

文章编号: 1673-3363-(2014)02-0236-07

基于主成分分析与距离判别分析法的 突水水源识别方法

宫凤强^{1,2}, 鲁金涛¹

(1. 中南大学资源与安全工程学院, 湖南 长沙 410083;

2. 深部金属矿产开发与灾害控制湖南省重点实验室, 湖南 长沙 410083)

摘要 矿井突水是采矿生产过程中威胁最大的自然灾害之一, 快速有效地判别矿井突水水源是采矿工程安全生产的重要保障。选取 7 种水化学成分指标作为突水水源识别的样本变量, 采用主成分分析与距离判别分析相结合的方法建立了突水水源判别模型。以淮南老矿区谢一煤矿不同水层的水化学特征资料中的 33 个为学习样本, 11 个为预测样本, 对该方法进行了检验和应用, 并与现有的灰色关联度判别模型、Bayes 判别模型的判别结果进行分析比较。研究结果表明: 基于主成分分析与距离判别方法的突水水源判别模型其回判准确率为 95%, 预测正确率为 91%, 为矿山突水水源的识别提供了一种新方法。

关键词 水源判别; 矿井突水; 距离判别分析; 主成分分析

中图分类号 TD 745 文献标志码 A

DOI:10.13545/j.issn1673-3363.2014.02.012

Recognition method of mine water inrush sources based on the principal element analysis and distance discrimination analysis

GONG Fengqiang^{1,2}, LU Jintao¹

(1. School of Resources and Safety Engineering, Central South University, Changsha, Hunan 410083, China;

2. Hunan Key Lab of Resources Exploitation and Hazard Control for Deep Metal Mines, Changsha, Hunan 410083, China)

Abstract Mine water inrush is one of the greatest natural disasters in the mining production process. Quick and effective identification of mine water inrush source is an important guarantee of safe production in mining engineering. Seven kinds of water chemical composition having been selected as the sample variables in water bursting source recognition, and a forecast model of water inrush source is built by combining distance discriminant analysis with principal component analysis. This model is tested and applied in the different water layer of Xieyi Coal Mine in Huainan with thirty-three training samples and eleven predicting samples, and it is compared with the gray correlation discrimination model and Bayesian discrimination model. The results show that the model shows 95% of accuracy, and the predicting correct rate is up to 91%, which provides a new method for mine water inrush identification.

Key words identification of water source; mine water inrush; distance discriminant analysis; principal component analysis

收稿日期: 2013-01-18

基金项目: 国家自然科学基金项目(41102170); 中国博士后科学基金项目(2011M500973, 2012T50702); 中央高校基本科研业务费专项资金项目(2011QNZT090); 中南大学前沿研究计划项目(2010QZZD001)

作者简介: 宫凤强(1979-), 男, 山东省潍坊市人, 博士后, 从事工程灾害预测与控制等方面的研究。

E-mail: fengqiang@126.com

Tel: 18175973819

矿井突(涌)水是在采矿生产过程中的一种破坏性巨大的自然地质灾害,矿井突水事故不仅会造成重大的人员伤亡,也会导致严重的财产损失。例如,2010年3月28日,山西省华晋焦煤王家岭煤矿发生小窑老空区透水事故,事故造成153人被困,经全力抢救,115人获救,另有38名矿工遇难,仅事故抢险搜寻耗资达1亿元^[1]。突水事故已成为采矿工程设计和施工组织中必须考虑的重要隐患之一,怎样准确地预判矿井突水水源,受到国内外相关学者的广泛关注^[2-6]。

为减少矿井突(涌)水给矿山生产带来的破坏,并降低对井下工作人员、设备威胁的风险,如何准确找出矿井突水的水源,并加以有效的防治是矿山工程中的一个重要的命题。由于矿井各水层的水化学成分数据能反映相应水层的本质特征,因此,根据矿井水样的水化学成分判别突水水源的方法有很多,如神经网络、灰色关联度理论、距离判别分析方法等^[2,6-7]。在这些判别方法中,大部分没有考虑水化学判别指标之间的信息叠加,导致水源误判的影响。为此本文借鉴主成分分析理论的思想对水源指标进行信息提炼,把多个彼此相关、信息重叠的指标变量通过适当的线性组合转化成为彼此独立,同时提取原指标变异信息并带有特定含义的综合成分,更加有效地描述各不同水源的特征^[7-8]。然后结合距离判别分析法^[9-11]对处理过的样本数据进行判别,建立基于主成分分析和距离判别分析的模型,对矿井突水水源进行识别。

1 理论与算法

1.1 主成分分析

主成分分析法^[8]是一种数据压缩和特征信息提取技术。在数据处理中,经常会遇到高维数据组。由于数据的维数高、变量较多,在这些变量之间往往存在一定的相关性,因此这些样本数据很难反映总本的主要信息,而且在用统计方法研究多变量问题时,变量太多会影响计算量和增加分析问题的复杂性。主成分分析通过投影的方式,将高维数据以尽可能少的信息损失投影到低维空间,使数据降维,数据结构实现简化。在简化过程中,主成分分析将多个相关变量以尽可能少的信息损失为原则进行综合变化。

1.1.1 主成分分析的数学模型

用原始数据矩阵 X 的 p 个变量 X_1, X_2, \dots, X_p 作线性组合 $Y=AX$, 即:

$$\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\dots \\ Y_p &= a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned} \quad (1)$$

满足 $a_{i1} + a_{i2} + \dots + a_{ip} = 1$; Y_i 与 $Y_j (i \neq j; i, j=1, 2, \dots, p)$ 之间不相关; Y_1 是 (X_1, X_2, \dots, X_p) 的一切线性组合方差最大的, Y_2 是与 Y_1 不相关的 X_1, X_2, \dots, X_p 的一切线性组合中方差最大的; Y_1, Y_2, \dots, Y_p 的方差之和等于 X_1, X_2, \dots, X_p 的方差之和。

1.1.2 主成分的求解

主成分的求解过程也就是求转换矩阵 A 的过程。求解主成分的一般步骤如下:

1) 首先将原有变量数据标准化,然后计算各变量之间的协方差矩阵。

2) 计算协方差矩阵的特征向量为 $\lambda_1, \lambda_2, \dots, \lambda_p$, 相应的单位特征向量为 T_1, T_2, \dots, T_p 。其中转换矩阵 $A=T'$, 即 A 的第 i 行就是 T_i 的第 i 大特征根对应的单位特征向量 T_i 。且第 i 个主成分 Y_i 的方差就等于 λ_i 。

3) 第 k 个主成分 Y_k 的方差贡献率为 $\eta_k = \lambda_k / \sum_{k=1}^p \lambda_k = 1$ 。若取 $m (m < p)$ 个主成分, 主成分 Y_1, Y_2, \dots, Y_m 的累积贡献率为 $\xi_m = \sum_{k=1}^m \lambda_k / \sum_{k=1}^p \lambda_k$ 。

4) 在选取主成分个数时,一般取决于累积方差贡献率。通常取 m 个主成分使得方差的累积贡献率达到 80% 以上^[5], 则对应的前 m 个主成分的样本信息量包含 p 个原始变量所能提供的绝大部分信息。

1.2 距离判别分析

距离判别法^[9-10]可通过对已知分类样本数据的识别学习,建立相应的判别规则,进而对未知样本进行分类判别。

1.2.1 马氏距离

设总体 $G = \{X_1, X_2, \dots, X_m\}^T$ 为 m 元总体(考察 m 个指标), 样本 $X = \{x_1, x_2, \dots, x_m\}^T$ 。令 $u_i = E(X_i) (i=1, 2, \dots, m)$, 则总体均值向量为 $u = \{u_1, u_2, \dots, u_m\}^T$ 。总体 G 的协方差矩阵为

$$\Sigma = \text{cov}(G) = E[(G-u)(G-u)^T] \quad (2)$$

样本 X 与总体 G 的马氏距离定义为

$$d^2(X, G) = (X-u)^T \Sigma^{-1} (X-u) \quad (3)$$

1.2.2 多个总体的距离判定

设有 k 个 m 元总体: $G_1, G_2, \dots, G_k (k > 2)$, 从中任意取 2 个总体 G_p, G_q , 考察新的样本 X 到总体

G_p 和 G_q 的马氏距离的平方差:

$$d^2(X, G_q) - d^2(X, G_p) = -2(W_q(X) - W_p(X))$$

式中:

$$W_q(X) = (\Sigma^{-1} u_p)^T X - 0.5 u_p^T \Sigma^{-1} u_p$$

$$W_p(X) = (\Sigma^{-1} u_q)^T X - 0.5 u_q^T \Sigma^{-1} u_q$$

由此有

$$d^2(X, G_q) - d^2(X, G_p) = W_p(X) - W_q(X)$$

$$d^2(X, G_q) < d^2(X, G_p) \quad W_p(X) < W_q(X)$$

一般情况下, 总体的均值向量 u_1, u_2, \dots, u_k 和公共协方差矩阵是未知的, 可以利用各总体的学习样本作估计。设 $X_{(t)}^{(q)} = \{x_{t1}^{(q)}, x_{t2}^{(q)}, \dots, x_{tm}^{(q)}\}^T$ ($q=1, 2, \dots, k; t=1, 2, \dots, n_q, n_q$ 为总体 G_q 的学习样本个数) 为来自总体 G_q 的学习样本, 则 u_q 的无偏估计为

$$\bar{X}^{(q)} = \left\{ \frac{1}{n_q} \sum_{t=1}^{n_q} x_{t1}^{(q)}, \dots, \frac{1}{n_q} \sum_{t=1}^{n_q} x_{tm}^{(q)} \right\}^T \quad (4)$$

学习样本的协方差矩阵 S_q 的估计为组内协方差矩阵 S_q :

$$S_q = \frac{1}{n_q - 1} \sum_{t=1}^{n_q} \left(X_{(t)}^{(q)} - \bar{X}^{(q)} \right) \left(X_{(t)}^{(q)} - \bar{X}^{(q)} \right)^T \quad (5)$$

当各总体的协方差矩阵相等时, 则此时总体的协方差矩阵的一个无偏估计为 $W_q(X)$

$$S = \frac{1}{\sum_{q=1}^k n_q - k} \sum_{q=1}^k (n_q - 1) S_q \quad (6)$$

以 $\bar{X}^{(q)}$ 和 S 分别代替 u_q 和 Σ , 从而可以得到 $W_q(X)$ ($q=1, 2, \dots, k$) 的估计为

$$\bar{W}_q(X) = \left(S^{-1} \bar{X}^{(q)} \right)^T X - 0.5 \left(\bar{X}^{(q)} \right)^T S^{-1} \bar{X}^{(q)} \quad (7)$$

因此, 多总体情况下的距离判别准则为: 若总体 G_{q0} 满足:

$$W_{q0}(X) = \max_{1 \leq q \leq k} W_q(X) \quad (8)$$

则 $X \in G_{q0}$ 。

1.2.3 判别准则的评价

为考察上述判别准则的优良性, 采用以训练样本为基础的回代估计法来计算误判率。以 2 个总体 G_1 和 G_2 为例^[3], 来自总体 G_i ($i=1, 2, \dots, n_i$) 的训练样 $X_{(t)}^{(i)} = \{x_{t1}^{(i)}, x_{t2}^{(i)}, \dots, x_{tm}^{(i)}\}^T$ ($t=1, 2, \dots, n_i$), 其中 n_i 为取自 G_i 的样本个数, 并且 2 个总体的容量分别为 n_1 和 n_2 。以全体训练样本作为 n_1+n_2 个新样本, 逐个代入已建立的判别准则中判别其归属, 这个过程称为回判。用 n_{12} 表示将属于总体 G_1 的样本误判

为总体 G_2 的个数, n_{21} 表示属于总体 G_2 的样本误判为总体 G_1 的个数。误判率 η 的回代估计为

$$\eta = \frac{n_{12} + n_{21}}{n_1 + n_2} \quad (9)$$

2 突水水源判别模型及应用

2.1 突水水源判别指标的确定

各含水层中水化学成分多, 考虑每一种水化学成分离子作为判别指标不太现实, 综合考虑离子对于不同水源的差异性及数据的有效性, 本文参考有关文献^[6]选取 $\text{Ca}^{2+}(X_1)$, $\text{Mg}^{2+}(X_2)$, $\text{K}^+ + \text{Na}^+(X_3)$, $\text{HCO}_3^-(X_4)$, $\text{SO}_4^{2-}(X_5)$, $\text{Cl}^-(X_6)$, $\text{TDS}(X_7)$ 共 7 种离子组分的含量作为预测矿井突水判别分析模型的水源指标。

2.2 突水水源判别分析模型的建立

以参考文献[6]中的谢一矿井突水信息为工程背景, 结合主成分分析和距离判别分析建立突水水源的识别模型。谢一矿井突水水源分为 3 类: 煤系砂岩裂隙含水层、石炭系太原群灰岩溶隙含水层、奥陶系灰岩溶隙含水层 (简称煤系、太灰、奥灰)。从大量的矿井突水水源中选取 33 个突水水样作为训练样本, 其中煤系水样 13 个, 太灰水样 13 个, 奥灰水样 7 个, 谢一矿水源的水化学数据实际资料如表 1。

2.2.1 主成分分析

对表 1 中的样本数据进行归一化处理, 然后进行主成分分析。其各项因素之间相关系数矩阵如表 2 所示。从表 2 可知输入因素中的几个因素彼此具有明确的相关性, 这必定会对矿井突水水源预测模型的精度造成影响, 容易造成水源误判。因此, 有必要对输入数据进行主成分分析。

表 3 为各成分方差贡献率及累计贡献率。选择主成分是通过特征值从大往小取, 特征向量也根据相应的特征值来选取, 特征值越大, 相对应的主成分数据越重要。选取的主成分数据一般占 80% 以上即可, 因此在本文中, 通过主成分分析方法处理后, 提取到的前 2 个主成分, 包含了原数据的 83.8% 的信息, 可以概括原始变量。同时与图 1 中的碎石图显示的信息相吻合。

由表 4 主成分因子载荷矩阵可知, Y_1 主要由原始变量 X_1, X_2, X_3, X_4, X_7 表征, 它反映了原始变量的 66.36% 的信息; Y_2 主要由原始变量 X_5, X_6 表征, 它反映了原始变量的 17.48% 的信息。

表 1 谢一矿井突水水源判别的变量资料
Table 1 Data of cases used in discriminating resource of Xieyi coal mine intrush

样品	水化学成分指标/(mg·L ⁻¹)							判别分类结果			
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	实际结果	灰色关联度	Bayes 判别	本文方法
1	4.61	1.34	237.95	533.9	25.53	12.14	548.52				
2	5.21	2.55	373.36	422.26	61.33	357.17	1010.75				
3	4.41	2.00	339.79	620.57	54.59	125.95	837.20				
4	3.81	2.19	235.42	443.62	49.28	66.88	579.39				
5	4.81	2.07	469.00	743.22	56.37	281.74	1182.6				
6	3.81	2.31	342.78	812.8	29.42	10.70	795.42				
7	4.81	3.04	333.67	761.53	28.71	5.97	757.00				
8	5.81	4.50	339.71	647.39	46.44	123.89	844.05				
9	2.40	0.61	277.61	652.88	15.60	16.87	639.53				
10	4.41	3.92	327.29	819.46	23.04	5.35	773.74				
11	2.00	1.22	352.13	755.53	14.54	12.35	760.01				
12	2.00	2.07	298.77	698.98	13.47	10.29	674.09				
13	7.62	2.67	376.28	856.07	71.97	43.22	929.85				
14	71.54	29.43	41.40	386.87	29.42	29.64	394.87				
15	91.58	26.27	16.10	379.54	28.71	19.14	371.57				
16	79.96	27.97	17.94	356.97	29.07	19.14	352.57				
17	85.77	26.14	21.16	407.00	29.42	36.84	402.83				
18	92.79	28.58	24.61	10.81	358.17	25.53	16.88				
19	89.98	24.07	10.81	358.17	25.53	16.88	346.36				
20	86.37	25.89	24.15	386.85	25.17	20.99	376.00				
21	88.38	24.07	31.05	389.9	27.85	27.37	393.67				
22	88.38	22.85	31.51	361.83	34.03	37.04	747.06				
23	94.19	29.54	19.32	370.37	34.39	44.86	407.49				
24	89.58	24.31	14.26	356.34	26.59	23.87	356.78				
25	90.58	24.56	19.78	369.13	26.59	28.81	374.89				
26	91.78	24.68	11.04	356.34	27.65	22.64	355.96				
27	58.72	24.93	28.52	331.93	19.50	8.05	305.69				
28	61.32	18.60	24.84	210.52	41.12	50.83	301.97				
29	97.39	29.43	27.14	330.12	40.41	91.58	451.01				
30	79.36	23.59	23.00	284.35	38.99	54.54	361.66				
31	52.30	29.30	22.08	223.32	98.29	59.68	313.31				
32	90.18	27.23	13.57	345.36	34.39	33.75	371.80				
33	48.70	35.51	28.05	354.51	22.34	6.17	318.03				

表 2 各水化学成分指标 Pearson 相关系数矩阵
Table 2 Pearson correlation coefficient matrix of each water chemical component

指标	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
X ₁	1.00						
X ₂	0.92	1.00					
X ₃	-0.93	-0.94	1.00				
X ₄	-0.79	-0.81	0.86	1.00			
X ₅	0.13	0.14	-0.07	-0.400	1.00		
X ₆	-0.30	-0.31	0.45	.085	-0.40	1.00	
X ₇	-0.77	-0.82	0.90	0.82	0.09	0.58	1.00

表 3 Y₁~Y₂ 各成分特征值、方差贡献率及累积贡献率
Table 3 Eigenvalues, contribution rate accumulative and contribution rate of principal component variance from Y₁~Y₂

因素	提取平方与载入		
	特征值	贡献率/%	累积贡献率/%
Y ₁	4.65	66.36	66.36
Y ₂	1.22	17.48	83.84

表 4 2 个主成分因子的荷载矩阵
Table 4 Loading matrix on 2 principal components

指标	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
Y ₁	-0.199	-0.203	0.211	0.194	-0.058	0.090	0.202
Y ₂	-0.008	-0.009	0.102	-0.283	0.643	0.555	0.062

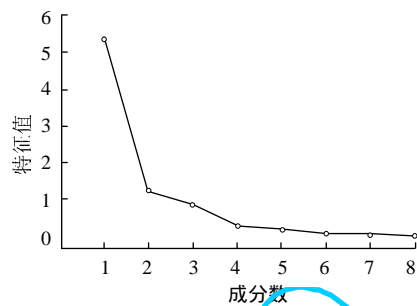


图 1 主成分分析碎石图

Fig.1 Principal component analysis of the macadam

根据主成分分析矩阵，表 5 中给出因子 Y_1 、 Y_2 与原始变量之间的关系，其因子表达式为：

$$Y_1 = -0.199X_1 - 0.203X_2 + 0.211X_3 + 0.194X_4 - 0.058X_5 + 0.090X_6 + 0.202X_7 \tag{10}$$

$$Y_2 = -0.199X_1 - 0.203X_2 + 0.211X_3 + 0.194X_4 - 0.058X_5 + 0.090X_6 + 0.202X_7 \tag{11}$$

根据式(10)，(11)对表 1 的数据进行标准化处理，然后对处理后数据进行主成分分析计算，计算的样本数据见表 5。

表 5 主成分计算后的数据

Table 5 The data after calculation of principal components

指标	1	2	3	4	5	6	7	8	9	10	11
Y_1	0.756	1.368	1.201	0.714	1.813	1.262	1.164	1.192	0.997	1.211	1.239
Y_2	-0.495	2.738	0.657	0.293	1.792	-0.729	-0.715	0.514	-0.689	-0.863	-0.802

指标	12	13	14	15	16	17	18	19	20	21	22
Y_1	1.060	1.394	-0.590	-0.689	-0.694	-0.598	-1.567	-0.691	-0.638	-0.595	-0.362
Y_2	-0.804	-0.034	-0.345	-0.441	-0.408	-0.332	3.676	-0.470	-0.469	-0.389	-0.131

指标	23	24	25	26	27	28	29	30	31	32	33
Y_1	-0.711	-0.678	-0.652	-0.699	-0.812	-0.824	-0.897	-0.868	-0.971	-0.933	-0.900
Y_2	-0.172	-0.401	-0.375	-0.401	-0.586	0.131	0.272	0.037	0.802	-0.257	-0.604

2.2.2 距离判别分析模型构建与检验

对表 1 中 33 个样本数据进行学习。以 I、II 和 III 这 3 类水源作为 3 个不同的总体，并假设 3 个水源总体的协方差矩阵相等。将上面通过主成分分析方法得到的第一主成分 Y_1 、第二主成分 Y_2 的 2 个指标作为距离判别分析模型的判别因子，按照本文提出的距离判别分析法进行建模。

在利用样本数据进行学习时，当输入变量个数确定时，样本数据的个数肯定会对判别分析模型的预测精度有所影响。目前为止，专门探讨样本数据个数对模型预测精度影响的论文尚不多见。为此，本文在前面利用 33 组样本数据进行主成分分析的基础上，简要探讨一下样本数据个数对判别分析模型精度的影响规律。在 33 组样本数据个数中，来自 I、II 和 III 水源的样本数分别为 13，13 和 7，为了体现均等性和随机性，在每个水源样本数据中分别依次从最后抽取 2，2 和 1 组数据，剩余的样本作为学习样本进行建模并对模型的预测精度进行检验，考察样本数据个数对预测精度的影响。按照抽样原则，进行 4 次抽样，进而分别得到 33，28，23，18 和 13 组样本数据。对上述 5 种情况的样本数据进行学习，并利用回代估计法考察对应模型的

误判率，结果如表 6，图 2 为所得误判率的比较图。

表 6 样本数据个数对判别模型预测精度的影响

Table 6 Effects of sample data number on the discriminant model prediction accuracy

抽样组数	样本序号	实际水源	判别水源
13	14	II	III
	15	II	III
	16	II	III
	17	II	III
	29	III	II
18	28	III	II
23	18	II	III
28	18	II	III
33	18	II	III

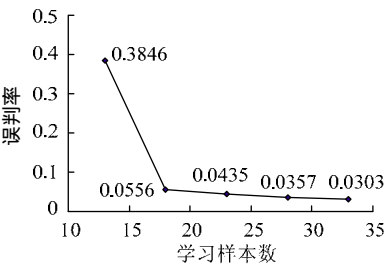


图 2 学习样本数对误判率的影响规律

Fig.2 Influence of the number of training samples to misjudgment rate

由表 6 可知，在样本数据为 13 组的情况下，所得判别模型对 13 组全体样本回代判别时，判错了 5 个样本，所得误判率为 0.384 6，显然精度非常低。之后随着学习样本个数的增加，在其余 4 种情况下，所得判别模型分别只判错了 1 个样本，依次所得误判率分别为 0.055 6，0.042 5，0.035 7 和 0.030 3。由图 6 可以看出，误判率在学习样本个数为 23 时已经很小。之后随着学习样本个数的增加，误判率呈现依次降低、逐渐趋向收敛的趋势。检验结果说明学习样本个数对判别模型的精度预测有较大影响，之后当学习样本个数达到一定容量时，判别模型的预测精度才能满足要求。同时也说明，在条件允许的情况下，应尽量获取最大数量的学习样本用于判别模型的构建，其所得判别模型的预测精度会更高。

需要说明的是，此处随机抽取的学习样本是利用 33 组原始样本进行主成分分析后所得数据。所得结果可能要优于直接利用原始数据考察样本数对判别模型预测精度影响的结果。

2.3 判别模型的应用

如 2.2.2 节中所述，将表 1 中全体训练样本逐个代入已建立的判别准则中进行回判，只有 18 号

水样奥灰水被误判成了太灰水，其判别正确率为 97%。误判原因可能由于太灰水样与奥灰水样在水质上比较相近，有较强的关联性，从而造成了误判。回判结果见表 1，同时与灰色关联度理论模型和 Bayes 理论模型相比较，灰色关联度理论判别模型回判错了 5 个，有 4 个奥灰水样误判成了太灰水样，有 1 个太灰水样被判成了奥灰，判别正确率为 85%。Bayes 水源判别模型回判错了 3 个，有 2 个奥灰水样误判成了太灰水样，有 1 个太灰水样误判成了奥灰，判别正确率为 91%。通过回判结果，总体认为本文的矿井水源识别模型的稳定程度高，而且比较可靠，能满足水源判别的实际要求。

根据学习好的主成分分析与距离判别模型对谢一矿井的 11 个待判水样样本^[15]进行判别，结果表明，除了 44 号太灰水误判成了奥灰水，其他预测结果与实际结果是一样的，预判准确率为 91%，能准确地对待测水样进行判别分类。预测结果见表 7，同时列出了灰色关联度理论、Bayes 判别理论^[4]对水样进行预测的结果，利用灰色关联度理论模型误判了 1 个，利用 Bayes 判别理论模型误判了 2 个。综合比较，可认为本文的判别模型具有更好的识别能力。

表 7 水源判别模型的测试样本
Table 7 Testing samples of water Inrush sources discriminant model

预测 样品	水化学成分指标/(mg·L ⁻¹)							判别分类结果			
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	实际结果	灰色关联度	Bayes 判别	本文方法
34	7.41	3.16	346.69	674.24	53.18	121.83	869.39				
35	3.81	2.31	294.04	638.24	46.8	59.27	725.35				
36	3.81	2.31	294.04	638.24	46.08	59.27	724.63				
37	1.4	1.46	233.58	444.2	42.9	52.07	553.51				
38	5.21	1.34	252.43	508.88	51.05	67.91	632.38				
39	63.93	30.39	48.53	389.9	28.01	30.05	395.86				
40	88.78	26.62	12.42	357.56	26.94	25.93	359.47				
41	85.77	28.57	17.48	363.66	28.72	30.05	372.42				
42	82.96	28.93	25.99	366.1	28.22	40.34	389.49				
43	101.2	31.62	17.01	381.97	48.22	36.84	425.88				
44	84.97	12.04	11.96	251.31	37.74	24.69	297.06				

3 结 论

1) 利用主成分分析法将突水水源信息的高维数据进行降维，提取了数据的主要信息，同时通过更少的变量有效地表征了矿区水源的水化学特性，然后结合距离判别理论进行水源判别，可以有效消除水源判别中评价指标之间的相互影响。当评价指

标越多时，其优越性更能凸显出来。
2) 通过本文结果表明，基于主成分分析的距离判别模型正确率高达 95% 以上，为提高水源判别提供了一种新的方法，为预防突水工作提供了决策依据。
3) 在水源判别模型的建立过程中，矿区水源的原始水质资料是基础，其选择相关的水质离子信息

是否有代表性、准确性,对于模型建立是非常重要的。因此需要大量、广泛地收集矿井水源相关的实测数据,建立完善的训练样本数据库,提高水源判别模型的精度。

参考文献:

- [1] 陈美智,周桂铨.煤矿水害事故分析及其防治技术[J].西部探矿工程,2012(3):201-206.
CHEN Meizhi, ZHOU Guiquan. Accident analysis and prevention technology of coal mine water disaster[J]. West China Exploration Engineering, 2012(3):201-206.
- [2] 姜谟男,梁冰.基于最小二乘支持向量机的煤层底板突水量预测[J].煤炭学报,2005,30(5):613-617.
JIANG Annan, LIANG Bing. Forecast of water inrush from coal floor based on least square support vector machine[J]. Journal of China Coal Society, 2005, 30(5): 613-617.
- [3] 杨天鸿,唐春安,谭志宏,等.岩体破坏突水模型研究现状及突水预测预报研究发展趋势[J].岩石力学与工程学报,2007,26(2):268-277.
YANG Tianhong, TANG Chun'an, TAN Zhihong, et al. State of the art of inrush models in rock mass failure and developing trend for prediction and forecast of ground-water inrush[J]. Chinese Journal of Rock Mechanics and Engineering, 2007, 26(2): 268-277.
- [4] 陈红江,李夕兵,刘爱华,等.煤层底板突水量的距离判别分析预测方法[J].煤炭学报,2009,34(4):487-491.
CHEN Hongjiang, LI Xibing, LIU Aihua, et al. Forecast method of water inrush quantity from coal floor based on distance discriminant analysis theory[J]. Journal of China Coal Society, 2009, 34(4): 487-491.
- [5] 李利平,李术才,石少帅,等.岩体突水通道形成过程中应力-渗流-损伤多场耦合机制[J].采矿与安全工程学报,2012,29(2):232-238.
LI Liping, LI Shucui, SHI Shaoshuai, et al. A multi-field coupling mechanism of seepage damage for the water inrush channel formation process of coal mine[J]. Journal of Mining & Safety Engineering, 2012, 29(2): 232-238.
- [6] 吕纯.谢一矿地下水化学特征及突水水源判别 Elman 神经网络模型[D].合肥:合肥工业大学,2009:34-35.
- [7] 王超,宋大钊,杜学胜,等.煤与瓦斯突出预测的距离判别分析法及应用[J].采矿与安全工程学报,2009,26(4):470-474.
WANG Chao, SONG Dazhao, DU Xuesheng, et al. Prediction of coal and gas outburst based on distance discriminant analysis method and its application[J]. Journal of Mining & Safety Engineering, 2009, 26(4): 470-474.
- [8] 范金城,梅长林.数据分析[M].北京:科学出版社,2002:27-42.
- [9] 宫凤强,李夕兵.距离判别分析法在岩体质量等级分类中的应用[J].岩石力学与工程学报,2007,26(1):190-194.
GONG Fengqiang, LI Xibing. Application of distance discriminant analysis method to classification of engineering quality of rock masses[J]. Chinese Journal of Rock Mechanics and Engineering, 2007, 26(1): 190-194.
- [10] 刘金海,冯涛,王卫军,等.急倾斜煤层顶煤可放性识别的距离判别方法及应用[J].煤炭学报,2008,33(6):601-605.
LIU Jinhai, FENG Tao, WANG Weijun, et al. Distance discriminant analysis method for distinguishing the difficulty degree of top-coal caving in steep seam and its application[J]. Journal of China Coal Society, 2008, 33(6): 601-605.
- [11] 朱卓慧,赵伏军,叶洲元.基于距离判别分析法的冲击低压预测研究[J].中国安全科学学报,2008,18(3):41-45.
ZHU Zhuohui, ZHAO Fujun, YE Zhouyuan. Prediction of rock burst in mining based on distance discriminant analysis method[J]. China Safety Science Journal, 2008, 18(3): 41-45.