

基于支持向量机的突发事件分级研究

商丽媛, 谭清美

(南京航空航天大学经济与管理学院, 江苏 南京 211106)

摘要: 突发事件的分级是应急管理的重要环节。将支持向量机方法应用于突发事件的分级问题, 阐述了基于支持向量机的突发事件分级方法的过程, 以对地震灾害进行分级作为实例, 采用多分类支持向量机中的一对一组方法进行实验。实验结果表明, 支持向量机应用于突发事件分级是可行而且有效的, 能够为决策者准确地进行分级决策提供科学依据。

关键词: 突发事件分级; 支持向量机; 多类分类

中图分类号: C931 **文献标识码:** A **文章编号:** 1004-6062(2014)01-0119-05

0 引言

近年来, 自然灾害、公共卫生和社会安全等领域的突发事件在各国频繁发生, 对人类的生存和社会发展构成了重大的威胁, 突发事件的应急管理越来越被世界各国所关注。我国历来自然灾害频发, 灾害种类多、频率高、分布地域广、造成损失严重。同时, 伴随着经济发展, 我国在生产安全、公共卫生和社会治安等领域暴露出来的问题日益突出。因此, 我国政府需要加强应急管理工作, 提高预防和处理突发公共事件的能力。突发公共事件一般具有突发性、不确定性、多样性、紧迫性和非常规性等特点。对各类型突发公共事件进行快速准确的识别, 将其划分为不同级别, 从而采取相应的应急措施, 是应急管理的重要环节, 是制定应急预案以及资源调配的基础^[1]。

分级是指将具有相同特征的一类事物, 根据某种特征的程度不同, 从高到低(或从低到高)分为不同的级别。目前, 国内外对突发事件的分级管理的研究已经有一定实践经验和理论成果。例如美国按照事件危险程度建立了五级国家威胁预警系统, 以五种颜色代表不同的风险和威胁程度。这五个等级是: 绿色(低风险状态—正常/常规级别); 蓝色(警戒状态—提高关注); 黄色(较高风险状态—常规威胁); 橙色(高风险状态—迫近威胁)和红色(严重状态—一定域威胁)^[2]。我国在综合考虑各类突发公共事件的性质、严重程度、可控性和影响范围等因素下建立了四级预案体系。2006年国务院发布的《国家突发公共事件总体应急预案》中, 将突发公共事件分为四级, 即Ⅰ级(特别重大)、Ⅱ级(重大)、Ⅲ级(较大)和Ⅳ级(一般)。在理论研究方面, 吴宗之、刘茂对重大事故应急预案分类、分级体系及其内容进行了总结和分析, 建议我国重大事故应急体系由五级四类预案组成^[3]; 薛澜、钟开斌系统地讨论了我国突发公共事

件分类、分级与分期制度, 提出了应急管理体制的框架^[4]; 吕欣驰阐述了城市火灾等级的划分, 提出将灾害事故划分为两个层次类别再进行分级^[5]; 钟茂华、陈宝智将神经网络应用于重大危险源分级研究^[6]; 杨静、陈建明将聚类分析和判别分析引入到突发事件的分类分级评价中^[7]; 刘佳、陈建明利用模糊决策的方法, 提出了应急管理的动态模糊分类分级模型^[8]。

纵观国内外学者对突发事件的分类分级法通常分三种类型: 一是根据事件发生后的影响和严重程度来划分, 属于事后评估; 二是以主观经验打分为基础, 评价级别标准和级别数一成不变, 属于静态的分级方法; 三是分类分级随着突发事件的发展不断进行调整, 属于动态的分级方法。近年来关于突发事件分级问题的理论研究已经取得了一定成果, 但在实际进行分级决策时, 由于突发事件的类型不同, 各类型的评估指标也不相同, 并且相关数据采集困难, 因此, 分级决策仍然主要依靠决策者的直觉和经验来确定。决策者依靠直觉和经验所做的决策主观性和随意性较强, 不能客观地反映突发事件的级别特征。为了得到科学的分级结果, 需要建立基于数据的客观分级模型, 为决策者提供科学依据。基于统计学习理论发展起来的支持向量机是处理分类问题的有效方法^[9], 本文将支持向量机方法应用于突发事件管理分级问题中, 描述了基于支持向量机的突发事件分级方法的过程, 并通过地震灾害分级实验, 说明了该方法可行性和科学性, 为准确地进行突发事件分级提供了有效方法和科学依据。

1 支持向量机分类算法简介

支持向量机(Support Vector Machine SVM)是Vapnik^[9]在1995年提出的一种基于统计学习理论的机器学习方法,

收稿日期: 2011-11-08 修回日期: 2012-10-12

基金项目: 国家自然科学基金资助项目(71073079); 江苏省高校哲学社会科学重点资助项目(08SJB7900029)

作者简介: 商丽媛(1981—), 女, 山东滨州人, 南京航空航天大学经济与管理学院, 博士研究生, 研究方向: 应急管理、物流管理等。

具有数学理论完备、全局优化性能好、泛化能力强、算法复杂度与特征空间维数无关等优点,在解决小样本、非线性及高维空间模式识别问题中表现出许多特有的优势,在许多领域得到成功的应用。支持向量机是一种有监督的分类和回归方法,可以有效处理分类问题,本文首次尝试将其应用于突发事件分级问题。

支持向量机的数学模型如下:在一个 n 维空间中输入两类数据,支持向量机在该空间中构造一个超平面来区分两类数据,这个超平面距离两类数据的边界最大。给定训练集 $\{x_i, y_i\}$, 其中 $i = 1, 2, \dots, l, y_i \in \{-1, 1\}, x_i \in R^D$ (x_i 包含 D 个特征)。超平面可以表示为 $w \cdot x + b = 0$, 其中 w 是超平面的法向量, $|b|/\|w\|$ 是超平面距原点的垂直距离。所有的训练样本满足条件:

$$y_i(x_i \cdot w + b) - 1 + \xi_i \geq 0, \xi_i \geq 0, \forall i \quad (1)$$

ξ_i 的引入是为了解决线性不可分的问题, ξ_i 称为松弛变量, $\xi_i \geq 0, i = 1, 2, \dots, l$ 。通过式(1),最大化 SVM 的边界等价

于解决如下的最优化问题: $\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i$

$$\text{s. t. } y_i(x_i \cdot w + b) - 1 + \xi_i \geq 0, \xi_i \geq 0, \forall i \quad (2)$$

式中: C 为惩罚系数,用来平衡松弛变量和分类边界的大小。(2)式是一个凸二次优化的问题,其解为

$$f(x) = \text{sign}\left(\sum_{i=1}^l \alpha_i y_i K(x_i, x) + b\right) \quad (K \text{ 为 Mercer 核}) \quad (3)$$

Mercer 条件的核函数有多项式函数、径向基函数和 Sigmoid 函数等,选用不同的核函数可构造不同的支持向量机。通过非线性核映射,使得 SVM 可以解决线性不可分的问题。多项式核函数、径向基(Radial Basis Function RBF)核函数和 Sigmoid 函数的表达式如下:

$$K(x_i, x_j) = (x_i^T x_j + 1)^p, i = 1, 2, \dots, N \quad (4)$$

$$K(x_i, x_j) = \exp(-\|wx_i - x_j\|^2 / (2\sigma^2)), i = 1, 2, \dots, N \quad (5)$$

$$K(x, x_i) = \tanh[a(x \cdot x_i) + b], i = 1, 2, \dots, N \quad (6)$$

Sigmoid 函数采用双曲正切函数 \tanh , 仅当 a 和 b 的取值适当时才满足 Mercer 条件。径向基函数是最常用的核函数,因为它对应的特征空间是无穷维的,有限的数据样本在该无穷维特征空间是线性可分的^[10]。

支持向量机(SVM)算法最初是由二值分类问题引发的,后来逐渐被用于多类分类问题。将 SVM 推广到多类 SVM 有两种方法:一是组合几个二值分类器去构造多类分类器;二是在一个优化公式中直接考虑所有数据进行全局优化^[11]。然而,不管是由二值分类器组合构造还是全局优化,解决多类 SVM 问题的公式中都含有与分类数目成比例的变量。因此,在数据数目相同的情况下,多类问题的计算要比二值分类问题的计算复杂得多。由于在优化模型中考虑多类问题方法中,目标函数较复杂,难以求解,因此在工程应用领域很少被采用。组合多个二值分类支持向量机实现多类分类通过将多类问题分解为多个两类问题,将一个复杂问题转化为若干个简单问题,因此应用较多。常用的组合多个二值分类

支持向量机有四种方法:一对多^[12]、一对一^[13]、导向无环图^[14]和二叉树^[15]。本文实验中选择一对一方法解决多类分类问题。

一对一组合对于 k 类问题构造 $\frac{k(k-1)}{2}$ 个二值分类器,

每个分类器实现 k 类中两类的区分,其训练样本来自两层数据,组合所有的二值分类器,构成一对一多类支持向量机。构造好 $\frac{k(k-1)}{2}$ 个分类器后,采用投票法进行决策,类别输出为得票数最多的那一类。如果两个类别具有相同的票数,选择类别标签小的一类作为样本的类别。一对一组合方法的优点是:每个二值分类器只涉及两类训练样本,能够缩短训练时间;每个二值分类器的训练样本是平衡的,训练精度较高。这种方法缺点是分类函数随着类别数的增加迅速增加,从而使得决策过程变慢^[16]。

2 基于支持向量机的突发事件分级方法

在应急管理中,由于突发事件分级方法多数属于以主观经验打分为基础,评价级别标准不变的静态分级方法,因此具有主观随意性,难以可靠、客观地反映真实情况。支持向量机是一种在有限数据集的情况下基于数据的机器学习方法,能够从一组观测数据集的数据出发,得到一些不能通过原理分析而得到的规律,进而利用这些规律对未来的数据或无法观测到的数据进行预测和分析,应用于分级问题,就是进行模式识别。现有研究表明,支持向量机方法能够有效解决模式识别问题。因此,采用支持向量机方法解决突发事件分级方法能够避免主观随意性,为决策者提供科学依据。

不同突发公共事件发生的原因、所导致危急状态的影响范围、产生社会危害的严重程度都有很大差异,因此,对于突发事件的分级需要建立在突发公共事件分类的基础上。突发事件的分类可以根据各自的类别属性确定,而突发事件的分级则需要建立对应于不同类别的分级指标。分级指标的确立需要根据该类型突发事件的共有特性建立,并且这些指标都能够收集到客观数据,所收集的数据能够用来建立利用支持向量机方法进行预测所需的属性训练集。根据历史指标数据,通过支持向量机方法进行客观分级,得到的分级结果是建立在真实数据基础上,从而避免了主观随意性。基于支持向量机的突发事件分级过程如图 1。

(1) 根据突发事件的类型,分析相关的若干因素,建立指标体系。

(2) 根据指标体系中的因素,收集相关的历史数据,构成 SVM 的训练样本集。

(3) 利用收集的 SVM 训练样本,根据 SVM 的分类算法对其进行分类学习,获取问题的固有规律。

(4) 找到训练集中的支持向量构造 SVM 决策函数。

(5) 将待分级对象的指标数据输入到决策函数中,从而得到分级结果。

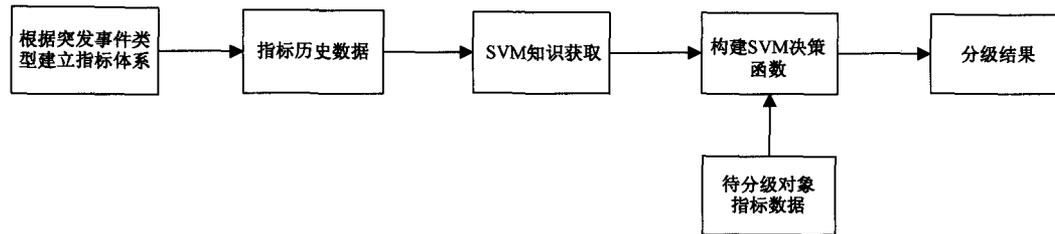


图1 突发事件分级方法过程

3 实验结果及分析

本文以对地震灾害进行分级作为实例来验证基于支持向量机的突发事件分级方法的可行性。根据国家地震科学数据共享中心网站提供的地震灾害相关数据,选取我国1995—2004年地震样本数据30个,其中24个作为训练数据,6个作为测试数据。选取的地震数据包含特征属性9项,分别为:里氏震级、震中强度、VI度区面积、受灾人数、人员死亡人数、人员受伤合计、民房毁坏/倒塌间数、房屋破损合计和

直接经济损失。其中原始数据的里氏震级用罗马数字表示,为简化计算,将其化成相应阿拉伯数字,如V为5, V+为5.5。支持向量机方法的输出为分类模式,即该实验中的地震灾害等级。我国的地震灾害等级分为:一般、中等、严重和特大四级。由于特大地震数据样本较少,本文仅选择部分等级为一般、中等和严重三个级别的数据,以1、2、3分别表示地震等级为严重、中等和一般。训练数据集如表1,用该数据作为训练样本,采用支持向量机方法对其进行学习。

表1 1995—2004部分地震灾害数据表

序号	地震日期	震中参考地名	里氏震级 (Ms)	震中烈度	VI度区面积 (平方米)	受灾人数 (万人)	人员死亡 (人)	人员受伤合计 (人)	民房毁坏/倒塌 (平方米)	房屋建筑破损 (平方米)	直接经济损失 (万元)	地震等级
1	19951024	云南武定	6	8	12380	101.4	58	13815	254520	9516450	74383	1
2	20000115	云南姚安	6.5	8	7834	69.4	7	2528	18732	4271699	106621	1
3	20010412	云南施甸	5.9	8	3414	66.9648	3	235	91000	2746800	50490	1
4	20030224	新疆巴楚	6.8	9	14505	65.9392	268	4853	2452193	5150522	139792	1
5	20030721	云南大姚	6.2	8	2292	32.2962	16	793	48410	7311598	59190	1
6	19950712	云南孟连	7	6	10400	57.7	11	136	870	2159830	20550	2
7	20041019	云南保山	5	6	443	39.8327	0	16	6227	2307089	21720	2
8	20040810	云南鲁甸	5.6	8	697	31.3556	4	597	78881	3104869	33226	2
9	19960319	新疆阿图什	7	6	9286	30	24	128	940924	9222821	35366.81	2
10	19970121	新疆伽师	6	6	2828	23.82	12	44	421677	1569959	37413.6	2
11	19970411	新疆伽师	7	7	2386.2	25.34	8	62	1694365	2985318	32003	2
12	19980110	河北张北	6	8	780	16.9	49	11439	2579377	3970054	78800	2
13	19980827	新疆伽师	7	6.5	1853	12.05	3	13	204946	628310	12507.85	2
14	19981119	云南宁蒗	6	6	2428	12.8	5	1487	14337	2270245	39114	2
15	19950722	甘肃永登	6	7	1300	22.4	10	592	31560	809750	6779	3
16	19960925	云南丽江	6	7	769	3.9	1	141	0	339004	3080	3
17	19961221	四川巴塘	6	7	2100	1.1	2	60	6375	323475	3998.5	3
18	19970301	新疆伽师	6	7	2153	13.1	1	6	4711	126201	8819.48	3
19	19990311	河北张北	5.6	7	36	1.2769	0	3	1755	25518	1105.5	3
20	19990415	甘肃文县	4.7	6	176	8.1339	1	30	0	46202	627.1	3
21	20000429	河南内乡	4.7	6	140	19	1	28	7840	860176	5680	3
22	20010312	云南澜沧	5	6	441	4.5449	0	6	2064	102362	5575	3
23	20010608	云南施甸	5.3	6	510	10.6128	1	15	0	195657	3660	3
24	20021020	内蒙乌珠	5	6	310	2.7	0	0	0	104730	800.76	3

资料来源:国家地震科学数据共享中心网

支持向量机最重要的是寻找一个合适的分类函数,分类函数的确定主要是对核函数的选择和对惩罚系数 C 的确定。本文在 LibSVM 工具包的基础上进行编程计算,选取多类分类支持向量机方法中的一对一组合方法,经过多次实验,选择了 RBF 核函数。惩罚系数 C 和核参数 σ^2 的选择主要采用网格搜索寻优法。其基本思路是:分别选定 C 和 σ^2 的一个区间,确定步长,在选定的区间内,让 C 和 σ^2 呈指数增长,通过

不断改变 C 和 σ^2 的组合,初步寻找二者的最佳组合,然后确定准确率出现最高的一个小区间,并逐步缩短步长重复搜索,直至搜索结果精度变化不大为止,所得结果即为最优化参数^[17]。本文经过计算得到惩罚系数 $C = 4$,核参数 $\sigma^2 = 1$,交叉验证精度达 80.7692%。用该结果模型对测试数据样本进行分级测试,结果如表 2 所示。

表 2 对测试样本的分级结果

序号	地震日期	震中参考地名	里氏震级 (Ms)	震中烈度	VI度区面积 (平方米)	受灾人数 (万人)	人员死亡 (人)	人员受伤合计 (人)	民房毁坏/倒塌 (平方米)	房屋建筑破损 (平方米)	直接经济损失 (万元)	地震等级	测试地震等级
1	20030816	内蒙巴林	5.9	8	10362	48.0869	4	1064	12190	70370	80649	1	1
2	20031025	甘肃民乐	6.1	8	2002	19.0101	10	46	376070	2182406	50140	1	1
3	20031115	云南鲁甸	5.1	7	453	23.6652	4	94	19960	3016195	19190	2	1
4	20040324	内蒙乌珠	5.9	7	17090	5.3	1	5	29409	646848	20272	2	2
5	20040907	甘肃岷县	5	7	302	4.0838	1	36	25205	658195	6600.189	3	3
6	20041226	云南双柏	5	6	493	3.5715	1	20	2858	512739	4070	3	3

由表 2 可知,测试分级结果与实际分级基本一致,6 个测试样本中仅有 1 个结果不同,用描黑斜体标注。计算得到的分类总精度为 83.34%,说明采用支持向量机对地震灾害进行分级是有效的。虽然基于支持向量机的分级方法还不是完全精确,但它是一种建立在数据基础上的客观方法,并且具有一定自学习能力,易于计算机操作,能够有效地避免分级过程中的主观性和随意性。

4 结束语

本文将支持向量机方法应用于突发事件的分级问题,描述了基于支持向量机的突发事件分级方法的过程,给出验证该方法可行性的地震灾害分级实验,采用多分类支持向量机中的一对一组合方法,对样本数据进行处理,通过测试,得到了比较符合实际的结果。实验结果表明,基于支持向量机的突发事件分级方法是可行而且有效的,同时它又是一种完全客观的方法,能够为决策者准确地进行突发事件分级提供科学依据。

参 考 文 献

[1] 朱子华,李严峰. 供应链中突发事件的分类、分级与分期研究[J]. 中国物流学术前沿报告(2006—2007):74~80.
 [2] 赵红,汪亮. 从美国联邦应急计划看美国国家应急管理运行机制[J]. 项目管理技术,2004,(1):26~30.
 [3] 吴宗之,刘茂. 突发事件管理中的“分类”“分级”与“分期”原则——《中华人民共和国突发事件应对法(草案)》的管理学基础[J]. 中国安全科学学报,2003,13(1):15~18.
 [4] 薛澜,钟开斌. 突发公共事件分类、分级与分期:应急体制的管理基础[J]. CPA 中国行政管理,2005,236(2):102~107.
 [5] 吕欣驰. 论城市火灾与灾害事故等级划分和灭火救援力量出动预案编制[J]. 消防科学与技术,2003,22(4):276~277.

[6] 钟茂华,陈宝智. 基于神经网络的重大危险源动态分级研究[J]. 中国安全科学学报,1997,7(2):6~9.
 [7] 杨静,陈建明,赵红. 应急管理中的突发事件分类分级研究[J]. 管理评论,2005,17(4):37~41.
 [8] 刘佳,陈建明,陈安. 应急管理中的动态模糊分类分级算法研究[J]. 管理评论,2007,19(3):38~43.
 [9] Vapnik V. The nature of statistical learning theory[M]. New York: Wiley, 1998.
 [10] 薛宁静. 多类支持向量机分类器对比研究[J]. 计算机工程与设计,2011,32(5):1792~1795.
 [11] Chao L C, Tong L I. Wafer defect pattern recognition by multi-class support vector machines by using a novel defect cluster index[J]. Expert Systems with Applications, 2009, 36(6):10158~10167.
 [12] Polat K, Günes S. A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems[J]. Expert Systems with Applications, 2009, 36(2):1587~1592.
 [13] Arindam C, kajal D, Dipak C. A comparative study of kernels for the multi-class support vector[C]. Los Alamitos: IEEE Computer Society, 2008:3~7.
 [14] Manikandan J, Venkataramani B. Study and evaluation of a multi-class SVM classifier using diminishing learning technique[J]. Neurocomputing, 2010, 73(10-12):16767~1685.
 [15] CHEONG S, SANG H O, LEE S Y. Support vector machines with binary tree architecture for multi-class classification[J]. Neural Information Processing, 2004, 2(3):47~51.
 [16] 白鹏,张喜斌等. 支持向量机理论及工程应用实例[M]. 西安:西安电子科技大学出版社,2008.
 [17] 张银德,童凯军,郑军,王道申. 支持向量机方法在低阻油层流体识别中的应用[J]. 石油物探,2008,47(3):306~310.

Emergency Classification Based on Support Vector Machine

SHANG Li-yuan, TAN Qing-mei

(College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: Natural disasters, and issues related to public health and social security have occurred frequently. Emergency classification is a method to identify emergency events rapidly and accurately, and divide them into different levels. The appropriate emergency measures are taken according to classification results. At present, the actual classification decisions still rely mainly on intuition and experience of decision-makers. In order to obtain the results of scientific classification, it needs to establish a classification model based on objective data and provides a scientific basis for decision-makers. Support vector machine (SVM) is developed based on statistical learning theory. It is an effective way to deal with classification problems. This paper presents the application of support vector machine in order to solve emergency classification problems.

Support vector machine algorithm is originally designed for binary classification problem, and then used for multi-category classification problem. There are two methods for extending SVM to multi-class SVM. The first method is to construct a multi-class classifier by combing several binary classifiers. The second method is to consider all data in an optimized formula for global optimization. There are four ways to solve multi-class recognition problems through a combination of multiple binary classifiers: one-against-all, one-against-one, directed acidic graph and binary tree. In this paper, one-against-one SVM is selected to solve emergency classification problems.

This paper describes the emergency classification process based on support vector machine. The process consists of five steps: (1) establish an index system based on emergency types and the analysis of relevant factors; (2) collect historical data of the index system to constitute the SVM training sample set; (3) use a training set to classify learning according to the SVM classification algorithm and obtain an inherent law; (4) find support vector of the training set to construct SVM decision function; and (5) enter the index data of classification object into the decision function and obtain the classification results.

The classification of earthquake disasters is an example to verify the feasibility of the emergency classification method based on support vector machine. According to an earthquake disaster database from the National Earthquake Science Data Sharing Center website, 30 earthquake sample data from 1995 to 2004 were selected, with 24 of them as training data, and others as the testing data. Selected seismic data contain nine characteristic properties: Richter scale, epicenter intensity, VI degree area, the number of people affected, the number of deaths, the number of injuries, the number of houses destroyed, the number of total housing damage and direct economic losses. According to the calculation result of LibSVM toolkit, experimental results show that support vector machine is effective to solve classification problems related to earthquake disasters.

In the present paper, an emergency classification method based on support vector is introduced. Earthquake classification is proposed as an example. The one-against-one method of multi-class support vector machine is applied to the experiment. Experimental results show that multi-class support vector machine is a feasible approach dealing with emergency classification problems.

Key words: emergency classification; support vector machine; multi-class classification

中文编辑: 杜 健; 英文编辑: Charlie C. Chen