

文章编号: 1673- 9469(2010) 04- 0030- 06

基于 RS- SVM 的建筑施工项目安全预警模型

李万庆, 安娟

(河北工程大学 经管学院, 河北 邯郸 056038)

摘要: 针对目前建筑施工项目 安全风险管理的现状, 应用 粗集理论对建筑施工项目的 安全因素进行预处理, 将处理后的信息 结构作为支持向量机的输入数据进行训练和预测, 构建建筑施工项目 安全风险预警模型, 并在小样本条件下, 与 BP 神经网络进行对比分析。结果表明, RS- SVM 预测模型的最小均方根误差为 0. 011 5, BP 神经网络的均方根误差为 0. 070 7, RS- SVM 预警模型的预测精度、泛化能力明显优越于 BP 神经网络学习方法。

关键词: 建筑施工; 粗集; 支持向量机; 安全预警

中图分类号: X947

文献标识码: A

Safety early- warning model of construction project based on rough set and support vector machine

LI Wan-qing, AN Juan

(School of Economic and Management, Hebei University of Engineering, Hebei Handan 056038, China)

Abstract: Aimed at the uncertainty of safety risk prediction in construction project currently, Rough set was used to pretreat the safety factors of construction project, and then the pretreatment results were put as the information structure of support vector machine (SVM) to training and prediction. So the construction project safety risk early- warning model was established, and the validity of the algorithm was proved by empirical analysis. The MSE of SVM is 0. 011 5, and the BP neural network is 0. 070 7. Experimental results revealed that the model of rough set - support vector machine (RS- SVM) can realize to delete the superfluous inputting information, to reduce the complexity and improve the interfere resistance. In small sample conditions, the early- warning model of RS- SVM is greatly superior to BP neural network in prediction precision and generalization ability.

Key words: construction project; rough set; support vector machine; safety early- warning

建筑施工项目是安全风险最密集、安全事故发生率最多的领域之一, 加强施工项目的安全管理与预警有着现实的政治、经济和社会意义^[1- 3]。目前国内外关于建筑施工项目安全预警的方法主要有趋势外推法、动态层次分析法、模糊综合分析法、故障树分析法、神经网络等方法。然而趋势外推模型虽然简单易懂, 但在建筑工程上的使用范围却很小; DAHP 模型中专家评估的风险要素带有很大的主观性; 模糊分析法的运算精度依然有待商榷; 神经网络虽具有很好的非线性逼近能力, 但常常出现过度训练或训练不足; 故障树分析法则

对事故类型复杂的系统处理效果欠佳。

粗集(rough set, RS)是一种处理模糊和不确定知识的数据分析理论, 能在保持知识库分类能力不变的条件下, 删除其中的不相关或不重要的知识, 提取满足条件的最小属性集^[4]。支持向量机(Support Vector Machine, SVM)是近年来在结构风险最小化基础之上发展起来的统计学习理论, 它可以按任意精度逼近非线性函数, 具有全局极小点和收敛速度快的优点。结合 RS 的属性约简和 SVM 的回归机理, 提出了基于 RS- SVM 的建筑施工项目安全风险预警模型, 并与 BP 神经网络进行

实证对比分析。

1 RS- SVM 模型构建的基础理论

1.1 RS 基本理论

在 RS 理论中,“知识”被认为是一种将现实或抽象的对象进行分类的能力。一个知识表达系统(决策信息表达系统,知识表达系统)由有限个研究对象组成,这些对象的知识通过指定对象的属性和属性值来描述,其表达式为

$$S \quad (U, V, A, F)$$
$$V \quad \bigcup_{a \in A} V_a$$

式中 S —知识表达系统; U —论域; A —属性集; a —属性; V_a —属性 a 的值域; F —信息函数,指定论域 U 中每一对象的属性值。

运用 RS 理论处理决策表时,要求决策表中的值用离散化数据表达,由于本文数据较集中故采用 C —均值聚类算法。对于定量指标中连续数据的离散分类,离散后的区间数值由小到大分别用数字 0、1、2、3 对应表示;对于定性指标,分别用 0、1、2 表示差、中、优;决策属性值以是否完成安全目标分为 0、1,0 表示未完成安全目标,1 表示完成安全目标。将项目的风险等级划分为 5 级,分别用 $[0, 0.3], (0.3, 0.5], (0.5, 0.7], (0.7, 0.9]$ 及 $(0.9, 1]$ 表示低风险、较低风险、一般风险、较高风险及高风险。

1.2 SVM 基本原理

SVM 是针对结构风险最小化原则提出的,改变了传统的经验风险最小化原则,具有很好的泛化能力^[5]。SVM 在处理非线性问题时,首先将非线性问题转化为高维空间中的线性问题,然后用一个核函数来代替高维空间中的内积运算,从而巧妙地解决了复杂计算问题,并且有效地克服了维数灾难及局部极小问题。

给定训练数据

$$\{(x_i, y_i), i = 1, 2, \dots, l\}$$

式中 x_i —第 i 个学习样本的输入值, $x_i = [x_i^1, x_i^2, \dots, x_i^d]^T, x_i \in R^d; y_i$ —对应的目标值, $y_i \in R$ 。

采用一个非线性映射将样本从原空间映射到维数为 K (K 可能是无穷大)的高维特征空间中,然后在高维特征空间中进行线性回归。设回归函数为

$$\min \left[\frac{1}{2} \| \omega \|^2 + C \frac{1}{l} \sum_{i=1}^l (\zeta_i + \zeta_i^*) \right]$$
$$s. \ t. \begin{cases} y_i - \omega \bullet \phi(x_i) - b \leq \varepsilon + \zeta_i \\ \omega \bullet \phi(x_i) + b - y_i \leq \varepsilon + \zeta_i^* \\ \zeta_i, \zeta_i^* \geq 0, i = 1, 2, \dots, l \end{cases}$$

通过拉格朗日变换得到其优化问题的对偶形式

$$\max_{\alpha, \alpha^*} \left[- \frac{1}{2} \sum_i \sum_j (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i \bullet x_j) - \sum_i \varepsilon (\alpha_i + \alpha_i^*) + \sum_i y_i (\alpha_i - \alpha_i^*) \right]$$
$$s. \ t. \begin{cases} \sum_i (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C/l] \end{cases}$$

式中 $K(x_i \bullet x_j)$ —核函数, $K(x_i \bullet x_j) = \phi(x_i) \bullet \phi(x_j)$; C —惩罚参数, $C > 0$; $(\alpha_i - \alpha_i^*)$ —拉格朗日乘子,对应的训练样本为支持向量, $(\alpha_i - \alpha_i^*) \neq 0$ 。

取 $w = \sum_i (\alpha_i - \alpha_i^*) \phi(x_i)$, 然后得到偏置项系数 b

$$b = \frac{1}{N_{Nsv}} \left\{ \sum_{0 \leq \alpha_i \leq C} [y_i - \sum_{x_j \in sv} (\alpha_j - \alpha_j^*) k(x_j, x_i) - \varepsilon] + \sum_{0 \leq \alpha_i^* \leq C} [y_i - \sum_{x_j \in sv} (\alpha_j - \alpha_j^*) k(x_j, x_i) + \varepsilon] \right\} \quad (1)$$

式中 N_{Nsv} —支持向量的个数。

通过训练和学习得到回归预测函数为

$$f(x) = \sum_{x_j \in sv} (\alpha_j - \alpha_j^*) k(x_i, x_j) + b \quad (2)$$

2 RS- SVM 模型的构建

用 RS 对施工项目安全因素进行预处理,减少 SVM 数据维数和降低系统结构的复杂性,即把粗集作为前置系统,再根据 RS 方法预处理后的信息结构,构成 SVM 信息处理系统,从而可构造出具有更好的抑制噪声干扰和良好泛化能力的预测模型。具体的流程如图 1 所示。

2.1 指标体系的选取

在人—机—环境系统理论、人机工程学事故致因理论的基础上,分析了可能影响施工项目安全的因素。并在遵循指标体系构建的科学性、系统性、动态性、客观性和可操作性原则的基础上,结合专家经验,构建了建筑施工项目安全预警指标体系(图 2)。

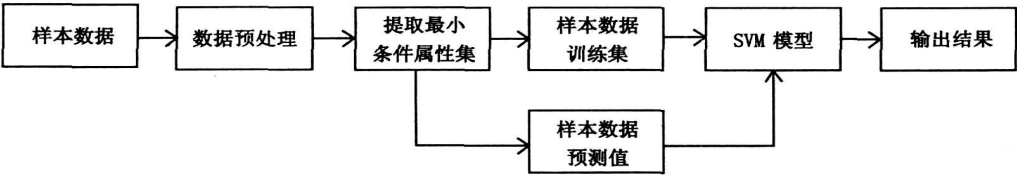


图1 RS-SVM安全预警系统
Fig.1 Safety early-warning system of RS-SVM

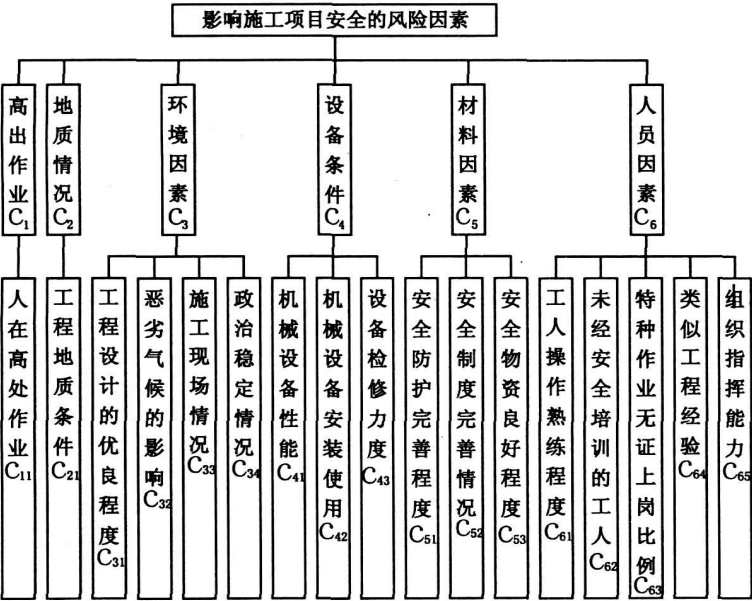


图2 施工项目安全风险因素指标体系
Fig.2 Safety risk factors index of construction project

2.2 建模步骤

- (1) 参考类似工程的项目数据建立训练样本集。
- (2) 采用 C- 均值聚类算法, 对样本数据进行离散, 然后用 RS 对样本集进行约简, 去除不完全数据, 提取最优属性集。
- (3) 利用交叉验证法法选择核函数参数 σ 和惩罚参数 C , 根据式(1)、式(2) 并通过 MATLAB 编程求解模型的参数(w, b)。
- (4) 建立最优预测模型。

2.3 预测误差的确定

采用相对误差函数(Error) 和均方根误差函数(MSE) 来检验模型的预测精度, 相对误差是评价预测模型对每一个样本的测试效果, 而均方根误差是预测模型整体效果的误差检验函数。

$$Error(n) = \frac{|x(n, true) - x(n, pred)|}{|x(n, true)|}$$

(3)

$$MSE = Sqrt\left\{\frac{1}{N} \sum_{n=1}^N [x(n, true) - x(n, pred)]^2\right\}$$

(4)

式中 $x(n, true)$ —第 n 个测试样本的实际值; $x(n, pred)$ —第 n 个测试样本的预测值。

2.4 安全预警等级的划分

将预测模型计算求得的预测值 y' 和实际值 y 相比较, 得到 $|y - y'|$ 。根据概率统计理论, $|y - y'|$ 落入 $[0, 2S]$ 的概率为 95.5%, 落入 $[0, 3S]$ 的概率为 99.7%, 其中 S 为模型标准差^[6]。据此, 可按以下几种情况对施工项目安全性态进行预警:

- (1) 正常: $|y - y'| \leq 2S$ 。
- (2) 基本正常: $2S < |y - y'| \leq 3S$, 且测值无趋势性变化。
- (3) 异常: $2S < |y - y'| \leq 3S$, 但测值有趋势性变化。
- (4) 险情: $|y - y'| > 3S$, 应分析其成因。

3 实证分析

3.1 数据采集

借鉴文献[7] 的数据进行预测研究, 数据如表 1 所示。

3.2 安全预警决策表约简

采用 C- 均值聚类算法对表 1 中的数据进行离散化(其中政治的稳定情况不具有分类能力已被删除), 利用 RS 将离散化后的数据在 MATLAB 7.0 中进行约简, 得到最优属性集 12 个指标, 约简冗余因素 5 个(表 2)。将 RS 约简后的最优属性集作为 SVM 的训练和测试数据, 输入条件属性和决策属性的历史数据, 取前 8 组作为训练样本来进

行机器学习, 从而得到预测模型, 余下的 4 组作为检验样本考察模型的预测和泛化能力。

3.3 核函数及参数的选择

SVM 采用满足 Mercer 条件的核函数来代替映射函数, 将输入变量 x 映射到高维特征空间。常用的核函数有线性核函数、多项式核函数、径向基核函数和神经网络核函数, 就函数拟合而言, 径向基函数(radial basis function, RBF) 能使 SVM 以任意精度逼近非线性函数曲线, 其函数拟合性优于多项式核函数。因此, 采用 RBF 核函数, 参数包括调整惩罚参数 C , 核宽度 σ , 和不敏感系数 ε 。对参数的选取采用交叉验证法, 以最小均方根误差 (MSE_{min}) 为参数选择标准。交叉验证如下:

表 1 建筑施工项目安全因素及历史数据

Tab. 1 Historical date and safety risk factors of construction project

项目	C ₁₁	C ₂₁	C ₃₁	C ₃₂	C ₃₃	C ₃₄	C ₄₁	C ₄₂	C ₄₃	C ₅₁	C ₅₂	C ₅₃	C ₆₁	C ₆₂	C ₆₃	C ₆₄	C ₆₅
1	0.30	82	中	中	优	95	差	82	优	中	0.14	0.12	中	0.6	68	中	稳定
2	0.38	64	差	差	差	84	差	72	中	优	0.15	0.14	中	0.6	50	中	稳定
3	0.20	73	中	优	优	75	中	93	中	中	0.11	0.15	优	0.9	84	中	稳定
4	0.34	83	中	中	优	80	中	91	差	差	0.12	0.05	中	0.9	82	优	稳定
5	0.22	74	中	中	差	92	优	94	中	中	0.10	0.12	中	0.8	88	优	稳定
6	0.32	94	优	差	优	80	中	95	中	差	0.20	0.13	中	0.6	80	优	稳定
7	0.35	92	中	中	优	90	中	78	中	中	0.00	0.00	中	0.7	81	优	稳定
8	0.34	91	中	中	优	84	差	64	优	中	0.15	0.14	中	0.6	71	中	稳定
9	0.40	75	中	中	中	62	差	69	优	中	0.02	0.04	差	0.4	72	中	稳定
10	0.35	95	中	中	中	94	优	86	优	差	0.13	0.11	优	1.0	90	优	稳定
11	0.25	82	优	差	中	65	中	90	优	中	0.21	0.21	中	0.8	84	中	稳定
12	0.38	96	优	差	优	75	中	94	中	差	0.22	0.12	中	0.7	82	优	稳定

表 2 建筑施工项目安全预警决策表

Tab. 2 Safety early- warning decision table of construction project

项目	C ₁₁	C ₃₁	C ₄₁	C ₄₂	C ₄₃	C ₅₁	C ₅₂	C ₆₁	C ₆₂	C ₆₃	C ₆₄	C ₆₅	决策属性值
1	0.7	0.3	0.3	0.3	0.5	0.3	0.5	0.7	0.7	0.5	0.5	0.3	0.420 3
2	0.9	0.7	0.7	0.3	0.5	0.5	0.5	0.7	0.7	0.5	0.3	0.7	0.521 8
3	0.9	0.9	0.9	0.9	0.3	0.5	0.9	0.9	0.9	0.9	0.3	0.9	0.745 9
4	0.7	0.7	0.9	0.7	0.5	0.7	0.5	0.7	0.9	0.7	0.5	0.7	0.694 5
5	0.3	0.7	0.5	0.1	0.3	0.5	0.1	0.3	0.3	0.1	0.5	0.3	0.361 2
6	0.7	0.9	0.9	0.9	0.5	0.5	0.9	0.7	0.9	0.7	0.3	0.9	0.702 9
7	0.9	0.7	0.7	0.5	0.5	0.5	0.7	0.7	0.9	0.9	0.5	0.7	0.592 8
8	0.7	0.9	0.7	0.5	0.1	0.1	0.1	0.3	0.5	0.5	0.5	0.3	0.464 5
9	0.5	0.7	0.9	0.3	0.7	0.3	0.1	0.3	0.1	0.1	0.1	0.3	0.359 4
10	0.1	0.3	0.3	0.1	0.3	0.3	0.1	0.3	0.1	0.1	0.3	0.3	0.262 2
11	0.3	0.7	0.9	0.5	0.7	0.9	0.3	0.7	0.7	0.5	0.5	0.7	0.762 3
12	0.7	0.7	0.5	0.3	0.7	0.5	0.3	0.7	0.9	0.3	0.1	0.5	0.532 1

(1) 固定 $\varepsilon = 0.01, C = 90$, 令 σ 在 $[0.01:0.05:10]$ 期间上变动, 计算 SVM 预测的均方根误差, 如图 3 所示。由图可见, 在区间 $[0.01, 1]$ 之间时, 模型训练的均方根误差显著降低, 在区间 $[1, 10]$ 之间则增长缓慢。当 $\sigma = 0.85$ 时, $MSE = 0.0137$ 为最小值。

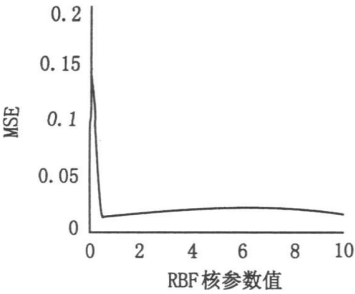


图3 $\varepsilon=0.01, C=90$ 时MSE随 σ 变化曲线
Fig. 3 The σ Variation curves of MSE when $\varepsilon=0.01, C=90$

(2) 固定参数 $C = 90, \sigma = 0.8$, 调整不敏感损失函数值 ε , 如图 4 所示。当 $\varepsilon = 0.003$ 时, 模型测试的均方根误差 $MSE = 0.0137$ 为最小值。

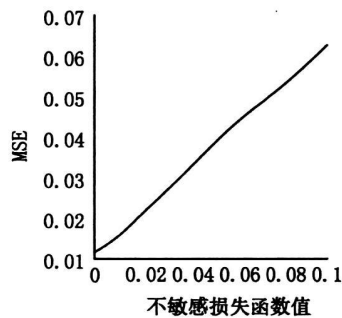


图4 $C=90, \sigma=0.8$ 时, MSE随 ε 变化曲线
Fig. 4 The ε Variation curves of MSE when $\sigma=0.8, C=90$

(3) 固定参数 $\sigma = 0.85$ 以及参数 $\varepsilon = 0.0003$, 使模型的惩罚参数 C 在区间 $(0, 10\ 000]$ 以步长 50 进行变动以测试模型的训练精度。程序的运行结果显示, 实例中惩罚参数 C 的变化对模型的预测精度没有显著影响。

这样, 当取参数 $C = 90, \sigma = 0.8, \varepsilon = 0.0003$ 时, 预测模型的均方根误差最小 $MSE = 0.0115$, 此时的 SVM 模型便为最优模型。

3.4 RS-SVM 模型的建立

根据选取的核函数和最优参数, 在 MATLAB 7.0 语言支持下, 利用式 (1) 计算出偏置项系数 $b = 0.53862$; 支持向量的个数 $N_{\text{Sv}} = 4$, 分别为第 3、8、11、12 个项目样本; 对应的拉格朗日乘子为 m_3

$0.372, m_8 = 0.015, m_{11} = 0.385, m_{12} = 0.044$ 。
将所求参数代入式 (2), 得到 RS-SVM 模型
$$f(x) = \sum (\alpha_i - \alpha_i^*) (x \cdot x_i' + 1) + 0.53862$$

式中 x —待测建筑施工项目的安全指标值向量;
 x_i —支持向量, $i = 3, 8, 11, 12$ 。

3.5 预测结果分析

为检验 RS-SVM 算法的优越性, 将其与 BP 神经网络进行对比分析。在 BP 神经网络模型中, 采用 3 层结构, 输入向量和输出向量同 RS-SVM, 将前 8 项作为训练样本, 后 4 项作为预测样本, 隐层选 20 个神经元。利用神经网络系统对学习数据反复训练, 得到当均方根误差为 0.0707 时, 实验结果最优。两种模型的实际值和预测值的对比曲线见表 3, 拟合预测效果见图 5。

表 3 两种模型的预测值与实际值比较
Tab. 3 Compare by predicted result and fact
of the two models

项目 编号	预测值		实际值	相对误差	
	BP	RS-SVM		BP	RS-SVM
9	0.498 0	0.369 4	0.359 4	0.385 6	0.027 8
10	0.237 8	0.272 2	0.262 2	0.093 0	0.028 1
11	0.751 1	0.752 3	0.762 3	0.009 4	0.013 1
12	0.541 6	0.522 1	0.532 1	0.017 8	0.018 7

由表 3 可以看出, 对于同一个工程项目样本, RS-SVM 的相对误差明显都低于 BP 神经网络, 这说明 RS-SVM 的预测精度要明显优于比 BP 神经网络的预测精度, 而且 RS-SVM 模型测试样本的相对误差控制在 3% 以内。

从图 5 两种模型的预测结果拟合情况可知, 基于 SVM 建立的安全预警模型的预测拟合效果要优于 BP 神经网络, 这主要是 BP 神经网络在处理有限样本数据的局限性所致。BP 神经网络是一种基于无限样本的渐近理论, 只有样本数量趋于无穷多时理论上才收敛于全局最优值, 也难免陷入局部极值。而 SVM 是建立在结构风险最小化的原则之上的, 能够根据有限样本信息, 在模型的复杂性和学习能力之间寻求最佳折衷, 得到的一定是全局最优解, 而且这些计算单元具有有限可控的学习能力。

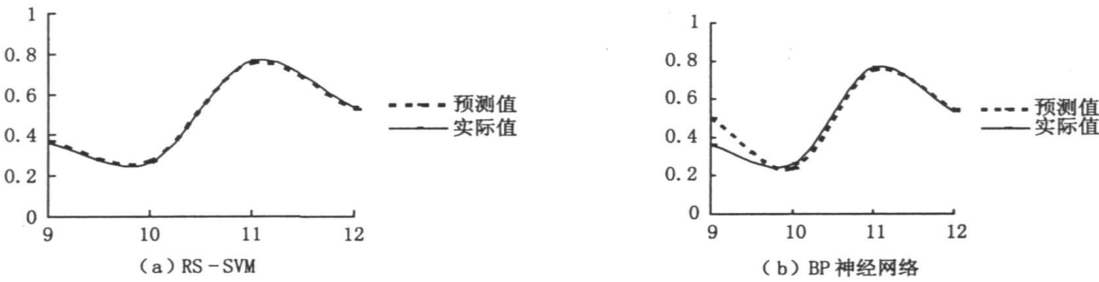


图5 两种模型拟合预测过程对比曲线

Fig.5 Fitting and forecasting comparison of the two models

4 结论

- 1) 简化了建筑施工项目安全风险预警模型的信息表达空间, 减小了SVM 构成系统的复杂性, 提高了容错及抗干扰能力。
- 2) BP 神经网络的均方根误差为0.070 7, RS-SVM 预测模型的最小均方根误差为 0.011 5, 可见在小样本条件下, RS- SVM 预警模型的预测精度、泛化能力明显优越于 BP 神经网络学习方法。

参考文献:

[1] 孙华山, 安全生产风险管理[M]. 北京: 化学工业出版社, 2006.

[2] 程杰. 建筑安装工程施工安全风险评价与管理[J]. 现代管理科学, 2002(9): 42- 43.

[3] 张仕廉, 董勇, 樊承仁. 建筑安全管理[M]. 北京: 中

国建筑工业出版社, 2005.

[4] 李万庆, 李继萍, 孟文清, 等. 基于粗糙集的载体桩质量核心影响因素分析[J], 河北工程大学学报(自然科学版), 2010, 27(1): 88- 91.

[5] 李万庆, 李海涛, 孟文清. 工程项目工期风险的支持向量机预测模型[J]. 河北工程大学学报(自然科学版), 2007, 24(4): 1- 4.

[6] 苏怀智, 温志萍, 吴中如. 基于 SVM 理论的大坝安全预警模型研究[J]. 应用基础与工程科学学报, 2009, 17(1): 41- 47.

[7] MENG W, LI W, LI J, et al. Risk prediction model for construction projects based on rough sets and artificial neural networks[C]// FIIME. 2nd International Conference on Future Information Technology and Management Engineering. US: IEEE Computer Society, 2009.

(责任编辑 马立)

(上接第 25 页)

3 结论

- 1) 利用废弃混凝土磨细粉作为水泥的掺和料是可行的, 随着废弃混凝土磨细粉掺量的增加, 水泥的胶砂强度越低。用掺量为 10% 废弃混凝土磨细粉可以生产出 42.5[#] 水泥, 掺量 20% 时强度要降低为 32.5[#] 水泥。
- 2) 激发剂 Na₂SiO₃ 的合适掺量是 1%, 激发剂 NaOH 的作用效果不明显。说明碱性激发剂对混凝土废弃物的激发效果不好, 硅酸盐类激发剂对废弃混凝土有较好的激发效果。

参考文献:

[1] CHARLSON A. Recycling and reuse of waste in the construction industry [J]. Structural Engineer, 2008, 86(4): 32- 37.

[2] POON C S, LAM C S. The effect of aggregate- to- cement ratio and types of aggregates on the properties of pre- cast concrete blocks [J]. 2008, 30(4): 283- 289

[3] 李九苏, 肖韩宁, 龚建清. 再生骨料水泥混凝土的级配优化试验研究[J]. 建筑材料学报, 2008, 11(1): 105- 110.

[4] LI JIU SU, XIAO HAN NING, GONG JIAN QING. Granular effect of fly ash repairs damage of recycled coarse aggregate [J]. Journal of Shanghai Jiaotong University: Science, 2008, 13(2): 177- 180.

[5] 王立久, 郑芳宇, 迟耀辉. 水泥生料组分混凝土设计理论及试验研究[J]. 大连理工大学学报, 2007, 47(2): 222- 227.

[6] SHUI ZHONG HE, XUAN DONG XING, WAN HUI WEI, et al. Rehydration reactivity of recycled mortar from concrete waste experienced to thermal treatment [J]. Construction and Building Materials, 2008, 22(8): 1723- 1729.

[7] 方开泰, 马长兴. 正交与均匀试验设计[M]. 北京: 科学出版社, 2001.

(责任编辑 刘存英)