

基于 RS—LS—SVM 的建筑物室内空气品质评价研究

于军琪¹,王 佳²

YU Jun-qi¹,WANG Jia²

1.西安建筑科技大学 信息与控制工程学院,西安 710055

2.北京市住宅建筑设计研究院有限公司 项目设计部,北京 100005

1.School of Information and Control Engineering,Xi'an University of Architecture and Technology,Xi'an 710055,China

2.Department of Project Design,Beijing Institute of Residential Building Design & Research Co.,Ltd,Beijing 100005,China

E-mail:xiaojia@163.com

YU Jun-qi,WANG Jia.Evaluation of building indoor air quality based on RS—LS—SVM.Computer Engineering and Applications,2009,45(16):235-237.

Abstract: An arithmetic for evaluation of indoor air quality was proposed based on composite of Rough Set(RS)—Least-Squares Support Vector Machine(LS-SVM).Six monitoring indexes of indoor air quality are utilized as the evaluation items,use the RS to reduce the properties of indoor air quality monitoring data,eliminate the redundant information,and use the reductive regulations to train the LS-SVM,achieve the satisfactory precision.The simulation results indicate that the method has better performance of convergence speed and nonlinear approaching,and gives a real-time and accurate result of evaluation of indoor air quality,it provides a scientific basis for building indoor environment monitor and pollution control.

Key words: Rough Set(RS);Least-Squares Support Vector Machine(LS-SVM);evaluation of indoor air quality

摘 要:提出一种基于粗糙集(Rough Set,RS)—最小二乘支持向量机(Least-squares-Support Vector Machine,LS-SVM)复合的建筑物室内空气品质评价方法,选取描述室内空气品质的六项监测指标作为评价因子,利用 RS 理论,对室内空气品质监测数据进行属性约简,消除冗余信息,用约简后的规则集对 LS-SVM 进行训练,使其达到满意精度。实验仿真表明:该复合方法具有良好的收敛速度与非线性逼近能力,能对室内空气品质进行实时、准确的评价,为建筑物室内空气品质监测、环境污染治理提供科学依据。

关键词:粗糙集;最小二乘支持向量机;室内空气品质评价

DOI:10.3778/j.issn.1002-8331.2009.16.068 **文章编号:**1002-8331(2009)16-0235-03 **文献标识码:**A **中图分类号:**TP391;TU023

1 前言

由于建筑装修中有机合成材料的广泛使用,使建筑物室内空气品质不尽人意,已经无形地影响了人们的健康、工作效率以及生活质量。如何构建一个健康、空气清新的生活工作环境,已备受关注;关于室内空气品质评价的研究,显得愈加迫切。该课题的研究,能使人们提早预防,尽快治理,为建设一个良好的工作、生活环境打下基础。

迄今为止,关于建筑物室内空气品质评价的研究方法主要包括基于模糊^[1]、灰色^[2]以及 BP 神经网络理论^[3]建立的综合评价方法。但基于模糊的方法信息损失多,权重赋值上具有较大的主观随意性;灰色聚类法在有些情形下分辨率不够;BP 算法则存在收敛速度较慢且容易陷入局部极小值,以上缺点使得这些方法在应用中受到一定的限制。

针对上述问题,提出 RS 与 LS-SVM 复合的室内空气品质评价方法。利用 RS 对采集的数据样本集进行处理,去除冗余信息,再用约简后的规则集训练 LS-SVM 进行室内空气品质评

价。这样不但降低了计算复杂性,而且提高了室内空气品质评价的准确性。

2 RS—LS—SVM 评价模型

图 1 所示为 RS—LS—SVM 复合评价模型的获得过程,首先利用 RS 理论对采集的建筑物室内空气品质数据进行约简处理,得到约简的规则集,并将其用于训练 LS-SVM 网络,训练好的 LS-SVM 就可以对室内空气品质进行评价。

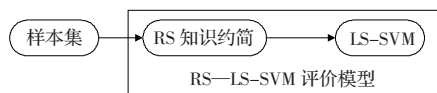


图 1 基于 RS—LS—SVM 的评价模型

2.1 RS 知识约简

自波兰 Z.Pawlak 教授提出 RS 理论后,因其在保持分类能力不变的前提下,不需要任何先验知识,通过知识约简,消除冗

基金项目:陕西省自然科学基金基础研究基金(the Natural Science Basic Research Foundation of Shaanxi Province of China under Grant No.SJ08F30);陕西省教育厅专项科研项目(No.08JK321)。

作者简介:于军琪(1969-),男,博士,教授,主研领域:建筑智能化与智能信息处理,数据融合等;王佳(1985-),女,硕士生,主研方向:建筑智能化。

收稿日期:2008-11-17 **修回日期:**2009-02-09

余属性,即可求出问题的分类规则集,因而被广泛应用到有关模式识别的各类评价问题中。

一个室内空气品质评价数据表可以映射为 RS 表示的知识系统 $S=\langle U, A, V, f \rangle$, 其中, U 为论域, $A=C \cup D$ 是属性集合, 子集 C 称为条件属性, D 称为决策属性, $C \cap D = \emptyset$, $V = \bigcup_{r \in R} V_r$ 是属性取值集合, V_r 表示属性 r 的值域。 $f: U \times R \rightarrow V$ 是一个信息函数。下面给出 RS 最佳知识约简算法的步骤^[9]:

输入: 具有条件属性集 $C=\{a_1, a_2, \dots, a_n\}$ 和决策属性 $D=\{Dec\}$ 的知识表达系统 S ;

- 步骤 1 计算 D 的 C 正域 $pos_c(D)$;
- 步骤 2 从 C 中剔除属性 $a_i, C_i=C-\{a_i\}$;
- 步骤 3 计算 D 的 C_i 正域 $pos_{c_i}(D)$;
- 步骤 4 若 $pos_{c_i}(D)=pos_c(D)$, 则去除 a_i , 反之保留 a_i ;
- 步骤 5 按以上步骤遍历所有属性。

输出: 条件属性集 C 对于决策属性 D 的一个相对约简。

2.2 LS-SVM 算法

LS-SVM 是将传统 SVM 中的不等式约束改为等式约束, 且将误差平方和损失函数作为训练集的经验损失, 这样就把解二次规划问题转化为求解线性方程组问题, 提高了求解问题的速度和收敛精度。

经 RS 约简后的规则集即可作为 LS-SVM 的输入输出模式来训练 LS-SVM, 设训练样本为: $(x_1, y_1), \dots, (x_k, y_k), x_i \in R^n$ 是第 i 个样本的输入模式, $y_i \in R$ 是对应于第 i 个样本的期望输出。首先用一个非线性映射 $\Psi(\cdot)$ 把样本从原空间 R^n 映射到特征空间 $\Phi(x_i)$, 在这个高维特征空间中构造最优决策函数:

$$y(x) = w \cdot \Phi(x) + b \tag{1}$$

其中 w 为权向量, b 为偏差量。利用结构风险最小化原则, 寻找最小化目标函数:

$$R = \frac{1}{2} \|w\|^2 + c \cdot R_{emp} \tag{2}$$

其中 $\|w\|^2$ 为控制模型的复杂度, c 为正规化参数, 控制对超出误差样本的惩罚程度。 R_{emp} 为误差控制函数, 也即 ε 不敏感损失函数。LS-SVM 优化目标损失函数为误差 ξ_i 的二次项, 故优化问题为:

$$\min J(w, \xi) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^k \xi_i^2 \tag{3}$$

$$\text{st: } y_i = \Phi(x_i) \cdot w + b + \xi_i, i=1, \dots, k \tag{4}$$

式中, ξ_i 为松弛因子。用拉格朗日法求解这个优化问题:

$$L(w, b, \xi, a, \gamma) = \frac{1}{2} \|w\|^2 + c \sum_{i=1}^k \xi_i^2 - \sum_{i=1}^k a_i (\Phi(x_i) \cdot w + b + \xi_i - y_i) \tag{5}$$

其中: $a_i (i=1, \dots, k)$ 是拉格朗日乘子。

根据优化条件:

$$\frac{\partial L}{\partial w} = 0 \quad \frac{\partial L}{\partial b} = 0 \quad \frac{\partial L}{\partial \xi} = 0 \quad \frac{\partial L}{\partial a} = 0 \tag{6}$$

可得到:

$$\begin{cases} w = \sum_{i=1}^k a_i \Phi(x_i) \\ \sum_{i=1}^k a_i = 0 \end{cases} \tag{7}$$

其中: $a_i = c \cdot \xi_i, w \cdot \Phi(x_i) + b + \xi_i - y_i = 0$ 。

定义核函数 $K(x_i, x_j) = \Phi(x_i)^T \cdot \Phi(x_j)$ 是满足条件的对称函

数, 则优化问题转化为求解线性方程

$$\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & K(x_1, x_1) + 1/c & \dots & K(x_1, x_k) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_k, x_1) & \dots & K(x_k, x_k) + 1/(-c) \end{bmatrix} \cdot \begin{bmatrix} b \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_k \end{bmatrix} \tag{8}$$

最后用最小二乘法求出 a 与 b , LS-SVM 也由此得名, 并且得到非线性预测模型:

$$f(x) = \sum_{i=1}^k a_i K(x, x_i) + b \tag{9}$$

2.3 RS-LS-SVM 评价模型

基于 RS-LS-SVM 模型的建筑物室内空气品质评价流程图如图 2 所示。

- (1) 分析采集的空气品质数据, 选择样本集;
- (2) 利用 RS 方法, 对样本集进行属性值约简, 形成知识表达系统;
- (3) 利用 RS 方法, 对知识表达系统进行属性约简, 形成约简后的规则集;
- (4) 建立 LS-SVM 模型, 选择适当的核函数及其参数, 用约简后的规则集训练 LS-SVM 模型, 直至取得满意效果;
- (5) 用测试样本集对 RS-LS-SVM 评价模型进行验证;
- (6) 整个评价结果分析。

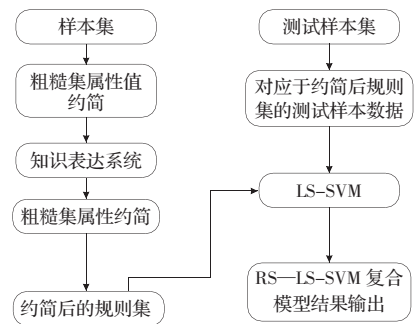


图 2 基于 RS-LS-SVM 的室内空气品质评价流程图

3 RS-LS-SVM 评价模型应用

为了给室内空气品质评价提供基准依据, 根据 GB/T18883-2002 中的标准值, 将室内空气品质分为 5 个等级, 取标准值为未污染值, 标准值的 1/2 为清洁, 超过标准值 1 倍为轻污染, 2 倍为中污染, 5 倍为重污染。文中在评价室内空气品质时, 选取甲醛、苯、二氧化氮(NO₂)、氨气(NH₃)、可吸入颗粒物(PM₁₀)、总挥发性有机化合物(TVOC)作为评价因子。随机抽取某写字楼的 14 个独立用户, 测试数据及评价标准分别见表 1 和表 2。

3.1 RS 知识约简

3.1.1 RS 属性值约简

根据表 2 的评价标准浓度值, 将表 1 中的 14 个采样点的实测值转化为相应的空气品质级别, 并用 $a_1, a_2, a_3, a_4, a_5, a_6$ 以及 Dec 分别代表 6 个条件属性甲醛、苯 NO₂、NH₃、PM₁₀、TVOC 以及决策属性模糊评判法评价结果, 即论域 $U=\{1, 2, \dots, 14\}$ 表示采样点, 条件属性 $C=\{a_1, a_2, a_3, a_4, a_5, a_6\}$, 决策属性 $D=\{Dec\}$ 。然后合并其相同行, 得到室内空气品质知识表达系统, 如表 3 所示。

3.1.2 RS 知识系统属性约简

根据计算最佳约简的算法对室内空气品质知识表达系统

表1 测试数据 (mg/m³)

采样点	甲醛	苯	NO ₂	NH ₃	PM ₁₀	TVOC	模糊评判法 评价结果
1	0.098	0.086	0.215	0.189	0.163	0.560	II
2	0.421	0.088	0.330	0.580	0.501	1.711	IV
3	0.263	0.173	0.602	0.350	0.405	1.512	III
4	0.291	0.162	0.652	0.327	0.416	1.263	III
5	0.543	0.092	0.281	0.557	0.488	1.673	IV
6	0.041	0.082	0.213	0.185	0.273	0.477	II
7	0.478	0.183	0.331	0.325	0.274	1.105	III
8	0.284	0.196	0.550	0.182	0.415	1.452	III
9	0.082	0.075	0.226	0.195	0.142	0.536	II
10	0.233	0.171	0.577	0.289	0.374	1.722	III
11	0.498	0.078	0.321	0.496	0.532	1.702	IV
12	0.223	0.138	0.603	0.311	0.396	1.352	III
13	0.087	0.072	0.221	0.188	0.157	0.471	II
14	0.342	0.215	0.423	0.294	0.223	0.741	III

表2 评价标准

空气品质	评价因子/(mg/m ³)					
	甲醛	苯	NO ₂	NH ₃	PM ₁₀	TVOC
清洁级 I	0.05	0.05	0.12	0.10	0.08	0.30
未污染级 II	0.10	0.10	0.24	0.20	0.15	0.60
轻污染级 III	0.20	0.20	0.48	0.40	0.30	1.20
中污染级 IV	0.30	0.30	0.72	0.60	0.45	1.80
重污染级 V	0.60	0.60	1.44	1.20	0.90	3.60

表3 室内空气品质知识表达系统

采样点	a ₁	a ₂	a ₃	a ₄	a ₅	a ₆	Dec
1	II	II	II	II	III	II	II
2	V	II	III	IV	V	IV	IV
3	IV	III	IV	III	IV	IV	III
4	I	II	II	II	III	II	II
5	V	III	III	III	III	III	III
6	IV	III	IV	II	IV	IV	III
7	II	II	II	II	II	II	II
8	V	IV	III	III	III	III	III

进行属性约简,每次去掉一个条件属性 a_i ,检查 D 在 $C=\{a_i\}$ 的正域是否等于 D 在 C 的正域,如相等,则为冗余属性,删除之。经计算表 3 中 a_3 、 a_6 是可省略的,删除后,得到属性约简后的规则集,如表 4 所示。

表4 约简后的规则集

采样点	a ₁	a ₂	a ₄	a ₅	Dec
1	II	II	II	III	II
2	V	II	IV	V	IV
3	IV	III	III	IV	III
4	I	II	II	III	II
5	V	III	III	III	III
6	IV	III	II	IV	III
7	II	II	II	II	II
8	V	IV	III	III	III

3.2 LS-SVM 设计

在设计 LS-SVM 时,使用 MATLAB 的 LS-SVMlab 工具箱创建 LS-SVM,选用 RBF 核函数,训练建模用函数 trainlssvm,测试用函数 simlssvm。以约简后的规则集中 8 组数据作为 LS-SVM 的输入,以其对应的空气品质级别作为网络的输出,见表 5。根据所选核函数进行相关参数的选择,如松弛变量(ξ)和惩

罚因子(C)等。并采用交叉验证优化参数函数 tunelssvm 来优化参数,最后取得参数 $[\gamma \ \sigma^2]=[0.7067 \ 0.013771]$,仿真过程如图 3 所示。

表5 LS-SVM 的输入输出模式

采样点	甲醛	苯	NH ₃	PM ₁₀	空气品质级别	对应输出
1	0.098	0.086	0.189	0.163	II	2
2	0.421	0.088	0.580	0.501	IV	4
3	0.263	0.173	0.350	0.405	III	3
4	0.041	0.082	0.185	0.273	II	2
5	0.478	0.183	0.325	0.274	III	3
6	0.284	0.196	0.182	0.415	III	3
7	0.082	0.075	0.195	0.142	II	2
8	0.342	0.215	0.294	0.223	III	3

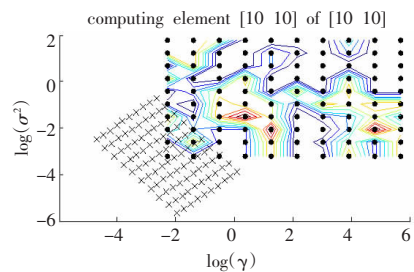


图3 参数优化过程图

3.3 仿真测试

以表 6 中对应于约简后规则集的测试样本数据,输入到训练好的 LS-SVM 网络,最终得到如下的评价结果:

测试结果为:Y=2 3 3 2

表6 测试数据

采样点	甲醛	苯	NH ₃	PM ₁₀	室内空气品质级别
1	0.088	0.079	0.192	0.174	II
2	0.500	0.173	0.307	0.251	III
3	0.355	0.270	0.311	0.215	III
4	0.066	0.073	0.201	0.144	II

可以发现识别正确率完全可以满足室内空气品质评价的要求。再将文中的复合方法与单一的 LS-SVM 相比,如表 7 所示,net1 为未用 RS 对样本数据进行约简,训练得到的 LS-SVM,识别正确率为 75%;net2 为文中方法得到的评价结果,识别正确率达 100%。结果表明,RS—LS-SVM 复合方法优化了网络结构,同时提高了评价结果的准确性。

表7 net1 和 net2 的识别结果

	训练时间	识别率/(%)
net1	1.25	75
net2	0.80	100

4 结语

文中基于 RS—LS-SVM 复合模型的建筑物室内空气品质评价方法,将 RS 对数据的约简能力、LS-SVM 算法良好的非线性逼近能力、运算精度和速度结合起来,比目前已有的评价方法具有良好性能。MATLAB 仿真结果表明:该方法具有良好的实时性、容错性和抗干扰能力,对建筑物空气品质能取得准确、科学的评价结果。