

文章编号: 2095-2163(2019)03-0089-04

中图分类号: TP181

文献标志码: A

基于支持向量机和 Bootstrap 的粮仓建筑气密性区间预测方法

刘震华¹, 张梦歌¹, 姜楠¹, 童沪琨², 李建平¹

(1 河南工业大学 土木建筑学院, 郑州 450001; 2 河南工业大学 信息科学与工程学院, 郑州 450001)

摘要: 粮仓建筑的气密性对于储粮安全具有重要影响, 本研究的目的在于提供一种可以用于在粮仓设计阶段进行粮仓气密性预测的方法, 以方便设计人员根据粮仓设计方案的气密性预测结果优化设计, 确保粮仓建成后能够符合气密性要求。本研究采用的样本数据集包括已建成粮仓的建筑特征变量和气密性检测结果, 将该样本集合随机划分为训练集合和测试集合, 利用训练数据集训练支持向量机回归模型, 并采用 Bootstrap 方法进行训练样本抽样, 从而实现了对粮仓气密性的区间预测。测试结果表明该方法具有良好的性能, 为改进粮仓设计方案提供了一种新工具。

关键词: 粮仓; 气密性; 区间预测; 支持向量机; Bootstrap

Method for interval prediction of granary airtightness based on Support Vector Machine and Bootstrap

LIU Zhenhua¹, ZHANG Mengge¹, JIANG Nan¹, TONG Hukun², LI Jianping¹

(1 College of Civil Engineering and Architecture, Henan University of Technology, Zhengzhou 450001, China; 2 College of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China)

[Abstract] Airtightness of granary has an important effect on grain storage safety, the aim of the paper is to provide a method which can be used in the prediction of the granary airtightness during the design stage. According to the prediction of granary airtightness, designers can improve the designs and make sure that the granaries in design will meet the airtightness requirements after they are built. In the paper, data including the architectural characteristic variables and airtightness test results of built granaries are taken as the sample data set, and the sample data set is randomly divided into the training set and the testing set. The training set is employed to train the Support Vector Machine regression model, and the Bootstrap method is adopted to sample with replacement so as to acquire the interval prediction of granary airtightness. The test results show that the method mentioned in the paper has good performance. In a word, the paper provides a new tool for improving the design schemes of granaries.

[Key words] granary; airtightness; interval prediction; Support Vector Machine; Bootstrap

0 引言

粮仓的气密性是指仓房对气体的密封性能^[1], 粮仓具有良好的气密性是安全储粮的重要保证。目前, 气控储粮是中国的一种重要储粮措施, 气控储粮是通过改变储粮环境的气体配比, 来达到杀虫、抑霉等效果, 这就要求粮仓具有良好的气密性^[2]。此外, 粮食含水量的变化, 粮仓的保温隔热性能, 环流熏蒸的实施等都与粮仓的气密性密切相关。因此, 粮仓的气密性需满足一定的要求, 以此来保证储粮效果^[3]。

粮仓气密性的检测方法有压力衰减实验法、示踪剂浓度衰减实验法和平衡压力—流量平衡实验法, 在国内外应用范围较广的是压力衰减实验

法^[4]。

粮仓的气密性与粮仓易漏气部位的密封性有很大关系^[5]。仓房的漏气部位归纳起来, 主要包括门窗设计、墙体设计、屋面设计、密封类孔洞等^[6]。

为了提高粮仓的气密性, 学者们进行了大量的研究。张来林等人^[7]提出从选用合理的仓房结构, 严把施工安装质量等方面出发, 注重建设高质量仓房, 为气密性达标的新仓建设提供了保障。陶金亚等人^[8]从选用气密好的设计方案入手, 通过加强建筑设计 with 仓储工艺的有机结合等措施, 提高了仓房的气密性。王薇^[9]通过对国内现有平房仓气密性能进行调研分析, 利用新材料、新技术, 研制出了一系列行之有效的仓房气密的新型构造措施, 在实际工程中得到了很好的应用。

基金项目: 国家粮食行业公益项目(201513001-03)。

作者简介: 刘震华(1998-), 男, 本科生, 主要研究方向: 土木工程建造; 张梦歌(1998-), 女, 本科生, 主要研究方向: 土木工程建造; 姜楠(1999-), 女, 本科生, 主要研究方向: 土木工程建造; 童沪琨(1998-), 男, 本科生, 主要研究方向: 软件应用技术; 李建平(1974-), 男, 博士, 副教授, 主要研究方向: 工程智能计算。

收稿日期: 2019-03-08

上述研究在提高粮仓气密性方面做出了一定贡献,但是并没有分析粮仓气密性影响因素和气密性之间的定量关系。本文提出了一种气密性评价方法,该方法利用支持向量机并结合 Bootstrap 方法对粮仓建筑设计方案的气密性进行区间预测,通过输入粮仓的一些设计参数值,可以得出粮仓的气密性区间预测结果,从而为粮仓的设计提供参考。

将人工智能技术应用于建筑设计已成为建筑行业发展的新趋势。文献[10]采用神经网络结合遗传算法来降低算法的复杂性,优化建筑设计。文献[11-12]研究分析了神经网络在民用建筑设计工程中的应用。同时,支持向量机被应用于预测建筑能耗^[13]。支持向量机也被用于快速、准确地预测工程造价^[14]。

由于获得大量已建成粮仓建筑的气密性数据难度较大,将少量的粮仓建筑气密性数据用于训练神经网络,进行点预测会有较大的误差。本研究采用支持向量机作为回归模型,为了能够估计出气密性真实值的变动范围,采用 Bootstrap 方法进行训练样本有放回抽样,实现了对粮仓建筑气密性的区间预测。

1 理论基础

1.1 支持向量机

支持向量机是一种基于统计学的机器学习算法,能够有效地处理小样本数据,可以避免局部最小值,具有较好的泛化能力^[15-16]。支持向量机可以完成分类、回归和分布估计等任务^[17]。关于支持向量机回归的标准形式见文献[17]。

1.2 Bootstrap 基本原理和方法

用训练数据通过回归模型预测结果变量值的时候,模型输出值会有一定的随机性,如果知道了模型输出值的概率分布就可以估计出模型输出值在一定置信概率水平下的波动范围,这就是区间预测。本研究采用 Bootstrap 方法对粮仓气密性进行区间预测,Bootstrap 方法由美国统计学家 Efron 系统地提出^[18]。Bootstrap 应用了有放回抽样技术,可以用于机器学习模型的区间预测^[19-20],Bootstrap 的基本方法如下:

有一个实际观测到的数据集 Φ , 含有 l 条观测数据 $\Phi = \{(x_1, z_1), \dots, (x_l, z_l)\}$, 其中 $x_i \in R^n$ 是原因变量值, $z_i \in R^1$ 是结果变量值。从这个数据集中有放回地随机抽取 l 个观测数据组成一个样本,称为抽样样本。在这个随机抽样中,原始数据集中的

观测数据有的只被抽到一次,有的超过一次,也有的没有被抽到。利用抽样样本,代入某个统计量计算公式 $f(\cdot)$, 得到统计量值 k_i 。如此反复抽样和计算 B 次,最后由统计量值组成一个数据集,利用这个数据集来反映估计值的概率分布和随机性特征。

2 评价模型的建立和算法

2.1 数据集

本研究采用的数据来自对华南、华中和东北地区的粮仓的调查结果,经过整理,共 104 条数据。数据集包括 18 个原因变量和 1 个结果变量。原因变量分别是:墙体结构层厚度、仓型、单仓进(卸)粮口个数、单仓容积、单仓建筑面积、单仓层高、单仓轴流风机口个数、门窗密封措施、墙体结构层类型、墙体防潮层高度、单仓环流熏蒸孔个数、单仓机械通风口个数、单仓门窗面积、单仓自然通风口个数、仓体通风形式、屋顶结构形式、墙体防潮层做法、粮仓地面做法,其中仓型、门窗密封措施、墙体结构层类型、仓体通风形式、屋顶结构形式、墙体防潮层做法、粮仓地面做法为种类变量,其他为连续变量;结果变量是:500 Pa 压力半衰期,为连续变量。

在数据预处理阶段,对连续变量数据进行标准化,标准化计算公式如下:

$$\hat{x}_{ij} = \frac{x_{ij} - \min_j(x_{ij})}{\max_j(x_{ij}) - \min_j(x_{ij})}. \quad (1)$$

利用公式(1),将连续变量数据映射到区间[0,1]。对于种类变量,采取 onehot 编码^[21]。

2.2 基于支持向量机和 Bootstrap 的区间预测方法

基于支持向量机和 Bootstrap 的区间预测算法如下:

输入 原始数据集 Φ , $\Phi = \{(x_1, z_1), \dots, (x_l, z_l)\}$, 其中 $x_i \in R^n$ 是原因变量值, $z_i \in R^1$ 是结果变量值, $l \in N^+$; 置信概率 p , $0 \leq p \leq 1$; Bootstrap 抽样次数 B , $B \in N^+$

输出 区间预测的结果 $[z_{i,low}^b, z_{i,up}^b]$

Step 1 数据集划分。将原始数据集随机划分为训练集合 Φ' 和测试集合 Φ'' , Φ' 中有 s 条数据, Φ'' 中有 k 条数据, $s + k = l$ 。

Step 2 数据预处理。对原因变量中的连续变量进行归一化,对于原因变量中的种类变量进行 onehot 编码。

Step 3 Bootstrap 抽样。从 Φ' 有放回地抽取 s 条数据,构成 Bootstrap 样本,共抽取 B 次得到 B 个

样本集合: $\Phi_1^b, \dots, \Phi_B^b$ 。

Step 4 训练回归模型。将 Step3 中得到的 B 个样本集合 $\Phi_1^b, \dots, \Phi_B^b$ 用于训练支持向量机,得到回归模型 $f_1^b(\cdot), \dots, f_B^b(\cdot)$ 。

Step 5 测试集合数据点预测。把测试集合中第 i 条数据代入回归模型 $f_1^b(\cdot), \dots, f_B^b(\cdot)$, 得到 B 个点预测值 $\{z_{i1}^b, \dots, z_{iB}^b\}$, 其中 $z_{ij}^b = f_j^b(x_i^v)$, $(x_i^v, z_i^v) \in \Phi^v$ 。

Step 6 测试集合数据区间预测。对于测试集合中的数据 (x_i^v, z_i^v) , 将 B 个点预测值 $\{z_{i1}^b, \dots, z_{iB}^b\}$ 由小到大排序, 得到 $\{\bar{z}_{i1}^b, \dots, \bar{z}_{iB}^b\}$, 令 $\alpha = 1 - p$, 令 $d = B \times \frac{\alpha}{2}, u = B \times (1 - \frac{\alpha}{2})$, 则可求得区间预测^[20,22]:

区间的上限 $z_{i,up}^b = \bar{z}_{iu}^b$, 区间的下限 $z_{i,low}^b = \bar{z}_{id}^b$ 。对于测试集合中的数据 (x_i^v, z_i^v) , 区间预测结果为 $[z_{i,low}^b, z_{i,up}^b]$, 即 $P(z_{i,low}^b \leq z_i^v \leq z_{i,up}^b) = p$ 。

Step 7 按照下式评价区间预测的质量^[23]:

(1) 范围概率 (Prediction interval coverage probability, $PICP$), 具体公式为:

$$PICP = \frac{1}{k} \sum_{i=1}^k \xi_i, \quad (2)$$

其中, $\xi_i = \begin{cases} 1, & z_{i,low}^b \leq z_i^v \leq z_{i,up}^b; \\ 0, & \text{其他.} \end{cases}$

(2) 归一化区间宽度 (PI-normalized averaged width, $PINAW$), 具体公式为:

$$PINAW = \frac{1}{k \cdot R} \sum_{i=1}^k (z_{i,up}^b - z_{i,low}^b), \quad (3)$$

其中, $R = \max_i(z_i) - \min_i(z_i)$ 。

(3) 综合指标 (Coverage width-based criterion, CWC), 具体公式为:

$$CWC = PINAW + \gamma(PICP) e^{\eta(\mu - PICP)}, \quad (4)$$

其中, $\gamma(PICP) = \begin{cases} 0, & PICP \geq \mu \\ 1, & PICP < \mu \end{cases}; \eta = 50; \mu =$

p 。一般来说, 在 $PICP \geq \mu$ 条件下, $PINAW$ 越小越优; CWC 是综合评价区间预测的质量的指标, CWC 越小说明区间预测质量越好。

如果评价指标能够满足要求, 进入 Step8, 如果不能满足要求, 需要调整区间预测模型。

Step 8 区间预测。将待评价的粮仓设计方案的数据 x^* 代入 Step4 得到的回归模型 $f_1^b(\cdot), \dots, f_B^b(\cdot)$, 得到 B 个点预测值 $\{z_1^*, \dots, z_B^*\}$, 按照 Step6 的方法进行区间预测。如果对粮仓气密性要求较严格, 可以根据区间预测下限判断是否满足要求。

3 计算结果

本研究在 104 条原始数据中, 随机选取 20 条数据组成测试集合, 另外 84 条数据构成训练集合, Bootstrap 抽样次数为 1 000 次, 置信概率为 0.9。作为对比, 采用同样的数据, 将 BP 神经网络作为回归工具, 利用上述算法进行了区间预测。BP 神经网络共 3 层, 输入层有 18 个神经元, 输出层有 1 个神经元, 中间层有 10 个神经元, 计算结果评价指标见表 1。

表 1 区间预测评价指标对照

Tab. 1 Comparison of prediction interval evaluation indexes

计算模型	评价指标		
	$PICP$	$PINAW$	CWC
支持向量机和 Bootstrap 方法	1	0.102 6	0.102 6
BP 神经网络和 Bootstrap 方法	1	0.350 5	0.350 5

由于样本数量较少, 基于支持向量机和 Bootstrap 的区间预测方法的表现优于基于 BP 神经网络和 Bootstrap 的区间预测方法。

4 结束语

因为获得大量粮仓气密性数据较为困难, 本文采用支持向量机结合 Bootstrap 方法建立了一种区间预测方法, 该方法可以采用较少的样本数据对粮仓建筑气密性进行区间预测, 计算结果表明该方法具有良好的性能, 可以为设计人员改进粮仓设计方案, 提高粮仓气密性提供帮助, 从而有助于提高粮食的储存质量, 减少粮食产后损失。

参考文献

- [1] 中华人民共和国国家质量监督检验检疫总局. GB/T25229-2010 粮油储藏平房仓气密性要求[S]. 北京: 中国标准出版社, 2010.
- [2] 李宝升, 李岩峰, 凌才青, 等. 气调储粮技术的发展与应用研究[J]. 粮食加工, 2015, 40(5): 71-74, 77.
- [3] 陈旭, 张峻岭. 粮仓气密性检测对储粮安全的重要性分析[J]. 粮油仓储科技通讯, 2017, 33(6): 30-33.
- [4] 张峻岭, 陈旭, 孙慧男. 论粮仓气密性对储粮安全的重要性[J]. 现代食品, 2017(18): 22-26.
- [5] 张来林, 李超彬, 赵英杰. 粮仓的气密性测试与气密改造[J]. 粮食储藏, 2003, 32(4): 15-18.
- [6] 刘佳, 戴春景, 许隽, 等. 粮食平房仓气密性影响因素及改进措施[J]. 粮食与食品工业, 2010, 17(4): 33-35.
- [7] 张来林, 赵英杰, 狄彦芳, 等. 粮食仓房气密性测试与 CM 系列仓用密封门窗应用报告[J]. 粮食储藏, 2000, 29(6): 18-24.
- [8] 陶金亚, 张来林, 黄浙文, 等. 粮食仓房的气密性分析[J]. 现代食品, 2016(16): 26-30.
- [9] 王薇. 粮食仓房的保温隔热、气密性技术研究及应用[D]. 天津: 天津大学, 2009.