

# 基于缺失数据估计的 PLSS 古建筑寿命预测

路 杨,吕 欣,周福娜,王佳瑜

(河南大学 计算机与信息工程学院,河南 开封 475004)

**摘 要:**影响木结构古建筑寿命的因素主要包括物理、化学和生物等多个方面,且影响过程是错综复杂的。这些因素之间可能非线性相关,相应观测数据特征提取的完整性决定了其寿命预测的精确性。文中引入基于样条变换的偏最小二乘预测方法,在更高维空间中把非线性预测模型的建立问题转化为线性预测模型的建立问题实现木结构古建筑寿命预测。针对观测数据缺失情况下,基于样条变换的非线性偏最小二乘方法预测精度较低的问题,文中给出一种基于缺失数据估计的非线性偏最小二乘预测方法,以更充分地抽取可用观测数据的特征信息,并将其应用于木结构古建筑寿命预测中。仿真和实验验证结果表明了该方法的有效性。

**关键词:**数据缺失;样条变换;偏最小二乘回归;古建筑;寿命预测

中图分类号:TP301

文献标识码:A

文章编号:1673-629X(2016)06-0195-05

doi:10.3969/j.issn.1673-629X.2016.06.044

## Prediction of Ancient Building Life for PLSS Based on Missing Data Estimation

LU Yang, LÜ Xin, ZHOU Fu-na, WANG Jia-yu

(School of Computer and Information Engineering, Henan University, Kaifeng 475004, China)

**Abstract:** The factors that affect the life of the ancient building of the wood structure include many aspects, such as physical, chemical and biological factors, and the process of impact is complicated. These factors may be nonlinear correlation, and the completeness of the corresponding observation data feature extraction determines the accuracy of the life prediction. In this paper, the PLS based on spline transformation is introduced, and the problem of building the nonlinear model in higher dimensional space is transformed into it of building the linear model to realize the life prediction of wood structure. In order to estimate the accuracy of the nonlinear PLS method based on the case of the missing data, a nonlinear PLS prediction method is given based on the missing data estimation to extract the feature information of the available observation data more adequately and used in the life prediction of wood structures. Simulation and experimental verification results show the effectiveness of the proposed method.

**Key words:** data missing; spline transformation; Partial Least-Squares Regression (PLSS); ancient building; life prediction

## 0 引 言

木构古建筑是我国的历史见证之一,对了解历史、传承民族文化都有积极、深远的意义;充分反映我国的科技水平,且是不可再生资源,具有极高的社会、经济和文化价值。在历经岁月沧桑的保存过程中,由于人为的和自然力的破坏,木构古建筑都不同程度地遭受损害,甚至毁灭。科学准确地预测木构古建筑的使用寿命,对重大木结构的管理维护具有重要的意义<sup>[1]</sup>。由于影响木构古建筑寿命的因素众多,影响过程错综复杂,难以用准确的数学表达式来描述,所以研究基于数据特征抽取技术的寿命预测方法具有重要意义。

关于木构古建筑论文多是遗产保护<sup>[2]</sup>、测绘方法技术<sup>[3]</sup>、修缮方法<sup>[4]</sup>、地震灾害<sup>[5]</sup>等。对木构古建筑寿命预测的算法较少,多集中于木结构累积损伤模型<sup>[6]</sup>、有限元分析模型<sup>[7]</sup>和 BP 神经网络<sup>[8]</sup>等。木结构累积损伤模型,由于结构的特点,技术问题较为复杂,目前各种技术标准和规范不能全部覆盖<sup>[9]</sup>。有限元分析,受到分析软件功能的限制,在选择单元属性以及屈服准则的时候,是不精确的<sup>[10]</sup>。集中在木构件自身的承受能力和结构特性。BP 神经网络算法本质上为梯度下降法,且至今尚无一种统一而完整的理论指导,一般只能由经验选定,具有不确定性。这些算法均是建立

收稿日期:2015-09-12

修回日期:2015-12-16

网络出版时间:2016-05-25

基金项目:国家自然科学基金资助项目(61203094,61174112);河南省高校科技创新人才支持计划(2012HASTIT005)

作者简介:路 杨(1972-),女,博士,教授,研究方向为模式识别、数据挖掘;吕 欣(1987-),女,硕士研究生,研究方向为故障诊断。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160525.1706.034.html>

在数据完整的基础上的。

文中引入基于样条变换的非线性偏最小二乘回归<sup>[11]</sup>(Partial Least-Squares Splines, PLSS)方法,跳出了之前木构古建筑寿命预测的纯结构特性和力学特性,建立在数据基础上对木构古建筑寿命进行预测。木构古建筑所测数据是复杂的非线性关系,样条变换将所测数据的非线性问题转换为拟线性问题,既解决了所测数据间的非线性问题,又保留了偏最小二乘<sup>[12]</sup>(Partial Least-Squares regression, PLS)算法的预报能力。木构古建筑由于传感器故障或离线维护,人为漏采集数据等造成数据缺失或不匹配等,坚持通过完整数据进行木构古建筑寿命预测,需要删除大量数据,不仅造成人力物力的严重浪费,而且预测结果具有较大预测误差。PLSS 算法不能解决采集数据的缺失问题,为对缺失参数数据进行合理且有效的估计<sup>[13]</sup>,并且科学地对木构古建筑寿命进行预测,文中提出一种基于缺失数据估计的 PLSS 预测方法并将其应用于木构古建筑寿命预测中,对采集数据中缺失数据进行估计,使采集数据得到充分利用,并且提高了寿命预测的准确度。

## 1 PLS 与 PLSS

### 1.1 PLS

自变量与因变量的数据表  $\mathbf{X} = [x_1, x_2, \dots, x_p]_{n \times p}$  和  $\mathbf{Y} = [y_1, y_2, \dots, y_q]_{n \times q}$ 。PLS 预测模型步骤为:

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{X}_1 \quad (1)$$

$$\mathbf{Y} = \mathbf{t}_1 \mathbf{r}_1^T + \mathbf{Y}_1 \quad (2)$$

其中,  $\mathbf{p}_1 = \mathbf{X}^T \mathbf{t}_1 / \|\mathbf{t}_1\|^2$ ,  $\mathbf{r}_1 = \mathbf{Y}^T \mathbf{t}_1 / \|\mathbf{t}_1\|^2$  为负荷向量;  $\mathbf{X}_1$ 、 $\mathbf{Y}_1$  为残差矩阵, 将  $\mathbf{X}_1$ 、 $\mathbf{Y}_1$  代替  $\mathbf{X}$ 、 $\mathbf{Y}$  进行求解。

$$\hat{\mathbf{X}} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \dots + \mathbf{t}_k \mathbf{p}_k^T \quad (3)$$

其中,  $k$  是通过交叉有效性确定的。

$$Q_h^2 = 1 - \text{PRESS}_h / \text{SS}_{h-1} \quad (4)$$

当  $Q_h^2 \geq (1 - 0.95^2) = 0.0975$ , 判定  $k = h$ , 停止迭代。

PLS 算法<sup>[14]</sup>对具有线性关系的变量,具有很好的预测能力,但是木构古建筑变量间是复杂的非线性关系,PLS 并不能很好地进行预测。

我国的木构古建筑中存在大量梁柱式的木结构,最典型的有殿堂式木结构,这些结构的存在对研究我国的历史和建筑技术发展具有非常重要的意义。影响木构古建筑寿命的因素主要有荷载因素、环境因素、结构材料因素、力学因素、微生物等,其中选取最重要的影响因素—虫蛀面积率、腐朽深度、应力水平、虫蛀面积的均值及其变异系数和开裂深度的均值及其变异

系数来进行寿命预测。木构古建筑寿命预测关注的质量指标仅有一个即剩余寿命,所以文中讨论  $q = 1$ ,  $p = 6$  情况下的木构古建筑数据特征抽取和预测问题。

### 1.2 PLSS

在实际应用中,由于因变量和自变量之间的关系常常不是线性的,它们之间存在着复杂的非线性关系。传统的 PLS 仅适用于线性模型,实际应用中造成较大偏差。PLSS 算法<sup>[15]</sup>是在 PLS 算法中加入样条变换,用分段拟合的思想构造样条函数,具有按需要裁剪以适应任意曲线连续变化的优点,样条函数具有光滑性、连续性。

Step1:采集自变量和因变量  $[X, Y]$ , 对自变量  $X$  进行样条变换,在高维空间中把非线性问题转化为拟线性问题。模型采用三次 B 样条函数作为变换基数。

$$\Omega_3\left(\frac{x_j - \xi_{j,l-1}}{h_j}\right) = \frac{1}{3! h_j^3} \sum_{k=0}^4 (-1)^k \binom{4}{k} (x_j - \xi_{j,l-1+k})_+^3 \quad (5)$$

其中,  $\xi_{j,l-1}$  为对变量  $x_j$  划分所增加的插入点;  $h_j$  为对变量  $x_j$  划分的分段长度。

记变量  $x_j$  的最小观测值为  $\min(x_j)$ , 最大观测值为  $\max(x_j)$ , 有

$$\xi_{j,l-1} = \min(x_j) + (l-1)h_j \quad (6)$$

$$h_j = \frac{\max(x_j) - \min(x_j)}{M_j} (l = 0, 1, \dots, M_j + 2) \quad (7)$$

其中,  $M_j$  为对变量  $x_j$  划分的分段个数。

全体自变量与因变量的非线性函数关系为:

$$y = \beta_0 + \sum_{j=1}^p \sum_{l=0}^{M_j+2} \beta_{j,l} \Omega_3\left(\frac{x_j - \xi_{j,l-1}}{h_j}\right) + \varepsilon \quad (8)$$

其中,  $y$  与  $\Omega_3\left(\frac{x_j - \xi_{j,l-1}}{h_j}\right)$  呈线性关系,可以变换为拟线性回归模型。每一维  $x_j$  进行三次 B 样条变换  $x_j \rightarrow Z_j$ , 原变量空间  $(X, Y)$  变换为  $(Z, Y)$ 。

Step2:用新的自变量和因变量  $(Z, Y)$  进行 PLS 算法预测,求  $\hat{y}$ 。

PLSS 预测是建立在数据完整的基础上的,数据有缺失的情况下预测精度较差。故文中提出基于缺失数据估计的 PLSS 预测算法。

## 2 基于缺失数据估计的 PLSS 预测

现有观测数据阵  $\mathbf{Y} = [y_1, y_2, \dots, y_q] \in R^{n \times q}$ ,  $\mathbf{X}_0 = [x_1, x_2, \dots, x_m] \in R^{n \times m}$ , 基于缺失数据估计的 PLSS 预测算法思路是:先将观测数据阵  $\mathbf{X}_0$  中缺失的部分通过算法迭代重构<sup>[16]</sup>,完成缺失数据,得到估计值  $\hat{\mathbf{X}}$ 。再通过样条变换,数据  $\hat{\mathbf{X}}$  进行扩维得到  $\mathbf{Z}$ , 将非线性问

题转化为拟线性问题。数据  $[\mathbf{Z}, \mathbf{Y}]$  进行 PLS 建模,尽可能多地提取主成分,转化为  $\mathbf{Y}$  关于  $\mathbf{Z}$  的非线性回归模型,并计算预测结果。具体算法步骤如下:

Step1:采集数据  $\mathbf{X}_0$ ,其中缺失部分赋该列均值得数据  $\mathbf{X}$ 。

Step2:基于现有信息建立粗略 PCA 模型。计算预测值  $\hat{\mathbf{X}}$ 。

$$\hat{\mathbf{X}} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \cdots + \mathbf{t}_k \mathbf{p}_k^T$$

其中,负荷向量  $\mathbf{p}_i \in R^{m \times 1}$ ;得分向量  $\mathbf{t}_i = \mathbf{X} \mathbf{p}_i, \mathbf{t}_i \in R^{n \times 1}$ 。

Step3:检查收敛性,计算  $\mathbf{X}$  和  $\hat{\mathbf{X}}$  中缺失数据相对预测误差。

$$\text{RPE} = \|\mathbf{X} - \hat{\mathbf{X}}\| / \|\mathbf{X}\| \quad (9)$$

当  $\text{RPE} \geq 10^{-5}$ ,  $\hat{\mathbf{X}}$  代替  $\mathbf{X}$  进行迭代直到收敛。

Step4:对  $\hat{\mathbf{X}}$  进行样条变换。每一维  $\hat{x}_j (j=1, 2, \cdots, m)$  进行三次 B 样条变换  $\hat{x}_j \rightarrow Z_j$ , 对  $\hat{x}_j$  做如下三次 B 样条变换:

$$z_{j,0} = \Omega_3\left(\frac{\hat{x}_j - \xi_{j,-1}}{h_j}\right), z_{j,1} = \Omega_3\left(\frac{\hat{x}_j - \xi_{j,0}}{h_j}\right), \cdots, z_{j,M_j+2} = \Omega_3\left(\frac{\hat{x}_j - \xi_{j,M_j+1}}{h_j}\right) \quad (10)$$

记  $Z_j = (z_{j,0}, z_{j,1}, \cdots, z_{j,M_j+2})$ , 所得新的数据为  $[\mathbf{Z}, \mathbf{Y}] = [Z_1, Z_2, \cdots, Z_m, \mathbf{Y}]$ 。

Step5:对  $[\mathbf{Z}, \mathbf{Y}] = [Z_1, Z_2, \cdots, Z_m, \mathbf{Y}]$  进行标准化处理,以去除量纲的影响。

$$\tilde{z}_{j,l} = \frac{z_{j,l} - \bar{z}_{j,l}}{s_{j,l}}, \tilde{y} = \frac{y_i - \bar{y}_i}{s_y} (l=1, 2, \cdots, M_j+2; j=1, 2, \cdots, p; i=1, 2, \cdots, n) \quad (11)$$

其中,  $\bar{z}_{j,l}, \bar{y}$  分别是  $z_{j,l}, y$  的样本均值;  $s_{j,l}, s_y$  分别是  $z_{j,l}, y$  的样本方差。

记处理后的自变量  $\tilde{\mathbf{Z}} = (\tilde{Z}_1, \tilde{Z}_2, \cdots, \tilde{Z}_p)$ , 因变量为  $\tilde{y}$ , 则原变量空间  $(\mathbf{X}, \mathbf{Y})$  变换为  $(\tilde{\mathbf{Z}}, \tilde{y})$ , 并建立新的线性回归模型。

$$\tilde{y} = \sum_{j=1}^p \sum_{l=0}^{M_j+2} \alpha_{j,l} \tilde{z}_{j,l} + \varepsilon \quad (12)$$

Step6:对  $(\tilde{\mathbf{Z}}, \tilde{y})$  进行 PLS 建模,求得回归系数  $\alpha_{j,l} (j=1, 2, \cdots, p; l=1, 2, \cdots, M_j+2)$ 。

$$\frac{y - \bar{y}}{s_y} = \sum_{j=1}^p \sum_{l=0}^{M_j+2} \alpha_{j,l} \tilde{z}_{j,l} + \varepsilon \quad (13)$$

得到

$$y = \beta_0 + \sum_{j=1}^p \sum_{l=0}^{M_j+2} \beta_{j,l} \tilde{z}_{j,l} + \varepsilon \quad (14)$$

其中

$$\beta_{j,l} = s_y \frac{\alpha_{j,l}}{s_{j,l}}, \beta_0 = \bar{y} - \sum_{j=1}^p \sum_{l=0}^{M_j+2} \beta_{j,l} \bar{z}_{j,l} \quad (15)$$

从而可建立起  $\mathbf{Y}$  关于  $\mathbf{X}$  的非线性回归模型进行木构古建筑寿命预测。

$$\hat{y} = \beta_0 + \sum_{j=1}^p f_j(x_j) + \beta_0 + \sum_{j=1}^p \sum_{l=0}^{M_j+2} \beta_{j,l} \Omega_3\left(\frac{x_j - \xi_{j,l-1}}{h_j}\right) \quad (16)$$

图1为基于缺失数据估计的 PLSS 预测框图。在木构古建筑中每个木构件的寿命均对该建筑的寿命有影响,不能机械地删除其中任何一个部件的数据,或者是删除某一项寿命预测重要指标,所以对缺失数据的估计对木构古建筑寿命预测具有实践意义。

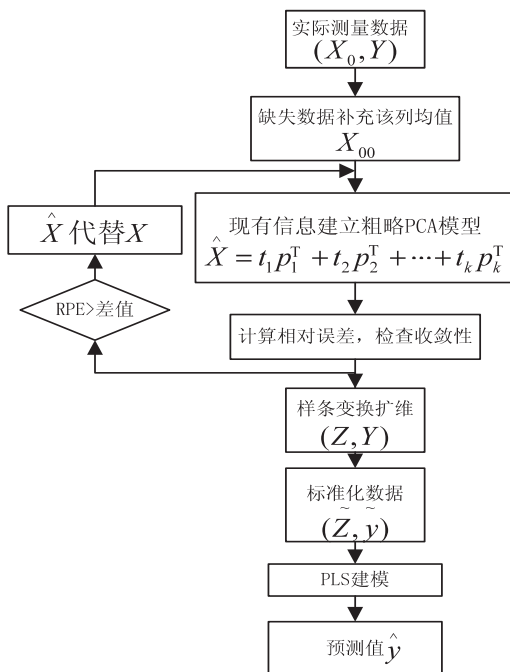


图1 基于缺失数据估计的 PLSS 预测流程图

### 3 仿真与实验分析

本节首先采用模拟数值仿真的基于缺失数据估计的 PLSS 算法,利用相对预测误差对文中的 PLSS 算法进行测试,然后利用实际采集数据的基于缺失数据估计的 PLSS 算法对木结构古建筑寿命进行预测。

#### 3.1 数据仿真实验

首先采用模拟数值验证的基于缺失数据估计的 PLSS 算法,并利用相对预测误差衡量预测算法的精度。观测数据矩阵  $\mathbf{X} = [x_1, x_2, x_3]_{n \times 3}$  和  $\mathbf{Y} = [y]_{n \times 1}$ , 样本个数均为  $n = 30$ 。

$$x_1 = \text{randn}(n, 1) \quad (17)$$

$$x_2 = \text{randn}(n, 1) \quad (18)$$

$$x_3 = 1.2x_1^2 + 3x_2 + 0.01\text{randn}(n, 1) \quad (19)$$

$$y = 1.5x_1 + x_2^2 - 0.4x_3 + 0.02\text{randn}(n, 1) \quad (20)$$

$x_3$  中第 21 到 30 的数据缺失,对缺失数据进行估计,并对重构后的数据进行样条变换,对数据进行扩维。并对基于缺失数据估计的 PLSS 预测、基于 PLSS 删除缺失数据项模型预测、基于 PLSS 缺失值补 0 后做预测以及 BP 神经网络预测方法进行比较。

图 2 给出了不同预测方法的仿真结果。其中,横坐标表示预测样本,纵坐标表示样本对应的预测输出。

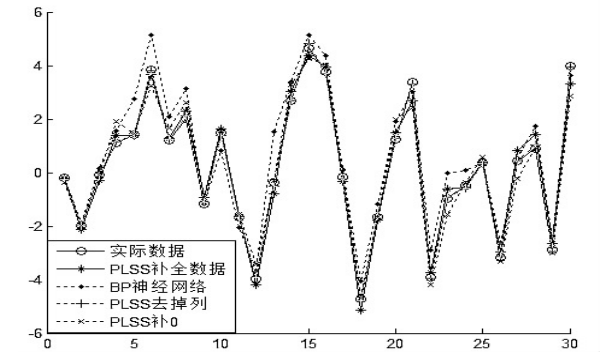


图 2 不同预测方法的仿真结果

图 2 表明,基于缺失数据估计的 PLSS 预测相比其他算法与实际值更吻合。

表 1 给出了不同预测方法所得相对预测误差比较,其中相对预测误差通过式(21)定义。

$$RPE = \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{\sum_{i=1}^n y_i} \tag{21}$$

表 1 不同预测方法预测结果精度比较

模型	RPE
PLSS 补全数据	0.121 4
BP 神经网络	0.297 4
PLSS 删除列	0.351 7
PLSS 补 0	0.166 5

表 1 表明,基于缺失数据估计的 PLSS 方法预测精度更高,预测值更接近于期望输出,且相对预测误差较小。

3.2 实验验证

本节采用历史建筑木结构殿堂式结构,采集到观测变量有 6 个,应力水平  $x_1$  和腐朽深度  $x_2$ 、虫蛀面积的均值  $x_3$ 、变异系数  $x_4$  和开裂深度的均值  $x_5$ 、变异系数  $x_6$ <sup>[1]</sup> 作为自变量  $X_0 = [x_1, x_2, \cdots, x_6]$ ;寿命预测年数作为因变量  $Y = [y_1]_{n \times 1}$ ,如表 2 所示。

搜集到的数据缺失  $x_6$  中第 21 到 28 行数据,对缺失数据进行估计,并对重构后的数据进行样条变换,取  $M_j = p(j = 1, 2, \cdots, p)$ ,进行自变量空间的 3 次 B 样条变换,经过变换后,自变量空间由  $p = 6$  维增至  $p = \sum_{j=0}^p (M_j + 3) = 54$  维。经变换后,新变量间存在线性相关性。对基于缺失数据估计的 PLSS 预测、基于 PLSS 删除缺失数据项模型预测、基于 PLSS 缺失值补 0 后做预测以

及 BP 神经网络预测方法进行比较。图 3 给出了不同预测方法的实测数据结果。其中,横坐标表示预测的样本编号,纵坐标表示样本对应的预测输出。

表 2 实际采集数据

编号	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$y_1$
1	0.003 074 3	30	0.2	0.25	15	0.4	778
2	0.002 515 4	25	0.05	0.12	5	0.3	1 985
3	0.026 108	52	0.05	0.6	20	0.33	736
4	0.091 176	44	0.01	0.005	11	0.45	975
5	0.003 074 3	21	0.08	0.1	20	0.5	1 614
6	0.002 515 4	40	0.12	0.3	30	0.42	968
7	0.026 108	20	0.02	0.55	50	0.34	2 706
8	0.006 735	35	0	0	30	0.33	2 677
9	0.005 512 4	50	0	0	25	0.2	2 054
10	0.057 475	40	0.02	0.14	18	0.5	1 109
11	0.197 82	20	0.03	0.23	12	0.22	1 466
12	0.006 731 9	47	0.04	0.2	20	0.21	1 206
13	0.005 509 5	23	0	0	10	0.41	4 819
14	0.057 481	42	0.08	0.33	5	0.36	839
15	0.006 732 5	34	0.01	0.15	50	0.54	2 221
16	0.005 510 7	45	0.02	0.24	10	0.30	1 668
17	0.057 474	15	0.05	0.252	35	0.24	2 144
18	0.197 82	35	0.04	0.1	30	0.36	469
19	0.006 731 9	43	0.01	0.21	20	0.22	1 768
20	0.005 509 5	32	0.02	0.11	40	0.25	2 239
21	0.057 482	19	0.3	0.16	10		616
22	0.003 074 3	20	0.1	0.15	15		1 455
23	0.002 515 4	40	0.1	0.24	5		1 065
24	0.026 108	24	0	0	20		2 765
25	0.091 176	21	0	0	11		2 421
26	0.003 074 3	35	0.25	0.005	20		627
27	0.002 515 4	21	0.3	0.1	30		680
28	0.026 108	32	0	0	50		1 980

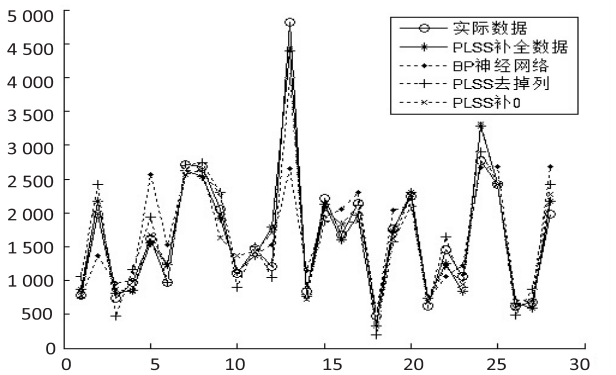


图 3 不同预测方法的实测数据结果

图 3 表明,基于缺失数据估计的 PLSS 木构古建筑



预测值相比较其他算法与实际寿命值更吻合。

表3给出几种不同预测算法的相对预测误差比较。

表3 实际采集数据预测方法精度比较

模型	RPE
PLSS 补全数据	0.094 1
BP 神经网络	0.190 1
PLSS 删除列	0.120 3
PLSS 补0	0.103 7

从表3中看出,基于缺失数据估计的 PLSS 方法对木构古建筑的寿命预测优于其他三种模型,预测值更接近于期望输出,预测精度更高。

4 结束语

木构古建筑实际采集数据存在数据缺失情况下,为了充分提取现有观测数据中的信息,建立了基于缺失数据估计的 PLSS 寿命预测方法来实现变量间非线性较强情况下的木构古建筑的寿命预测。经仿真测试或实验验证均表明了该方法的有效性,可以更好地为木构古建筑视情维护提供决策依据。

参考文献:

[1] 王雪亮. 历史建筑木结构基于可靠度理论的剩余寿命评估方法研究[D]. 武汉:武汉理工大学,2008.

[2] 刘泽谦,付 宏,王 明. 隆里古镇建筑文化遗产的保护与开发[J]. 产业与科技论坛,2015(12):242-243.

[3] 曹 勇. 现代科技在古建筑测绘中的应用[J]. 现代物业:上旬刊,2011(3):54-56.

[4] 姜丛梅. 结合古建筑修缮工程实例浅谈古建筑保护修缮方

法[J]. 江苏建筑,2015(3):57-61.

[5] 周 乾,闫维明,杨小森,等. 汶川地震导致的古建筑震害[J]. 文物保护与考古科学,2010,22(1):37-45.

[6] 王 阳,杨 娜. 基于累积损伤理论的古木构件剩余寿命计算方法[J]. 北京交通大学学报:自然科学版,2015,39(1):45-51.

[7] 王 阳. 殿堂式古建木构有限元分析与可靠度评估[D]. 北京:北京交通大学,2014.

[8] 路 杨,李鹏珊,翟盼盼. 改进 BP 神经网络在木构古建筑中的寿命预测[J]. 计算机技术与发展,2014,24(5):207-210.

[9] 李 瑜. 古建筑木构件基于累积损伤模型的剩余寿命评估[D]. 武汉:武汉理工大学,2008.

[10] 岳元元. 清代徽派木结构古民居塑性有限元分析[J]. 安徽建筑,2015,22(3):143-145.

[11] 王惠文,吴载斌,孟 洁. 偏最小二乘回归的线性与非线性方法[M]. 北京:国防工业出版社,2006:191-215.

[12] 王惠文. 偏小二乘回归方法及应用[M]. 北京:国防工业出版社,1999:150-171.

[13] Nelson P R C, Taylor P A, MacGregor J F. Missing data methods in PCA and PLS: score calculations with incomplete observations[J]. Chemometrics and Intelligent Laboratory Systems, 1996, 35(1):45-65.

[14] 刘 强,尹 力. 一种简化递推偏最小二乘建模算法及其应用[J]. 北京航空航天大学学报,2003,29(7):640-643.

[15] 吴晓华,陈德钊. 化学计量学非线性偏最小二乘算法进展评述[J]. 分析化学,2004,32(4):534-540.

[16] Muteki K, MacGregor J F, Ueda T. Estimation of missing data using latent variable methods with auxiliary information[J]. Chemometrics and Intelligent Laboratory Systems, 2004, 781: 41-50.