

陈小芳, 戚洪飞, 俞岗. 基于支持向量机的地震灾害自动成图方法[J]. 灾害学, 2021, 36(1): 117–121. [CHEN Xiaofang, QI Hongfei and YU Gang. Automatic Mapping Method of Earthquake Disaster Based on Support Vector Machine[J]. Journal of Catastrophology, 2021, 36(1): 117–121. doi: 10.3969/j.issn.1000-811X.2021.01.022.]

基于支持向量机的地震灾害自动成图方法^{*}

陈小芳¹, 戚洪飞², 俞 岗¹

(1. 广东省地震局地震监测与减灾技术重点实验室, 广东 广州 510070; 2. 中地数码科技有限公司, 湖北 武汉 430000)

摘要: 地震灾害风险评估中有许多定性图件, 目前常用的成图方法是根据研究区的离散点钻孔结果人工勾画, 费时费力; 或对离散点钻孔结果强制赋予0、1数值, 然后进行插值成图, 其从逻辑上不能真实反映场地的客观情况。基于支持向量机技术, 该文提出了使用分类工具SVC, 将不同结果的离散点钻孔在空间上进行合理分类的自动成图方法。相较于传统方法, 该方法效率高, 成本低。

关键词: 地震灾害; 风险评估; 钻孔; 定性灾害图件; 自动生成

中图分类号: X43; X915.5; P315 **文献标志码:** A **文章编号:** 1000-811X(2021)01-0117-05

doi: 10.3969/j.issn.1000-811X.2021.01.022

我国是世界上地震活动强烈和地震灾害严重的国家之一^[1]。一旦发生地震, 往往会带来十分严重的人员伤亡和巨大的经济损失^[2], 如2008年汶川地震^[3]。人们对地震灾害风险的正确评估和科学管理成为实现减灾的最佳途径^[4]。本文所说地震灾害的定性图件包括场地土类型分布图、场地类别分布图、砂土液化分布图、软土震陷分布图, 以及城市用地各种图件等。目前常用的成图方法是根据研究区的离散点钻孔的结果人工勾画, 费时费力。当遇到研究区面积大, 如达到上千平方公里, 且离散点钻孔分布不均匀时, 弊端更加显现。还有一种方法是对离散点钻孔结果强制赋予0、1数值, 然后进行插值, 即针对离散点钻孔插值(采用距离反比)生成栅格, 根据栅格分类输出为矢量图的成图方法, 其从逻辑上不能真实反映场地的客观情况。为了解决现有技术中的缺陷, 本文提出了基于SVC(Support vector classification)的自动成图方法, 其具有完善的将SVC作为分类工具的成图流程(图1), 实现了自动生成定性地震灾害图件的目标。

1 技术要点

SVM(Support Vector Machine, 支持向量机)是与相关的学习算法有关的监督学习模型, 可以分析数据、识别模型, 用于分类和回归分析。根据输入的数据不同可做不同的模型。当将SVM用于分类分析时, 其输入标签为分类值, 此时其也被称为SVC(Support vector classification)。将SVC作为分类工具被应用于许多领域, 如对岩相的识别^[5], 地震体波震相的识别与拾取^[6], 围岩力学参数识别^[7], 交通流量预测^[8]等。SVC不仅可用

于线性分类(图2), 而且还可用于非线性分类。

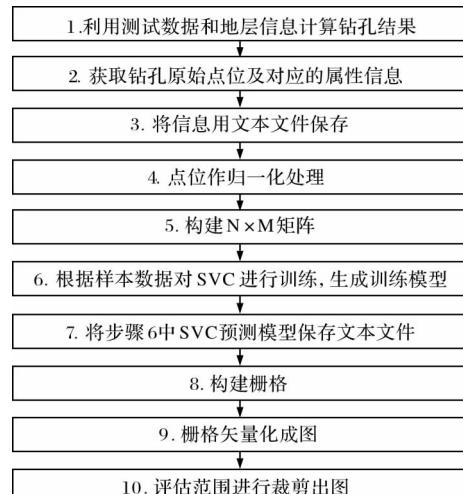


图1 基于SVC自动生成定性图件流程

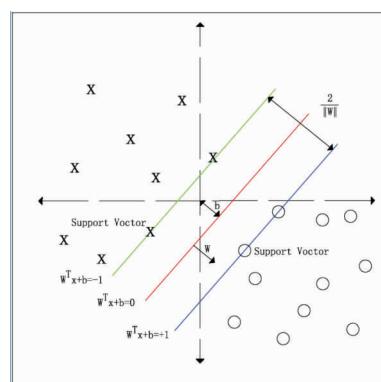


图2 线性可分SVC示意图

* 收稿日期: 2020-06-23 修回日期: 2020-08-10

基金项目: 广东省科技计划项目(2017B030314082); 广东省防震减灾现代化试点省项目(2060301)

第一作者简介: 陈小芳(1966-), 女, 汉族, 广东广州人, 高级工程师, 主要从事城市震害预测与灾害防御研究.

E-mail: 3211290447@qq.com

图 2 中绿色、蓝色直线分别为支持向量所在的边界，落在上面的点为支持向量，而非支持向量的点，则有 $y_i(w^T x_i + b) > 1$ 。红色线为超平面，用其可将样本划分出来。两个支持向量边界之间的距离用 $\frac{2}{\|w\|}$ 表示， $\frac{2}{\|w\|}$ 越大，分类的确信度也越大，即 $\frac{1}{2} \|w\|^2$ 要尽可能小。在此函数上加上一个惩罚函数，并且带上其限制条件，得到优化的 SVC 模型：

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m \xi_i, \text{ s.t. } y^i(w^T x^i + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, \dots, m. \quad (1)$$

经过拉格朗日转换，求对偶问题，得到预测函数：

$$f(x) = w^T x + b = \sum_{i=1}^m \alpha_i y^{(i)} x^{(i)T} x + b. \quad (2)$$

式中： m 是数据集个数， α 是拉格朗日乘子法引入的一个系数，针对数据集中的每个样本 x^i ，都有对应的 α_i 。 x^i 是数据集中第 i 个样本的输入，它是一个向量， y^i 对应输出标签，值为 $y \in [-1, 1]$ 。

ξ_i 表示对第 i 个样本点的分类错误概率， $\sum_{i=1}^m \xi_i$ 是总的样本误差。 C 是惩罚系数， C 值大时，对误分类的惩罚增大； C 值小时，对误分类的惩罚减小。要使 $\frac{1}{2} \|w\|^2$ 尽可能小，同时使误分类点的个数尽量少， C 即是调和两者的系数。

当 $C \rightarrow \infty$ 时，不允许出现分类误差的样本出现，则是过拟合。

当 $C \rightarrow 0$ 时，不再关注分类是否正确，只要求间隔越大越好，便无法得到一个有意义的解，且算法不会收敛，则是欠拟合。

现实中不存在线性的可分超平面，但是映射到更高维度就可分了。核函数就是用一定规则把这些无法进行线性分割的样本映射到更高纬度的空间里，然后找出超平面。令 $\varphi(x)$ 为将 x 映射后的特征变量，则划分超平面所对应的模型，即映射预测模型为：

$$f(x) = w^T x + b = w^T \varphi(x) + b. \quad (3)$$

经过相似性函数转换后，核函数： $k(x^{(i)}, x^{(j)}) = x^{(i)T} x^j = \varphi(x^{(i)})^T \varphi(x^j)$ ， $x^{(i)T} x^j$ 是两个向量内积， $\varphi(x^{(i)})^T \varphi(x^j)$ 是低维度的样本 $x^{(i)}$ 和 x^j 映射到特征空间之后的内积，它的物理含义是衡量高维度的两个向量的相似性。

我们提出的基于向量机的自动成图方法使用高斯核函数，其优点在于可以把特征向量映射到无限多维空间里，并且容易计算，参数也比较容易选择，这样极大简化了计算复杂度。当点密集且杂乱时，使用高斯核函数最佳。高斯核函数：

$$K(x^{(i)}, x^{(j)}) = \exp\left(-\frac{(x^{(i)} - x^{(j)})^2}{2\sigma^2}\right) = \exp(-gamma(x^{(i)} - x^{(j)})^2), \text{ 带宽 } \sigma > 0. \quad (4)$$

因 $k(x^{(i)}, x^{(j)}) = x^{(i)T} x^j = \varphi(x^{(i)})^T \varphi(x^j)$ ，经变换，可得高斯核函数对应的特征变换函数：

$$\varphi(x) = \sum_{i=0}^{\infty} \exp(-x^2) \sqrt{\frac{2^i}{i!}} x^i. \quad (5)$$

前面无限多项的累加器 $\sum_{i=0}^{\infty}$ ，其物理意义就是

把特征向量转换到无限多维向量空间里，即高斯函数可以把输入特征扩展到无限多维空间里。参照式(2)，得到高斯预测函数：

$$f(x) = \sum_{i=1}^m \alpha_i y^{(i)} k(x^{(i)}, x^{(j)}) + b = \sum_{i=1}^m \alpha_i y^{(i)} \varphi(x^{(i)})^T \varphi(x^j) + b. \quad (6)$$

式中： α_i 只在支持向量对应的样本处不为 0，其他的样本为 0。因此，预测函数是中心点在支持向量处的高斯函数的线性组合，其线性组合的系数为 $\alpha_i y^{(i)}$ 。

当用高斯核函数训练 SVC 模型时，有两个参数必须考虑： C 惩罚系数和 $gamma$ 。 C 表示对误差的宽容度， C 越高，说明越是不能容忍误差的出现， C 越小，表示越容易欠拟合，泛化能力变差。 $gamma$ 是在选择高斯函数后才出现的，表示数据映射到新的特征空间后的分布， $gamma$ 值越大，支持向量就越少， $gamma$ 值越小，支持向量就越多。支持向量的个数影响训练与预测的速度。

利用上述模型，我们将 SVC 作为分类工具开发出完善的成图模块，体现了 SVC 在样本点数量较少，且不密集情况下，以及评估范围达到上千平方公里，且离散点钻孔不均匀分布时的技术优势，其成图流程见图 1。在自动生成定性地震灾害分布图时，其可以对定性的离散点钻孔进行更加合理地分类，而不是简单的使用具体数值进行插值而分类。插值分类方法依赖于离散点的密集程度，往往很生硬，且容易弱化掉一些不突出的点，把本该归为 A 类的点归为 B 类。而基于 SVC 的方法，通过构建矩阵，训练 SVC 模型，以预测矩阵点的属性值，将预测后的矩阵栅格矢量化获得准确的评估图。这个成图过程可以修正插值分类的缺陷，为工程人员以及其他专业人员提供准确的分布图。此外，本成图方法不再使用人工勾画，提高了成图效率与精度，降低了人工成本。

2 具体实施步骤

步骤 1：计算钻孔结果。

①首先将钻孔数据按照基本信息数据表、地层信息数据表、土样试验数据表、岩石试验数据表、标贯数据表、波速测试数据表规范整理入库。

②利用上述钻孔相关信息，结合以下相关规范判别钻孔场地结果。

场地类别：根据《GB 50011-2010 建筑抗震设计规范》之 4.1.6 条^[9]，依据岩石的剪切波速或土的等效剪切波速进行分类。

场地土类型：根据《GB 50011-2010 建筑抗震设计规范》之 4.1.3 条^[9]，依据场地土的等效剪切波速进行分类。

砂土液化：根据《GB 50011-2010 建筑抗震设计规范》之 4.3.4、4.3.5 条^[9]，采用标准贯入试验判别法判别地面下 20 m 范围内土是否液化和土的液化等级。

软土震陷：①根据《JGJ83-2011 软土地区岩土工程勘察规程》之 6.3.4 条^[10]，依据临界等效剪切波速 v_{se} (m/s)，地基主要受力层深度内软土厚度 > 3 m 或地基土等效剪切波速值 < 90 m/s 进行判别；②根据《GB 50021-2001 岩土工程勘察规范》(2009 年版)之 5.7.11 条^[11]，依据承载力特征值 f_a (kPa) 或等效剪切波速 v_{sr} (m/s) 值判别。

表 1 场地土类型判别结果

钻孔 ID	项目编号	工程名称	工程地点	X 坐标	Y 坐标	孔号	等效剪切波速 $v_{se}/(m/s)$	场地土类型
107446	2003C0179	* * * * * * * * * * * * * * * *		431 517.58	2 563 434.22	37	134.52	软弱土
107593	2003c0262	* * * * * * * * * * * * * * * *		432 086.77	2 562 501.29	7	164.91	中软土
110760	2006C0271	* * * * * * * * * * * * * * * *		432 060.04	2 563 376.18	5	629.71	坚硬土或软质岩石
107112	2002c0480	* * * * * * * * * * * * * * * *		431 716.09	2 562 059.79	12	245.74	中软土
107622	2003C0293	* * * * * * * * * * * * * * * *		431 440.75	2 562 315.81	4	138.04	软弱土
.....

表 2 软土震陷判别结果

钻孔 ID	项目编号	工程名称	工程地点	X 坐标	Y 坐标	孔号	软土厚度/m	等效剪切波速 $v_{se}/(m/s)$	是否震陷
106057	2001c0300	* * * * * * * * * * * * * * * *		431 227.84	2 562 781.96	5	1.45	176.06	不震陷
110336	2006C0032	* * * * * * * * * * * * * * * *		431 873.59	2 562 911.87	9	1.50	206.54	震陷
107622	2003C0293	* * * * * * * * * * * * * * * *		431 440.75	2 562 315.81	4	2.15	138.04	震陷
109054	2004c0539	* * * * * * * * * * * * * * * *		432 163.90	2 563 697.01	5	2.20	185.78	不震陷
107646	2003c0317	* * * * * * * * * * * * * * * *		431 773.83	2 563 488.39	1	3.00	204.34	震陷
.....

将判别结果写入判别表中, 供成图使用。如表 1、表 2 分别为场地土类型、软土震陷的判别结果。以上判别方法均已编程集成到系统中。

步骤 2: 基于系统, 从表 1、表 2……中获取钻孔原始点位坐标及对应的判别结果, 即属性值。点位坐标通常以 X 轴和 Y 轴表示, 属性值通常以 Z 轴表示。这些属性信息包括①场地土类型分布图对应的研究区每个离散点钻孔的场地土类型: 岩石、坚硬土或软质岩石、中硬土、中软土、软弱土; ②场地类别分布图对应的研究区每个离散点钻孔的场地类别: I₀、I₁、II、III、IV; ③砂土液化分布图对应的研究区每个离散点钻孔是否液化: 液化、不液化, 或液化等级: 轻微液化、中等液化、严重液化; ④软土震陷分布图对应的研究区每个离散点钻孔是否震陷: 震陷、不震陷, 以及其它类似的定性灾害构件的属性信息。

步骤 3: 将步骤 2 获取的信息用文本文件保存, 格式为每行记录 X、Y、Z, 其中 Z 为分类, 必须是整型数值(INT)。

步骤 4: 将钻孔点位的 X、Y 轴数据作归一化处理。由于步骤 2 中收集的钻孔点的坐标值是实际坐标, 数值较大, 导致后期 SVC 训练模型出现异常, 并且分类效果欠佳, 故此处做归一化处理, 默认范围是 [-1, 1], 更加便于程序调整网格间距。

步骤 5: 以钻孔点的外包矩形为边界和归一化之后设置的间距进行插值构建 N × M 个矩阵预测点。其属性值需要预测, 即为预测值 Z, N 和 M 的范围需要经验确定, 一般 2 000 以内可在效率与结果逼真度上取得平衡。N、M 越大, 对应的网格越密, SVC 预测越接近可信数据, 但是计算量也更大, 耗时越多。优选地, N = 2 000, M = 2 000。

步骤 6: 将步骤 4 中获得的数据作为训练样本, 导入下面程序中, 对步骤 5 数据进行训练, 获得 N × M 个矩阵点的预测值, 生成 SVC 预测模型。在该步骤中, 使用 python 脚本代码, python 脚本经

pyinstaller 打包成 exe, 在 .NET 环境中以进程方式调用, 此方法极大地简化了系统集成的复杂度。

Clf = SVC(kernel='rbf', gamma=50, class_weight='balanced', C=1.0, random_state=0)

其中'rbf'为高斯核函数。C、gamma 值用户可根据 SVC 训练结果进行调整。

步骤 7: 将步骤 6 中经过训练之后的预测模型保存到文本文件中(txt 格式)。

步骤 8: 根据步骤 7 文本文件中的 X、Y、Z 数据构建栅格。

步骤 9: 将步骤 8 中构建的栅格矢量化, 生成矢量分布图。由于步骤 6 的矩阵坐标和对应的属性值数据都通过了 SVC 模型训练, 数据准确, 所以将步骤 8 中的栅格矢量化, 相当于囊括了整个评估范围的点, 从而获得了整个评估范围的相应属性的评估图片。

步骤 10: 用评估范围进行裁剪, 最终生成指定范围内的分布图。

将上述步骤编程集成, 即构成系统中自动成图功能模块。成图时只需点击成图相关菜单即可生成定性评估图片。

运行环境: Python: 3.6, Matplotlib: 3.0.0, Pyshp: 1.2.12, Sklearn: 0.19.0。

3 应用示例

以广州市部分城区为例, 应用本方法生成场地土类型分布图和软土震陷分布图。该区域包括越秀区、荔湾区、天河区、海珠区, 以及白云区、萝岗区、番禺区、黄埔区的部分区域, 共计 518 km²。收集的钻孔深度在 20 ~ 30 m 不等, 均打穿第四系至基岩以下 2 ~ 3 m。第四系主要为全新世三角洲相和陆相冲积土, 晚更新世三角洲相和陆相冲积或坡积土, 以及少量晚更新世残积土。

具体示例操作如下:

以收集到区域内 517 个有效钻孔为离散点, 说明自动生成场地土类型分布图的具体步骤。

步骤 1: 运用成图系统计算各钻孔的场地土类型, 计算结果见表 1。

步骤 2: 运用成图系统选取成图范围内钻孔, 获取钻孔点位数据及属性信息。属性信息, 即分类值 Z : 岩石、坚硬土或软质岩石、中硬土、中软土、软弱土。

步骤 3: 将步骤 2 获得的数据信息以文本文件 (txt 格式) 保存, 每行记录 X , Y , Z 。其中 X , Y 分别代表钻孔的具体位置坐标, Z 为分类, 必须为整形数值 (INT), 比如 1/2/3/4/5 分别代表岩石/坚硬土或软质岩石/中硬土/中软土/软弱土。

步骤 4: 将步骤 2 中的钻孔坐标 X , Y 做归一化处理, 避免大数据范围下插值过于密集超出计算时长。

步骤 5: 以钻孔点位的外包矩形为边界和归一化之后设置的间距进行插值, 构建 $N \times M$ 个矩阵点。 N 和 M 的范围需要经验确定, 一般 2 000 以内可在效率与结果逼真度上取得平衡。 M , N 的值越大, 对应的网格越密, SVC 评估越接近可信数据, 但是计算量也更大。在本实施例中, N , M 取值均取 2 000。

步骤 6: 将步骤 4 获取的归一化之后的场地土类型结果的钻孔点数据作为训练样本, 导入下面程序中, 对步骤 5 数据进行训练, 获得 $N \times M$ 矩阵预测点的属性值, 生成 SVC 预测模型。针对场地土类型数据, 由于点密集且分布不均匀, 选择高斯核函数 ($kernel = 'rbf'$), 经过反复调参, 确定 $gamma$ 取 50, C 取 1.0, 效果和时间成本相对最优。由于样本分布不均匀, 为了避免部分数据被当成噪声点弱化掉, 采用均衡权重 ($class_weight = 'balanced'$)。其他参数使用默认值, 样本数据 SVC 训练代码如下:

```
Clf = SVC(kernel = 'rbf', gamma = 50, class_weight = 'balanced', C = 1.0, random_state = 0)
```

步骤 7: 将步骤 6 生成的 SVC 预测模型结果保存到文本文件中 (txt 格式)。

步骤 8: 根据步骤 7 文本文件中的 X , Y , Z 数据构建栅格, 得到 $2 000 \times 2 000$ 的栅格。

步骤 9: 对 $2 000 \times 2 000$ 的栅格进行矢量化。

步骤 10: 用评估范围对结果进行裁剪, 生成场地土类型分布图 3。

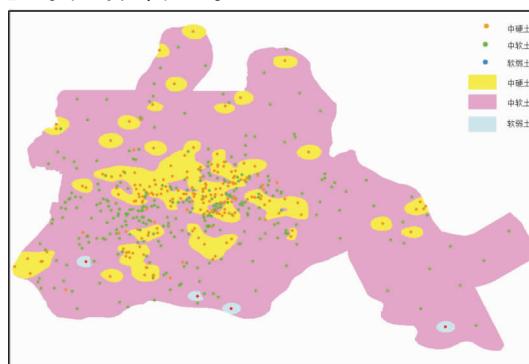


图 3 场地土类型分布图

同样地, 用此方法, 以收集到区域内 5 766 个有效钻孔为离散点, 其中整型分类值 Z 以 2、5 分别代表震陷/不震陷, 自动生成的软土震陷分布图 4、图 5。

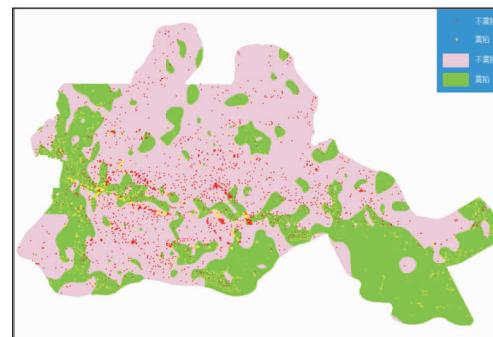


图 4 软土震陷分布图

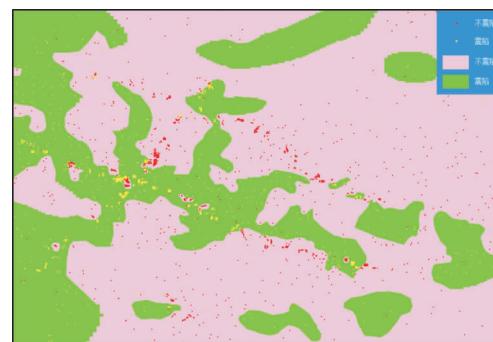


图 5 局部放大后的软土震陷分布图

将地形、地貌、水系等图层与生成的上述两种图件叠加, 结合专家经验发现, 基于 SVC 方法生成的成果图较为可靠。

以上步骤已编程集成。在成图时, 只需选取参与成图的钻孔, 点击相关预评估的灾害菜单, 即可自动生成相关图件。

如生成场地土类型分布图时, 点击图 6 中“场地土类型”菜单, 进入“场地土类型分布图”界面, 点击右下角“生成分区图…”, 即可生成场地土类型分布图。

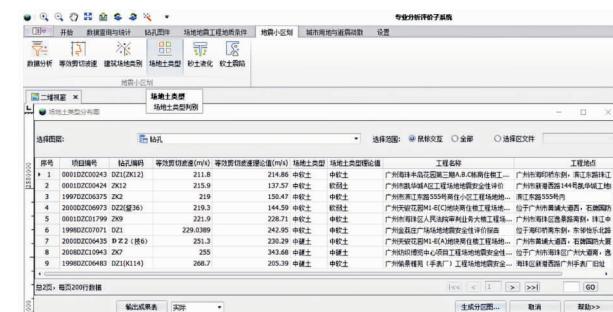


图 6 成图程序界面

5 讨论

(1) 基于 SVC 的自动成图方法, 可以通过训练获取预测点的结果, 生成 SVC 预测模型, 进而对定性的离散点进行更加合理地分类, 提高了成图效率, 降低了项目成本, 缩短了建设周期。

我国改革开放 40 a 来, 各地完成了大量基建项目, 积累了丰富钻孔数据, 将这些数据经过专业人员标准化处理, 导入系统, 运用系统相关功能, 即可快速获得灾害分布图。评估面积越大, 其优越性则更加显现出来。

截止目前, 广州市已完成地震小区划的面积占广州市总面积只有 11.2%, 需要开展地震灾害评估的地方还很多。近年来, 我们应用该方法已开展多区域地震灾害评估, 均取得较好效果。未来拟在更多区域的地震灾害评估中得到应用, 如深圳、中山、珠海、惠州、东莞、佛山等。其它省亦可应用此法开展地震灾害评估。

(2) 可对其它方法产出的结果进行验证或修正。如运用本方法评估时使用的钻孔密度大, 达到 $3 \sim 5$ 孔/ km^2 , 或更高, 高于地震小区划的规定, 其结果更加精准。因此, 可对完成过地震小区划的成果进行修正。

(3) 产出的成果可用于区域地震安全性评价。

参考文献:

[1] 张培震. 中国地震灾害与防震减灾[J]. 地震地质, 2008, 30

- [2] 张弓强. 试论地震灾害对经济社会发展的影响[J]. 灾害学, 2012, 27(1): 121–124.
- [3] 郑通彦, 李洋, 侯建盛, 等. 2008 年中国大陆地震灾害损失述评[J]. 灾害学, 2010, 25(2), 112–118.
- [4] 徐中春. 中国地震灾害风险综合评估[D]. 北京: 中国科学院大学, 2011.
- [5] 汪益宁, 闫荣堃, 罗佳洁, 等. 基于支持向量机的致密储层岩相识别——以徐家圈子断陷下白垩统沙河子组为例[J]. 长江大学学报(自然科学版), 2016, 13(29): 33–38.
- [6] 蒋一然, 宇杰远. 基于支持向量机的地震体波震相自动识别及到时自动拾取[J]. 地球物理学报, 2019, 62(1): 361–373.
- [7] 王复明, 李晓龙, 苗丽, 等. 基于小波支持向量机的围岩力学参数辨识[J]. 水利发电学报, 2010, 29(3): 184–190.
- [8] 郑义彬, 赖伟伟. 基于支持向量机的高速短时交通流量预测[J]. 工程与建设, 2020, 34(2): 201–204.
- [9] GB 50011–2010 建筑抗震设计规范[S]. 北京: 中国建筑工业出版社, 2010.
- [10] JGJ 83–2011 软土地区岩土工程勘察规程[S]. 北京: 中国建筑工业出版社, 2010.
- [11] GB 50021–2001 岩土工程勘察规范(2009 年版)[S]. 北京: 中国建筑工业出版社, 2009.

Automatic Mapping Method of Earthquake Disaster Based on Support Vector Machine

CHEN Xiaofang¹, QI Hongfei² and YU Gang¹

(1. Key Laboratory of Earthquake Monitoring and Disaster Mitigation Technology, Guangdong Earthquake Agency, Guangzhou 510070, China; 2. Wuhan ZONDY CYBER T&S CO., LTD, Wuhan 430000, China)

Abstract: There are many qualitative maps in the risk assessment of earthquake disaster. At present, they are commonly manual sketched based on the borehole results of discrete points in the study area, which is time-consuming and laborious. Or the borehole results of discrete points are forced to be given values of 0 and 1, and then the interpolation mapping is carried out. It can't really reflect the objective situation of the site logically. Based on the technology of support vector machine, this paper proposes a automaticmapping method by using the classification tool SVC to classify the discrete points with different results reasonably in space. Compared with the traditional method, this method has high efficiency and low cost.

Key words: earthquake disaster; risk assessment; borehole; qualitative disaster map; auto generate