

文章编号: 1671- 251X(2010) 10- 0036- 03

基于 KPCA- SVM 的煤与瓦斯突出预测方法

李大锋, 赵帅, 杨岱平

(义马煤业集团股份有限公司千秋煤矿, 河南 义马 472300)

摘要:提出了一种基于 KPCA- SVM 的煤与瓦斯突出预测方法。该方法首先通过 KPCA 方法对影响煤与瓦斯突出的相关指标进行特征提取, 然后利用 SVM 方法对煤与瓦斯突出进行分类预测。实例结果表明, 该方法对煤与瓦斯突出预测的准确率明显高于直接运用 SVM 方法的煤与瓦斯突出预测准确率, 且运算速度快, 识别能力强, 同时根据该方法建立的煤与瓦斯突出分类预测模型具有较好的稳定性和有效性。

关键词:煤与瓦斯突出; 核主成分分析; 支持向量机; 特征提取; 分类预测

中图分类号:TD713. 2 **文献标识码:**A

Forecasting Method of Coal and Gas Outburst Based on KPCA-SVM

LI Da feng, ZHAO Shuai, YANG Dai ping

(Qianqiu Coal Mine of Yima Mining Group Co. , Ltd. , Yima 472300, China)

Abstract: The paper proposed a forecasting method of coal and gas outburst based on KPCA-SVM. The method firstly used KPCA to select features of correlative indexes influencing coal and gas outburst, then used SVM to make classified forecasting coal and gas outburst. The example forecasting result showed that the correctness rate of forecasting coal and gas outburst by the method is obviously higher than the one by SVM, and the method has quick operating speed and strong identification ability. Meanwhile, the forecasting model of coal and gas outburst on the basis of the method has good stability and validity.

Key words: coal and gas outburst, KPCA, SVM, feature selection, classified forecasting

0 引言

众所周知, 煤与瓦斯突出是煤矿生产所面临的最严重的灾害, 而我国又是煤与瓦斯突出灾害发生最严重的国家之一。煤与瓦斯突出严重威胁着煤矿工人的生命安全, 如西山煤电(集团)有限责任公司屯兰矿在 2009 年 2 月 22 日发生的瓦斯爆炸事故共造成 78 人死亡。关于煤与瓦斯突出预测的方法很多, 如 D.K 指标法、钻屑量指标法等。但事实上影响煤与瓦斯突出的因素很多, 并且错综复杂, 特别是当今有关煤与瓦斯突出监测的技术日益更新, 各种监测传感器能够提供多方位、多角度的数据信息, 考虑的指标因素也逐渐增多, 这给准确预测煤与瓦斯突出提供了更多信息资料, 但同时也带来了新的问

题^[1]: 如何筛选这些信息, 如何有效提取特征, 以提高预测的准确性。针对该问题, 本文首先根据以往的分析指标, 再加入顶板岩性、底板岩性等指标, 运用 KPCA (Kernel Principal Component Analysis, 核主成分分析) 方法, 根据指标的累计贡献率对所选指标属性进行特征提取, 最后利用 SVM (Support Vector Machines, 支持向量机) 模型来对煤与瓦斯突出进行评估预测。

1 指标的选取

本文主要根据以往的专家系统分析方法及其它相对成熟的预测方法所采用的指标来进行分析, 同时增加了顶板岩性、底板岩性等指标, 主要概括为以下 11 个指标^[2]:

(1) 煤层瓦斯压力。瓦斯压力越高, 则煤体中的瓦斯含量越大, 在煤体破裂时单位面积裂隙上涌出的瓦斯量越多, 裂隙中就越可能积聚起较高的瓦

收稿日期: 2010- 06- 20

作者简介: 李大锋(1980-), 男, 河南洛阳人, 助理工程师, 硕士,

现主要从事矿井自动化方面的工作。E-mail: ldfbz@yahoo.com.cn

斯压力,从而越有可能撕裂煤体,并将撕裂形成的球盖状煤壳抛向巷道。

(2) 软分层煤厚^[3]。软分层煤厚 H 为

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

式中: R_i 为煤壳曲率半径; Φ 为煤壳所对的中心角。

从式(1)可看出,当煤与瓦斯突出阵面沿软分层发展时,软分层煤厚越小,形成的煤壳沿垂直煤层方向上的曲率半径及所对的中心角越小,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。

(3) 另外,还选取了地质构造、透气性系数、煤层倾角、构造煤的类型、瓦斯放散初速度、开采垂深、煤层厚度、顶板岩性、底板岩性等9个与煤与瓦斯突出有关的因素。

2 基于KPCA的特征提取

近年来,随着SVM的研究逐步展开,关于核方法的研究受到越来越多的重视。最新的理论研究成果表明,通过与核方法有机融合而形成的KPCA方法不仅适合处理非线性问题,而且能提供更多的信息。

KPCA的基本思想是将核方法应用到主成分分析中。首先通过变换 φ 实现输入空间 X 到特征空间 F 的映射,即将输入空间的样本点 x_1, x_2, \dots, x_l 变换为特征空间的样本点 $\varphi(x_1), \varphi(x_2), \dots, \varphi(x_l)$; 然后在特征空间中使用KPCA,即求解特征值:

$$\lambda \mu_i = \tilde{C} \mu_i \quad i, j = 1, 2, \dots, l \quad (2)$$

式中: $\tilde{C} = \frac{1}{l} \sum_{i=1}^l [\varphi(x_i) \varphi(x_i)^T]$, 为特征空间中

的样本协方差矩阵; λ 为 \tilde{C} 的一个非零特征值; μ_i 为 λ 所对应的特征向量。

3 KPCA-SVM 预测模型

KPCA-SVM 预测模型的基本思想是先用KPCA进行特征提取,然后用SVM进行分类。KPCA-SVM方法更适合处理非线性的过程数据,更能反映过程的本质特性^[4]。

下面简单介绍KPCA-SVM的算法步骤。

(1) 非线性主成分计算

① 选择合适的核函数,计算内积矩阵 $K(x_i, x)$ 。非线性主成分分析法只需要在原空间中计算用作内积的核函数,无需知道非线性映射函数 $\varphi(x)$ 的形式,也无需计算非线性变换。

② 在特征空间 F 中计算内积矩阵 $K(x_i, x)$ 的特征值及其对应的特征向量,并对其进行标准化处理。

③ 利用式(3)计算样本的投影,建立新的样本集。

$$V\phi(x) = \sum_{i=1}^m [\alpha_i \phi(x_i) \phi(x)] = \sum_{i=1}^m [\alpha_i K(x_i, x)] \quad (3)$$

式中: V 为向量 α_i 组成的矩阵。

(2) 建立训练模型

采用SVM算法中的一对多算法来建立煤与瓦斯突出的训练模型。

(3) 分类预测

根据步骤(1)得到的新样本集,用训练模型对其进行分类评估,并输出结果和分类精度。

4 应用实例

选取湖南某矿作为样本进行分析验证。该矿位于湖南省中东部,整个矿井共有9个采区,1~5采区已经开采结束,现在正开采6、7两个采区。从1982年12月至今,该矿的1~5采区已发生21次煤与瓦斯突出事故,被确定为煤与瓦斯突出矿井。选取1~7采区的82个区域作为样本来分析预测,其中1~5采区的50个区域作为训练样本,5~7采区的32个区域作为预测样本,一共将煤与瓦斯突出预测分为突出区域、危险区域和不突出区域3个类别。考虑因素包括地质构造、煤层倾角等11个指标,对指标数据进行标准化处理,从而消除量纲的影响,所有指标变量的方差为1,均值为0。去掉不相关或弱相关的属性^[5],并进行属性概化,根据KPCA-SVM算法,选择核函数 $K(x, y) = \exp(-\|x - y\|^2 / \sigma^2)$,通过实验设定 σ 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。利用统计软件SPSS10.0^[6]对11个指标进行KPCA计算,得到特征值及其贡献率和累计贡献率,如表1所示。

从表1可看出,利用KPCA对数据集的指标属性进行归约,可得出前6个主成分的累计贡献率为99.92%,大于90%。同时,第七个主成分的特征贡献率仅为0.071%,考虑到要尽可能地减少噪声数据对结果的影响,选取前6个主成分代替原来的11个指标来预测煤与瓦斯突出区域。建立KPCA-SVM模型,并对14个样本进行训练。经过KPCA对指标进行特征提取后的预测结果如表2所示。

表2仅列出了部分(6个)预测样本的分类结果,选择的是前6个主成分进行分析预测。为了得

表 1 11 个指标的特征值、特征值的贡献率及累计贡献率

	煤层瓦斯 压力	煤层 厚度	透气性 系数	煤层倾角	瓦斯放散 初速度	顶板岩性 指数	软分层 煤厚	开采 垂深	地质构造	构造煤的 类型	底板岩性 指数
特征值	5.154	2.674	1.245	1.024	0.476	0.107	0.004	0	0	0	0
贡献率/%	52.490	24.380	11.040	6.250	4.380	1.380	0.071	0	0	0	0
累计贡献率/%	52.490	76.870	87.910	94.160	98.540	99.920	100	100	100	100	100

表 2 经过 KPCA 特征提取后部分样本预测结果

样本序号	煤层瓦斯压 力/MPa	煤层厚度/m	透气性 系数	煤层倾角/(°)	瓦斯放散 初速度 /(m·s ⁻¹)	顶板岩性 指数	预测类别	实际类别
1	3.42	4.2	0.25	5.5	7.28	7.4	突出	突出
2	1.68	3.7	0.37	8.6	5.72	15.3	危险	突出
3	4.25	4.1	0.68	4.9	3.45	8.2	危险	危险
4	0.96	5.7	0.12	11.3	8.36	6.7	突出	突出
5	2.03	3.8	0.74	5.8	2.97	9.1	不突出	不突出
6	3.57	1.7	0.65	9.2	5.16	14.5	突出	突出

到更好的预测模型,笔者一共选用了 3 种不同数量的主成分个数来建立模型,当选取不同的主成分时,得到不同的分类准确率,如表 3 所示。

表 3 不同主成分个数的预测准确率

样本数/个	主成分数/个	准确率/%
32	3	72.37
32	5	81.25
32	6	91.26

从表 3 可看出,当主成分个数为 3 和 5 时,分类效果较好,分别达到 72.37%和 81.25%;而以 6 个主成分来进行预测时候,预测结果更是达到 91.26%。这说明主成分选取的个数对预测结果有着重要的影响,经过 KPCA 进行特征提取后的主成分个数越多越好,因此,要根据方差贡献率来确定最佳主成分个数。最后,对未用 KPCA 和使用 KPCA 后的 SVM 分类预测结果进行对比,结果如表 4 所示。

表 4 经 KPCA 特征提取前后预测结果

样本数目	准确率/%	
	SVM	KPCA-SVM
8	72.31	89.9
8	74.98	87.6
8	80.32	96.3
32	76.39	91.26

从表 4 可看出,首先通过 KPCA 方法对指标进行特征提取,再运用 SVM 模型对煤与瓦斯突出进

行分类预测,其分类准确率可达到 91.26%,明显要好于未经过 KPCA 特征提取而直接运用 SVM 方法进行分类预测的结果。

5 结语

基于 KPCA-SVM 方法建立的分类器能够很好地对煤与瓦斯突出进行分类预测,且准确率明显高于未经 KPCA 进行特征提取而直接运用 SVM 来分类预测所得到的准确率,同时,该方法运算速度快,识别分类能力强,说明所建立的分类预测模型具有较好的稳定性和有效性,具有很大的推广应用价值。

参考文献:

[1] 李大锋,赵帅,吴峰.基于 ICA-SVM 的煤与瓦斯突出预测研究[J].工矿自动化,2009(10):36-38.

[2] 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.

[3] 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.

[4] 王秉仁,刘兆阳,张家伟,等.基于支持向量机的旋转机械非线性故障诊断研究[J].煤矿机械,2005(2):122-123.

[5] 房瑞明.支持向量机理论及应用[J].系统工程,2006(4):59-62.

[6] 安金龙,王正欧,马振平.一种新的支持向量机多类分类方法[J].信息与控制,2004(3):262-267.