

刘燕萍, 董伟, 黄毕双, 等. 基于两阶段支持向量机的群体建筑物震害预测方法[J]. 华南地震, 2016, 36(2): 107-113. [LIU Yanping, DONG Wei, HUANG Bishuang, et al. Seismic Damage Prediction Method of Building Groups Based on a Two-stage Support Vector Machine[J]. South China Journal of Seismology, 2016, 36(2): 107-113.]

## 基于两阶段支持向量机的群体建筑物震害预测方法

刘燕萍<sup>1, 2</sup>, 董伟<sup>3</sup>, 黄毕双<sup>4</sup>, 李晋<sup>1, 2</sup>, 杨芳<sup>1, 2</sup>

(1. 广东省地震局, 广州 510070; 2. 中国地震局地震监测与减灾技术重点实验室, 广州 510070;

3. 中国科学技术大学, 合肥 230026; 4. 香港城市大学, 香港 999077)

**摘要:** 传统群体建筑物震害预测, 多是采用与现有的建筑物震害数据类比的方法。由于特殊地质环境和特定地震情景的影响, 外加人工统计的误差干扰, 现有数据中存在相当数量的异常数据指标。这些数据噪声将严重影响群体建筑物震害预测的准确度。引入一种新型两阶段支持向量机方法, 首先为正常数据和异常数据赋予不同的权重, 接着用加权支持向量机方法建立群体建筑物震害预测模型。通过对汶川地震中 640 栋建筑进行交叉验证发现, 提出的两阶段支持向量机方法不仅能有效识别出异常数据点, 而且能快速准确地预测建筑物震害结果, 可以用于实际的城市建筑物震害预测工作。

**关键词:** 群体建筑物; 震害预测; 支持向量机; 交叉验证

中图分类号: TU311.3

文献标志码: A

文章编号: 1001-8662 (2016) 02-0107-05

DOI: 10.13512/j.hndz.2016.02.017

## Seismic Damage Prediction Method of Building Groups Based on a Two-stage Support Vector Machine

LIU Yanping<sup>1, 2</sup>, DONG Wei<sup>3</sup>, HUANG Bishuang<sup>4</sup>, LI Jin<sup>1, 2</sup>, YANG Fang<sup>1, 2</sup>

(1. Earthquake Administration of Guangdong Province, Guangzhou 510070, China;

2. Key Laboratory of Earthquake Monitoring and Disaster Mitigation Technology, CEA, Guangzhou 510070, China; 3. University of Science and Technology of China, Hefei 230036, China;

4. City University of Hong Kong, Hong Kong 999077, China)

**Abstract:** In conventional seismic damage prediction of building groups, the researchers always make a simple analogy with existing data of damaged buildings. Due to the influence of special geological environment, particular earthquake scenarios, together with the errors in handiwork statistics, there are a certain amount of outliers in dataset. The random noise in the dataset will have a serious impact on prediction accuracy. Thus, this paper introduces a two-stage support vector machine method. In the first step, the authors add different weight

收稿日期: 2015-10-06

基金项目: “十二五”国家科技支撑计划项目(2015BAK18B01)、广东省科技计划项目(2015A020217007)、深圳市科技创新委员会项目(ZDSYS20140509155229805、JCYJ20140630144136828)

作者简介: 刘燕萍 (1963-), 女, 助理工程师, 主要从事震害预测相关工作研究。

E-mail: Liyp23@163.com.

values to normal data and outliers respectively. Then a weighted support vector machine is proposed to build the prediction model of building groups. By using a cross-validation approach, the paper empirically tests the proposed model on 640 buildings in Wenchuan earthquake. The results show that the proposed method can not only effectively detect the outliers, but also make a fast accurate prediction. It is capable to be applied to the actual seismic damage prediction of urban buildings.

**Keywords:** Building groups; Seismic damage prediction; Support vector machine; Cross-validation

## 0 引言

地震是威胁人民生命财产安全的严重自然灾害之一。尤其当地震发生在城市时,其强突发性和大破坏力势必会造成巨大的人员伤亡和经济损失。其中,建筑物的严重破坏或毁坏是造成人员伤亡和经济损失的最主要原因。因此,合理而准确的预测出建筑物在不同等级的地震作用下的破坏程度,可以找出城市中建筑物的抗震薄弱环节,指导城市抗震防灾规划的编制,进而有目的的对城市建筑进行抗震加固。此外,震害预测工作对震后有目标的应急救援起着至关重要的作用,可以将地震造成的损失降到最低。

在以往文献中,建筑物震害预测分为单体建筑物震害预测和群体建筑物震害预测两种<sup>[1]</sup>。崔玉红和邱虎<sup>[2]</sup>总结了单体建筑物震害预测方法,主要包括:历史震害统计法、专家评估法、半经验半理论法、结构理论计算方法等。其中比较成熟的方法有时程分析法、有限元分析法、Push-Over方法等<sup>[3]</sup>。然而单体建筑物震害预测主要是针对高层、特种结构以及一些对抗震救灾起重大作用的建筑物<sup>[1]</sup>,不适用于城市一般建筑物的群体震害预测。

群体建筑物震害预测是指对一个区域(城市或地区)的建筑群在预期地震作用下的破坏进行估计,给出每个预测单元中各类建筑物在每个震害破坏等级中所占的比例<sup>[4]</sup>。群体建筑物震害预测方法有:模糊类比法<sup>[5-8]</sup>、震害矩阵模拟法<sup>[9-10]</sup>、地震动参数法<sup>[11]</sup>、单元破坏度指数法<sup>[11]</sup>、判别分析法<sup>[12]</sup>。然而,这些群体建筑物震害预测方法大多需要对影响因子设定权重,预测过程需要人为干预,因而预测结果存在依赖性和波动性。近年来,人工智能算法由于自身的计算自动化和高准确度等优点,渐渐地被广泛应用于群体建筑物震害预测<sup>[13]</sup>。常用的算法包括神经网络算法<sup>[14-16]</sup>,和支持向量机算法(Support Vector Machine, SVM)<sup>[17-18]</sup>。这些人工智能算法在实际的震害预测工作中,首

先选取一批有实际震害结果的建筑物作为训练样本,再用训练好的分类器对新的一批建筑物进行预测。因而这些训练样本的选择将直接关系到预测结果的准确性。由于特殊地质环境和特定地震情景的影响,外加人工统计的误差干扰,在实际震后收集的训练样本中存在相当数量的异常样本。这些数据噪声将严重影响群体建筑物震害预测的准确度。

为此,本文提出一种新型两阶段支持向量机算法,首先识别出训练样本中的异常数据,自动对其赋予一个低权重值,降低异常点对预测结果的影响;然后建立一个加权的支持向量机模型,快速准确的给出震害预测结果。

## 1 两阶段支持向量机方法

在介绍本文的两阶段支持向量机方法前,我们先简单了解下一般支持向量机算法的分类原理。

### 1.1 一般支持向量机方法

支持向量机算法是一种结合 VC 维理论(Vapnik-Chervonenkis dimension)和结构风险最小化理论的一种机器学习分类算法。该算法通过构造分类函数将两类不同的样本尽可能的区分开来,在解决小样本、非线性和高维空间识别问题上具有显著优势<sup>[19]</sup>。

假设  $S$  是一个包含  $n$  个训练样本  $(y_1, x_1), \dots, (y_n, x_n)$  的集合,对于每个训练样本,  $x_i \in \mathbf{R}^N$  ( $\mathbf{R}$  为实数域),  $y_i \in \{-1, 1\} (i=1, 2, \dots, n)$  是其类别标签。支持向量机算法就是要找出一个最优超平面  $w^T x + b = 0$ , 从而对于任一样本  $x$ , 我们可以按照(1)式中定义的决策函数对其分类。

$$f(x) = \text{sign}(w^T x + b) = \begin{cases} 1, & \text{if } y_i = 1 \\ -1, & \text{if } y_i = -1 \end{cases} \quad (1)$$

其中  $w \in \mathbf{R}^N$ ,  $b \in \mathbf{R}$ 。如果集合  $S$  是线性可分的,那么存在参数  $(w, b)$  使得所有样本点满足  $y_i (w^T x + b) \geq 0$ 。如果集合  $S$  是非线性可分的,那么定义  $z =$

$\varphi(x)$  ( $z \in \mathbf{R}^M$ ,  $M \geq N$ ), 表示用非线性函数  $\varphi(\cdot)$  将特征向量  $x$  从低维空间  $\mathbf{R}^M$  投影到更高维空间  $\mathbf{R}^M$ , 进而在高维空间中构建最优超平面。此时, 决策函数由式(1)转变成式(2)。

$$f(x) = \text{sign}(w^T z + b) \quad (2)$$

式(2)中,  $w \in \mathbf{R}^M$ ,  $b \in \mathbf{R}$ 。该最优超平面不仅要使两类样本正确的分开, 使训练错误率最小, 而且要使两类样本到该超平面的间隔最大。为了得到  $w$  和  $b$  的最优值, 可以将求最优超平面式(2)的分类问题转化成一个求最小值的优化问题式(3)。

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \quad (3)$$

$$s.t. \quad y_i(w^T z_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i=1, 2, \dots, n$$

其中非负松弛变量  $\xi_i$  表示对应的点  $z_i$  不满足式

(2), 因此  $\sum_{i=1}^n \xi_i$  可以看作是对误分类的一个测度。 $C$  为惩罚系数( $C > 0$ ),  $C$  越大, 说明对误分类的惩罚力度越大, 误分类越少, 但模型泛化能力下降, 反之相反。通过调节  $C$  获得最佳的分类器式(2)。式(3)可以通过构建 Lagrange 函数并将其转

化为对偶形式(4)进行求解。

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j z_i^T z_j \quad (4)$$

$$s.t. \quad \sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i=1, 2, \dots, n$$

其中  $\alpha_i$  为 Lagrange 乘子, 若  $\alpha_i > 0$ , 则对应的  $z_i$  为支持向量。令  $z_i^T z_j = \varphi(x_i)^T \varphi(x_j) = K(x_i, x_j)$ ,  $K(\cdot)$  为定义的核函数, 通常使用径向基函数  $\exp(-\gamma \|x - x_i\|^2)$ ,  $\gamma$  影响样本的分类间隔, 当  $\gamma$  越小时, 分类间隔越窄, 结构风险越大。最终的最优分类超平面可以通过求解式(5)获得。

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (5)$$

$$s.t. \quad \sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i=1, 2, \dots, n$$

对应的最终决策函数为  $f(x) = \text{sign}(w^T z_i + b) = \text{sign}(\sum_{i=1}^n \alpha_i y_i K(x_i, x) + b)$ 。

## 1.2 识别异常样本

上述最优分类超平面的构建易受到异常样本

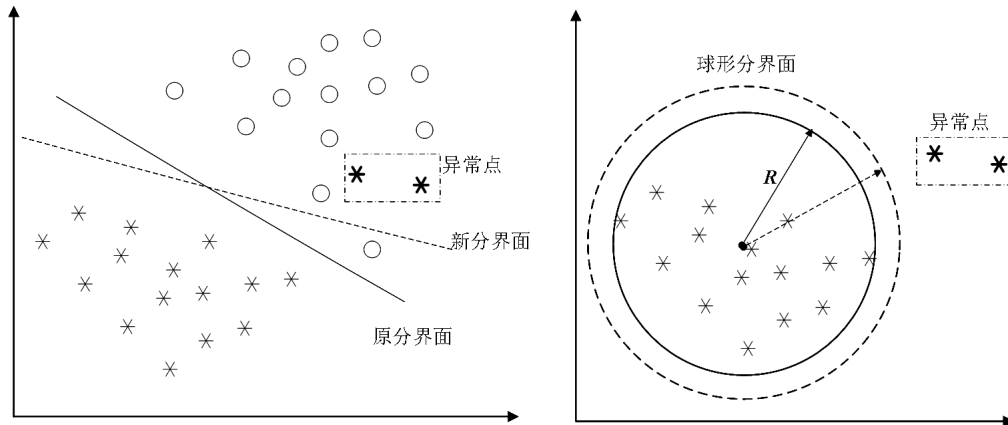


图1 异常点识别示意图

Fig.1 Schematic diagram of the identification of outliers

(噪声数据) 的影响, 导致分类结果不理想。图1中左图是以“\*”和“o”表示的两类样本, 当出现虚框中的两个“\*”类异常样本时, 使得最优分界面向异常样本倾斜, 导致“o”类样本出现误分类。尤其是在震后由于特殊地质环境和特定地震情景的影响, 外加人工统计的误差干扰, 收集的数据中存在相当数量的异常样本。为了识别出异常样本, Tax 和 Duin<sup>[20]</sup>在每类样本集所构成的空间外找到一中心为  $a$ , 半径为  $R$  的超球面, 要求在最

小化球半径的同时最大程度上包含该类的所有样本  $x_i$ 。如图1中右图所示, 实线球形分界面是满足要求的最优分界面。最终落在超球面外面的样本点就是异常样本。本文在该思想的基础上用核函数将特征向量投影到更高维空间, 在高维空间中构造球形分界面。该方法可用模型(6)表示<sup>[20]</sup>。

$$\max_{\eta} R^2 + D \sum_{i=1}^m \eta_i \quad (6)$$

$$s.t. \quad \|\varphi(x_i) - a\|^2 \leq R^2 + \eta_i, \eta_i \geq 0, i=1, 2, \dots, m$$

式(6)的构造和式(3)一致。假设该类样本有  $m$  个样本点, 当非负松弛变量  $\eta_i$  大于零时, 表示对应的样本点落在超球面外围, 因此  $\sum_{i=1}^m \eta_i$  可以看作是对异常样本量的一个测度。 $D$  为惩罚系数 ( $D>0$ ), 用来控制超球面的体积和允许落在球外面的样本量之间的折中。 $D$  越大, 说明允许落在球外面的样本量越少。通过调节  $D$  获得最佳的包围样本点的超球面。按照式(3)~(5)相同的流程, 式(6)可转化成其对应的偶问题式(7)进行求解。

$$\begin{aligned} \max_{\beta} \quad & \sum_{i=1}^m \beta_i K(x_i, x_i) - \sum_{i=1}^m \sum_{j=1}^m \beta_i \beta_j K(x_i, x_j) \\ \text{s.t.} \quad & \sum_{i=1}^m \beta_i = 1, \quad 0 \leq \beta_i \leq D, \quad i=1, 2, \dots, m \end{aligned} \quad (7)$$

$$d(x_k) = \|\varphi(x_k) - a\| = \sqrt{K(x_k, x_k) - 2 \sum_{i=1}^m \beta_i K(x_k, x_i) + \sum_{i=1}^m \sum_{j=1}^m \beta_i \beta_j K(x_i, x_j)} \quad (8)$$

$$\mu(x_i) \begin{cases} 1 - \frac{d(x_i)}{R} + \varepsilon, d(x_i) \leq R \\ \lambda, 0 < \lambda < \varepsilon, d(x_i) > R \end{cases} \quad (9)$$

而提高分类器的准确度。

### 1.3 加权支持向量机模型

在上一节获得样本权重后, 训练样本集表示为  $(y_1, x_1, \mu_1), \dots, (y_n, x_n, \mu_n), i=1, 2, \dots, n$ 。Lin 和 Wang<sup>[21]</sup>提出了一种样本加权支持向量机方法, 直接在式(3)中添加样本权重, 如式(10)所示。

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^n \mu_i \xi_i \quad (10)$$

$$\text{s.t. } y_i(w^T z_i + b) \geq 1 - \xi_i, \xi_i \geq 0, i=1, 2, \dots, n$$

其中  $C$  为惩罚因子,  $z = \varphi(x)$ 。当  $\mu_i$  很小时, 将相应的样本  $x_i$  看成是异常样本, 减少其对最优分类面的影响。采用式(3)~(5)同样的方法, 式(10)转化成为式(11)。

$$\begin{aligned} \max_a \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \text{s.t.} \quad & \sum_{i=1}^n \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq \mu_i C, \quad i=1, 2, \dots, n \end{aligned} \quad (11)$$

对应的最终决策函数为  $f(x) = \text{sign}(\sum_{i=1}^n \alpha_i y_i K(x_i, x_j) + b)$ 。通过对比模型(5)和模型(11), 可以看出一般支持向量机模型(5)是加权支持向量机模型(10)在  $\mu_i=1$  时的特例。

同时球心  $a = \sum_{i=1}^m \beta_i \varphi(x_i)$ , 是所有样本点的线性组合。 $K(\cdot)$  依然采用径向基函数  $\exp(-\gamma \|x - x_i\|^2)$ 。当  $\beta_i > 0$  时, 对应的样本为支持样本; 当  $0 < \beta_i < D$  时, 对应的样本落在球面上; 当  $\beta_i = D$  时, 对应的样本位于球面的外围, 即异常样本。该类样本集中任一样本  $x_k$  到球心的距离为  $d(x_k)$ 。

其中, 球半径  $R$  是球面上任一支持样本到球心的距离。由于正常样本落在最小超球面的内部, 而异常样本在球面的外围, 因而我们可以根据样本点到球心的距离为每个样本点赋予不同的权重。

其中是小于 1 的常数, 确保权重不为零, 分类超球面内的样本点被赋予权重。是充分小的常数, 使得异常样本被赋予一个充分小的权重, 这样极大的消除异常样本对构建分类器的影响, 从

## 2 震害预测模型的建立

2008 年 5 月 12 日汶川发生 8.0 级强震。笔者单位参与的地震科考队在陇南武都区、天水秦州区、文县城关镇、碧口镇、成县、康县、徽县、西和县详细地调查了房屋的震害情况, 取得了大量实际地震下的房屋震害样本。该房屋震害样本共 912 个建筑物, 包括 559 个砖混结构, 228 个钢筋混凝土框架结构, 57 个土木结构, 40 个底框架结构。本文采用真实房屋震害样本来建立模型, 比用单体模型预测的结果来建立模型更加符合真实情况, 预测结果更加准确。

### 2.1 震害影响因子的确定

根据宏观震害经验和相关研究<sup>[1,7]</sup>中采用的震害影响因子, 并考虑实地调查时的条件限制, 本文确定了 7 个主要的震害因子, 即用途, 结构类型, 建设年代, 层数, 结构平面型式, 设防烈度, 所处烈度。每个震害因子的分类如表 1 所示。其中, 结构平面型式是指建筑物的规整性, 所处烈度为地震发生时房屋所处烈度区。本文以这 7 个震害影响因子作为模型的输入, 而模型的输出为建筑物的震害结果。本文采用三种震害结果分类,

第一是按照尹之潜<sup>[24]</sup>分为 5 类，即基本完好，轻微破坏，中等破坏，严重破坏，毁坏；另外两种采用 Yüçemen 等人<sup>[10]</sup>提出的生命安全分类和立即居住分类。生命安全分类根据震害结果是否会危及到建筑物中人员的生命安全来分为两类，第一类包括基本完好、轻微破坏和中等破坏，第二类包括严重破坏和毁坏。立即居住根据震害对房屋质量影响、房屋是否能立即入住来分为两类，第一类包括基本完好和轻微破坏，第二类包括中等破坏、严重破坏和毁坏。

2.2 数据预处理

为了方便代入支持向量机模型中计算，本文

按照表 1 中的量化值对影响因子进行量化。由于本文将每个影响因子划分为多个子类别，而每个类别之间没有数值上的程度差异，因此本文将每个类别变量视为名义变量，采用哑变量对每个类别进行量化。由于实际数据收集过程中存在数据缺损，本文最终选取了数据集中、数量比较多的三种结构类型的房屋，即砖混结构(420 个)、钢筋混凝土框架结构(183 个)和底框架结构(37 个)，总计 640 栋建筑。样本中各影响因子的取值分布也列于表 1 中。对于五类分类的震害结果从基本完好到毁坏依次量化为 1, 2, 3, 4, 5。对于生命安全分类和立即居住分类，将一类结果量化为 1，另一类量化为-1。

表 1 震害影响因子分类及量化值

Table 1 Damage factors and quantized value of damage factors

类型	影响因子	量化值	样本分布/%	类型	影响因子	量化值	样本分布/%
用途	其它	(1,0,0,0,0,0,0)	0.47	层数	1 至 3 层	(1,0,0,0)	1.09
	厂房	(0,1,0,0,0,0,0)	0.16		4 至 6 层	(0,1,0,0)	42.66
	学校	(0,0,1,0,0,0,0)	6.56		7 至 9 层	(0,0,1,0)	48.91
	医院	(0,0,0,1,0,0,0)	5.00		10 层及以上	(0,0,0,1)	7.34
	住宅	(0,0,0,0,1,0,0)	49.06	平面型式	不规则	(1,0)	14.69
	商业	(0,0,0,0,0,1,0)	31.25		规则	(0,1)	85.31
	办公	(0,0,0,0,0,0,1)	7.50	设防烈度	未设防	(1,0,0,0)	19.22
结构类型	砖混结构	(1,0,0)	65.63		7 度设防	(0,1,0,0)	15.94
	底框架结构	(0,1,0)	5.78		8 度设防	(0,0,1,0)	64.84
	混凝土框架结构	(0,0,1)	28.59		9 度设防	(0,0,0,1)	0.00
建设年代	1979 年以前	(1,0,0,0)	2.03	所处烈度	6 度	(1,0,0,0)	34.06
	1980-1989	(0,1,0,0)	10.94		7 度	(0,1,0,0)	54.53
	1990-2000	(0,0,1,0)	39.22		8 度	(0,0,1,0)	9.69
	2001 年以后	(0,0,0,1)	47.81		9 度	(0,0,0,1)	1.72

2.3 模型的建立与评估

本文采用 5 折交叉验证的方法来训练支持向量机模型和测试样本，并采用 GridSearch 方法<sup>[25]</sup>对模型参数 和 进行寻优。首先将总体样本的顺序随机排列,再将样本平均划分为五部分，以组合的方式每次取其中的四部分(512 个样本)作为训练样本来训练模型，用剩下的样本(128 个样本)作为未知类别测试样本来测试模型的预测准确率。最后将五次预测准确率的平均值作为模型预测准确率的度量。在训练支持向量机模型时也采用 5 折交叉验证，将训练样本划分为五部分，最后以五次计算中效果最好的参数来预测测试样本类别。

另外，在两阶段支持向量机模型的第一阶段权重计算中，本文 取 0.1， 取 0.001，以此来消除

异常样本对第二阶段加权支持向量机模型构建的影响。表 2 给出了三种分类方式下的模型预测准确率，并将一般支持向量机方法和本文提出的方法进行对比。根据计算结果，我们得出以下规律

(1) 两种方法在生命安全分类和立即居住分类上的准确率都远远高于 5 类分类准确率。该规律说明在震害结果的五类中都存在大量的异常样本，导致 5 类分类准确率不高。

(2) 本文所提方法在立即居住分类上相对生命安全分类比一般方法更好。由于立即居住分类是主要划分轻微破坏和中等破坏，而生命安全分类是主要划分中等破坏和严重破坏，由此说明在人工判定震害结果时，轻微破坏和中等破坏更难区别，两类之间异常样本多，正是本文所提方法

消除了部分异常样本影响，使得在立即居住分类上取得更好效果。

(3) 本文所提方法在 5 类分类和立即居住分类上比一般 SVM 方法有更好的训练和测试准确率，在生命安全分类上二者获得同等程度的训练

和测试准确率。经过反复测试，均发现这个规律，不仅说明支持向量机模型能够适用于群体建筑物震害预测，而且验证了本文所提两阶段加权支持向量机方法比一般方法更有优越性，可以应用于实际的震害预测工作中。

表 2 分类结果对比  
Table 2 Comparison of the classification results between two methods

方法	5 类分类		生命安全分类		立即居住分类	
	训练	测试	训练	测试	训练	测试
一般 SVM	61.91% <sup>①</sup> (4, 0.063) <sup>②</sup>	52.34%	92.38%	91.41%	91.58%	81.25%
两阶段加权 SVM	67.72% (0.3, 1), (2, 0.5) <sup>③</sup>	53.54%	92.38%	91.41%	92.60%	85.16%

注：① 本文给出的准确率是 5 折交叉验证后的平均值；② 这是一般 SVM 模型的参数；③ 前一个是两阶段模型第一阶段的参数，后一个是第二阶段的参数。

在本文所提方法的第一阶段识别异常样本过程中，模型(7)中参数  $D$  是对异常样本的惩罚系数，越大，说明允许落在超球面外面的异常样本越少。由于本文采用的数据集是地震后人工采集的建筑物震害样本，数据集中存在部分异常样本。为了将异常样本识别出来，可以通过调整参数  $D$ ，强制使部分样本落在最优超球面外面。图 2 给出了固定时，针对同一套训练样本和测试样本，不同取值下的最终 5 类分类准确率的变化。由图 2 可以看出，当  $D$  较小时，被视为异常样本的数目过多，预测准确率不高；当  $D$  较大时，对异常样本的惩罚

过大，预测准确率也不高；只有当  $D$  取值适中，将部分样本视为异常样本，消除其影响，取得更好的准确率。因此，在实际的城市建筑物震害预测工作中，可以人为对采样的建筑物设置一定的异常样本比率，从而使得模型预测的更加准确。

3 结语

本文针对城市群体建筑物震害预测的已知样本中存在大量异常样本的问题，提出了一种新型两阶段加权支持向量机方法。利用一般支持向量机方法在解决小样本、非线性和高维空间识别问

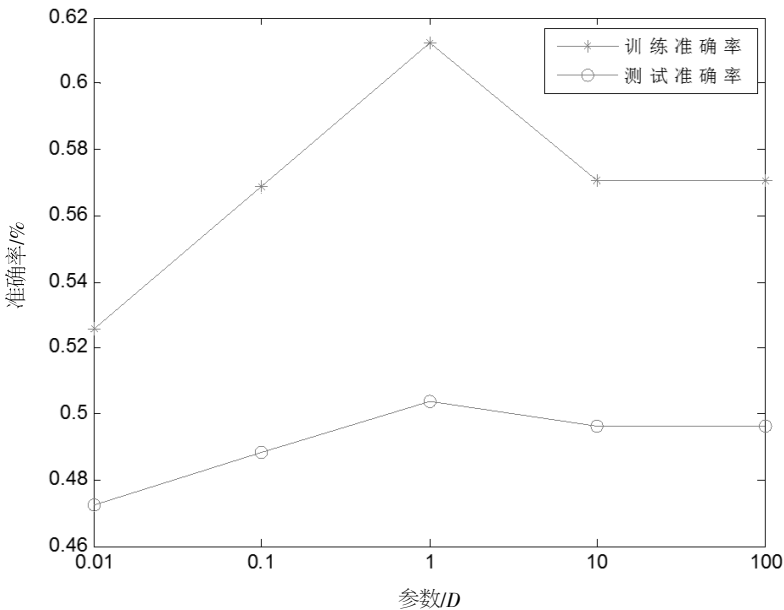


图 2 预测准确率的变动图  
Fig.2 Variation of predicting accuracy with the change of D

题上的显著优势, 再通过对正常样本和异常样本赋予不同的权重, 消除异常样本影响, 最后通过加权支持向量机模型实现群体建筑物震害预测。通过对汶川地震中的实际震例测试证明本文所提方法不仅能有效识别异常数据样本, 而且比一般支持向量机算法更具优越性, 可以用于实际的震害预测工作中。

本文研究的一个不足之处是没有考虑样本类别分布及影响因子取值分布不均匀对两阶段支持向量机模型预测效果的影响, 以后在收集到大量的各类样本后可以逐个考虑样本分布对模型计算结果的影响。未来的研究还可以在本文所提两阶段加权支持向量机方法的基础上, 考虑多种结构类型的建筑物样本, 并当每种类型数据样本很多时, 可以按结构类型分库, 使得预测更加准确。另外, 由于震后难以快速获得场地类别数据, 本文没有把场地类别纳入影响因子。然而场地类别对建筑物震害有显著影响, 以后的研究可以考虑把场地类别作为影响因素, 使得震害预测结果更加准确。

## 参考文献:

- [1] 高杰, 冯启民, 张海东. 城市群体建筑物震害模拟方法研究[J]. 震灾防御技术, 2007, 2 (2): 193-200.
- [2] 崔玉红, 邱虎. 国内外单位建筑物震害预测方法研究述评[J]. 地震研究, 2001, 24 (2): 175-182.
- [3] 魏巍, 冯启民. 几种 push-over 分析方法对比研究[J]. 地震工程与工程振动, 2002, 22 (4): 66-73.
- [4] 胡少卿. 建筑物的群体震害预测方法研究及基础设施经济损失预测方法探讨[D]. 中国地震局工程力学研究所, 2007.
- [5] 郭小东, 苏经宇, 马东辉, 等. 城市建筑物快速震害预测系统[J]. 自然灾害学报, 2006, 15 (3): 128-134.
- [6] 王志涛, 苏经宇, 马东辉, 等. 群体建筑物震害特征类比预测方法与应用[J]. 北京工业大学学报, 2008, 34 (8): 842-847.
- [7] 孙柏涛, 孙得璋. 建筑物单体震害预测新方法[J]. 北京工业大学学报, 2008, 34 (7): 701-707.
- [8] 李升才, 王毅恒. 类比预测法在城市群体建筑物震害预测中的应用[J]. 地震工程与工程振动, 2011, 31 (5): 166-171.
- [9] 孙柏涛, 胡少卿. 基于已有震害矩阵模拟的群体震害预测方法研究[J]. 地震工程与工程振动, 2006, 25 (6): 102-108.
- [10] 张桂欣, 孙柏涛. 多因素影响的建筑物群体震害预测方法研究[J]. 世界地震工程, 2010, 1 (1): 26-30.
- [11] 陶正如, 陶夏新. 基于地震动参数的建筑物震害预测[J]. 地震工程与工程振动, 2004, 24 (2): 88-94.
- [12] Yüçemen, M Ö zcebe, G, Pay, A. Prediction of potential damage due to severe earthquakes [J]. Structural Safety, 2004, 26 (3): 349-366.
- [13] Tesfamariam, S, Liu, Z. Earthquake induced damage classification for reinforced concrete buildings [J]. Structural safety, 2010, 32 (2): 154-164.
- [14] 汤皓, 陈国兴. 灰关联与人工神经网络在建筑物震害预测中的应用[J]. 地震工程与工程振动, 2006, 26 (3): 57-59.
- [15] Reyes, J, Morales-Esteban, A, Martínez-Álvarez, F. Neural networks to predict earthquakes in Chile [J]. Applied Soft Computing, 2013, 13 (2): 1 314-1 328.
- [16] De Lautour, O R, Omenzetter, P. Prediction of seismic-induced structural damage using artificial neural networks [J]. Engineering Structures, 2009, 31 (2): 600-606.
- [17] 刘勇健, 刘义建, 张伯友. 基于粗糙集—支持向量机的震害预测模型及应用[J]. 地震研究, 2008, 31 (3): 289-295.
- [18] Chen, C-S, Cheng, M-Y, Wu, Y-W. Seismic assessment of school buildings in Taiwan using the evolutionary support vector machine inference system[J]. Expert Systems with Applications, 2012, 39 (4): 4 102-4 110.
- [19] Vapnik, V N. The nature of statistical learning theory[M]. New York: Statistics for engineering and information science. Springer-Verlag, 2000.
- [20] Tax, D M, Duin, R P. Support vector data description [J]. Machine learning, 2004, 54 (1): 45-66.
- [21] Lin, C-F, Wang, S-D. Fuzzy support vector machines [J]. IEEE Transactions on Neural Networks, 2002, 13 (2): 464-471.
- [22] 尹之潜. 地震灾害及损失预测方法[M]. 北京: 地震出版社, 1995.
- [23] 何萍, 王挺. 基于抗震性能普查数据的城区群体建筑震害预测初探[J]. 华南地震, 2014, 34 (1): 21-25.