

基于粗糙集—支持向量机的震害预测模型及应用^{*}

刘勇健^{1,3}, 刘义建², 张伯友³

(1. 广东工业大学, 广州 510006 2. 湖南省地质矿产勘查开发局 407 队, 湖南怀化 418000

3. 中国科学院广州地球化学研究所, 广州 510640)

摘要: 结合粗糙集理论的属性约简与支持向量机的分类功能, 建立了基于粗糙集与支持向量机的建筑物震害预测模型。该模型首先运用粗糙集理论, 建立决策表, 进行属性离散、属性重要性排序、属性约简和分类规则的提取, 然后用所提取的关键成分训练支持向量机。该模型不但能有效降低建筑物震害影响因子数据维数及支持向量机的复杂程度, 提高训练速度和分类精度, 而且还能对各因子的影响程度进行排序。最后, 通过实例验证了该模型的性能。

关键词: 粗糙集; 支持向量机; 属性约简; 震害预测; 条件信息熵

中图分类号: P315.9 **文献标志码:** A **文章编号:** 1000-0666(2008)03-0289-07

0 引言

地震对人类社会造成了重大的人员伤亡和巨大的经济损失。20 世纪的 100 年间, 我国共发生 650 余次 6 级以上的破坏性地震, 死亡约 59 万人, 全球死亡 20 万人以上的两次地震都发生在我国, 2008 年四川汶川 8.0 级地震撼动了大半个中国, 伤亡十分惨重。据资料统计, 我国有 67% 的大城市位于地震烈度 VII 度及以上的地震区, 大部分地区受到地震灾害的威胁 (尹之潜等, 2003)。为了减轻地震灾害损失, 许多专家学者在震害预测方面做了大量研究, 提出了多种震害预测方法 (刘恢先, 1986; 阎维明, 2006; 汤皓, 2006; 赵艳林等, 2002; 缪升等, 2000), 如历史地震统计法、半经验半理论法、模糊类比法、结构算法等, 近年来又出现了神经网络方法。建筑物震害一方面与地震特性相关, 另一方面又与建筑结构特性有关。震害影响因素众多, 而且各因素之间具有随机性、模糊性、不确定的相关性及统计资料的不完备性, 因此, 建筑物震害预测是一个高度复杂的非线性系统。

20 世纪 60~70 年代, Vapnik (1993, 1995, 1998) 等人开始致力于统计学习理论 (Statistical Learning Theory 简称 SLT) 的研究, 并于 90 年代提出了支持向量机 (Support Vector Machine 简称 SVM)。它能很好地解决小样本、非线性、高维数

和局部极小等问题, 并已成功地应用于函数模拟、模式识别和数据分类等领域, 成为继经典神经网络研究之后的又一研究热点。1982 年波兰科学家 Pawlak 提出的粗糙集理论 (Rough Sets 简称 RS) 能较好地处理不准确或不完整知识信息及冗余信息 (张文修, 2001)。本文中笔者运用粗糙集理论对建筑物震害的影响因子进行约简, 剔除了与决策信息不相关的 (或影响甚微的) 属性, 进而简化 SVM 的输入变量, 将两种算法有机地结合起来, 建立了基于粗糙集—支持向量机的建筑物震害智能预测模型, 然后通过实例验证了该模型的有效性。

1 粗糙集理论

粗糙集理论是分析不完整、不精确信息系统的有力工具。粗糙集方法把研究对象视为一个信息系统, 通过不可分辨关系和不可分辨类确定问题的近似域。根据知识表达中不同属性的重要性, 明确有用信息和冗余信息, 简化输入信息的表达空间维数。在保持分类能力不变的前提下, 通过知识约简, 直接从训练数据中导出分类规则 (张文修, 2001)。

1.1 信息系统及条件信息熵

一个信息系统可表示为

$$S = (U, R, V, f). \quad (1)$$

* 收稿日期: 2007-11-08.

基金项目: 广东省自然科学基金 (6021462) 和广东省岩土工程重点学科基金联合资助

其中, U 为对象的非空有限集合, 即论域; R 为属性的非空有限集合; $V=\bigcup_{R} V_i$, V_i 为属性 R 的值域; $f:U \times R \rightarrow V$ 是一个信息函数, 指定 U 中各对象的属性唯一值. 对信息系统 S 若属性集 $R=\bigcup D$ 子集 C 和 D 分别称为条件属性集和决策属性集, $C \cap D \neq \phi$, 则该信息系统称为一个决策系统或决策表.

在粗糙集理论研究中, Dunsch等 (1998) 从信息论的角度, 建立了知识与信息熵的关系, 引入了信息熵和条件熵概念. 信息熵是信源总体的平均不确定性的量度. 定义知识 (属性集合) P 的熵 $H(P)$ 和知识 (属性集合) Q 相对于知识 (属性集合) P 的条件信息熵 $H(Q/P)$ 为:

$$H(P)=-\sum_{i=1}^n P(X_i) \lg P(X_i), \tag{2}$$

$$H(Q/P)=-\sum_{i=1}^n P(X_i) \sum_{j=1}^m P(Y_j/X_i) \lg P(Y_j/X_i). \tag{3}$$

式中, $P(X_i)$ 为 P 在论域 U 上的划分 $X=(X_1, X_2, \dots, X_n)$ 上的概率, $P(X_i)=|X_i|/|U|$, $i=1, 2, \dots, n$. 条件概率 $P(Y_j/X_i)=|Y_j \cap X_i|/|X_i|$, $i=1, 2, \dots, n$, $j=1, 2, \dots, m$.

这样, 就可以利用每个条件属性对决策属性的条件信息熵大小来判断该条件属性与决策属性的影响程度. 由 RS 理论的信息论观点 (阎维明, 2006), 设 C 是 U 的一个条件属性集, D 是决策属性集, 且 $C \cap D \neq \phi$, $C \subseteq D$ 则属性 C 是 C 相对于决策属性 D 不必要的充要条件是

$$H(D/C)=H(D/(C-\{C\})). \tag{4}$$

C 相对于决策属性 D 独立的充要条件是对于 C 中任意属性 C 都有

$$H(D/C) \neq H(D/(C-\{C\})). \tag{5}$$

1.2 基于条件信息熵的相对约简算法

属性重要性是知识约简中的一个关键概念, 设 $A \subseteq C$ 则对于任意属性 $a \in (C-A)$ 重要性

$$\gamma = S_{\text{IF}}(a, A, D) = H(D/A) - H(D/A \cup \{a\}). \tag{6}$$

$S_{\text{IF}}(a, A, D)$ 的值越大, 说明在已知 A 的条件下, 属性 a 对于决策 D 就越重要.

但它对结果的影响往往很小. 本文结合震害预测的实际情况, 设 δ 为一个很小的数, 当满足

$$\gamma = S_{\text{IF}}(a, A, D) = H(D/A) - H(D/A \cup \{a\}) \leq \delta \tag{7}$$

该属性可相对约简. 该算法的主要步骤如下:

(1) 计算决策表中条件属性 C 和决策属性 D 的信息熵 $H(C)$ 和 $H(D)$.

(2) 计算决策表中决策属性 D 相对于条件属性 C 的条件信息熵 $H(D/C)$.

(3) 给定一个较小的值 δ 令 $B=C$ 对于条件属性集 C 中的任一属性 $C_i \in C$ 重复以下计算过程:

① 计算决策属性集 D 相对于属性集 $(C-C_i)$ 的条件熵 $H(D/(C-C_i))$, 即 $H(D/B-C_i)$;

② 计算 C_i 的重要性 $\gamma = S_{\text{IF}}(C_i, (B-C_i), D) = H(D/(B-C_i)) - H(D/C)$;

③ 若 $\gamma = S_{\text{IF}}(C_i, (B-C_i), D) = H(D/(B-C_i)) - H(D/C) \leq \delta$ 时, 令 $B = B - C_i$, 重复 (3), 否则终止计算.

(4) 得到 B 为 C 相对于 D 的近似相对约简.

2 支持向量机

支持向量机 (SVM) 是一种基于 VC (Vapnik-Chervonenkis) 理论的创造性机器学习方法, 由 Vapnik 和他的合作者 (1993) 提出. SVM 实现了结构风险最小化, 同时还最小化了经验风险与 VC 维的界. 分析问题时, 设样本集 $\{(x_i, y_i)\} (i=1, 2, \dots, m)$, 其中, x_i 为输入向量, y_i 为目标输出向量, m 为样本容量. 大多数情况下输入与输出之间呈非线性关系, 将样本 x 通过非线性映射 $\phi(x)$ 映射到高维特征空间 (Hilbert 空间), 并在该空间进行线性回归, 就可获得在原空间非线性回归的效果, 其表达式为

$$f(x, w) = w \phi(x) + b = (w, \phi(x)) + b \tag{8}$$

式中, w 为权矢量, b 为偏差, $(w, \phi(x))$ 表示内积. SVM 采用结构风险最小原理, 其中风险用 Vapnik 提出的不敏感损失函数来度量. ϵ 定义为

$$L(y - f(x)) = |y - f(x)|_\epsilon = \begin{cases} 0 & |y - f(x)| \leq 0 \\ |y - f(x)| - \epsilon & |y - f(x)| > \epsilon \end{cases} \tag{9}$$

为了确定系数 w 和 b 考虑到允许误差的情况, 引入松弛变量 $\xi_i \geq 0$ 和 $\xi_i^* \geq 0$ 求解以下最小化问题

$$\min \phi(w, \xi, \xi^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*), \tag{10}$$
$$y_i - w\phi(x_i) - b \leq \epsilon + \xi_i, \quad \xi_i \geq 0$$
$$w\phi(x_i) + b - y_i \leq \epsilon + \xi_i^*, \quad \xi_i^* \geq 0$$

式中, 第一项为正规化项; 第二项为经验误差项; C 为正规化常数, 它确定了经验误差项与正规化项之间的某种平衡关系。SVM 通过构造核函数 $K(x_i, x_j)$ 来克服维数灾难。当函数满足 Mercer 条件时, 某一映射空间的内积运算就存在 $K(x_i, x_j) = \phi(x_i)\phi(x_j)$, 即线性问题的内积运算用核函数来代替, 高维空间的内积运算就可通过输入空间的函数实现。运用对偶原理并引入拉格朗日乘子 α 和 α^* 后, 决策函数的形式为

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{11}$$

这样 (10) 式最优化问题就转化为对偶问题:

$$\max W(\alpha, \alpha^*) = -\frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j)$$
$$+ \sum_{i=1}^k y_i (\alpha_i - \alpha_i^*) - \epsilon \sum_{i=1}^k (\alpha_i - \alpha_i^*)$$
$$\sum_{i=1}^m (\alpha_i - \alpha_i^*) = 0 \quad C \geq \alpha_i \geq 0 \quad i = 1, 2, \dots, n \tag{12}$$

在 (12) 式中只有少数 α_i, α_i^* 不为 0 它们所对应的点为支持向量, 回归函数 $f(x)$ 可以由支持向量完全表征。其中 $K(x_i, x_j)$ 为满足 Mercer 条件的内积函数, 即核函数。常用的核函数有多项式核函数、Gauss 函数、Sigmoid 函数和径向基核函数。支持向量机法采用结构风险最小化原则, 具有很好的推广能力, 求解一个凸二次优化问题, 就能保证找到全局最优解。

尽管 SVM 方法具有如前所述的优点, 但亦存在一些不尽人意之处, 如当处理的数据量过大时, 冗余信息的存在使得 SVM 的训练时间过长、速度变慢。RS 理论能很好地处理不准确或不完整知识信息及冗余信息。因此, 可将两者有机结合起来, 用于建筑物震害预测中。

3 基于粗糙集—支持向量机的建筑物震害预测模型

3.1 宏观震害经验及震害影响因子

随着我国经济的飞速发展, 近年来新建了大量高层建筑, 但还存在一定比例老旧房屋和一些抗震性能较差的房屋, 其中多层砖房问题尤为突出。笔者以多层砖房为研究对象, 从狭义震害预测角度分析了建筑物的易损性, 即对结构在确定的地震强度作用下, 发生某种破坏程度的概率或可能性进行了预测。多层砖房在地震作用下, 其破坏的部位、特点以及具体形态多种多样, 但概括起来, 可依据其破坏的程度将震害划分为 5 个等级 (表 1)。大量的宏观震害经验表明, 多层砖房的震害程度与下列因素密切相关。

表 1 多层砖房破坏等级的划分标准

破坏状态	破坏等级	震 害 描 述
基本完好	I	墙体完好, 少数非主体结构有局部轻微破坏, 或个别突出部分偶有轻微破坏裂纹
轻微破坏	II	非主体结构局部有明显的破坏, 或少数墙板有轻微裂纹, 少数墙体偶有明显裂纹
中等破坏	III	主体结构或连接部分多处发生明显裂缝, 或填充墙体、附属建筑等破坏严重, 甚至倒塌
严重破坏	IV	主体结构破坏严重, 墙体开裂, 并有明显的滑移、错位, 部分墙体外闪, 个别墙板塌落
毁坏	V	外墙体近乎全部倒塌, 承重结构部分倒塌, 需拆除重建

(1) 地震特性。在以往的研究中, 通常采用烈度指标表示地震特性, 而本文选取地面加速度峰值作为地震特性指标。震害经验表明, 对于强震而言, 地面运动力特别是地面运动水平加速度

峰值是造成建筑物破坏的主要因素; 而且以峰值加速度和特征周期为参数的中国地震区划图已颁布实施。

(2) 场地与地基条件。大量宏观震害资料表

明,地震破坏程度不仅与地基有关,而且与建筑物的刚度有关。松软的地基上,地基的沉降会使其上部结构遭到局部破坏,而强烈振动是导致坚实地基上刚性建筑物破坏的主要原因。如果把刚性建筑物修建在松软地基上,则地基能吸收地震的部分能量,从而起到隔震的效果。在松软地基上的柔性房屋,其破坏原因是地震波产生低频率共振而造成的。由于地震重复性荷载的作用,地基因液化和强度骤然降低而失去稳定性,建筑物基础的竖向和水平位移以及地基的不均匀沉降,往往是松软地基上建筑物遭受破坏的主要原因。一般说来,基岩上的地震烈度约比一般土层要低1度,而软弱土层上的地震烈度比一般土层要高近1度;从水文地质方面来说,地下水浅的地区比地下水深的地区效应大,破坏重;从地形地貌方面来说,地形地貌复杂的地区地震效应大,破坏严重(刘本玉,2001)。

(3) 建筑物层数与层高。房屋的层数与层高不仅影响房屋的空间刚度,而且使房屋受到的地震力的大小亦大为不同,层数越多,层高越高,其震害就越重。

(4) 房屋的整体性。在多层砖房中,水平承重构件为楼屋盖。历次震害表明,现浇钢筋混凝土、预制混凝土与木楼屋盖的地震表现是有差异的,其原因是现浇楼屋盖整体性最好,预制楼屋盖次之,而木盖最差。横墙承重或纵横墙承重的房屋比纵墙承重的房屋抗震能力要强;用钢筋混凝土构造柱和圈梁来增强多层砖房的抗倒能力效果显著。

(5) 建筑物体型。建筑体型对建筑结构的抗震性能有着显著的影响,当建筑形体不规则,墙体位置不均匀,建筑物刚度中心与荷载作用点有较大间距时,会产生扭转。当在水平地震力作用下剪切与扭转组合时,其所受应力加大,震害显著加重。特别是平面突出部分与主体中心部分震动不同步时破坏更明显。同样地,多层砖房在立面竖向的变化,即刚度在竖向的不均匀或突变也使震害更为严重。

(6) 施工质量。施工质量对房屋抗震性能的影响是显而易见的。施工质量好、达到抗震设计标准的房屋抗震能力就好;施工质量差的房屋其抗震性能必定不好。建筑年代的影响也需要综合考虑。

(7) 含砖墙率。它表达了房屋的刚度、强度、空旷程度等特征。含砖墙率越大房屋的空间刚度越大,抗剪能力越强,震害就越轻。

(8) 承重墙体砂浆标号。多层砖房的地震破坏主要是由于砖墙首先开裂,这是由砂浆强度较低造成的。

在用特征向量建立预测模型时,除了要考虑各因子对房屋结构的抗震性能的影响外,还需考虑参数的容易获取性和关联性(或独立性)。

3.2 基于粗糙集—支持向量机的震害预测模型

基于粗糙集—支持向量机的震害预测模型由两部分构成(图1),预测模型将粗糙集网络作为前置系统,支持向量机作为后置的信息识别系统。首先运用粗糙集理论获取分类知识,建立决策表,进行属性离散化、属性重要性排序、属性约简和分类规则的提取;然后用所提取的关键成分训练支持向量机。该方法可有效降低建筑物震害影响因素数据维数及支持向量机的复杂程度。通过粗糙集方法去掉冗余信息后,减少了信息表达的特征数量,而简化支持向量机的复杂程度,可缩短训练时间。支持向量机能很好地消除粗糙集对噪音敏感的缺陷,解决预测中的小样本、非线性、高维数和局部极小等问题。

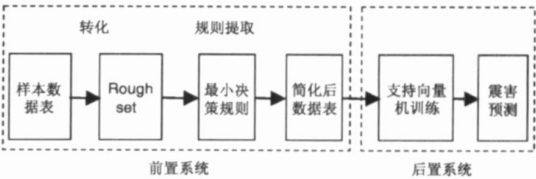


图 1 粗糙集—支持向量机训练过程

3.3 实例及分析

笔者收集了1966年云南东川地震、1978年唐山大地震、1988年云南澜沧—耿马地震、1995年云南武定地震和1997年云南丽江地震、2001年云南施甸地震、2007年云南保山隆阳地震中的多层砖房震害资料(刘恢先,1986;缪升等,2000;白良等,1989;汤皓,2006;非明伦,2002,2007),选择其中96个样本用于训练基于粗糙神经网络的多层砖房震害模型。

本文选择以下震害影响因子:房屋层数、房屋质量、砂浆标号、砖墙面积率、房屋整体性、场地条件、地震峰值加速度作为模型的输入。对于定性指标参考专家经验赋值。①房屋质量、施工质量优

取为 10、中取 8、差取 6 同时考虑建筑年代进行房屋折旧，每 5 年减 0.5；② 房屋的整体性：楼盖和屋盖，现浇取 5、预制取 4、木屋盖取 3、有圈梁加 1、有构造柱再加 1、有地下室、筏基加 1、房屋开裂减 1；③ 场地条件：I 类场地取 1、II 类取 2、II 类取 3、IV 类取 4。将给定地震下可能的建筑物的震害程度为模型的输出震害作为系统的最终输出，把震害程度划分为 5 级：基本完好、轻微破坏、中等破坏、严重破坏、毁坏，所收集的样本资料如表 2（限于篇幅仅列出部分样本）。

表 2 多层砖房震害样本的输入与输出

序号	房屋层数 c_1	房屋质量 c_2	砂浆标号 c_3	砖墙面积率 c_4	房屋整体性 c_5	场地条件 c_6	峰值加速度 c_7	震害类型 d
1	2	10	25	8.2	8	3	0.20	III
2	2	6	10	10.0	8	4	0.25	IV
3	2	8	20	9.0	12.0	1	0.10	I
4	2	10	35	9.7	9.5	2	0.15	II
5	4	10	25	7.6	7.5	1	0.20	III
6	5	6	10	3.8	6.5	4	0.25	V
7	4	8	10	9.0	6.0	3	0.30	V
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
96	5	8	10	9.0	9.0	1	0.15	II

震害预测信息系统 $S = (U, R, V, \delta)$ 中，将研究区域内的全部建筑物组成的集合作为论域 $U = \{U_1, U_2, \dots, U_6\}$ ，震害影响因子作为条件属性 $C = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7\}$ ，给定地震加速度峰值下建筑物可能的破坏程度为决策属性 $D = \{d_1, d_2, d_3, d_4, d_5\}$ ，其中 d_1 = 基本完好、 d_2 = 轻微破坏、 d_3 = 中等破坏、 d_4 = 严重破坏、 d_5 = 毁坏，属性集 $R = C \cup D, C \cap D \neq \emptyset$ 。决策表的表头为各属性，其每一行表示论域中的一个事例或称一条决策规则，每一列表示属性及属性值。各因子变化范围及其离散化值如表 3。按表 3 离散化后形成初始决策如表 4。

表 3 属性值及其离散化表

V	c_1	c_2	c_3	c_4	c_5	c_6	c_7	d
0	[2, 3)	10	[30, 35]	[12.0, 15.0]	[10, 12)	1	[0.1, 0.15)	I
1	[3, 4)	8	[20, 25]	[9.0, 12.0]	[8, 10)	2	[0.15, 0.20)	II
2	[4, 5]	6	[5, 15]	[6.0, 9.0)	[6, 8]	3	[0.20, 0.25)	III
3		4		[3.0, 6.0]		4	[0.25, 0.30)	IV
4							[0.30, 0.35]	V

表 4 初始决策表

U_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	d
u_1	0	0	1	3	1	2	2	2
u_2	0	3	2	1	1	1	3	3
u_3	1	0	1	2	0	0	2	0
u_4	0	0	0	1	1	1	1	1
u_5	2	0	1	2	2	0	2	2
u_6	2	2	2	3	2	0	3	4
u_7	2	1	2	2	2	0	4	4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
u_{96}	2	1	1	2	1	0	1	1

利用本文前述的重要性概念，分别分析建筑震害影响因子（条件属性）的重要性。

(1)按决策属性 D对 U进行划分得: $U/D = \{Y_1, Y_2, Y_3, Y_4, Y_5\}$ ，其中， $Y_1 = \{U_3, U_8, \dots, U_{90}\}_{23}$ ， $Y_2 = \{U_4, U_9, \dots, U_{96}\}_{31}$ ， \dots ， $Y_5 = \{U_6, U_{16}, \dots, U_{98}\}_9$ ，下标表示组中所含元素的个数。则可得 $P(Y_1) = 23/96$ $P(Y_2) = 31/96$ \dots ， $P(Y_5) = 9/96$

(2)按条件属性 $C = \{C_1, C_2, C_3, C_4, C_5, C_6, C_7\}$ 对 U进行划分得 $U/C = \{X_1, X_2, \dots, X_{81}\}$ ，可计算出 $X_1 = \{U_1\}_1$ ， $X_2 = \{U_2\}_2$ ， \dots ， $X_{81} = \{U_{81}\}_{81}$ ，则可得 $P(X_1) = 1/96$ $P(X_2) = 1/96$ \dots ， $P(X_{81}) = 1/96$ 。计算条件信息熵得 $H(D/C) = -\sum_{i=1}^{81} P(X_i) \sum_{j=1}^5 P(Y_j/X_i) \lg P(Y_j/X_i) = 0$

(3)按式（8）计算各因子的重要性系数 γ_i ，得 $\gamma_1 = 0.023$ $\gamma_2 = 0.009$ $\gamma_3 = 0.018$ $\gamma_4 = 0.049$ $\gamma_5 = 0.034$ $\gamma_6 = 0.025$ $\gamma_7 = 0.0127$ ，各因子的重要性排序为： $\gamma_7 > \gamma_4 > \gamma_5 > \gamma_6 > \gamma_1 > \gamma_3 > \gamma_2$ ，若取 $\epsilon = 0.02$ 得该决策表的条件属性集 C相对于 D的一个近似相对简约 $B = \{C_1, C_4, C_5, C_6, C_7\}$ 。

将经过条件信息熵简约后的 5个关键特征因子 $\{C_1, C_4, C_5, C_6, C_7\}$ 作为支持向量机的输入向量，而以建筑物震害等级 D作为输出。用对决策表进行约简后得到的规则作为样本数据进行训练。选择径向基函数作为核函数，即

$$K(x, x_i) = \exp(-|x - x_i|^2 / \sigma^2).$$

(13)

式中， x, x_i 为训练输入， σ 为核函数宽度。

支持向量机的训练参数为 σ 和 C 。 σ 的大小影响样本的输出响应区间，当 σ 越小时，响应区间越窄，得到的分类面经验风险越小，表现为函数曲线光滑，但结构风险越大。惩罚因子 C表征对错误的惩罚程度，C越大表示对错误分类的惩罚越大。我们通过试算，取 $\sigma = 0.1$ ， $C = 1000$ 。

为了验证模型的性能，另选 7个样本进行测试，计算结果如表 5所示，除 6号样本的预测结果有偏差外，其余 6个样本的计算结果与实测结果完全吻合。出现误判的原因可能是由于建模所采用的原始样本中缺乏与 6号相近的样本。因此，选样本建模时要特别注意原始样本的代表性。

表 5 7个样本的预测结果

序号	房屋层数 c_1	房屋质量 c_2	砂浆标号 c_3	砖墙面积率 c_4	房屋整体性 c_5	场地条件 c_6	峰值加速度 c_7	目标输出 d	计算结果 d'
1	3	8	20	6.8	8.0	4	0.15	III	III
2	2	10	25	4.5	7.5	1	0.10	II	II
3	4	6	10	3.6	7.5	4	0.25	IV	IV
4	4	8	25	8.5	9.5	2	0.20	III	III
5	2	9	25	10.3	7.0	2	0.30	III	III
6	4	6	10	6.7	7.0	4	0.30	V	IV
7	5	10	10	11.5	8.0	2	0.15	I	I

4 结论

(1)建立了基于粗糙集—支持向量机的建筑物震害预测模型，该模型综合了粗糙集理论在知识获取方面的能力和支持向量机很强的分类能力。降低了数据维数及支持向量机的复杂度，提高了训练速度和分类精度。实例研究表明，运用本模型对多层砖房的震害预测与实际震害基本吻合，明显提高了建筑物震害预测的可靠性。该模型不仅能预测震害，还能对各因子对建筑物抗震性能的影响程度进行排序，因此对建设工程抗震设计具有重要参考价值。

(2)震害预测工作涉及到前期大量数据的收集及后期的数据处理，将属性约简用于震害预测工作，将大大减少数据收集与处理的工作量。这一研究思路和方法同样适合于其它结构类型建筑物的震害预测。

参考文献:

白良. 1989 澜沧—耿马地震多层砖房的震害及机理简析 [J]. 云南工学院学报, 4 (3): 35—43.
非明伦, 崔建文, 赵永. 2002 施甸地震震害分析 [J]. 地震研究, 192—199.
非明伦, 周光全, 余庆坤, 等. 2007 保山隆阳 5.0级地震永昌城区房屋震害统计与分析 [J]. 地震研究, 30 (1): 77—82.

刘本玉, 江见鲸. 2001. 多层砖房震害因子的选择及其与震害程度的灰色关联序分析 [J]. 地震研究, 151—156.

刘恢先. 1986. 唐山大地震震害 [M]. 北京: 地震出版社.

缪升, 刘本玉, 叶燎原. 2000. 多层砖房震害预测的人工神经网络方法与实践 [J]. 世界地震工程, 16 (1): 21—27.

汤皓, 陈国兴. 2006. 基于 BP 神经网络模型的多层砖房震害预测方法 [J]. 地震工程与工程振动, (4): 141—146.

阎维明, 乔亚玲, 何浩祥. 2006. 粗糙集理论在震害预测中的应用 [J]. 自然灾害学报, 15 (3): 147—151.

尹之潜. 1996. 地震灾害及损失预测方法 [M]. 北京: 地震出版社.

尹之潜, 赵直, 杨淑文. 2003. 建筑物易损性和地震损失与地震加速度谱值的关系 (上) [J]. 地震工程与工程振动, 23 (4): 195—200.

张文修, 吴伟志. 2001. 粗糙集理论与方法 [M]. 北京: 科学出版社.

赵艳林, 杨绿峰. 2002. 多层砖房震害的灰色预测 [J]. 世界地震工程, 18 (1): 56—61.

Dunish L, Gedga G. 1998. Uncertainty measures of rough set prediction [J]. Artificial Intelligence, 106: 109—137.

Pawlak Z. 1982. Rough sets [J]. International Journal of Information and Computer Science, 11: 341—356.

Vapnik V N. 1993. Three fundamental concepts of the capacity of learning machines [J]. Physics (A), 200: 538—544.

Vapnik V N. 1995. The Nature of statistical learning theory [M]. New York: Springer Verlag.

Vapnik V N. 1998. Statistical learning theory [M]. New York: Wiley.

Seismic Damage Prediction Model Based on Rough Set and Support Vector Machine and Its Application

LU Yong-jian³, LIU Yi-jian², ZHANG Bo-you¹

(1. Guangdong University of Technology, Guangzhou 510006, Guangdong, China)

(2. No. 407 Brigade, Geology and Mineral Resources Exploration and Development Bureau of Hunan Province, Huaihua 418000, Hunan, China)

(3. Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou 510640, Guangdong, China)

Abstract

A model for seismic damage prediction of building is proposed based on two integrated intelligent algorithms: attribute reduction algorithm of rough sets and support vector machines. At first, rough set theory is used to acquire the knowledge of classification, which includes decision table construction, attribute discretization, attribute importance ranking, attribute reduction and rule abstract. Then, the key components are extracted as the input of support vector machine. The method can reduce the dimensions of the data and the complexity and raise the efficiency of training and the accuracy of prediction. The effect extent to the earthquake resistance performance of these factors can be obtained in the model. Finally, it is validated applying the method to forecast seismic damage of multistory masonry buildings.

Key words: rough set; support vector machine; attribute reduction; seismic damage prediction; conditional information entropy



温瑞智 中国地震局工程力学研究所信息技术与工程材料室主任、研究员,中国地震局新世纪百人计划人选。1994年毕业于大连理工大学工程力学系,获硕士学位。1999年毕业于中国地震局工程力学研究所,获博士学位,随即入哈尔滨工业大学力学博士后流动站和北京工业大学土木工程博士后流动站工作。中国地震学会会员、中国地理信息系统协会会员、加拿大土木工程协会会员、加拿大地震工程协会会员。主要从事强震动观测、断层对地震动影响、信息技术在防灾减灾领域中的应用等研究。



潘纪顺 中国地震局地球物理勘探中心高级工程师。2002年毕业于中国科学技术大学地球与空间科学学院固体地球物理学专业,获理学硕士学位。2007年毕业于中国地震局地球物理研究所固体地球物理学专业,获理学博士学位。主要从事浅层地震勘探、地震宽角折射、反射勘探、高分辨率折射地震勘探、火山监测、桩基检测等工作,主要研究方向是地震数据处理、地震成像与反演。



杜方 四川省地震局研究员。1992年毕业于武汉测绘科技大学,获硕士学位。主要从事现今地壳运动及动力学、前兆机理与地震预测方法等研究。



刘勇健 广东工业大学副教授。1993年毕业于成都理工学院油气地质与勘探专业,获工学硕士学位。现在中国科学院广州地球化学研究所地质构造专业攻读博士学位。主要从事岩土工程的教学与研究工作。



刘琼仙 云南省地震局高级工程师。1988年毕业于江苏工学院计算机及应用专业,获工学学士学位。主要从事强震动观测网络及信息的研究工作。



耿杰 山东省地震局副局长。1985年毕业于山东科技大学地质系。主要从事地震分析预报与研究工作。



胡先明 四川省地震局高级工程师。1981年毕业于国家地震局天水地震学校地球物理专业,1988年毕业于中国科学技术大学地球物理专业。主要从事水库诱发地震研究和地下水观测等工作。



王爱国 中国地震局地震预测研究所兰州基地副研究员。1994年毕业于兰州大学地质系,2002年获中国地震局兰州地震研究所硕士学位。主要从事地震地质、工程地震及数值模拟等方面的研究工作。