

基于粒子群优化极限学习机的排水管道结构状况评价

郑茂辉¹, 刘少非^{1,2}, 柳娅楠², 李浩楠²

(1. 同济大学 上海防灾救灾研究所, 上海 200092; 2. 同济大学 土木工程学院, 上海 200092)

摘要: 基于极限学习机(ELM)和粒子群优化(PSO)算法, 建立一个新型排水管道结构性状况评价模型。采用 PSO 算法优化 ELM 中的输入权值矩阵和隐含层偏置, 改善网络参数随机生成带来的分类精度偏低的问题。以上海市洋山保税港区排水管网为例, 对分类器模型进行训练测试, 并与 ELM 分类结果进行对比分析。结果表明, PSO-ELM 算法以较少的隐含层神经元节点获得更高的分类精度, 参数优化提高了模型拟合能力, 对于城市排水管道结构性状况分类、判断具有可行性和有效性。

关键词: 排水管道; 结构性状况评价; 极限学习机; 粒子群优化

中图分类号: TU992

文献标志码: A

Structural Condition Assessment of Urban Drainage Pipes Based on Particle Swarm Optimization-Extreme Learning Machine

ZHENG Maohui¹, LIU Shaofei^{1,2}, LIU Ya'nan², LI Haonan²

(1. Shanghai Institute and Disaster Prevention of Relief, Tongji University, Shanghai 200092, China; 2. College of Civil Engineering, Tongji University, Shanghai 200092, China)

Abstract: Structural condition assessment of drainage pipes has been a major concern for asset managers in maintaining the required performance of urban drainage systems. This paper proposed a neural network model combining extreme learning machine (ELM) and particle swarm optimization (PSO) to classify the structural condition status of urban drainage pipes. Besides, in an attempt to look for better classification performance, it used the PSO algorithm to optimize the input weight matrix and the hidden layer offset of ELM. Moreover, it validated the PSO-ELM model by using the dataset supplied from the Yangshan Bonded Port Area in Shanghai. Furthermore, it compared the predictive performance of PSO-ELM with ELM on the same dataset.

The result shows that the PSO-ELM can achieve a higher classification accuracy by utilizing less neuron nodes in the hidden layer, and improve the fitting capability of ELM. The method proposed has feasibility and effectiveness for structural condition assessment of urban drainage pipes.

Key words: drainage pipes; structural condition assessment; extreme learning machine (ELM); particle swarm optimization (PSO)

随着城市地下管道设施规模的不断扩大和服役年限的增长,管道老化和运行安全问题得到越来越多的研究和关注^[1-2]。城市排水管道状况评价是制定管网养护计划和修复计划的重要依据,也是城市安全监管的重要内容之一。

排水管道状况包括结构性状况和功能性状况两方面,它们分别描述管道内部结构性缺陷类型、等级,以及管道过流受阻等水力状况^[3]。CCTV (closed circuit television, 管道内窥检测)是目前国内外排水管道检测评估先进、有效的技术手段,其关键内容就是对管道影像数据信息的判读,包括确定内部缺陷种类、位置表述、等级和状况评估计算。由于 CCTV 检测成本高昂,通常只是针对城市特定区域或运行年限较长的排水管道设施。而对于其他缺乏检测资料的城市管道网络,其健康状况诊断和病害趋势预判等则亟待更多深入的研究。近些年,国内外学者采用随机统计和机器学习方法研究建立了多种管道状况模型,如多元回归模型^[4-6]、马尔科夫链模型^[7-9]、神经网络模型^[10-12],等等。鉴于管道老化因素众多、过程复杂,数据驱动的神经网络模型无需深入理解老化机理,只要通过数据样本监督学习即可建立管道状况同相关特征变量的关联模式,无疑具有更好的适用性和应用前景^[10]。

收稿日期: 2019-03-13

基金项目: 国家重点研发计划(2016YFC0802400, 2017YFC0803300)

第一作者: 郑茂辉(1976—),男,副研究员,理学博士,主要研究方向为城市综合防灾。E-mail: zmh@tongji.edu.cn



论文
拓展
介绍

极限学习机(extreme learning machine, ELM)是Huang等^[13]基于广义逆矩阵理论提出的一种单隐层前馈神经网络的监督型学习算法。本文充分利用ELM计算速度快、泛化性能好的特点,将粒子群优化(particle swarm optimization, PSO)^[14]方法与其相结合,选择最优的网络参数,建立一个适用于城市排水管道状况评价的分类器模型,并以上海市洋山港保税区排水管网为例进行实证研究,验证模型方法的合理性。

1 方法

1.1 PSO-ELM算法

设 n, L, m 分别为输入层、隐含层和输出层的节点数。给定 N 组任意的样本 $(x_i, t_i) \in \mathbf{R}^n \times \mathbf{R}^m$, ELM算法的输出函数可以表示为

$$f_L(x) = \sum_{i=1}^L \beta_i g(\omega_i \cdot x_j + b_i) = o_j, \quad j=1, \dots, N \quad (1)$$

式中: $g(x)$ 为激励函数; $\omega_i = [\omega_{i1} \ \omega_{i2} \ \dots \ \omega_{in}]^T$,表示隐含层第 i 个神经元与输入层的连接权值; $\beta_i = [\beta_{i1} \ \beta_{i2} \ \dots \ \beta_{im}]^T$,表示隐含层第 i 个神经元与输出层的连接权值; b_i 是隐含层神经元的阈值; $\omega_i \cdot x_j$ 表示 ω_i 和 x_j 的内积。

假定训练样本数量 N 与隐含层神经元节点数 L 相等,则对于任意给定的 β_i 和 ω_i , ELM能零误差逼近学习样本^[13],即

$$\sum_{i=1}^L \beta_i g(\omega_i \cdot x_j + b_i) = t_j, \quad j=1, \dots, N \quad (2)$$

将其以矩阵形式表达为

$$H\beta = T \quad (3)$$

其中

$$H = \begin{bmatrix} h(X_1) \\ \vdots \\ h(X_N) \end{bmatrix} = \begin{bmatrix} g(\omega_1 \cdot x_1 + b_1) & \dots & g(\omega_L \cdot x_1 + b_L) \\ \vdots & & \vdots \\ g(\omega_1 \cdot x_N + b_1) & \dots & g(\omega_L \cdot x_N + b_L) \end{bmatrix}_{N \times L}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, \quad T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m}$$

对ELM的网络训练等价于如下优化问题:

$$\|H\hat{\beta} - T\| = \min_{\beta} \|H\beta - T\| \quad (4)$$

当 $g(x)$ 无限可微时,随机给定网络参数 ω_i 和 b_i ,

ELM模型训练过程可看作求解线性系统 $H\beta = T$ 关于 $\hat{\beta}$ 的最小二乘解

$$\hat{\beta} = H^+ T \quad (5)$$

式中: H^+ 为隐含层输出矩阵 H 的Moore-Penrose广义逆。

ELM模型随机生成输入权值矩阵和隐含层神经元的阈值,无需迭代调整就可以解析计算得出网络输出值。不过参数随机产生使结果存在一定的随机性^[15],容易产生较差的分类效果。本文结合PSO算法,优化ELM网络参数 ω_i 和 b_i ,以改善ELM算法输出权值 β_i 不稳定的缺点。

PSO算法是一种模拟鸟群觅食行为的群体智能优化算法,其核心是粒子更新速度和位置,即

$$v_i^{t+1} = \omega v_i^t + c_1 r_1 (p_i^t - x_i^t) + c_2 r_2 (g_i^t - x_i^t) \quad (6)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (7)$$

$$\omega = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{t_{\max}} \times t \quad (8)$$

式中: p_i 表示粒子的个体最优位置; g_i 表示搜索到的群体最优位置; ω_{\max} 和 ω_{\min} 分别为惯性权重的最大值和最小值; c_1 和 c_2 均为学习因子,也称为加速常数; r_1 和 r_2 为 $[0, 1]$ 范围内的均匀随机数; t_{\max} 和 t 分别为最大迭代次数和当前的迭代次数。

将ELM输入权值 ω_i 和隐含层神经元偏置 b_i 作为PSO算法的粒子,以极限学习机训练集的均方误差作为适应度函数,计算每个粒子的适应度值,通过比较不断更新粒子的速度和位置,直至达到最小误差或最大迭代次数,得到经过PSO优化的ELM网络参数。

1.2 模型构建与评价

影响排水管道结构性状况的因素较多。为了简化网络模型,根据文献^[6],选取管材、管龄、管径、埋深、管长、坡度和所在道路类型7项影响因素,作为管道结构性状况评价的输入向量,以CCTV检测结果中管道修复指数(RI)所对应的结构性状况等级作为网络输出,利用PSO-ELM算法建立分类器,通过样本数据训练学习确定影响因素与结构性状况之间的非线性关系。表1给出的管道修复指数及其对应的结构性状况等级、修复建议^[3]。其中,等级I ($RI \leq 1$)表示基本完好,等级II ($1 < RI \leq 4$)表示轻微破坏,等级III ($4 < RI \leq 7$)表示中等破坏,等级IV ($RI > 7$)表示严重破坏。

根据上述输入、输出向量建立排水管道结构性状态评价的PSO-ELM网络拓扑模型,并在Matlab R2016b软件平台中实现。模型仿真时将采用试验

表1 管道修复指数与等级划分

Tab.1 Pipeline rehabilitation index (RI)

等级	RI	修复建议及说明
I	RI≤1	结构条件基本完好,不修复
II	1<RI≤4	结构在短期内不会发生破坏现象,但应做修复计划
III	4<RI≤7	结构在短期内可能发生破坏,应尽快修复
IV	RI>7	结构已经发生破坏或即将发生破坏,应立即修复

方法设定不同的神经元节点数 L 和激励函数 $g(x)$,对标准ELM和PSO-ELM模型训练测试结果进行比较分析。

混淆矩阵分析方法是评价分类器性能好坏最直接有效的方法。假定 n_{ij} 表示被分类为 j 类的 i 类样本数,则分类准确率以正确分类的样本数与总样本数 N 的比值来表示

$$P = \frac{\sum_{i=1}^K n_{ii}}{N}, K=4 \quad (9)$$

定义 R_i 是第 i 类样本的查全率

$$R_i = \frac{n_{ii}}{\sum_{j=1}^K n_{ij}} \quad (10)$$

定义 G 为样本所有类别查全率的几何平均值

$$G = \left(\prod_{i=1}^K R_i \right)^{\frac{1}{K}} \quad (11)$$

指标 G 适于评价不平衡数据集上分类器性能,其基本思想是使每一分类正确率尽可能大的同时,保持各类之间的平衡。

2 试验研究

2.1 试验数据

以上海市浦东新区洋山港保税港区公共排水管道为例,采用标准ELM模型以及优化后的PSO-ELM模型开展试验研究,预测、判定排水管道结构性状况等级。

试验区基础数据包含两部分:一是排水管道GIS(地理信息系统)数据,包含管材、管龄、管径、埋深、管长、坡度和所在道路类型等模型输入的特征数据;二是试验区2018年10月份排水管道CCTV检测数据,检测管道里程20.3 km。基于管段唯一性标识建立GIS数据和CCTV检测数据对应关系,提取有效样本数据共559条,采用4折交叉验证法将样本集平均分为4份,每次验证抽取1份作为测试集,余下3份作为训练集,重复进行4次,取4次测试验证结果的平均值作为模型的误差估计。另外,为避免各指

标量纲和数量级不同造成的不平衡性,采用最大最小法对原始数据进行了量纲一化处理,使得处理后的数据分布范围在 $[0,1]$,公式如下:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (12)$$

式中: x 为实测值; x_{\min} 为样本数据的最小值; x_{\max} 为样本数据的最大值。

2.2 网络参数优化

利用标准ELM构建管道状态分类评价模型时,仅需要设定隐含层神经元节点数 L 和激励函数 $g(x)$ 的构造形式就可以求算输出权值矩阵。实际工程应用中隐含层神经元节点数 L 一般远小于样本数 N , L 过小网络预测误差较大, L 过大则会增加模型预测的时间、空间成本,容易出现过拟合现象。图1给出sigmoid、sin和hardlim 3种常见激励函数下隐含层神经元节点数 L 对分类器能力的影响。其中,sigmoid函数和sin函数的分类器性能接近, L 增至300时分类准确率相对较高,约为60%。因此,对于未优化的ELM管道状况分类评价模型,激励函数选择整体性能较优的sigmoid函数,隐含层神经元节点数 L 设定为300。

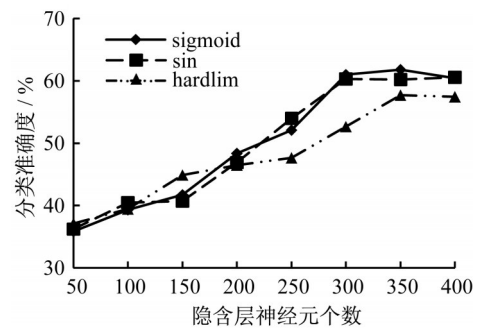


图1 3种激励函数下隐层节点数对ELM分类性能的影响
Fig.1 Effect of number of hidden nodes on ELM performance in different activation functions

为方便比较,同样选择sigmoid激励函数,对PSO-ELM分类器性能进行仿真分析。经测试比较,设定PSO种群数 $m=40$, $c_1=c_2=1.5$, $\omega_{\max}=0.95$, $\omega_{\min}=0.4$, $t_{\max}=200$ 。如图2所示, $L=70$ 时,PSO-ELM分类准确率约为61%; $L=130$ 时,分类准确率达到90%以上。可见,采用PSO优化ELM的网络参数,能够以更少的隐层神经元节点获得更高的分类准确率。需要说明的是,在复杂性程度上PSO-ELM比标准ELM增加了网络参数寻优的步骤,但PSO算法可以较为快速地获取到最优参数,能够有效避免ELM可能存在的结果不稳健的缺点。

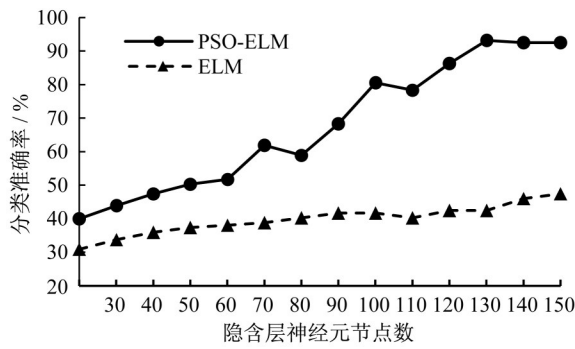


图2 PSO-ELM隐层节点数影响

Fig.2 Effect of number of hidden nodes on PSO-ELM performance

2.3 分类结果分析

通过多次试验后,确定 PSO-ELM 和 ELM 的隐层神经元节点数分别为 130 和 300 时分类效果最佳。如表 2 所示,经 PSO 优化 ELM 网络参数后,不光分类准确率 P 显著提升, G 数值也由 58.36% 增至 91.0%。通常,城市排水管道的结构性状况是非均衡分布的,除了整体准确率,不同分类准确度、查全率及相互之间的平衡也是分类器的重要评价指标。

表2 PSO-ELM 和 ELM 模型分类性能比较

Tab.2 Comparison of performance of PSO-ELM and ELM

模型	激励函数	隐层节点	P / %	G / %
PSO-ELM	sigmoid	130	92.81	91.00
标准 ELM	sigmoid	300	61.15	58.36

表 3 以混淆矩阵形式给出 PSO-ELM 和 ELM 在测试集上的分类结果对比。其中,结构性损坏比较严重的 III 类管道样本 17 条, PSO-ELM 准确预测 15 条,查全率 88.2%; IV 类管道样本 23 条,准确预测 21 条,查全率 91.3%。ELM 分类器的相应查全率则分别为 64.7% 和 43.5%。结果可见,相比较 ELM, PSO-ELM 对于非均衡数据集显然具有更优的分类预测能力。

表3 分类结果混淆矩阵

Tab.3 Confusion matrix for PSO-ELM

类型	PSO-ELM 模型				ELM 模型			
	I	II	III	IV	I	II	III	IV
I	44	3	0	0	28	19	0	0
II	0	49	3	0	5	36	11	0
III	0	2	15	0	0	6	11	0
IV	0	0	2	21	0	1	12	10

图 3 给出了测试集上逐个样本的 PSO-ELM 分类结果与实测值的对照。结果也表明, PSO-ELM 具备较高的分类精度和拟合能力,能够较好满足排水

管道结构性状况评价的应用要求。

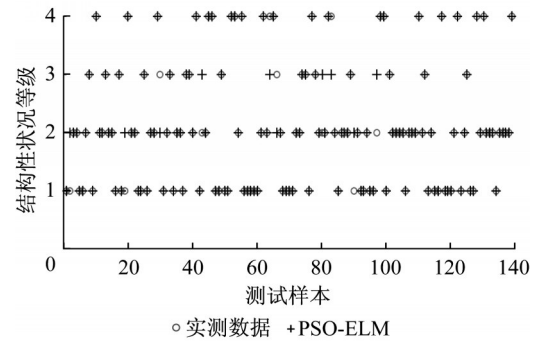


图3 PSO-ELM分类测试结果

Fig.3 Predicted versus observed result for PSO-ELM

3 结论

(1) 提出了 PSO-ELM 排水管道结构性状况评价模型,采用 PSO 算法优化 ELM 输入连接权值和隐层偏置,避免参数随机初始化造成的分类结果不稳定、准确率偏低的弊端。

(2) 利用 CCTV 检测样本集对 PSO-ELM、ELM 进行仿真测试,试验结果表明, PSO-ELM 能够以较少的隐层神经元节点达到更高的分类精度,提高了 ELM 模型的拟合能力。

(3) 本文为城市排水管道结构性状况评价提供了一个新型的数据驱动的建模方法,具有较好的可行性和适用性。后续将利用更多的样本数据训练优化模型,进一步提高预测精度和泛化能力。

参考文献:

[1] DUCHESNE S, BEARDSSELL G, VILLENEUVE J P, et al. A survival analysis model for sewer pipe structural deterioration [J]. Computer-Aided Civil and Infrastructure Engineering, 2013, 28 (2): 146.

[2] 刘克会, 江贻芳, 邓楠, 等. 城市地下管线主要风险因素分析 [J]. 工程勘察, 2013(9): 51.

LIU Kehui, JIANG Yifang, DENG Nan, et al. Analysis on main risk factors of urban underground pipelines [J]. Geotechnical Investigation & Surveying, 2013(9): 51.

[3] 中华人民共和国住房和城乡建设部. 城镇排水管道检测与评估技术规程: CJJ 181—2012[S]. 北京: 中国建筑工业出版社, 2012.

Ministry of Housing and Urban-Rural Development of the People's Republic of China. Technical specification for inspection and evaluation of urban sewer: CJJ 181—2012[S]. Beijing: China Architecture & Building Press, 2012.