
- [14] FARKAS K, HOSSMANN T, LEGENDRE F, et al. Link quality prediction in mesh networks[J]. *Computer Communications*, 2008, 31(8): 1497-1512.
- [15] SRINIVASAN K, DUTTA P, TAVAKOLI A, et al. Understanding the causes of packet delivery success and failure in dense wireless sensor networks[C]//International Conference on Embedded Networked Sensor Systems. 2006: 419-433.
- [16] BACCOUR N, PUCCINELLI D, VOIGT T, et al. Overview of link quality estimation[M]. Berlin: Springer, 2013.
- [17] 赵腾, 王林童, 张焰, 等. 采用互信息与随机森林算法的用户用电关联因素辨识及用电量预测方法[J]. *中国电机工程学报*, 2016, 36(3): 604-614.
- ZHAO T, WANG L T, ZHANG Y, et al. Relation factor identification of electricity consumption behavior of users and electricity demand forecasting based on mutual information and random forests[J]. *Proceedings of the CSEE*, 2016, 36(3): 604-614.

[作者简介]



刘琳岚（1968- ），女，湖南东安人，南昌航空大学教授、硕士生导师，主要研究方向为无线传感器网络、分布系统等。



高声荣（1994- ），男，江西南康人，南昌航空大学硕士生，主要研究方向为无线传感器网络。



舒坚（1964- ），男，江西南昌人，南昌航空大学教授、硕士生导师，主要研究方向为无线传感器网络、软件工程等。