

基于投影寻踪技术的矿井突水水源判别模型

——以新庄孜煤矿为例

钱家忠, 杜奎, 赵卫东, 周小平, 马雷

合肥工业大学资源与环境工程学院, 合肥, 230009

内容提要:投影寻踪是一种降维处理技术,它可以将高维分析问题通过投影方向转化为低维问题分析。应用该方法的关键在于寻求最佳投影方向,这可以转化为一个复杂的非线性优化问题,结合 Matlab 的遗传算法工具箱进行优化求解。本文以淮南新庄孜煤矿为例,建立突水水源判别投影寻踪模型,并与模糊综合评判模型、神经网络模型、灰色聚类模型进行分析比较。结果表明:投影寻踪判别模型能够有效地判别突水水源,比模糊综合评判、神经网络模型、灰色聚类模型具有更高的准确性,为矿井突水水源判别提供了一个新途径。

关键词:遗传算法;投影寻踪模型;新庄孜煤矿;突水水源

矿井突水是煤矿生产的重大灾害之一,对矿井突水水源的识别是疏干、降压、注浆等防治水工作的基础。判别突水水源包括地下水化学、同位素、水温、水位动态预测和分析等方法(马雷等,2009;桂和荣和陈陆望,2004)。而利用水化学成分判别水源具有快速、准确、经济的优点。目前,依据水化学成分识别水源的方法很多,如模糊综合评判法(李世峰和李耀华,2006)、物元可拓法(张瑞钢等,2009)、灰色聚类法(杨永国和黄福臣,2007)、神经网络(钱家忠等,2010)等。然而上述方法都存在有一定的缺陷,如模糊综合评判法需要对各指标赋予不同的权重,而权重的确定常带有一定的主观随意性;物元可拓法在计算关联度时常以区间中点为最优,从而遗漏了重要的约束条件,导致评价结果与实际情况存在差异;用灰色理论进行突水水源判别,当原始数据序列波动较大且信息过于分散时,预测精度将会降低。神经网络法自身存在收敛速度慢的缺点。

投影寻踪(projection pursuit,简称PP)方法是处理和分析高维数据(尤其是高维非正态数据)的一类新兴的统计方法,对监测数据不作正态等任何假定,能充分利用高维数据中的所有信息(Li Dingding and Stengos Thanasis, 2001; Zhang Chi and Dong Sihui, 2009),但由于存在目标函数优化困难的原因,使其在应用上受到一定的限制(叶浩等,2005)。

为此,本文尝试将投影寻踪算法和 Matlab 遗传算法工具箱相结合,提出了突水水源判别的投影寻踪模型,并将其应用于具体的工程实例中进行验证,为矿井突水水源判别提供了一个新途径。

1 基本原理

投影寻踪的基本思想是:利用计算机技术,把高维数据通过某种组合,投影到低维(1~3维)子空间上。然后通过优化投影指标函数,求出能反映原高维数据结构或特征的投影向量,在低维空间上对数据结构进行分析,以达到研究和分析高维数据的目的(叶浩等,2005;李祚泳,1997)。

1.1 数据归一化处理

设第*i*个样本第*j*个指标值为 $x^*(i,j)$; $i=1,2,\dots,n$; $j=1,2,\dots,p$; n,p 分别为样本的个数(样本容量)和指标的数目。本文采用如下公式:

$$x(i,j) = \frac{x^*(i,j) - x_{\min}(j)}{x_{\max}(j) - x_{\min}(j)} \quad (1)$$

其中 $x_{\max}(j)$ 、 $x_{\min}(j)$ 分别为第*j*个指标值的最大值和最小值; $x(i,j)$ 为指标特征值归一化的序列。

1.2 构造投影指标函数 $Q(a)$

PP方法就是把*p*维数据 $\{x(i,j) | i=1,2,\dots,p\}$ 综合成以 $a = \{a(1), a(2), a(3), \dots, a(p)\}$ 为投影方向的一维投影值 $z(i)$

$$z^*(i) = \sum_{j=1}^p a(j)x(i,j) \quad (i = 1,2,\cdots,n)$$

(2)

然后根据 $z^*(i) - y(i)$ 的散点图建立适当的数学插值模型。其中 a 为单位长度向量,投影指标函数 $Q(a)$ 可构造 (Friedman and Turkey, 1974)

$$Q(a) = S_z D_z$$

(3)

其中 S_z 为投影值 $z(i)$ 的标准差; D_z 为投影值 $z(i)$ 的局部密度。

$$S_z = \sqrt{\frac{\sum_{i=1}^n [z(i) - E(z)]^2}{n - 1}}$$

(4)

$$D_z = \sum_{i=1}^n \sum_{j=1}^n [R - r(i,j)] \cdot u[R - r(i,j)]$$

(5)

$E(z)$ 为 a 投影方向投影值的平均值; R 为局部密度的窗口半径(密度窗宽),其值可以根据数据特征来确定; $r(i,j)$ 表示样本之间的距离, $r(i,j) = |z(i) - z(j)|$; $u(t)$ 为一单位阶跃函数,当 $t \geq 0$ 时,其值为 1,当 $t < 0$ 时其函数值为 0。

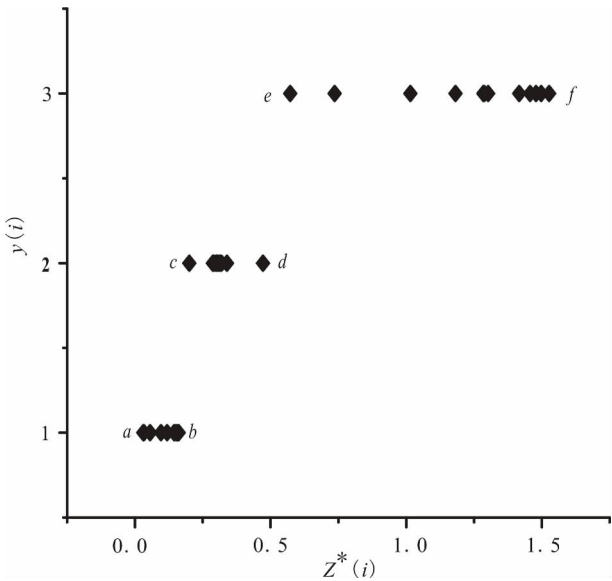


图 1 突水水源判别建模样本投影值与经验层位散点图
Fig. 1 Scatter dots of modeling samples projection value and the experience layer for water-inrush source discrimination

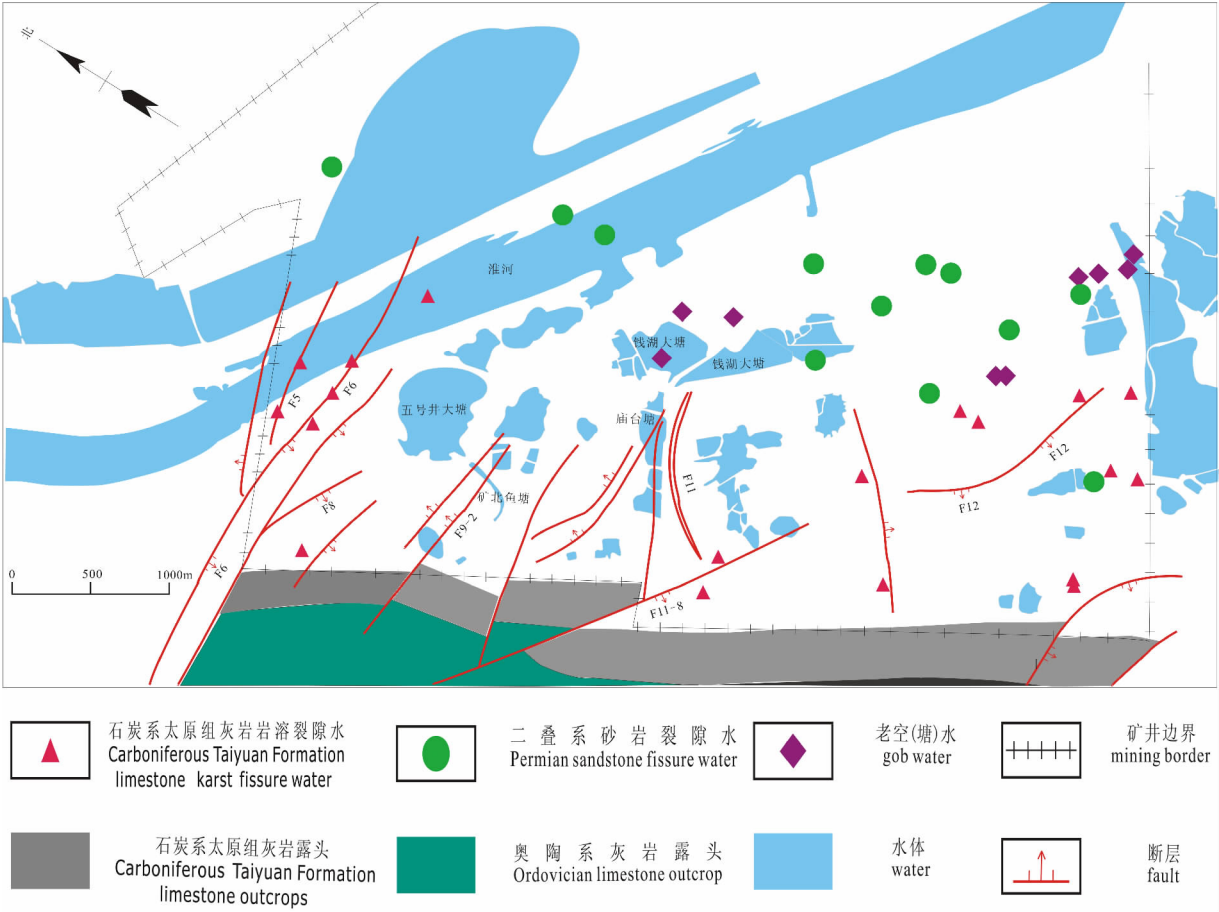


图 2 新庄孜煤矿取样点分布示意图
Fig. 2 Distribution of water samples in Xinzhuangzi mine

1.3 优化投影指标函数

投影寻踪方法成功应用的关键在于投影指标函数的优化。不同的投影方向反映不同的数据结构特征,最佳投影方向就是最大可能暴露高维数据某类特征结构的投影方向,因此可以通过求解投影指标函数最大化问题来估计最佳投影方向,即最大化目标函数:

$$\text{Max: } Q(a) = Z_i D_z$$

(6)

约束条件

$$\sum_{j=1}^p a^2(j) = 1$$

(7)

这是一个以 $\{a(j) \mid j = 1, 2, \cdots, p\}$ 为优化变量的复杂非线性优化问题,用传统的优化方法处理较难,可以引入现代优化算法,本文在 Matlab 运行环境下,使用遗传算法工具箱(雷英杰等,2005)对投影寻踪模型中的最佳向量进行优化求解。

1.4 投影寻踪判别模型

把求得的最佳投影方向 a^* 代入 (2) 式后可得各样本点的投影值 $z^*(i)$ 。根据第 i 个样本的计算的投影值 $z^*(i)$ 与经验等级 $y(i)$, 建立 $z^*(i) - y(i)$ 的散点图(图 1), 根据散点图建立投影寻踪判别模型。

2 实例分析

2.1 矿区水文地质概况

新庄孜井田位于八公山东麓,地貌属山前斜坡和冲积平原,淮河从上斜穿而过。井田内主要含水层(组)由第四系砂层孔隙含水层、二叠系砂岩裂隙含水层、石炭系太原组灰岩岩溶裂隙含水层、奥陶系灰岩岩溶裂隙含水层和寒武系灰岩岩溶裂隙含水层组成。通过新庄孜煤矿历史上发生的 21 起水害(突水)事故来看,矿井水害可分为以下四种类型:① 煤层顶板砂岩裂隙含水层水害;② 煤系底板太原群、奥陶系灰岩岩溶含水层水害;③ 老空(塘)水害;④ 地表水害。其中第四系砂层孔隙含水层:平均组厚 29.05m,流砂层平均厚 16.18m。砂层局部直接覆盖在煤系地层之上,形成砂层水对煤系地层的直接补给区。二

叠系砂岩裂隙含水层(以下简称“煤系水”):主要分布在 A3、B8、B11b 和 C13 煤层顶板,为砂岩脉状裂隙含水层,无直接补给来源,静储量疏干型特征明显。太原组灰岩岩溶裂隙含水层(以下简称“太灰水”)其中太原组 C3- I 组灰岩岩溶含水层地层总厚 40m,距 A1 煤层 16m 左右,是本矿开采 A 组煤的

表 1 突水水源判别建模样本指标投影值及经验层位
Table 1 Modeling sample index projection values and experience layer for water-inrush source discrimination

样本序号	离子浓度 (mg/L)						投影值	经验层位
	Ca ²⁺	Mg ²⁺	K ⁺ + Na ⁺	HCO ₃ ⁻	Cl ⁻	SO ₄ ²⁻		
1	78.18	16.43	26.30	255.06	30.56	62.77	0.1187	1
2	95.33	26.61	14.49	367.31	22.41	25.90	0.1533	1
3	101.98	21.40	14.03	395.23	22.41	16.88	0.1541	1
4	93.67	26.50	20.24	401.52	26.52	2.04	0.1574	1
5	97.79	27.59	14.03	387.46	28.36	15.64	0.1585	1
6	62.32	34.38	37.95	398.78	25.09	16.25	0.1558	1
7	88.18	28.08	20.24	389.29	24.11	25.52	0.1567	1
8	83.57	32.43	24.38	389.29	26.94	28.81	0.1613	1
9	85.77	29.18	25.07	397.83	25.17	20.93	0.1591	1
10	98.44	25.35	17.71	387.21	35.12	20.58	0.1579	1
11	6.37	5.84	49.45	54.92	40.77	1.23	0.0306	1
12	28.33	3.87	106.72	9.35	125.92	41.56	0.0559	1
13	93.39	13.74	29.21	394.17	28.01	39.10	0.1461	1
14	4.21	12.52	32.43	54.38	31.91	1.03	0.0331	1
15	85.77	30.02	21.16	38.29	27.65	24.49	0.0962	1
16	71.44	27.59	1.92	401.44	24.46	23.87	0.1444	1
17	2.4	1.9	452.18	1081.83	10.64	5.35	0.3041	2
18	1.80	2.31	463.95	1138.53	14.18	4.53	0.3177	2
19	4.81	2.19	609.50	817.63	16.66	15.64	0.3003	2
20	2.00	1.46	494.73	932.95	2.13	4.32	0.2874	2
21	2.40	2.43	442.29	1100.80	9.93	5.35	0.3059	2
22	4.41	2.31	721.51	1635.86	25.17	2.47	0.4719	2
23	2.00	2.31	522.23	1175.19	16.31	4.96	0.3391	2
24	94.99	32.47	375.36	410.66	587.41	20.58	0.2920	2
25	7.41	1.70	773.84	15.77	24.11	5.56	0.2003	2
26	2.40	2.19	429.68	1054.98	14.89	2.68	0.2944	2
27	11.20	4.13	450.11	1091.59	34.03	78.41	0.3226	2
28	18.45	11.43	453.59	846.92	40.86	255.60	0.3133	2
29	483.59	291.11	245.53	247.12	45.38	2565.71	1.0157	3
30	273.35	145.80	168.98	327.05	35.53	1292.22	0.5721	3
31	110.22	51.19	2486.60	1090.98	46.44	4739.57	1.4563	3
32	111.22	52.04	2123.59	1090.37	46.44	3987.58	1.2847	3
33	112.02	57.15	2518.78	1035.46	45.38	4879.93	1.4784	3
34	153.91	102.63	2442.46	811.53	41.13	5122.46	1.5266	3
35	144.89	91.44	2434.41	762.71	46.09	5131.42	1.4980	3
36	102.60	80.38	2111.86	882.92	37.93	4229.19	1.3024	3
37	88.18	60.56	1964.27	846.31	54.24	3814.71	1.1817	3
38	18.64	15.56	1263.07	1170.31	34.35	1698.26	0.7358	3
39	239.68	99.83	1838.51	586.37	39.35	4295.05	1.2879	3
40	105.81	48.03	2415.56	1122.71	46.80	4542.83	1.4163	3

主要充水因素之一。地表水:淮河从井田内斜穿而过,流经矿区段 4km,河床宽 250 ~ 300m,淮河水位一般在 +17 ~ +18m,最高洪水位为 +25.43m。另外,本井田范围内小井遍布,开采浅部露头煤和断层边角零星块段,许多小井井口位于塌陷积水坑(塘)附近,给大井安全生产造成威胁,老空(塘)水成为矿井充水因素之一。因此,开展新庄孜煤矿基于水化学特征的矿井突水水源快速判别研究意义重大。

2.2 投影寻踪突水水源判别模型

2.2.1 建模数据

此次研究共选取 55 个水样分布如图 2 所示,由于地表水水样很少,因此只选取了太灰水样、煤系水样、老空(塘)水样作为本次研究主要水样。用于建模的样本 40 个(表 1),其中老空(塘)水样 12 个、煤系砂岩水样 12 个、太原组灰岩水样 16 个。剩余 15 个样本(表 2)作为预测样本。判别指标选取 Ca^{2+} 、 Mg^{2+} 、 $\text{K}^{+} + \text{Na}^{+}$ 、 HCO_3^{-} 、 SO_4^{2-} 、 Cl^{-} 6 大常规离子。

2.2.2 建立投影寻踪水源判别模型

根据选取的 55 个样本直接代入公式(1)进行归一化处理,编写目标函数文件和约束函数文件,调用 Matlab 的遗传算法工具箱,并在 GUI 中完成相关参数设置(周正武等,2006),同时对 6 个水质指标进行非线性优化。经过优化,得到相应的最佳投影方向:

$$a(1) = 0.2615,$$
$$a(2) = 0.3687,$$

$$a(3) = 0.6315,$$
$$a(4) = 0.2870,$$
$$a(5) = 0.0350,$$
$$a(6) = 0.5597。$$

将归一化处理后的数据代入公式(2)按上述最佳投影方向 a 进行投影计算,即可计算出反映 6 个离子指标综合特征信息的投影值(表 1),并以投影值为横坐标,经验层位为纵坐标得到二者的散点图,如图 1。

从图 1 看出 $z^{*}(i)$ 与 $y(i)$ 的图形成阶梯形,各含水层所得投影值 $z^{*}(i)$ 并不重合,由此说明了投影寻踪模型充分放映了各含水层六大离子信息的差别,实现了在低维空间上分析高维数据。因此,确定各含水层投影值范围依次为:太灰水(ab 段) [0.0306, 0.1613]、煤系水(cd 段) [0.2003, 0.4719]、老空(塘)水(ef 段) [0.5721, 1.5266]。对于评价样本,归一化处理后按照最佳投影方向求得投影值 $z^{*}(i)$,根据 $z^{*}(i)$ 落在某个含水层所对应的投影值范围之内,或者与之最接近,然后根据相邻相似原理,判断评价样本的归属层位。例如,第 1 个评价样本的投影值为 0.1601,落在太灰水(ab 段) [0.0306, 0.1613] 之内,可以判定第 1 个评价样本的所属层位为太灰水。同理可以判定其余评价样本的所属层位,评价结果见表 2 所列。

2.2.3 四种判别模型的比较

针对新庄孜矿 40 个建模样本 15 个预测样本分

表 2 突水水源判别预测样本指标投影值及判别层位

Table 2 Forecast sample index projection values and discriminate layer for water-inrush source discrimination

水样编号	原归层位	Ca ²⁺	Mg ²⁺	K ⁺ + Na ⁺	HCO ₃ ⁻	Cl ⁻	SO ₄ ²⁻	投影值	判别层位
		(mg/L)							
41	太灰水	58.32	30.63	63.02	421.02	24.11	28.4	0.1601	太灰水
42	太灰水	11.42	8.02	20	67.12	19.94	20.99	0.0315	太灰水
43	太灰水	69.34	20.91	50.58	381.69	17.09	30.66	0.1435	太灰水
44	太灰水	7.41	2.67	234.73	617.53	16.14	10.7	0.1722	太灰水
45	太灰水	75.75	19.45	51.04	403.46	17.57	23.67	0.1483	太灰水
46	太灰水	6.41	4.25	234.96	626.96	15.19	8.44	0.1751	太灰水
47	老空(塘)水	92.58	25.89	1688.43	1336.27	52.12	2727.67	1.0382	老空(塘)水
48	老空(塘)水	55.91	29.18	1478.03	1418.04	36.16	2172.01	0.9225	老空(塘)水
49	老空(塘)水	24.85	12.52	815.46	1501.02	45.03	570.04	0.5586	老空(塘)水
50	老空(塘)水	245.49	209.4	304.39	637.63	57.08	1472.29	0.7475	老空(塘)水
51	煤系水	2.8	207	50.27	1153	10.19	0.62	0.4766	煤系水
52	煤系水	102.50	45.22	408.25	1185.62	56.26	267.48	0.4522	煤系水
53	煤系水	19.84	12.56	268.06	655.93	84.11	108.46	0.2214	煤系水
54	煤系水	3.21	0.49	365.08	858.55	18.79	3.91	0.2418	煤系水
55	煤系水	20.24	12.56	721.43	1831.83	33.33	118.75	0.5410	老空(塘)水(误判)

别采用投影寻踪模型、模糊综合评判模型、人工神经网络模型以及灰色聚类模型进行预测判别和回带判别,结果见表 3。从表 3 中可以看出投影寻踪判别模型的预测准确率均要高于另外三种模型,实现了投影寻踪模型在矿井突水水源判别中的应用。

3 结论

- (1) 应用高维降维技术——投影寻踪建立了新庄孜煤矿突水水源判别的 PP 模型,将突水评价指标作为多个投影参数来寻求最佳投影方向,并由最佳投影指标投影值的大小来确定突水水源的判别范围,避免人为赋权的干扰。
- (2) 本文采用 Matlab 的遗传算法工具箱与投影寻踪分类模型相结合,利用开放性函数和 GUI 集成化、模块化的特点,避免了用传统编程语言的复杂工作,解决了高维数据全局寻优难题,减少了寻优工作量,使模型能够得以更为广泛的应用。
- (3) 将投影寻踪分类模型应用于新庄孜煤矿水源判别分析是可行而有效的。为煤矿及时准确地查找突水成因和突水水源提供科学的依据。

参 考 文 献 / References

桂和荣,陈陆望. 2004. 皖北矿区主要突水水源水文地质特征研究. 煤炭学报,29(3):323 ~ 327.
雷英杰,张善文,李续武,周创明. 2005. Matlab 遗传算法工具箱及应用. 西安:西安电子科技大学出版社,95 ~ 102.

表 3 新庄孜煤矿突水水源快速判别准确率对比表
Table 3 Comparative accuracy of discriminating the water-bursting source quickly of Xinzhuangzi coal mine

模型	模糊综合评判		投影寻踪模型	人工神经网络		灰色聚类法	
	回判	预测	预测	回判	预测	回判	预测
结果	87.5%	73.33%	93.33%	91.12%	88%	95%	66.67%

李世峰,李耀华. 2006. 模糊综合判别矿井突(涌)水水源. 煤炭工程,(9):71 ~ 73.
李祚泳. 1997. 投影寻踪技术及其应用进展. 自然杂志,19(4):224 ~ 227.
马雷,钱家忠,赵卫东,张春雷,潘婧. 2009. 多元对应分析法在潘一矿地下水化学特征分析中的应用. 煤炭学报,34(11):1517 ~ 1521.
钱家忠,吕纯,赵卫东,潘婧. 2010. Elman 与 BP 神经网络在矿井水源判别中的应用研究. 系统工程理论与实践,30(1):145 ~ 150.
杨永国,黄福臣. 2007. 非线性方法在矿井突水水源判别中的应用研究. 中国矿业大学学报,36(3):284 ~ 286.
叶浩,钱家忠,黄夕川,李友龙,董洪信. 2005. 投影寻踪模型在地下水水质评价中的应用. 水文地质工程地质,5(4):9 ~ 10.
张瑞钢,钱家忠,马雷,覃华. 2009. 可拓识别方法在矿井突水水源判别中的应用. 煤炭学报,34(1):33 ~ 38.
周正武,丁同梅,田毅红,王晓峰. 2006. Matlab 遗传算法优化工具箱(GAOT)的研究与应用. 机械研究与应用,6(19):69 ~ 71.
Friedman J H, Turkey J W. 1974. A projection pursuit algorithm for exploratory data analysis. IEEE Trans on Computer, 23(9):881 ~ 890.
Li Dingding, Stengos Thanasis. 2001. The partially linear regression model: Monte Carlo evidence from the projection pursuit regression approach. Economics Letters, 75(2):11 ~ 16.
Zhang Chi, Dong Sihui. 2009. A new water quality assessment model based on projection pursuit technique. Journal of Environmental Sciences, 21(Supplement 1):154 ~ 157.

Discrimination Model of Mine Water-inrush Source Based on Projection Pursuit Method: A Case Study in Xinzhuangzi Coal Mine

QIAN Jiazhong, DU Kui, ZHAO Weidong, ZHOU Xiaoping, MA Lei

School of Resources and Environmental Engineering, Hefei University of Technology, Hefei, 230009

Abstract: Projection pursuit is a dimension-reduced data treatment method. Multi-dimensional question can be converted into low-dimension question according to projection direction. The key of using this method is to search the best projection direction, which can be solved by transforming a nonlinear optimization question, combined with Matlab genetic algorithm toolbox for the optimal solution. In this paper, Xinzhuangzi Mine was taken as an example, establishing projection pursuit model to discriminate water-inrush source, compared with fuzzy comprehensive evaluation, artificial neural network and gray clustering model. The result indicated that projection pursuit discrimination model has higher discrimination accuracy than fuzzy comprehensive evaluation, artificial neural network and gray clustering model, which provides a new way to discriminate water-inrush source.

Key words: genetic algorithm; projection pursuit model; Xinzhuangzi coal mine; water-inrush source