

基于模糊支持向量回归机的 WSNs 链路质量预测

舒 坚 汤 津 刘琳岚 胡 刚 刘 松

(南昌航空大学软件学院 南昌 330063)

(liulinlan@nchu.edu.cn)

Fuzzy Support Vector Regression-Based Link Quality Prediction Model for Wireless Sensor Networks

Shu Jian, Tang Jin, Liu Linlan, Hu Gang, and Liu Song

(School of Software, Nanchang Hangkong University, Nanchang 330063)

Abstract In wireless sensor networks (WSNs), link is a key element to achieve interconnects and multi-hop communication. Link quality is the fundamental of upper protocols, such as topology control, routing, and mobile management. The effective link quality prediction (LQP) can not only improve networks throughput and decrease node energy consumption, but also prolong network life time. In this paper, we give a concrete analysis about the related works on WSNs link quality prediction. A novel model, fuzzy support vector regression (FSVR), is proposed to predict link quality, which makes the impact of noise and outliers get high accuracy. The link quality samples are collected from three different scenarios. Taking the character of data distribution in unstable links into consideration, a kernel fuzzy c -means (KFCM) algorithm as an unsupervised learning algorithm, is applied to cluster the training set automatically in terms of partition coefficient and exponential separation (PCAES). The membership degree of samples is obtained to get fuzzy set for FSVR. The chaos particle swarm optimization (CPSO) algorithm is employed on each cluster in order to choose the suitable parameter combination for the model. The experimental results show that compared with the empirical risk-based BP neural network prediction methods, the proposed prediction model achieves higher accuracy and better generalization ability.

Key words wireless sensor networks (WSNs); link quality prediction (LQP); support vector regression (SVR); kernel fuzzy c -means (KFCM); chaos particle swarm optimization (CPSO)

摘 要 在无线传感器网络中,链路是实现节点互连和多跳通信的基本元素,链路质量是拓扑控制、路由协议和移动管理的基础,准确的链路质量预测不仅可以提高整个网络的数据吞吐率,降低节点能耗,还可延长整个网络的工作时间.在分析现有链路质量预测方法的基础上,提出一种基于模糊支持向量回归机(fuzzy support vector regression, FSVR)的链路质量预测模型,以降低噪声与孤立点对预测性能的影响.通过收集不同场景下的链路质量样本,考虑不稳定链路中数据分布的特点,该模型采用无监督模糊核聚类算法(kernel fuzzy c -means, KFCM)自动划分样本集,并获得样本隶属度;采用混沌粒子群优化算法(chaos particle swarm optimization, CPSO)选择子模型参数.实验结果表明,与基于经验风险的BP神经网络相比,基于模糊支持向量回归机的链路质量预测模型具有更好的预测精度和泛化能力.

关键词 无线传感器网络;链路质量预测;支持向量回归机;模糊核聚类;混沌粒子群

中图法分类号 TP393

收稿日期:2014-07-23;修回日期:2014-11-26

基金项目:国家自然科学基金项目(61262020,61363015);航空科学基金项目(2012ZC56006)

通信作者:刘琳岚(liulinlan@nchu.edu.cn)

无线传感器网络(wireless sensor networks, WSNs)是一种由大量部署在监控区域的节点以自组织方式形成的网络^[1]。节点间通过协作方式感知、采集和处理感知对象的信息,并将数据发送给基站,在军事、环境监测和智能家居等领域有较为广阔的应用前景。

节点通过无线电射频进行通信,容易受到电磁波自身的多径、损耗和邻居干扰的影响,使得链路具有方向性^[2]、不规则性和非对称性^[3],导致“灰区”的普遍存在,影响了链路的稳定性。链路是实现节点互连和多跳通信的基本元素,较高的链路质量对于建立高效的拓扑控制机制、设计上层路由协议起着关键作用^[4],良好的链路质量预测机制能够有效感知链路变化状况,提高网络通信效率与可靠性。

1 相关研究

近年来,许多学者对链路质量预测方法进行了研究与探索,主要包括:

1) 基于链路特性的预测方法。文献[5]研究信噪比(signal to noise ratio, SNR)的特性,将期望的包成功接收率(packet success rate, PSR)与 SNR 通过 sigmoid 函数进行曲线拟合,提出通过计算 SNR 值与拐点的距离能快速计算当前的链路质量;文献[6]证实了文献[5]的结论,指出链路中 SNR 值波动较大,基于 SNR-PSR 的曲线并不精确,提出对 SNR 值进行卡尔曼滤波,通过预测和校准的迭代过程减少误差;文献[7]分析码片错误率,得到 SNR-PRR(packet receive rate, PRR)对应的曲线,由于无法获知丢包下物理层参数的特性且硬件中存在校准误差^[8],因此较难得到准确的映射关系。

2) 基于概率估计的预测方法。文献[9]指出应当动态获取链路的信息,采用滑动窗口加权指数平均(window mean with exponential weighted moving average, WMEWMA)机制计算 30 m 内的包接收率,并对其进行平滑处理,分析 PRR 会造成延时且开销较大;文献[10]利用中等质量链路短暂的稳定性传输数据,指出节点间包接收率具有一定相关性^[11-13],实验表明根据 3~6 个数据包就能判断下一时刻的链路质量。

3) 基于智能学习理论的预测方法。文献[14]研究链路质量与网络模型问题,提出基于 SNR 模式匹配的交叉相关预测机制(cross-correlation to predict, XCoPred),由于链路质量由多种因素决定,所以该

机制对链路质量的描述不够全面;文献[15]提出一种基于监督学习理论的链路质量预测方法,基于决策树对样本进行离线训练,实验结果证明机器学习在预测中有较大的优势,该方法中物理层度量对硬件较为依赖^[16],尤其当 RSSI 值接近无线芯片接收阈值时,容易出现长尾现象,进而导致欠学习;文献[17]提出一种 4C 链路质量预测方法,综合考虑物理层与链路层的参数^[18],采用 Logistic 回归方法预测包接收率;文献[19]利用模糊逻辑思想,设计出反映链路波动性、突发性的指标,采用贝叶斯神经网络训练样本,实现对链路质量的多维评估。上述 2 种分析方法均以经验风险最小化为目标,建立在大量的训练样本基础上,易造成较大延时与过拟合。

支持向量机(support vector machine, SVM)是 Vapnik 等人提出的一种基于 VC 维理论和结构风险最小化原则的学习机器,能在一定程度上解决“维数灾难”和“过拟合”等问题,在解决非线性、高维数等问题中有不可比拟的优势。目前,SVM 已经应用于链路质量预测^[20]、手写字体识别^[21]、语音识别^[22]、遥感图像分析^[23]等领域,SVM 参数的选取成为制约其应用的主要因素,常用的参数选取方法,如网格搜索法存在着一定的盲目性,遗传寻优法则较难获得全局最优解。本文采用混沌粒子群算法(chaos particle swarm optimization, CPSO)^[24]优化参数,该算法结合混沌特性,模拟鸟类觅食过程,综合个体与群体信息寻找最优解,具有参数少、收敛速度快、搜索范围广、计算复杂度低等特点。

尽管 SVM 有出色的学习性能和泛化能力,但是对噪声和孤立点过分敏感的缺点容易引起过拟合,因此,文献[25]提出模糊支持向量机(fuzzy support vector machine, FSVM),通过构建隶属度函数,使不同样本对建立目标函数的贡献程度不同,从而提高模型的精度,其中隶属度函数的设计影响最优超平面的确定,是 FSVM 的关键问题;文献[26]将基于时间度量的模糊集与支持向量回归机结合,利用模糊支持向量回归机(fuzzy support vector regression, FSVR)的非线性逼近能力以预测项目完成的时间,而 WSNs 链路质量仅在短期内具有一定的时间相关性^[6],以此确定隶属度的方式并不适用。聚类是数据挖掘、模式识别、机器学习等领域的主要研究内容之一,在识别数据的分布结构方面具有极其重要的作用,基于核函数的模糊核聚类算法^[27]利用非线性变换将样本映射到高维特征空间,根据样本间的紧密度进行模糊聚类(fuzzy c-means,

FCM),对噪声和孤立点样本有较强鲁棒性.此外,由于链路质量的时空性^[6],通常将其通信区域划分成几个不同等级,单一的回归模型不能满足预测的需要,本文采用基于样本空间的模糊核聚类算法(kernel fuzzy c-means, KFCM)划分链路、确定隶属度.

综上所述,基于机器学习的 WSNs 链路质量预测方法在性能上优于传统的统计学方法,而 SVM 能在小样本下达到较好的泛化性.本文综合物理层和链路层参数,采用 KFCM 对链路质量划分等级,针对不同等级基于 CPSO-FSVR 建立相应的子模型,获得较准确的链路质量预测模型.

2 基于 CPSO 的 FSVR 预测模型

2.1 模糊支持向量回归机

支持向量机中,给予所有样本固定的惩罚因子 C ,但不同样本对超平面的贡献不同,因此模糊隶属度 u 的引入可以区分噪声、孤立点和有效样本,减小误差,提高预测精度.

对于训练集 $T' = \{(\mathbf{x}_1, y_1, u_1), (\mathbf{x}_2, y_2, u_2), \dots, (\mathbf{x}_l, y_l, u_l)\}$, $0 \leq u_i \leq 1$, $\mathbf{x}_i \in \mathbb{R}^n$ 为输入样本, $y_i \in \mathbb{R}$ 为输出值. FSVR 的回归问题等同于求解以下二次规划问题:

$$\min \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l u_i (\xi_i + \xi_i^*) \right\}, \quad (1)$$

$$\text{s. t. } (\mathbf{w} \cdot \mathbf{x}_i) + b - y_i \leq \epsilon + \xi_i, i = 1, \dots, l, \quad (2)$$

$$y_i - (\mathbf{w} \cdot \mathbf{x}_i) - b \leq \epsilon + \xi_i^*, i = 1, \dots, l, \quad (3)$$

$$\xi_i > 0, \xi_i^* > 0, i = 1, \dots, l, \quad (4)$$

其中, C 为惩罚因子, Cu_i 值越小,在优化问题中起的作用越小; ξ_i, ξ_i^* 为松弛变量.引入 Lagrange 乘子 $\alpha_i^*, \alpha_i, \eta_i^*, \eta_i$ 求解上述问题,建立 Lagrange 方程如下:

$$\begin{aligned} L = & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l u_i (\xi_i + \xi_i^*) - \\ & \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) - \\ & \sum_{i=1}^l \alpha_i (\epsilon + \xi_i + y_i - (\mathbf{w} \cdot \mathbf{x}_i) - b) - \\ & \sum_{i=1}^l \alpha_i^* (\epsilon + \xi_i^* - y_i + (\mathbf{w} \cdot \mathbf{x}_i) + b), \end{aligned} \quad (5)$$

对 L 进行偏导置 0,构造式(1)的对偶问题:

$$\max \left\{ -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(\mathbf{x}_i \cdot \mathbf{x}_j) \right.$$

$$\left. - \epsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) + \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \right\}, \quad (6)$$

$$\text{s. t. } \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0, 0 \leq \alpha_i^* \leq C \cdot u_i. \quad (7)$$

求解式(6)(7),得 α_i^*, α_i ,由此计算 $\mathbf{w}$ 和偏置量 b ,构造出决策函数式(8):

$$y = g(\mathbf{x}) = \sum_{i=1}^l (\alpha_i^* - \alpha_i) K(\mathbf{x}_i, \mathbf{x}) + b, \quad (8)$$

其中, K 为核函数,通过表达式 $K(\mathbf{x}_i, \mathbf{x}) = (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}))$ 可将样本映射至高维空间,解决样本的线性不可分问题,降低计算复杂度.常见的核函数包括径向基核函数(radial basis function, RBF)、多项式核函数(Poly)与感知器核函数(sigmoid).本文选用 RBF 核函数,其插值能力较强,善于处理维数较高的样本,具有收敛域宽、稳定性高、参数少等特点.

2.2 混沌粒子群优化算法

FSVR 中待定参数包括:惩罚因子 C 、不敏感性参数 ϵ 、RBF 中的核宽度 σ .其中,惩罚因子 C 用于权衡经验风险和机器学习的复杂度, C 过大或过小易造成过学习或欠学习; ϵ 指示经验风险的大小; σ 确定核函数的结构,决定训练模型的复杂性,当 σ 很小时,样本线性可分,但容易产生过拟合,泛化性降低,随着 σ 增大,越多的样本位于间隔内或者错分.因此,参数组合的优劣直接影响预测性能.

粒子群优化算法(particle swarm optimization, PSO)在进行参数选取时,将随机初始化粒子群 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, $\mathbf{x}_i = (C_i, \epsilon_i, \sigma_i)$,每个粒子代表一组潜在解,粒子间相互合作,以式(9)(10)在 d 维空间中不断更新全局和个体的位置与速度,最后确定的位置即为最优参数组合:

$$v_{id} = \mathbf{w} \times v_{id} + c_1 \times \text{rand}() (p_{id} - \mathbf{x}_{id}) + c_2 \times \text{rand}() (p_{gd} - \mathbf{x}_{id}), \quad (9)$$

$$\mathbf{x}_{id} = \mathbf{x}_{id} + v_{id}, \quad (10)$$

其中, v_{id} , p_{id} , p_{gd} 分别代表粒子的速度、个体最优位置和全局最优位置; c_1, c_2 为加速因子; $\text{rand}()$ 随机生成 $(0, 1)$ 之间的数; $\mathbf{w}$ 为惯性权重,用于控制历史速度对当前速度的影响,进而控制粒子的全局与局部搜索能力.惯性权重大利于全局搜索,惯性权重小则利于局部搜索,合适的惯性权重将会在全局与局部探索中取得平衡,减少寻优过程中的迭代次数.式(9)中 $\mathbf{w} \times v_{id}$ 为粒子的记忆能力, $c_1 \times \text{rand}() (p_{id} - \mathbf{x}_{id})$ 反映个体的思考能力, $c_2 \times \text{rand}() (p_{gd} - \mathbf{x}_{id})$ 为粒子群的社会认知能力.式(10)依据当前速度和位置更新粒子的搜索路径.

当个体极值信息占优势时,PSO 会陷入局部最优解.为提高粒子的全局搜索能力,本文结合混沌运动的遍历性和随机性特点,采用 CPSO 选取 FSVR 参数.典型的混沌系统方程如下:

$$\mathbf{x}_{n+1} = \gamma \cdot \mathbf{x}_n (1 - \mathbf{x}_n), \quad (11)$$

其中, $\mathbf{x}_n \in (0, 1)$ 且 $\mathbf{x}_n \notin \{0.25, 0.5, 0.75\}$; γ 为控制参数,当 $\gamma=4$ 时,周期分叉结束,进入混沌区,此时可以不重复地经历空间中的所有状态.因此,可以利用混沌扰动防止某些粒子在迭代过程中出现停滞,即对当前最佳适应度的粒子进行 Logistic 方程映射,产生 $\mathbf{x}_0$; 然后根据式(11)进行 m 次迭代生成混沌序列 $\mathbf{x}^k = (\mathbf{x}_1^k, \mathbf{x}_2^k, \dots, \mathbf{x}_m^k)$, k 代表粒子群迭代的次数;最后将序列中的最优位置逆映射回原解空间,并随机替换某一粒子继续寻优,使邻域的粒子迅速分离,提高了优化性能.

3 模糊核聚类

模糊核聚类算法通过核函数将原始空间的非线性信息映射到特征空间中,以线性的方式在特征空间中进行算法的分析与计算,从而得到原始空间的聚类划分.

令样本集合 $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ 属于 $\mathbb{R}^s$, s 是样本的维数, n 为样本的个数. KFCM 算法的问题描述如下:

$$\min J_{\text{KFCM}}(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \cdot d_{ij}^2, \quad (12)$$

$$\text{s. t.} \quad \sum_{i=1}^c \sum_{j=1}^n u_{ij} = n, \quad (13)$$

$$\sum_{j=1}^n u_{ij} > 0, 1 \leq i \leq c, \quad (14)$$

其中,目标函数 J 为关于 $(\mathbf{U}, \mathbf{V})$ 的凸优化问题; $m > 1$ 为模糊系数,控制类间的分享程度; $\mathbf{U}$ 为模糊矩阵, u_{ij} 为 j 样本属于 i 类的隶属度; $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_c]$ 是由聚类中心向量构成的矩阵; d 代表样本与聚类中心的距离, c 为聚类数目. 利用欧氏距离计算样本间的紧密度,定义非线性映射函数 $\Phi: \mathbf{x} \rightarrow \Phi(\mathbf{x}) \in F$, $\mathbf{x} \in X$, 则特征空间中的距离为

$$\begin{aligned} d_{ij} &= \|\Phi(\mathbf{x}_j) - \Phi(\mathbf{v}_i)\| = \\ &= \sqrt{\|\Phi(\mathbf{x}_j) - \Phi(\mathbf{v}_i)\|^2} = \\ &= \sqrt{2 - 2K(\mathbf{x}_j, \mathbf{v}_i)}, \end{aligned} \quad (15)$$

其中, K 为高斯核函数,求解 J 可得对应的迭代方程如下:

$$u_{ij} = n \cdot \left[\sum_{k=1}^c \sum_{l=1}^n \left(\frac{d_{ij}}{d_{lk}} \right)^{\frac{2}{m-1}} \right]^{-1}, \quad (16)$$

$$\mathbf{v}_i = \frac{\sum_{j=1}^n u_{ij}^m \cdot \mathbf{x}_j}{\sum_{j=1}^n u_{ij}^m}, \quad i = 1, 2, \dots, c. \quad (17)$$

针对 KFCM 中确定聚类数目的问题,聚类有效性指标可用于确定最优的划分. PCAES (partition coefficient and exponential separation)^[28] 能在噪声情况下依旧保持聚类数目不变、可靠性高;文献[29]指出对于处理高维的数据,PCAES 识别性能更优. 本文综合考虑类内紧密性、类间分离性、可靠性和计算复杂度,采用 PCAES 作出分类决策,其函数定义如式(18)所示:

$$\begin{aligned} V_{\text{PCAES}} &= \sum_{i=1}^c \sum_{j=1}^n u_{ij}^2 / u_M - \\ &= \sum_{i=1}^c \exp(-\min_{k \neq i} \{\|\mathbf{v}_i - \mathbf{v}_k\|^2 \cdot \beta_T\}), \end{aligned} \quad (18)$$

其中, $u_M = \min_{1 \leq i \leq c} \left\{ \sum_{j=1}^n u_{ij}^2 \right\}$, $\beta_T = \frac{\sum_{i=1}^c \|\mathbf{v}_i - \bar{\mathbf{v}}\|^2}{c}$, $\bar{\mathbf{v}} = \sum_{j=1}^n \mathbf{x}_j / n$, V_{PCAES} 的最大值对应最好的类别划分. 因此,通过求解函数值能够获得最佳的聚类数目. KFCM 算法描述如下:

算法 1. KFCM.

Step1. 初始化聚类数目 $i=2$, 设定最大聚类数目 $c_{\max}$ 、模糊系数 m 、最大迭代次数 $k_{\max}$ 、收敛精度 ϵ ;
Step2. 初始化聚类中心 $\mathbf{V}_1, \mathbf{U}_1$;
Step3. 更新聚类数目 $i=i+1, 2 \leq i \leq c_{\max}$, 令迭代次数 $k=1$, 初始化聚类中心 $\mathbf{V}_i^1$;
Step3. 1. 根据式(16)计算模糊矩阵 $\mathbf{U}_i^k$, 令 $k=k+1$;
Step3. 2. 根据式(17)计算聚类中心 $\mathbf{V}_i^k$;
Step3. 3. 重复 Step3. 1, Step3. 2, 直到满足 $\|\mathbf{V}_i^k - \mathbf{V}_i^{k-1}\| \leq \epsilon$ 或迭代次数等于 $k_{\max}$;
Step4. 根据式(18)计算 V_{PCAES} , 重复步骤 Step3, 直到找到 V_{PCAES} 的最大值.

4 链路质量预测模型

4.1 参数选择

国内外学者对物理层参数进行了大量的研究^[30-32], 指出基于硬件的预测方法能够从节点中直接获取信息, 快速反映当前的链路质量而无需额外

的计算,但是硬件校准会影响信号的测量^[8];基于概率估计机制的信息量太少且链路变化频繁,以此建立的模型适用性不高.因此,本文综合考虑物理层与链路层信息,采用均值 RSSI、均值 LQI、均值 SNR 作为特征参数,以 PRR 为输出变量.

4.2 链路划分

WSNs 的链路不稳定性普遍存在,在不稳定链路中,链路质量发生变化的概率高达 90%^[12].实时性要求高的 WSNs 需要节点快速的传播数据,而分析不稳定链路能够挖掘潜在的链路,减少节点的能量损耗.因此,分析不稳定链路成为近几年研究的热点^[8,12,18],根据链路中呈现的不同特性,将链路通信范围划分为连接区、过渡区、非连接区,并划分相应的链路等级.但是,这种以确定的 PRR 阈值评价链路质量的方法缺乏一定的理论依据,如文献[4]与文献[12]定义链路质量等级的标准不一致,且链路特性随场景的不同而异,特定的阈值不具适用性.因此,对中等链路质量进行详细分析具有较大的实际意义.本文根据不同场景中的样本,从数据的紧密性、相似性出发,利用无监督算法 KFCM 将样本自动划分成不同类别,继而对样本子空间采用不同的建模方式,由此加强对不稳定链路的分析.

4.3 预测模型建立

实际场景中,受无线电波多径、损耗、邻居或其他无线设备的干扰,易出现噪声、孤立点,为减少其对预测性能的影响,本文采用 FSVR 建立预测模型,通过对不同样本赋予对应的隶属度降低其对超平面的作用.链路质量预测流程如图 1 所示:

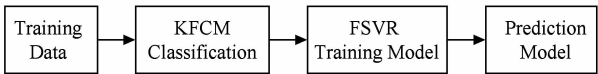


Fig. 1 Flow diagram of predictor.

图 1 预测流程示意图

对于训练样本 $t_i=(x_i,y_i)$, $x_i=(\overline{RSSI_i},\overline{SNR_i},\overline{LQI_i})$, $y_i=(PRR_i)$,采用 KFCM 算法结合聚类有效性指标 PCAES 细粒度化样本空间,得到样本间关联度相对较高的子类($cluster_1,cluster_2,\cdots,cluster_n$),构建新的训练样本 $t_i=(u_i,x_i,y_i)$,其中 u_i 为 KFCM 中的模糊隶属度,该值大小与样本点距离聚类中心远近呈正相关关系.因此,进行 FSVR 建模时,可利用 $C\cdot u_i$ 平衡超平面的间隔大小与错分样本数目, $0<a_i<C\cdot u_i$ 为支持向量,则 $C\cdot u_i$ 取值较小的样本被错分的可能性大,对超平面影响小.

本文采用 CPSO 算法对每个子模型的 FSVR 参数进行优化.对应粒子 $x_i=(C_i,\epsilon_i,\sigma_i)$,选取 $c_1=c_2=2$,以均值误差作为适应度函数,设惯性权重为随时间线性下降的函数,以保证在算法初期有较好的全局搜索,后期能在局部区域寻优.函数如式(19)所示:

$$w=w_{\max}-\frac{w_{\max}-w_{\min}}{iter_{\max}}\times k.$$
 (19)

实验中,利用混沌初始化粒子,随机生成 x_0 ,根据式(11)迭代, $n=30$,选择 $m=20$ 个粒子作为初始值,以此扩大搜索范围,降低算法对初始值的敏感性,采用混沌扰动帮助粒子跳出局部最优解,以搜索较优的 FSVR 参数组合(C,ϵ,σ),构建决策方程,并建立预测子模型,其流程如图 2 所示:

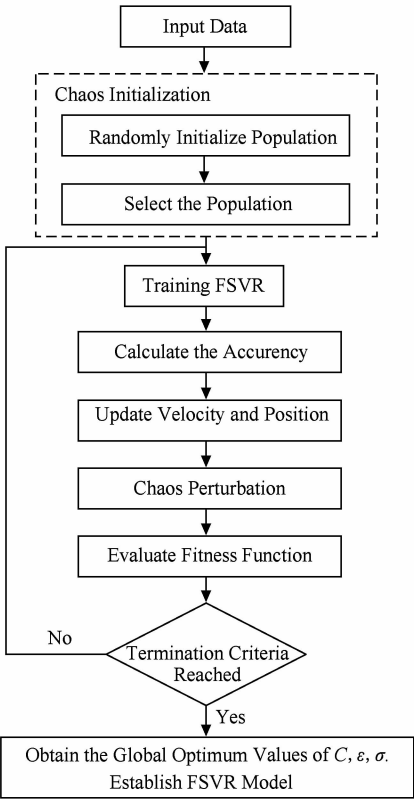


Fig. 2 Flow chart of FSVR algorithm.

图 2 FSVR 算法流程图

当接收到新的样本时,根据聚类中心向量计算与样本相似度最高的聚类,并利用该聚类空间中的 FSVR 模型对样本回归拟合,得到当前的链路质量.该预测模型通过在原数据集中发现数据的紧密性等结构信息,避免了人为划分等级产生的误差,通过对过渡区链路的细化分析提高了预测精度,考虑了噪声和孤立点对有效样本的影响与参数选择,使其在不同的场景具有较好的适用性.

5 实验设计与分析

实验选用美国 Crossbow 公司的 TelosB 节点,该系列节点遵循 IEEE802.15.4 协议,通信范围大约 100 m. 通过在 Tiny OS 系统下运行 NesC 程序以提取样本的各维数据,基于无线传感器网络可视化分析平台(wireless sensor networks visual data analysis platform, WSNs-VDAP)分析不同实验场景下的链路质量,在 MATLAB 仿真平台上实现基于 FSVR 的链路质量预测模型. WSNs-VDAP 平台由南昌航空大学物联网技术研究所设计,主要提供了传感数据图形化显示、链路质量数据分析等功能.

5.1 数据收集

实验考虑了环境中噪声、邻近信道的干扰、障碍物和无线电波的多径效应对链路质量的影响,设定实验场景为公路、广场和操场,如图 3~5 所示:



Fig. 3 Road.
图 3 公路实验场景

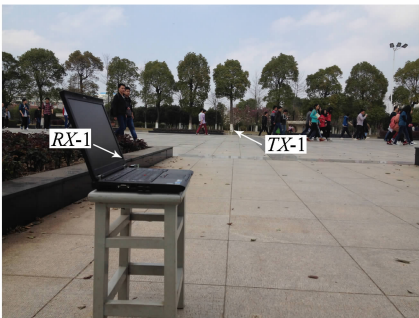


Fig. 4 Square.
图 4 广场实验场景

实验中,分别部署一个发送节点 TX-1,以每秒 10 个包的速率向 Sink 节点 RX-1 发送数据,Sink 记录 RSSI 和 LQI,计算 SNR 值,并将其通过串口发送至 PC(相应的收发节点在图 3~5 中分别以箭头标识). WSNs-VDAP 记录样本的 ID、接收时间,

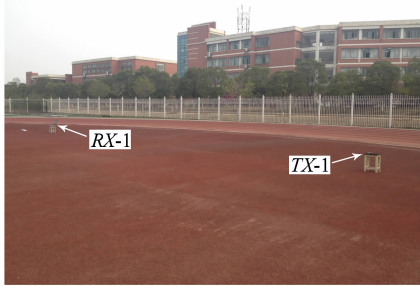


Fig. 5 Playground.
图 5 操场实验场景

计算链路中 RSSI,LQI,SNR 的 100 个包的均值.

公路实验场景中,节点收集上班高峰期链路质量数据.由于障碍物、噪声会使链路无线信号产生反射、折射、衍射等,造成信号的损耗,链路质量呈现较大的突发性;图 4 节点被部署在较为开阔的广场,收集学生上下课时间段的数据.由于受到障碍物与学生携带的无线电子设备的邻近信道干扰,链路的波动较大,链路质量较差;空旷的操场中,链路相对稳定.为了保证数据的可靠性,本文通过连续几天的测量获得不同场景下的 PRR 序列,其部分结果如图 6~8 所示.

图 6 显示较少车辆经过公路时,链路通信状况表现良好,但随着往来车辆数目的改变,噪声与障碍物直接影响 PRR 变化,其范围甚至从最低达到最高,说明链路突发性较高;图 7 显示学生上、下课时,人员密度较大,持续的人员走动对链路影响十分显著,尤其在序号 200~400 过程中,PRR 一直处于较低状态,随着人流量的减少,后期链路质量有上升趋势;图 8 显示尽管操场上偶尔有人员走动,但环境较为空旷,不受其他信号干扰,接收概率达 0.9 以上,

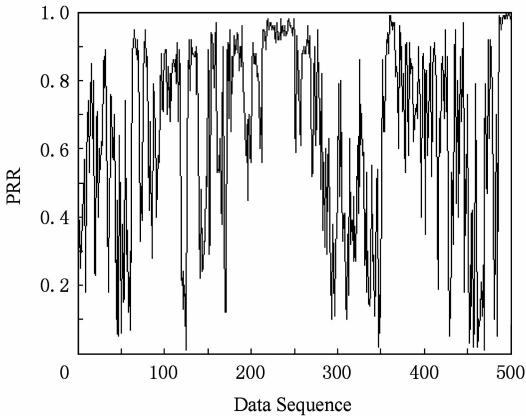


Fig. 6 Link quality in a road environment.
图 6 公路中链路质量

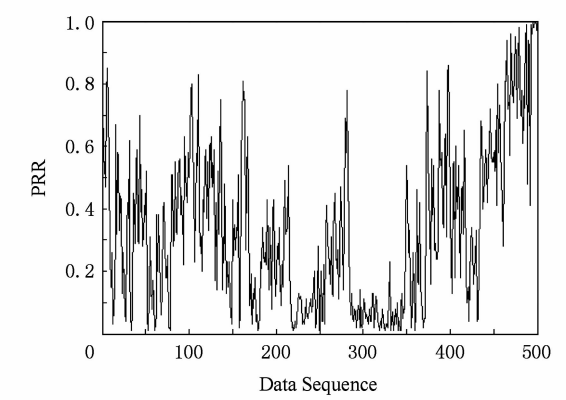


Fig. 7 Link quality in a square environment.
图 7 广场中链路质量

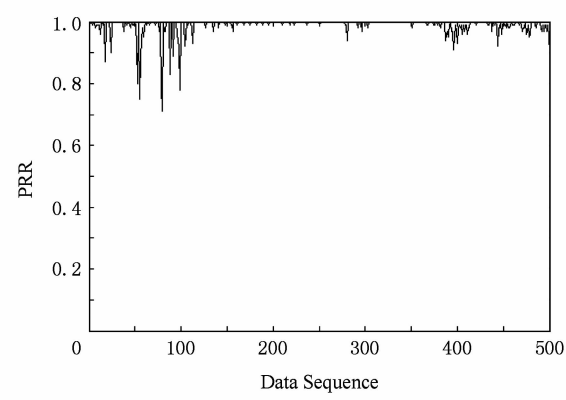


Fig. 8 Link quality in a playground environment.
图 8 操场中链路质量

链路整体表现稳定. 采用统计软件(statistical product and service solutions, SPSS)对测得的样本数据进行统计分析,结果如表 1~3 所示:

Table 1 Distribution of Sample in a Playground Environment
表 1 操场中样本分布情况

Parameters	Mean	Standard Deviation	95% Confidence Interval	
			Lower	Upper
PRR	0.96	0.07	0.95	0.96
RSSI	-70.62	7.58	-71.13	-70.11
SNR	24.34	8.48	23.78	24.91
LQI	105.18	3.08	104.97	105.38

Table 2 Distribution of Sample in a Road Environment
表 2 公路中样本分布情况

Parameters	Mean	Standard Deviation	95% Confidence Interval	
			Lower	Upper
PRR	0.65	0.27	0.63	0.67
RSSI	-84.14	7.25	-84.60	-83.69
SNR	12.08	6.35	11.68	12.47
LQI	97.92	9.39	97.33	98.50

Table 3 Distribution of Sample in a Square Environment
表 3 广场中样本分布情况

Parameters	Mean	Standard Deviation	95% Confidence Interval	
			Lower	Upper
PRR	0.36	0.25	0.34	0.37
RSSI	-89.27	1.95	-89.43	-89.12
SNR	5.34	1.29	5.23	5.44
LQI	81.37	6.93	80.81	81.93

由表 1~3 可知,不同场景下特征参数 RSSI, SNR,LQI 与 PRR 之间均表现出一定的相关性,即随着 PRR 的降低呈现下降的趋势. 广场中 RSSI 均值小于阈值 -87 dBm,通信质量明显降低,但在其他场景中,RSSI 的标准差较大、变化范围广;LQI 对环境的改变相对较为灵敏;SNR 则能很好区分出不同等级的链路质量. 由此说明,综合不同的参数能更好地预测链路质量.

5.2 实验过程

不同场景表现出来的链路特性说明需要对链路进行划分,在粒度更小的集合中进行分析以提高预测精度. 本文采用 KFCM 聚类算法,设置最小聚类数目为 3,最大为 10,由聚类有效性指标确定 8 个子空间,并在该空间中选取 200 个数据作为训练集,分别建立子模型. 然后,从原始样本空间中随机选定某一连续时间段的 50 个样本作为测试集,设置 $C=[0,150],\sigma=[0.01,10],\epsilon=[0,1]$. 目前使用较广的 FSVR 参数优化方法有网格搜索法、遗传算法^[33] (genetic algorithm, GA),本文分别与基于该 2 种方法的 FSVR 和 BP 神经网络方法在不稳定链路、稳定链路中进行性能对比,结果如图 9~10 所示:

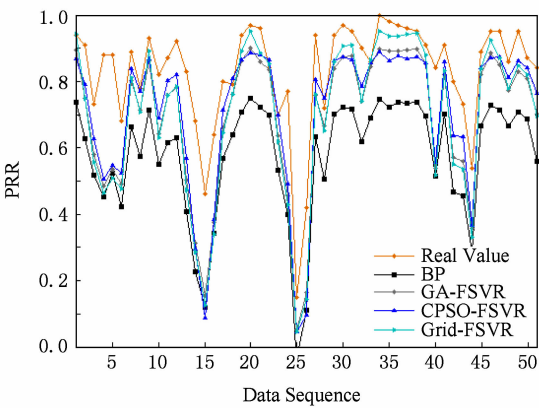


Fig. 9 Performance of the different LQPs in unstable link.
图 9 不稳定链路下性能对比

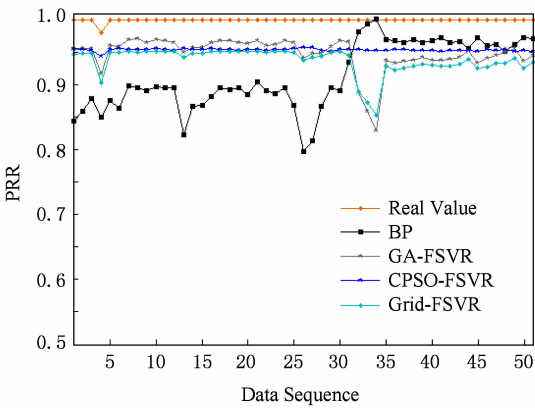


Fig. 10 Performance of the different LQPs in stable link.

图 10 稳定链路下性能对比

在不稳定链路中,链路受环境的影响较大,突变的概率大,与 BP 神经网络预测模型相比,FSVR 能更准确捕获链路的动态变化. BP 神经网络是基于经验风险的机器学习方法,需要大量样本作为训练集才能获得相应的预测精度;而 FSVR 以结构风险最小化原则为基础,具有较好的泛化性,且与模糊集的结合能进一步降低噪声与孤立点的影响,提高预测性能.如图 9 所示,对于样本点 15 和 25,任何一种模型都太过于敏感,本文下一步工作将继续研究链路质量的时间相关性以衡量稳定性与灵敏性.如图 10 所示,稳定链路中,本文采用 KFCM 聚类对样本细粒度化,预测性能高于 BP 神经网络方法.

如图 9~10 所示,无论是在恶劣的通信环境下,还是在受干扰较小的场景中,基于遗传算法模型,其性能高于基于网格搜索算法模型,说明启发式优化算法能更好地解决实际问题.但相较于 CPSO,这 2 种算法的误差均较为明显,尤其在图 10 的样本点 30~35 处,基于 CPSO-FSVR 模型能始终较好感知链路情况,主要归因于 CPSO 算法的搜索范围广、对初始值不敏感,混沌扰动能够避免粒子过早停滞,提高全局搜索能力.由此说明,FSVR 模型中的参数选取直接影响模型的学习能力和预测精度,优于 GA 和网格搜索方法.

本文选用均方根误差衡量预测的性能,结果如表 4~5 所示.

由表 4~5 可知,基于 FSVR 的预测模型的预测准确率均高于 BP 神经网络,其中 CPSO-FSVR 模型的均方根误差最小,可见,基于 CPSO-FSVR 的预测模型能够在小样本下有效预测链路质量,在保证通信可靠性的同时,能减少能量损耗、延长网络寿命.

Table 4 Comparison of Prediction Accuracy in Unstable Link

表 4 不稳定链路的预测准确率对比	
Prediction Model	MSE
Grid-FSVR	0.127 9
GA-FSVR	0.117 0
CPSO-FSVR	0.101 5
BP	0.284 0

Table 5 Comparison of Prediction Accuracy in Stable Link

表 5 稳定链路的预测准确率对比	
Prediction Model	MSE
Grid-FSVR	0.017 2
GA-FSVR	0.013 6
CPSO-FSVR	0.008 5
BP	0.040 3

6 总 结

本文分析了 WSNs 链路质量预测的研究现状,指出当前链路质量预测机制存在的问题,通过综合考虑物理层参数和链路层参数对链路质量的影响,提出了一种基于 FSVR 的 WSNs 链路质量预测模型,该模型考虑了中等链路对通信的影响,并对不同场景下的数据进行划分,引入模糊集弱化噪声与孤立点的影响,设计 CPSO 算法优化模型参数.实验结果表明,本文提出的预测模型具有良好的预测性能.

参 考 文 献

[1] Li Jianzhong, Gao Hong. Survey on sensor network research [J]. Journal of Computer Research and Development, 2008, 45(1): 1-15 (in Chinese)
(李建中, 高宏. 无线传感器网络的研究进展[J]. 计算机研究与发展, 2008, 45(1): 1-15)

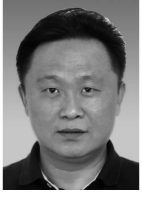
[2] Srinivasan K, Dutta P, Tavakoli A, et al. Understanding the causes of packet delivery success and failure in dense wireless sensor networks [C] //Proc of the 4th Int Conf on Embedded Networked Sensor Systems. New York: ACM, 2006: 419-420

[3] Zhao J, Govindan R. Understanding packet delivery performance in dense wireless sensor networks [C] //Proc of the 1st Int Conf on Embedded Networked Sensor Systems. New York: ACM, 2003: 1-13

[4] Srinivasan K, Levis P. Rssi is under appreciated [C] //Proc of the 3rd Workshop on Embedded Networked Sensors. New York: ACM, 2006: 1-5

- [5] Woo A, Culler D. Evaluation of efficient link reliability estimators for low-power wireless networks, UCB/CSD-03-1270 [R]. Berkeley, CA: University of California, Berkeley, 2003
- [6] Lal D, Manjeshwar A, Herrmann F, et al. Measurement and characterization of link quality metrics in energy constrained wireless sensor networks [C] //Proc of Global Telecommunications Conf. Piscataway, NJ: IEEE, 2003: 446-452
- [7] Zuniga M, Krishnamachari B. Analyzing the transitional region in low power wireless links [C] //Proc of the 1st Annual IEEE Communications Society Conf on Sensor and Ad Hoc Communications and Networks. Piscataway, NJ: IEEE, 2004: 517-526
- [8] Zuniga M, Irzynska I, Hauer J, et al. Link quality ranking: Getting the best out of unreliable links [C] //Proc of Distributed Computing in Sensor Systems and Workshops. Piscataway, NJ: IEEE, 2011: 1-8
- [9] Woo A, Tong T, Culler D. Taming the underlying challenges of reliable multihop routing in sensor networks [C] //Proc of the 1st Int Conf on Embedded Networked Sensor Systems. New York: ACM, 2003: 14-27
- [10] Senel M, Chintalapudi K, Lal D, et al. A kalman filter based link quality estimation scheme for wireless sensor networks [C] //Proc of Global Telecommunications Conf. Piscataway, NJ: IEEE, 2007: 875-880
- [11] Becher A, Landsiedel O, Wehrle K. Towards short-term wireless link quality estimation [C] //Proc of the 30th IEEE Real Time Systems Symp. New York: ACM, 2008
- [12] Srinivasan K, Kazandjieva M A, Agarwal S, et al. The β -factor: Measuring wireless link burstiness [C] //Proc of the 6th ACM Conf on Embedded Network Sensor Systems. New York: ACM, 2008: 29-42
- [13] Cerpa A, Wong J L, Potkonjak M, et al. Temporal properties of low power wireless links: Modeling and implications on multi-hop routing [C] //Proc of the 6th ACM Int Symp on Mobile Ad Hoc Networking and Computing. New York: ACM, 2005: 414-425
- [14] Farkas K, Hossmann T, Legendre F, et al. Link quality prediction in mesh networks [J]. Computer Communications, 2008, 31(8): 1497-1512
- [15] Wang Y, Martonosi M, Peh L S. Predicting link quality using supervised learning in wireless sensor networks [J]. ACM SIGMOBILE Mobile Computing and Communications Review, 2007, 11(3): 71-83
- [16] Renner C, Ernst S, Weyer C, et al. Prediction accuracy of link quality estimators [G] //LNCS 6567: Proc of the 8th European Conf on Wireless Sensor Networks. Berlin: Springer, 2011: 1-16
- [17] Liu T, Cerpa A E. Foresee(4C): Wireless link prediction using link features [C] //Proc of the 10th Int Conf on Information Processing in Sensor Networks. Piscataway, NJ: IEEE, 2011: 294-305
- [18] Cerpa A, Wong J L, Kuang L, et al. Statistical model of lossy links in wireless sensor networks [C] //Proc of the 4th Int Symp on Information Processing in Sensor Networks. Piscataway, NJ: IEEE, 2005: 81-88
- [19] Guo Zhiqiang, Wang Qin, Wan Yadong, et al. A classification prediction mechanism based on comprehensive assessment for wireless link quality [J]. Journal of Computer Research and Development, 2013, 50(6): 1227-1238 (in Chinese)
(郭志强, 王沁, 万亚东, 等. 基于综合性评估的无线链路质量分类预测机制[J]. 计算机研究与发展, 2013, 50(6): 1227-1238)
- [20] Haruki O. A link availability predictor for wireless sensor networks [R]. Palo Alto, California: Stanford University, 2010
- [21] Pal U, Jayadevan R, Sharma N. Handwriting recognition in Indian regional scripts: A survey of offline techniques [J]. ACM Trans on Asian Language Information Processing, 2012, 11(1): 1-35
- [22] Liu X F, Zhang X Y, Guo C. Speech recognition based on Mary support vector machine [J]. Journal of Computational Information System, 2011, 7(10): 3560-3566
- [23] Mountrakis G, Im J, Ogole C. Support vector machines in remote sensing: A review [J]. Journal of Photogrammetry and Remote Sensing, 2011, 66(3): 247-259
- [24] Gandomi A H, Yun G J, Yang X S, et al. Chaos-enhanced accelerated particle swarm optimization [J]. Communications in Nonlinear Science and Numerical Simulation, 2013, 18(2): 327-340
- [25] Liu C F, Wang S D. Fuzzy support vector machines [J]. IEEE Trans on Neural Networks, 2002, 13(2): 464-471
- [26] Cheng M Y, Hoang N D, Roy A F V, et al. A novel time-dependent evolutionary fuzzy SVM inference model for estimating construction project at completion [J]. Engineering Applications of Artificial Intelligence, 2012, 25(4): 744-752
- [27] Wu Z D, Xie W X, Yu J P. Fuzzy C-means clustering algorithm based on kernel method [C] //Proc of the 5th Int Conf on Computational Intelligence and Multimedia Applications. Piscataway, NJ: IEEE, 2003: 49-54
- [28] Saad M F, Alimi A M. Validity index and number of clusters [J]. International Journal of Computer Science Issues, 2012, 9(1): 52-57
- [29] El-Melegy M, Zanaty E A, Abd-Elhafiez W M, et al. On cluster validity indexes in fuzzy and hard clustering algorithms for image segmentation [C] //Proc of 2007 IEEE Int Conf on Image Processing. Piscataway, NJ: IEEE, 2007: VI-5-VI-8

- [30] Boano C A, Zuniga M A, Voigt T, et al. The triangle metric: Fast link quality estimation for mobile wireless sensor networks [C] //Proc of the 19th Int Conf on Computer Communications and Networks. Piscataway, NJ: IEEE, 2010: 1-7
- [31] Liu T, Cerpa A E. Data-driven link quality prediction using link features [J]. ACM Trans on Sensor Networks, 2014, 10(2): 37-72
- [32] Baccour N, Koubaa A, Youssef H, et al. F-LQE: A fuzzy link quality estimator for wireless sensor networks [G] // LNCS 5970: Proc of the 7th European Conf on Wireless Sensor Networks. Berlin: Springer, 2010: 240-255
- [33] Istvan K, Laszlo S, Janos C F, et al. PSO and GA optimization methods comparison on simulation model of a real hexapod robot [C] //Proc of the 9th Int Conf on Computational Cybernetics. Piscataway, NJ: IEEE, 2013: 125-130



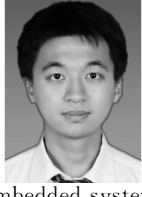
Shu Jian, born in 1964. Received his MSc in computer networks from Northwestern Polytechnical University. Professor in Nanchang Hangkong University. Senior member of China Computer Federation. His main research interests include wireless sensor networks, embedded system and software engineering (shujian@nchu.edu.cn).



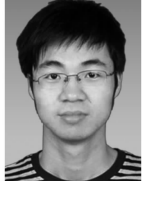
Tang Jin, born in 1989. Undergraduate student in Nanchang Hangkong University. Student member of China Computer Federation. Her main research interests include wireless sensor networks (tang_jin34@163.com).



Liu Linlan, born in 1968. Received his BSc in computer application from the National University of Defence Technology. Professor in Nanchang Hangkong University. Senior member of China Computer Federation. Her main research interests include software engineering and distributed system.



Hu Gang, born in 1991. Undergraduate student in Nanchang Hangkong University. Student member of China Computer Federation. His main research interests include wireless sensor networks and embedded system on programmable chips(hgmaizi927@163.com).



Liu Song, born in 1990. Undergraduate student in Nanchang Hangkong University. Student member of China Computer Federation. His main research interests include wireless sensor networks and embedded system(liusong090451@163.com).