

文章编号:1671-251X(2009)10-0036-03

基于 ICA - SVM 的煤与瓦斯突出预测研究

李大锋, 赵 帅, 吴 峰

(义煤集团千秋煤矿, 河南 义马 472300)

摘要:影响煤与瓦斯突出的因素很多且错综复杂,而现有的煤与瓦斯突出预测方法所考虑的影响煤与瓦斯突出的因素较少,针对大量指标属性时无法进行有效的分析处理。针对该问题,文章提出了一种基于 ICA - SVM 的煤与瓦斯突出预测方法。该方法首先通过独立成分分析方法对所选指标属性进行特征提取,然后利用支持向量机模型对煤与瓦斯突出进行预测分析。预测结果表明,基于 ICA - SVM 的煤与瓦斯突出预测方法的预测准确率高,且运算速度快,识别分类能力强,效果较好。

关键词:煤与瓦斯突出; 预测; 独立成分分析; 支持向量机; 特征提取

中图分类号:TD713 **文献标识码:**B

0 引言

我国是世界上煤与瓦斯突出灾害发生最严重的国家之一。煤与瓦斯突出严重影响了煤矿开采,给国家经济发展带来重大损失,并极大地危害着采煤矿工人的生命安全,特别是2004年10月20日郑煤集团大平煤矿发生的特大瓦斯爆炸事故造成147人死亡,2009年2月22日西山煤电屯兰矿发生的瓦斯爆炸事故造成78人死亡。因此,研究煤与瓦斯突出预测问题对于保障煤矿安全生产、职工生命安全至关重要。目前,有关煤与瓦斯突出的预测方法有很多,在煤与瓦斯突出预测实际应用中的D、K指标和钻销量指标等均属单指标法或简单综合指标^[1],所考虑的影响煤与瓦斯突出的因素较少。事实上,影响煤与瓦斯突出的因素很多,且错综复杂:既有地质的因素,又有人为的因素;既有区域因素,又有局部因素;既有确定性因素,又有不确定性因素。同时,随着各种技术,特别是有关煤与瓦斯突出监测技术的不断发展,各种监测传感器从多方位、多角度提供大量的数据信息,所考虑的影响煤与瓦斯突出的指标也逐渐增多,给准确预测煤与瓦斯突出提供了更多的信息资料,但同时也带来了新的问题。如何筛选这些信息,有效提取其特征以提高预测的准确性,就成了本文要解决的问题。本文首先根据以往煤与

瓦斯突出预测方法中的分析指标,加入顶板岩性、底板岩性等几个指标,然后通过运用ICA(独立成分分析)法,对所选指标属性进行特征提取,将噪音数据影响降到最小,最后利用SVM(支持向量机模型)对煤与瓦斯突出进行评估预测。

1 影响煤与瓦斯突出的指标属性选取

煤与瓦斯突出是发生在煤矿井下的复杂的地质灾害,影响煤与瓦斯突出的因素主要有瓦斯放散初速度、构造煤类型、软分层煤厚、煤层瓦斯压力、围岩透气性等。笔者选取以下16个指标分析预测^[2]。

煤层瓦斯压力:原始瓦斯压力越高、煤体内的瓦斯含量越大,煤体破裂时单位面积裂隙上涌出的瓦斯量就越多,裂隙中就越可能积聚起较高的瓦斯压力,从而越可能撕裂煤体,并将撕裂形成的球盖状煤壳抛向巷道。

围岩的透气性系数:围岩的透气性系数越大,越有利于煤层中瓦斯的泄漏。

构造煤的类型:构造煤是煤与瓦斯突出的必要条件,不同类型构造煤有不同的突出危险性。

瓦斯放散初速度:煤样放散瓦斯快慢的程度用P值表示,其大小与微孔隙结构、孔隙表面性质和孔隙大小有关,且随构造煤类型的增高而增大。

软分层煤厚:当突出阵面沿软分层发展时,在垂直煤层方向上有:

$$H = 2R_i \sin \alpha_i \quad (1)$$

式中: H 为软分层煤厚; R_i 为煤壳曲率半径; α_i 为煤壳所对的中心角。

从式(1)可看出,软分层煤厚越小,煤体撕裂后

收稿日期:2009-05-26

作者简介:李大锋(1980-),男,硕士,2008年毕业于厦门大学自动化专业,现在义煤集团千秋煤矿机电科从事矿井自动化方面的工作。E-mail:ldfbz@yahoo.com.cn

形成的球盖状煤壳在沿垂直煤层方向上的曲率半径及所对中心角越小,煤壳越不容易失稳抛出,煤与瓦斯突出灾害也越不容易发生。

此外,笔者还选取了巷道类型、开采垂深、作业方式、测井视电阻率、测井咖吗咖吗值、测井咖吗值、测井曲线识别的物理分层数、煤层厚度、煤层倾角、顶板岩性、底板岩性等与煤与瓦斯突出有关的因素,限于篇幅,本文不再一一详述。

2 基于 ICA 的特征提取

ICA 方法是近几年才发展起来的一种新方法^[3]。该方法的基本思路是以非高斯信号为研究对象,在独立性假设的前提下,对多路观测信号进行盲源分离。在满足一定的条件下,该方法能够从多路观测信号中较好地分离出隐含的独立源信号,其目的是将观察到的数据进行某种线性分解,使其分解成统计独立的成分。最早提出 ICA 概念的是法国学者 HERAUL J 和 JUTTER C,当时他们对 ICA 给出了一种相当简单的描述,认为 ICA 是从线性混合信号里恢复出一些基本的源信号的方法。

引理 1(数据处理不等式):假设 x 和 c 分别代表输入特征和输出类别的随机变量,对于任意关于 x 的确定的函数 $T(x)$, $T(x)$ 与输出类 c 之间的互信息为 x 和 c 之间的互信息的上界:

$$I(T(x);c) \leq I(x;c) \tag{2}$$

假设有 N 维标准输入特征向量 $x = [x_1, \dots, x_N]^T$ 和二元类 $c \in \{-1, 1\}$ 。进行特征提取的目的就是从包含最大类信息的 x 中提取出 M ($M \leq N$) 维新特征向量 $f_a = [f_1, \dots, f_M]^T$,如图 1 所示。其中选取出的新特征向量 f_a 与输出类 c 之间的互信息为 $I(f_a; c)$,与表示原始特征 x 和输出类 c 之间的互信息 $I(x; c)$ 紧密相关。

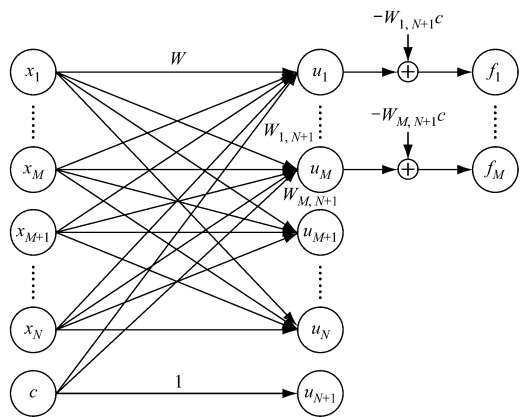
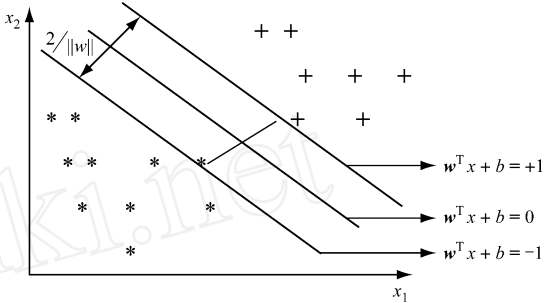


图 1 基于 ICA 算法的特征提取示意图

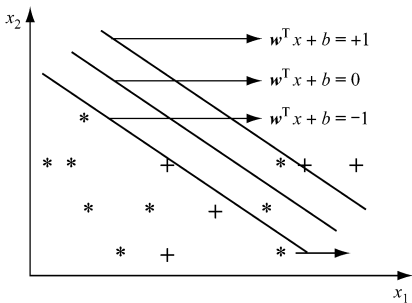
然后将输入特征空间 x 分散到 2 个线性子空间:一个由包含类标签 c 的最大化信息的 $f_a = [f_1, \dots, f_M]^T$ 界定,另一个则由尽可能独立 c 的 $f_b = [f_{M+1}, \dots, f_N]^T$ 界定,从而达到特征提取的目的。

3 煤与瓦斯突出预测的 SVM 模型

SVM 方法是从线性可分情况下的最优分类超平面 (Optimal Hyperplane) 提出的^[4]。假定 N 个训练样本集 $D = \{(x_i, y_i)\}_{i=1}^N$, $x_i \in \mathbf{R}^n$ 是第 i 个输入数据, $y_i \in \{-1, +1\}$ 是第 i 个输入数据的所属类别。如果 y_i 能被一个超平面 $w^T x + b = 0$ 无错误地分开,并且距离超平面最近的向量与超平面之间的距离是最大的,则这个向量集合被这个最优超平面分开,如图 2 所示。



(a) 线性可分示意图



(b) 线性不可分示意图

图 2 线性可分与不可分示意图

没有错误的分开可保证经验风险最小(为 0),使 Margin(分类间隔)的距离最大。要求分类超平面对所有样本能正确分类,必须满足:

$$\begin{cases} w^T x_i + b \geq +1 & y_i = +1 \\ w^T x_i + b \leq -1 & y_i = -1, \quad i = 1, 2, \dots, N \end{cases} \tag{3}$$

式(3)可由式(4)表示:

$$y_i (w^T x_i + b) \geq 1, \quad i = 1, 2, \dots, N \tag{4}$$

容易验证,最优分类超平面就是满足式(3)并且使得 $\|w\| = \|w\|^2$ 最小化的超平面。两类样本中离分类面最近的点且平行于最优超平面的训练样本,也就是使 $\|w\| = \|w\|^2$ 等号成立的那些样本

称为支持向量(Support Vectors, SV)。

由于支持向量机基本理论只考虑二类分类问题,而煤与瓦斯突出区域并没有绝对界线,因此,需要采用一对多分类方法^[5],即构造多个二类分类器并组合起来完成多类分类,其中的每一个分类器都把其中的一类同其余的各类划分开来,然后据此推断某个输入 x 的归属^[6]。一对多分类方法对于 x 类问题需要构造 k 个支持向量机二类分类器,第 i 个 SVM 用第 i 类中的训练样本作为正的训练样本,而将其它样本作为负的训练样本,分类时将待分类样本划分到具有最大分类函数值的那类,从而将样本分类为突出区域、不突出区域及危险区域。

4 应用实例

笔者共选取山西省 20 个高度危险的煤与瓦斯突出矿井进行分析预测,将其中 12 个矿井作为训练样本,剩余 8 个矿井作为预测样本,共分为突出、危险和不突出 3 类来对 SVM 模型进行验证。分析因素为前文提出的 16 个指标属性,通过对数据的预处理,去掉不相关或弱相关的属性,并进行属性概化,根据上面给出的算法,选择核函数 $K(x, y) = \exp(-\|x - y\|^2 / 2)$ ^[7],通过实验设定 γ 值,由条件属性集的协方差矩阵,计算得到不同数量的特征属性。当 $\gamma = 3, 6, 7$ 时,得到的属性个数依次为 6、7、10,其方差贡献率分别为 94.7%, 92.1%, 87.6%。特征提取后的指标属性及预测结果如表 1 所示。

表 1 特征提取后的指标属性及预测结果表

样本 序号	瓦斯 压力 /MPa	放散速度 /(mL · g ⁻¹ · s ⁻¹)	地质 构造	煤层 厚度 /m	煤层 倾角 /(°)	顶板 岩性 指数	底板 岩性 指数	预测 类别	实际 类别
1	2.58	6.37	6	3.5	6.2	6.5	14.2	突出	突出
2	0.67	2.79	4	4.7	4.7	14.2	11.7	危险	突出
3	3.16	3.17	5	8.2	12.4	11.7	8.9	危险	危险
4	1.24	6.67	2	3.9	9.3	9.5	6.7	不突出	不突出
5	0.98	5.42	1	5.7	8.9	7.9	10.2	突出	突出
6	4.26	1.69	5	2.4	2.4	13.4	11.7	突出	突出
7	3.49	3.29	4	2.9	6.7	8.5	8.6	不突出	不突出
8	4.51	5.13	3	6.4	4.8	6.9	6.9	突出	突出

表 1 为运用独立成分分析法进行特征提取后所得的预测结果,将其余未经过特征提取而直接运用

支持向量机进行分类后的预测结果作比较,结果如表 2 所示。

表 2 运用独立成分分析法与支持向量机分类的
预测结果比较表

样本序号	SVM	ICA - SVM	实际类别
1	危险	突出	突出
2	突出	危险	突出
3	突出	危险	危险
4	不突出	不突出	不突出
5	突出	突出	突出
6	危险	突出	突出
7	不突出	不突出	不突出
8	突出	突出	突出

5 结语

本文从多个因素对煤与瓦斯突出进行预测研究,考虑到其影响因素的多样性、复杂性及非线性,笔者首先运用 ICA 法对指标属性进行特征提取、属性约减,以减少噪声数据对预测效果的影响,然后运用 SVM 的一对多方法对样本分类训练,并对所得预测模型进行验证。从预测结果可看出,经过 ICA 对指标属性进行特征提取后,其预测的准确率达到 87.5%,明显高于未经特征提取而直接进行分类的方法所得到的准确率 62.5%,且该方法运算速度快,识别分类能力强,具有很大的推广应用价值。

参考文献:

[1] 林柏泉,崔恒信. 矿井瓦斯防治理论与技术[M]. 徐州:中国矿业大学出版社,1998.

[2] 胡千庭. 高效防治煤与瓦斯突出技术的研究[J]. 淮南工业学院学报,2002,22(4):11-14.

[3] 谷雨,郑锦辉,孙剑,等. 基于独立成分分析和支持向量机的入侵检测方法[J]. 西安交通大学学报,2005(8).

[4] 房瑞明. 支持向量机理论及应用[J]. 系统工程,2006(4).

[5] 安金龙,王正欧. 一种新的支持向量机多类分类方法[J]. 信息与控制,2004(3).

[6] 张冰,孔锐. 一种支持向量机的组合核函数[J]. 计算机应用,2007,27(1).

[7] CRISTIANINI N, SHAWE-TAYLOR J. 支持向量机导论[M]. 李国正,王猛,曾华军,译. 北京:电子工业出版社,2004.