

支持向量机在城市震后火灾损失预测中的应用*

王威¹, 苏经宇¹, 韩阳², 马东辉¹, 王志涛¹

(1. 北京工业大学 抗震减灾研究所, 北京 100022; 2. 河南工业大学 防灾减灾工程研究所, 河南 郑州 450052)

摘要: 针对地震火灾的复杂性和多变性的特点, 在综合分析了各种震后火灾损失预测方法的基础上, 提出了基于支持向量机的城市震后火灾损失预测模型。以地震火灾的历史资料为依据, 建立学习样本和测试样本, 并与 BP 神经网络预测方法相比较, 验证了其可行性, 也为其它自然灾害的损失预测提供了简单、有效的方法。

关键词: 地震火灾; 损失预测; 支持向量机(SVM); BP 神经网络

中图分类号: P315.9; TP391 **文献标识码:** A **文章编号:** 1000-811X(2008)01-0006-04

0 引言

城市遭受大地震后因房屋损坏、煤气管道与供电设施等生命线的破损而形成火源及其蔓延成灾的现象极为普遍, 因火灾造成损失往往远超过地震本身的破坏损失, 所以怎样对此作出预测与对策有重大的经济与实用意义^[1,2]。一些学者对地震火灾损失的预测与评估做了许多工作, 探讨了地震火灾损失评估的方法, 主要以下两类: ①确定性方法。根据以往的地震评估经验, 对地震损失的评估采用现场抽样统计和调查的方法^[3-5]。②不确定性方法。根据地震区地震的强度、房屋结构等有关资料, 综合考虑各个因素的影响, 利用一定的数学物理理论(如概率模型^[6,7]、模糊推理法^[8]、人工神经网络理论^[9-11]等), 给出判定地震火灾损失的半定量分析方法。一般而言, 现场调查法是目前最直观、最可靠的方法, 但存在工作量大、资料琐碎、计算周期长等缺陷, 不能及时适应紧急救灾、资源配置、城市抗震防灾规划等多方面的要求。因此, 针对地震火灾损失预测的情况及其影响因素的多样性、复杂性、非线性和非确定性, 不确定性的研究方法能够充分地考虑各种影响因素。对于不确定性的分析方法, 人工神经网络应用最为广泛。但人工神经网络是基于启发式的, 没有相当完备的理论基础, 不能很好

地控制训练后的网络推广, 有时不能达得全局最优, 且基于经验风险最小化(ERM)准则, 容易导致神经网络的过拟合问题。支持向量机(support vector machine, 简称 SVM)方法根据结构风险最小化(SRM)准则, 尽量提高了学习机的泛化(预测)能力, 较好地解决了小样本、非线性、高维数和局部极小点等实际问题, 其基本思想是通过非线性变换将输入空间变换到一个高维空间, 在这个新空间中求解一有约束的凸二次规划问题, 可以得到唯一的全局最优解。上述特点使得 SVM 方法成为一种优秀的学习算法, 目前已经成功地推广到模式识别、函数拟合等领域^[12]。由于地震火灾损失预测的复杂性, 在地震火灾损失与其影响因素之间存在着很强的非线性关系, 而 SVM 方法可以很好地完成输入数据与输出数据之间的非线性映射, 同时提高泛化能力。基于此, 本文建立了基于支持向量机方法建立了城市震后火灾损失预测模型, 并予以应用分析。

1 震后火灾损失预测的 SVM 模型原理

通过历史地震火灾损失数据获得样本集 $\{(x_1, y_1), \dots, (x_i, y_i)\}$, $x_i \in R^n$, $y_i \in R^1$, 其中 x_i 是影响地震火灾损失的各个因素, y_i 为损失值。使之通过训练后, 用一非线性回归函数 $f(x)$ 描述 x_i 与 y_i

* 收稿日期: 2007-08-05

基金项目: 国家十一科技支撑计划重点项目(2006BAJ06B01, 2006BAJ04A03, 2006BAJ13B04); 国家自然科学基金资助项目(50678060); 北京工业大学研究生科技基金资助课题(ykj-2007-1071)

作者简介: 王威(1981-), 男, 河南沈丘人, 博士研究生, 主要从事城市防灾减灾、系统安全评估等研究。

E-mail: weiweikunlun2008@126.com

之间的非线性映射, 对于实测值以外的 x_i 通过 $f(x)$ 找出对应的 y_i , 即所需预测地震火灾损失值。考虑线性回归, 设线性回归函数 $f(x) = wx + b$, 假设所有样本数据点集在精度 ε 下无误差地用 $f(x)$ 拟合^[13-14], 即

$$|y - wx_i - b| \leq \varepsilon, (i = 1, \dots, l). \quad (1)$$

对于存在拟合误差的情况, 引入松弛变量 $\xi_i \geq 0$, $\xi_i^* \geq 0$, 则线性函数回归问题变为:

$$\text{Min} \phi(w, b, \xi_i, \xi_i^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*), \quad (2)$$

$$\left. \begin{aligned} y_i - wx_i - b &\leq \varepsilon + \xi_i \\ wx_i + b - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i &\geq 0, \xi_i^* \geq 0 \end{aligned} \right\}, \quad (3)$$

式中, $C > 0$, 表示对超出误差 ε 的样本的惩罚程度。

目标函数是严格上凹的二次型, 约束函数是下凹的, 这是一个严格凸规划。按照最优化理论中凸二次规划的解法, 可以转化为下列对偶问题:

$$\begin{aligned} \max W(a, a^*) \mid_{w, b, \xi_i, \xi_i^*} = & -\frac{1}{2} \sum_{i,j=1}^l (a_i - a_i^*)(a_j - a_j^*)(x_i x_j) - \sum_{i=1}^l (a_i + a_i^*) \varepsilon + \sum_{i=1}^l (a_i - a_i^*) y_i, \end{aligned} \quad (4)$$

$$\sum_{i=1}^l (a_i - a_i^*) = 0, 0 \leq a_i, a_i^* \leq C, i = 1, \dots, l. \quad (5)$$

对于非线性函数回归问题, 可以通过一非线性映射 ($x \rightarrow \varphi(x)$) 把输入数据映射到一高维特征空间, 然后在该高维特征空间进行线性回归, 从而取得在原空间非线性回归的效果。根据上述思想, 非线性函数回归问题可以转化在约束式(5)下求解下述非线性规划问题:

$$\begin{aligned} \max W(a, a^*) \mid_{w, b, \xi_i, \xi_i^*} = & -\frac{1}{2} \sum_{i,j=1}^l (a_i - a_i^*)(a_j - a_j^*) \times (\varphi(x_i) \varphi(x_j)) - \sum_{i=1}^l (a_i + a_i^*) \varepsilon + \sum_{i=1}^l (a_i - a_i^*) y_i. \end{aligned} \quad (6)$$

令 $K(x_i, x_j) = \varphi(x_i) \varphi(x_j)$, 则式(6)为

$$\begin{aligned} \max W(a, a^*) \mid_{w, b, \xi_i, \xi_i^*} = & -\frac{1}{2} \sum_{i,j=1}^l (a_i - a_i^*)(a_j - a_j^*) K(x_i, x_j) - \sum_{i=1}^l (a_i + a_i^*) \varepsilon + \sum_{i=1}^l (a_i - a_i^*) y_i. \end{aligned} \quad (7)$$

此时, 函数 $f(x)$ 可表示为

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) K(x, x_i) + b. \quad (8)$$

其中 $a_i - a_i^* \neq 0$ 对应的 x_i 为支持向量; 偏置量 b 可以通过 KKT 条件计算; $K(x_i, x_j)$ 称为核函数。核函数的引入, 巧妙地解决了因 $\varphi(x)$ 未知而 w 无法显示表达的问题。

$K(x_i, x_j)$ 是对称正实数函数, 同时满足 Mercer 条件。目前研究最多的核函数主要有 3 类:

①多项式核函数 $K(x_i, x_j) = (x_i x_j + 1)^d$, $d = 1, 2, \dots$; ②高斯基核函数 $K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \sigma^2)$; ③Sigmoid 核函数 $K(x_i, x_j) = \tanh[\beta(x_i x_j) + c]$ 。

核函数的选取需要一定的先验知识。应用现成的优化软件包, 上述的二次规划问题不难求解。有关文献提出了一些用于大规模样本点的 SVM 算法, Cortes 和 Vapnik 提出了 Chunking 算法, Platt 提出了序贯最小优化算法 (Sequential minimal optimization, SMO)^[13] 等。

2 城市震后火灾损失预测的实例分析

2.1 震后火灾损失影响因素的确定

综合国内外的研究现状, 城市地震火灾损失主要从三方面分析^[15,16]: 地震起火分析、火灾蔓延建模和震后消防扑救分析。通过分析, 参考文献[9]考虑的因素主要有地震烈度、人口、地区总面积、城镇面积、建筑物面积、救火站与救火车数、供水(以不受地震影响为1)、着火次数、火源与救火站平均距离及平均反应时间(min)、要求最少救火车数等参数在 0 与 16.09(km/h-1) 风速下的燃烧土地与建筑物面积以及损失值(以保险金人民币计)。

2.2 学习样本和测试样本的选择

从文献[9]收集了 50 组的实际资料作为样本, 对地震火灾损失预测模型, 随机地选取前 45 个样本作为预测模型的学习样本(表 1)。其余 5 个样本实例用于测试建立的模型。采用改进的 SVM——最小二乘 SVM^[12], 选用 RBF 核函数。一般而言, SVM 的性能与所选用的核函数的类型关系不大, 而核函数的参数 σ^2 和误差惩罚因子 C 是主要因素(C 表示对错分样本的惩罚, C 越大表示惩罚越大; 而核函数参数 σ^2 表示了模型的复杂度, σ^2 越大表示模型复杂度越低, 因此 C 和 σ^2 的取值应该同方向变化)。由于 LS-SVM 求解速度很快, 采用交

叉验证的方法来选择参数 C 和 σ^2 ，最终确定 $C = 450$ 、 $\sigma^2 = 6\,720$ 。

表 1 部分震后火灾损失样本数据

序号	烈度	房屋密度	人口密度	救火车密度	供水	反应速率	风速	损失
1	7.8	1.881	1.232	0.045	0.5	264	10	22.9
2	8	0.596	0.894	0.077	0.8	226	10	4.9
3	8	0.435	0.652	0.04	0.8	264	10	3
4	7	0.574	0.793	0.053	0.8	264	0	2.4
5	8	0.274	0.324	0.017	0.8	377	0	0.8
6	8	0.274	0.324	0.017	0.8	293	10	1.8
7	7.3	0.515	0.721	0.018	0.8	377	0	2.2
8	8.5	0.874	0.305	0.049	0.8	377	0	7
9	8.5	0.874	0.305	0.049	0.8	293	10	7.9
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
43	8.8	0.139	0.007	0.029	1	452.6	0	0.817
44	8.8	0.139	0.007	0.029	1	452.6	10	1.798
45	8.5	0.17	0.001	0.029	1	452.6	0	0.79
46	7.8	1.881	1.232	0.045	0.5	352	0	16.7
47	8	0.596	0.894	0.077	0.8	264	0	4.2
48	7	0.574	0.793	0.053	0.8	226	10	2.9
49	7.3	0.515	0.721	0.018	0.8	293	10	2.5
50	8.5	0.425	0.253	0.037	1	377	0	1.1

2.3 预测结果及其分析

利用上面建立的预测模型对另外的后 5 组未曾参加学习的数据进行预测，同时将预测结果与文献[9]的 BP 神经网络模型预测对比，结果见表 2。

表 2 SVM 和 BP 预测结果对比表

序号	实 际 损失值	SVM 模型		BP 神经网络模型	
		预测值	绝对误差	预测值	绝对误差
1	16.7	17.09	-0.39	17.80	-1.1
2	4.2	3.58	0.62	2.99	1.21
3	2.9	3.20	-0.30	4.62	-1.72
4	2.5	2.84	-0.34	3.92	-1.42
5	1.1	1.70	-0.60	1.83	-0.73

从表 2 中可以看出：

(1) 在同样的训练集均方误差的情况下，BP 神经网络模型预测测试集样本的绝对误差均大于由 SVM 模型得出的预测结果，说明支持向量机模型泛化(预测)能力要优于 BP 神经网络模型。

(2) 预测的变形值与实测值基本接近，数据范围是合理的。同时也说明支持向量机可以很好地

表达城市震后火灾损失与其影响因素之间的非线性映射关系，并且该模型具有广泛的适用性，因此，采用支持向量机方法进行震后火灾损失预测是可行的，为其提供了一种新途径。

3 结语

支持向量机作为最近发展起来的一种机器学习算法，因为有严格的理论基础和良好的性能，受到了数学、计算机等许多研究领域的重视，是目前国际上研究的热点问题。通过本文对持向量机在城市震后火灾损失预测中的研究可以得出以下几点结论：①震后火灾损失受很多因素的影响，并且它们之间是高维的、高度非线性的、复杂的关系，用传统的建模方法很难处理，本文的支持向量机方法很好地处理了这种关系；②利用支持向量机进行建模比较简单，不需要对数据进行多的预处理，并且由于支持向量机算法对应一凸二次规问题，找到的解为全局最优解；③城市防灾减灾中的很多问题用传统的数学、力学方法很难

给出一个确定的模型, 用支持向量机方法可以很好地表达这一类问题, 因此支持向量机在城市防灾减灾中具有广泛的应用前景。

参考文献:

- [1] 苏幼坡, 苏经宇, 苏春生, 等. 城市生命线系统震后恢复的基础理论与实践[M]. 北京: 地震出版社, 2002.
- [2] 苏幼坡, 刘瑞兴, 马亚杰. 城市生命线震后恢复曲线与恢复过程优化的影响因素分析[J]. 灾害学, 2000, 15(4): 49-54.
- [3] 林均歧. 地震火灾损失评估研究[J]. 自然灾害学报, 2000, 9(4): 111-114.
- [4] 张宝红, 陈宏德. 地震火灾实例调查[J]. 自然灾害学报, 1994, 3(4): 39-48.
- [5] 刘欣, 谢庆胜. 基于 GIS 技术的快速震害评估方法[J]. 灾害学, 2002, 17(3): 26-29.
- [6] 尹之潜, 杨淑文. 地震损失分析与设防标准[M]. 北京: 地震出版社, 2004.
- [7] 李杰. 生命线工程抗震——基本理论与应用[M]. 北京: 科学出版社, 2005.
- [8] 化彬, 江见鲸. 模糊推理法在地震火灾评判中的应用[J]. 自然灾害学报, 1997, 6(2): 22-27.
- [9] 陆伟民, 董伟民. 利用神经网络预测城市震后火灾损失[J]. 同济大学学报, 1995, 23(1): 15-20.
- [10] 王海荣, 王明学. 基于径向基函数网络的地震火灾损失预测[J]. 地震学报, 2007, 29(1): 95-101.
- [11] Stefano A, Sabia D, Sabia L. Probabilistic neural networks for seismic damage mechanisms prediction [J]. Earthquake Engng Struct Dyn, 1999, 28(8): 807-821.
- [12] 邓乃扬, 田英杰. 数据挖掘中的新方法——支持向量机[M]. 北京: 科学出版社, 2004.
- [13] 范量, 韩阳, 王威, 等. 单桩竖向极限承载力预测的 SVM 模型[J]. 河南科学, 2007, 25(1): 85-87.
- [14] Burge CJC. A tutorial on support vector machines for pattern recognition [J]. Data Mining and Knowledge Discovery, 1998, (2): 121-167.
- [15] 赵思健, 任爱珠, 熊利亚. 城市地震次生火灾研究综述[J]. 自然灾害学报, 2005, 15(2): 57-67.
- [16] 王国权, 马宗晋, 周锡元, 等. 国外几次震后火灾的对比和研究[J]. 自然灾害学报, 1999, 8(3): 72-79.

Application of Support Vector Machine in the Prediction of Losses in Fires Following Earthquakes

Wang Wei¹, Su Jingyu¹, Han Yang², Ma Donghui¹ and Wang Zhitao¹

(1. Institute of Earthquake Resistance and Disaster Reduction, Beijing University of Technology, Beijing 100022;
2. Institute of Engineering Disaster Prevention and Mitigation, Henan University of Technology, Zhengzhou 450052)

Abstract: According to complexity and multiplicity of the post-earthquake fire, the method of support vector machines is proposed to predict the loss of earthquake fire based on the comprehensive analysis of other different methods. On the basis of historical data and documents of earthquake fire, the applicability and validity of the model is manifested through testing and discussion. A simple and available method is provided for the prediction of losses of other natural disaster.

Key words: earthquake fire; loss prediction; support vector machine (SVM); BP neural network