

文章编号:1007-5321(2018)02-0044-06

DOI:10.13190/j.jbupt.2017-024

基于改进最小二乘支持向量回归机的链路质量预测

舒 坚, 贾晨浩, 陶 娟

(南昌航空大学 软件学院, 南昌 330063)

摘要: 为了准确地预测链路质量,提出基于改进最小二乘支持向量回归机的无线传感器网络链路质量预测模型. 采用粗糙集理论约简链路质量参数,以提取出有效反映链路质量的特征参数;利用遗传算法优化最小二乘支持向量回归机的惩罚因子和核函数宽度. 实验结果表明,与 Experts Advice 预测模型相比,提出的预测模型具有更高的精度.

关 键 词: 无线传感器网络; 链路质量预测; 遗传算法; 最小二乘支持向量回归机

中图分类号: TN929.53

文献标志码: A

Link Quality Prediction for Sensor Network Based on Improved LS-SVR

SHU Jian, JIA Chen-hao, TAO Juan

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

Abstract: In order to predict the link quality accurately, a link quality prediction model was proposed to predict link quality for sensor networks based on improved least square support vector regression machine (LS-SVR). The rough set (RS) was introduced to reduce the link quality metrics so as to extract the effective characteristic metrics of the link quality. And the genetic algorithm (GA) was employed in LS-SVR to optimize the penalty factor and kernel width. Experiments show that compared with the experts advice-based prediction model, the proposed prediction model achieves better accuracy.

Key words: wireless sensor networks; link quality prediction; genetic algorithm; least squares support vector regression machine

无线传感器网络 (WSNs, wireless sensor networks) 是由能量受限的节点通过无线自组织方式形成的分布式系统,节点可随机部署在监测区域,应用于军事、医疗、环境监测^[1]等领域. 选择高质量的链路传输数据,可减少节点能量损耗、提高网络数据吞吐率. 有效的链路质量预测模型能够准确地预判链路状况,为路由协议 (CTP、AODV、QHCR^[2] 等) 选择稳定、可靠的链路提供依据,从而提高网络的整体性能.

WSNs 链路质量预测方法主要包括基于通信链路特性、基于概率估计和基于智能学习 3 类方法. 基于通信链路特性的预测方法涉及的链路评估物理层参数包括接收信号强度指标 (RSSI, received signal strength indicator)、链路质量指标 (LQI, link quality indicator)、信噪比 (SNR, signal-to-noise ratio) 等,涉及的数据链路层参数有包接收率 (PRR, packet reception ratio) 等;基于概率估计的预测方法从概率论的角度对包接收率进行预测;基于智能学

收稿日期: 2017-07-21

基金项目: 国家自然科学基金项目 (61762065, 61363015, 61501218); 江西省自然科学基金资助重点项目 (20171BAB202009, 20171ACB20018); 江西省研究生创新专项资金项目 (YC2016-S356)

作者简介: 舒 坚 (1964—), 男, 教授, E-mail: shujian@nchu.edu.cn.

习的预测方法主要有 FLI^[3]、Experts Advice^[4]、FS-VR^[5]、Gillbert-Elliot^[6]等方法。

采用卡尔曼滤波和粗糙集理论(RS, rough set)对链路质量参数集进行预处理;采用遗传算法(GA, genetic algorithm)优化最小二乘支持向量回归机(LS-SVR, least squares support vector regression)的惩罚因子和核函数宽度,得到链路质量预测模型(GA-SVR, least squares support vector regression based on genetic algorithm)。

1 链路质量参数的预处理

为了避免外界因素对链路质量参数的干扰,采用卡尔曼滤波(Kalman filter)方法对原始数据集进行降噪,形成决策表;采用RS对决策表进行约简,得到最小条件属性集,作为链路质量预测的样本集。

1.1 降噪

采用文献[7]中的降噪方法,对实验采集到的物理层参数RSSI、LQI和SNR分别进行降噪处理,其中对室内走廊中原始LQI降噪处理的效果如图1所示。

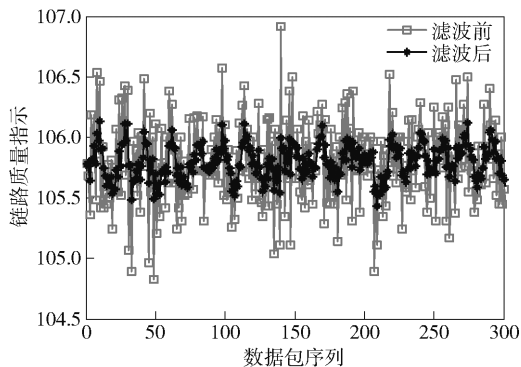


图1 降噪前、后LQI的对比

由图1可知,降噪后的LQI波动范围有明显的减小,降低了噪声对参数的影响。

1.2 知识约简

信息系统 $S = \langle U, A, V, f \rangle$, 其中 U 为有限个对象的集合, $U = \{x_1, x_2, \dots, x_n\}$, n 为实测数据的组数, $x_i = (v_{RSSI}^i, v_{LQI}^i, v_{SNR}^i, v_{PRR}^i)$ ($i = 1, 2, \dots, n$) 为第 i 组对应属性集 A 的实测数据; A 为非空有限属性集合, $A = C \cup D$, $C \cap D = \emptyset$, $A = \{RSSI, LQI, SNR, PRR\}$, 子集 C 为条件属性集合, $C = \{RSSI, LQI, SNR\}$, D 为决策属性集合, $D = \{PRR\}$; $V = \bigcup_{a \in A} V_a$, V_a 为属性 $a \in A$ 的值域; $f: U \times A \rightarrow V$ 为信息函数, 实现对 U 中每个对象的所有属性进行赋值。

基于知识约简的粗计算是粗糙集理论得以实际应用的关键。设 $X \subset U$ 是实体对象上的非空子集, $P \subseteq A$ 是属性集合的子集, 则 X 的下近似集 $\underline{P}(X)$ 、上近似集 $\overline{P}(X)$ 、依赖度 $\gamma_P(X)$ 及边界区域 $Bnd_P(X)$ ^[8] 分别为

$$\underline{P}(X) = \{Y \in U \mid P: Y \subseteq X\} \quad (1)$$

$$\overline{P}(X) = \{Y \in U \mid P: Y \cap X \neq \emptyset\} \quad (2)$$

$$\gamma_P(X) = |\underline{P}(X)| / |X| \quad (3)$$

$$Bnd_P(X) = \overline{P}(X) - \underline{P}(X) \quad (4)$$

其中 $|X|$ 为集合 X 的势, 即集合中所含元素个数。

设室内走廊信息系统 S 上的一个分类为 ψ ^[8], 经约简后的最小条件属性子集 $R = \{RSSI, LQI\}$, 与原始条件属性集 $C = \{RSSI, LQI, SNR\}$ 有相同的分类质量, 即 $\gamma_R(\psi) = \gamma_C(\psi)$ 。类似地, 得到校内树林信息系统中的约简结果为 $\{RSSI, SNR\}$, 校内广场信息系统中的约简结果为 $\{RSSI, SNR\}$ 。

2 链路质量预测

LS-SVR 利用二次损失函数取代传统支持向量机中不敏感的损失函数, 将传统支持向量机中的不等式约束条件改写成等式约束。LS-SVR 具有运算简单、收敛速度快的优点^[9], 采用 LS-SVR^[10] 建立链路质量预测模型。

2.1 LS-SVR

给定约简后的样本集 T 为

$$T = \{(x_1, y_1), \dots, (x_m, y_m)\} \in (R^n \times R) \quad (5)$$

其中: $x_i \in R^n$, $y_i \in R$, $i = 1, 2, \dots, m$ 。

LS-SVR 通过非线性映射 ϕ , 将低维不可分的数据样本集投射至高维特征空间, 构造最优线性回归函数。

$$y(x) = w^T \phi(x) + b \quad (6)$$

其中: w 为权重向量, $\phi(x)$ 为 x 从输入空间(低维空间)投射至高维空间的线性映射函数, b 为偏置项。根据结构风险最小化的原理, 综合考虑函数复杂度和拟合误差, LS-SVR 为

$$\left. \begin{aligned} \min_{w, b, \xi} J(w, \xi) &= \frac{1}{2} w^T w + \frac{g}{2} \sum_{i=1}^m \xi_i^2 \\ \text{s. t. } y_i &= w^T \phi(x_i) + b + \xi_i, i = 1, 2, \dots, m \end{aligned} \right\} \quad (7)$$

其中: g 为惩罚因子, ξ_i 为 x_i 的松弛系数, $\xi = (\xi_1, \xi_2, \dots, \xi_m)$ 。为求得目标函数的最小值, 构造拉格朗日(Lagrange)函数为

$$L(\mathbf{w}, b, \xi, \alpha) = J(\mathbf{w}, \xi) - \sum_{i=1}^m \alpha_i [\mathbf{w}^T \phi(\mathbf{x}_i) + b + \xi_i - y_i] \quad (8)$$

其中: $\alpha = (\alpha_1, \dots, \alpha_m)$. 为求得最优解, 可根据 KKT (Karush-Kuhn-Tucker) 条件^[10]将 \mathbf{w} 和 ξ_i 消去, 则上述求解的优化问题可转化为线性方程

$$\begin{pmatrix} 0 & \mathbf{Q}^T \\ \mathbf{Q} & \mathbf{H} + \mathbf{g}^{-1} \mathbf{I} \end{pmatrix} \begin{pmatrix} b \\ \alpha^T \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{y} \end{pmatrix} \quad (9)$$

其中: $\mathbf{Q} = (1, \dots, 1)^T$, $\mathbf{y} = (y_1, \dots, y_m)^T$, $\mathbf{H}(\mathbf{x}, \mathbf{x}') = \phi^T(\mathbf{x})\phi(\mathbf{x}')$, \mathbf{I} 为单位矩阵, 对式(9)求解, 可得 α 和 b .

基于 LS-SVR 的链路质量预测模型为

$$y(\mathbf{x}) = \sum_{i=1}^m \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b \quad (10)$$

通过核函数 $K(\mathbf{x}, \mathbf{x}_i)$ 将具有高维非线性特点的数据样本映射至高维特征空间中, 在低维空间中求得最优回归模型. 采用径向基核函数作为理想的选择, 决策函数 $y(\mathbf{x}) = \sum_{i=1}^m \alpha_i K(\mathbf{x}, \mathbf{x}_i) + b$ 可表示为

$$y(\mathbf{x}) = \sum_{i=1}^m \alpha_i \exp \left\{ -\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2z^2} \right\} + b \quad (11)$$

其中: z 为核函数宽度. LS-SVR 在二维空间中的拟合效果如图 2 所示.

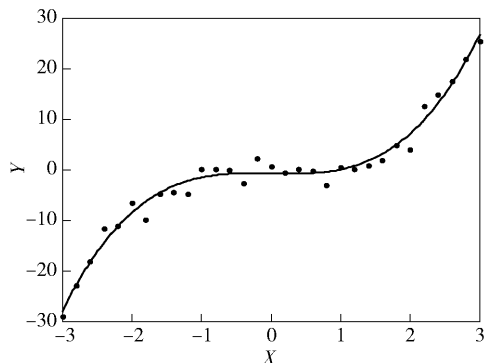


图 2 二维空间中的拟合效果

2.2 参数优化

GA 具有优化精度高、收敛速度快的优点. 采用 GA 对 LS-SVR 中的惩罚因子 g 和核函数宽度 z 进行优化.

采用均方误差 (MSE, mean square error)^[11] 统计预测值和观测值之间的误差为

$$e = \frac{1}{m} \sum_{i=1}^m (f(\mathbf{x}_i) - y_i)^2 \quad (12)$$

为提高预测精度, 在适值函数中考虑 MSE, 适值函数的定义为

$$\text{Fit}(g, 2z^2) = \frac{N}{e} \quad (13)$$

其中: N 为设定的最大遗传代数. 遗传算法中的选择算子 O_k 为

$$O_k = \frac{\text{Fit}(g, 2z^2)}{\sum \text{Fit}(g, 2z^2)} \quad (14)$$

2.3 GA-SVR

GA-SVR 链路质量预测模型分为训练阶段和测试阶段.

训练阶段, 设定迭代次数为 200、精度为 1×10^{-6} , 利用式(14)求得室内走廊、校内树林和校内广场 3 个场景中的 $(g, 2z^2)$ 分别为 $(1.313\ 8, 0.303\ 0)$, $(0.761\ 6, 15.000\ 0)$ 和 $(1.156\ 6, 9.566\ 5)$; 利用约简后的样本集训练 LS-SVR 模型, 得到回归机的相关参数 α 和 b .

测试阶段, 将未知状态的特征数据处理成数字特征向量形式. 由 1.2 节的约简结果可知, 室内走廊环境下的输入向量为 $(v_{\text{RSSI}}^i, v_{\text{LQI}}^i)$, 校内树林环境下的输入向量为 $(v_{\text{RSSI}}^i, v_{\text{SNR}}^i)$, 校内广场环境下的输入向量为 $(v_{\text{RSSI}}^i, v_{\text{SNR}}^i)$, 输出值均为 v_{PRR}^i , 采用式(11)预测链路质量.

3 实验设计及分析

实验场景如图 3 所示, 选用 CrossBow 公司的 TelosB 节点收集数据, 在自主研发的无线传感网络链路质量测试平台 (WSNs-LQT, wireless sensor networks link quality testbed) 上进行分析.

3.1 实验参数

实验参数设置如表 1 所示.

表 1 链路质量测试参数设置情况

参数属性	值
发送功率/dBm	0
信道	26
探测包数量/个	30
发包速率/(个·s ⁻¹)	5
测试周期/s	10
收发节点距离/m	20

3.2 相关性分析

相关性分析可以了解参数间的密切程度, 其中皮尔逊 (Pearson) 相关系数是对定距变量的数据进行计算, 如式(15)所示.



(a) 室内走廊



(b) 校内树林



(c) 校内广场

图 3 实验场景

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (15)$$

其中： r 为相关系数， \bar{x} 、 \bar{y} 分别为变量 x 、 y 的均值， x_i 、 y_i 分别为变量 x 、 y 的第 i 个观测值。

采用统计软件 SPSS 计算降噪后各个链路质量参数间的 Pearson 相关系数,3 个场景下的相关性分别如表 2~4 所示。

表 2 室内走廊各度量参数间的相关性

参数	RSSI 均值	LQI 均值	SNR 均值	PRR
RSSI 均值	1	0.973	0.951	0.884
LQI 均值	0.973	1	0.934	0.901
SNR 均值	0.951	0.934	1	0.854
PRR	0.884	0.901	0.854	1

表 3 校内树林各度量参数间的相关性

参数	RSSI 均值	LQI 均值	SNR 均值	PRR
RSSI 均值	1	0.882	0.783	0.790
LQI 均值	0.882	1	0.688	0.799
SNR 均值	0.783	0.688	1	0.639
PRR	0.790	0.799	0.639	1

表 4 校内广场各度量参数间的相关性

参数	RSSI 均值	LQI 均值	SNR 均值	PRR
RSSI 均值	1	0.626	0.690	0.448
LQI 均值	0.626	1	0.424	0.309
SNR 均值	0.690	0.424	1	0.308
PRR	0.448	0.309	0.308	1

室内走廊环境下,RSSI、LQI 和 SNR 与 PRR 之间的相关系数均高于校园树林环境和校园广场环境。与室内走廊环境相比,校内广场环境下 LQI 与 PRR 间的相关系数下降了 0.592;与校内树林环境相比,广场环境下 LQI 与 PRR 间的相关系数下降了 0.490,LQI 受环境影响较大。因此,选择 RSSI 和 LQI 为室内走廊环境下的链路质量预测样本集;校内树林环境和校内广场环境下选择 RSSI 与 SNR 为链路质量预测的样本集。

3.3 结果分析

3 个场景下,PRR 的实测值 (MPRR, measured packet reception ratio) 与 GA-SVR 预测模型的预测值分别如图 4~6 所示。

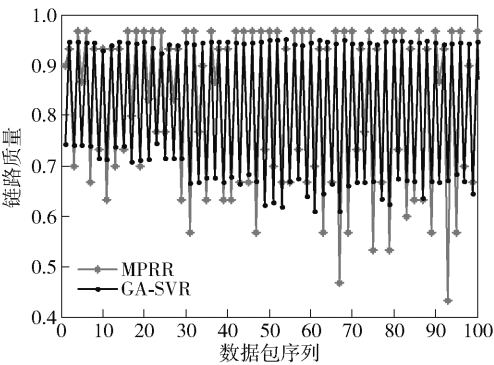


图 4 室内走廊误差对比

采用 MSE 评价链路质量预测的效果,3 个场景下 GA-SVR 预测模型的 MSE 如表 5 所示。

GA-SVR 预测模型能比较好地贴近实测 PRR,当 PRR 观测值变化幅度较大时,GA-SVR 预测模型能很好地捕捉到这一变化,具有很高的预测精度。

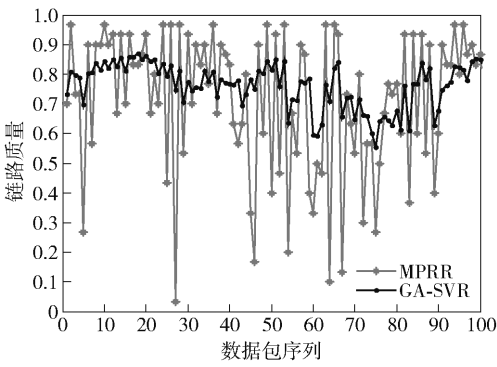


图 5 校内树林误差对比

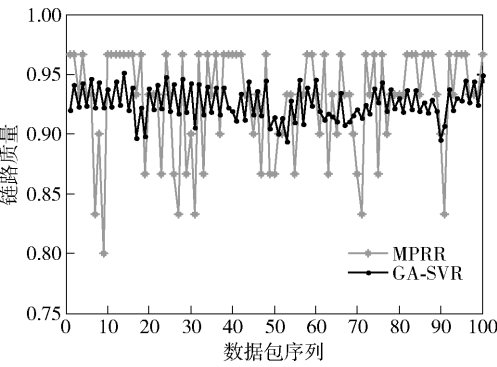


图 6 校内广场误差对比

表 5 3 个场景下的 MSE

类型	均方误差
室内走廊	0.004 0
校内树林	0.037 8
校内广场	0.001 8

3.4 与其他方法的对比

GA-SVR 预测模型与 Experts Advice 预测模型^[5]的预测性能对比如图 7 所示。

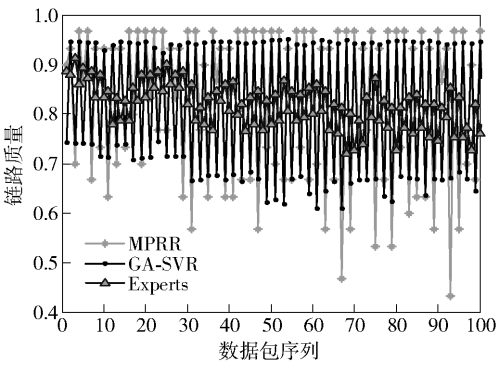


图 7 预测性能对比

由图 7 可知,GA-SVR 预测模型的预测结果更贴近实测 PRR 值,2 种方法 MSE 如表 6 所示,显

然,GA-SVR 预测模型的预测精度高于 Experts Advice 预测模型。

表 6 GA-SVR 模型与 Experts Advice 模型 MSE

类型	均方误差 (GA-SVR)	均方误差 (Experts Advice)
室内走廊	0.004 0	0.034 0
校内树林	0.037 8	0.058 5
校内广场	0.001 8	0.002 7

4 结束语

针对随机部署的无线传感器网络,采用卡尔曼滤波方法对链路质量参数进行降噪,形成决策表;采用粗糙集理论对决策表进行约简,得到最小样本集;提出 GA-SVR 链路质量预测模型. 与 Experts Advice 预测模型相比,GA-SVR 预测模型具有更高的预测精度. 下一步工作将对链路质量参数进一步研究,以找出能更全面、准确刻画链路状态的参数。

参考文献:

[1] 潘成, 张和生. 无线传感器网络快速数据收集的聚集调度方法[J]. 北京邮电大学学报, 2016, 39(4): 87-91.
Pan Cheng, Zhang Hesheng. Fast data collection of wireless sensor networks by aggregation scheduling [J]. Journal of Beijing University of Posts and Telecommunications, 2016, 39(4): 87-91.
[2] Amjad M, Afzal M K, Umer T, et al. QoS-aware and heterogeneously clustered routing protocol for wireless sensor networks [J]. IEEE Access, 2017, 39(99): 1-12.
[3] 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.
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.
[4] Marinca D, Minet P. On-line learning and prediction of link quality in wireless sensor networks [C]//2014 IEEE Global Communications Conference (GLOBECOM). Austin: IEEE Press, 2014: 1245-1251.
[5] 舒坚, 汤津, 刘琳岚, 等. 基于模糊支持向量回归机的 WSNs 链路质量预测 [J]. 计算机研究与发展, 2015, 52(8): 1842-1851.

- Shu Jian, Tang Jin, Liu Linlan, et al. Fuzzy support vector regression-based link quality prediction model for wireless sensor networks [J]. Journal of Computer Research and Development, 2015, 52(8): 1842-1851.
- [6] Bildea A, Alphand O, Rousseau F, et al. Link quality estimation with the Gibert-Elliot model for wireless sensor networks [C] // 2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (2015 PIMRC). Hong Kong: IEEE Press, 2015: 2049-2054.
- [7] Shu Jian, Tao Juan, Liu Linlan, et al. CCI-based link quality estimation mechanism for wireless sensor networks under non-perceived packet loss [J]. The Journal of China Universities of Posts and Telecommunications, 2013, 20(1): 1-10.
- [8] 赵斌, 何涇沙, 张伊璇, 等. 访问控制中基于粗糙集的授权规则指示发现 [J]. 北京邮电大学学报, 2016, 39(2): 48-52.
- Zhao Bin, He Jingsha, Zhang Yixuan, et al. Knowledge discovery of authorization rule based on rough set in trust based access control [J]. Journal of Beijing University of Posts and Telecommunications, 2016, 39(2): 48-52.
- [9] Eberts M, Steinwart I. Optimal learning rates for least squares SVMs using Gaussian kernels [C] // 2011 Annual Conference on Neural Information Processing Systems (2011NIPS). Granada: Advances in Neural Information Processing Systems, 2011: 1539-1547.
- [10] Ning Kefeng, Liu Min, Dong Mingyu, et al. Robust LS-SVR based on variational Bayesian and its application [C] // 2014 International Joint Conference on Neural Networks (IJCNN). Beijing: IEEE Press, 2014: 2920-2926.
- [11] Tang Yizhou, Zhou Jiawen. The performance of PSO-SVM in inflation forecasting [C] // 2015 12th International Conference on Service Systems and Service Management (ICSSSM). Guangzhou: IEEE Press, 2015: 1-4.