

# 基于PCA-Bayes综合判别方法的祁东矿煤层顶板突水水源判别研究

彭涛声, 胡友彪, 琚棋定, 胡泰丰

安徽理工大学地球与环境学院, 安徽 淮南

收稿日期: 2022年5月24日; 录用日期: 2022年6月27日; 发布日期: 2022年7月5日

## 摘要

祁东煤矿煤层顶板压架突水事故频发, 准确、快速地识别突水水源是防止突水再次发生的关键。为此, 从祁东煤矿二类含水层中提取了28个训练水样以及潘二矿三类含水层的11个水样作为验证样本, 以 $\text{Ca}^{2+}$ 、 $\text{Mg}^{2+}$ 、 $\text{Na}^+ + \text{K}^+$ 、 $\text{HCO}_3^-$ 、 $\text{Cl}^-$ 、 $\text{SO}_4^{2-}$ 为评价变量。采用主成分分析法消除训练样本中的冗余离子变量, 然后结合Bayes法建立模型, 判别结果表明: 新生界松散含水层(I)水样判别正确率为100%, 煤系砂岩水(II)水样判别正确率为66.67%, 太原组灰岩水(III)水样判别正确率为100%, 模型的综合精度为90.91%。与单一Bayes判别方法相比, 该方法具有准确率高、速度快等优点, 基于该模型判别结果, 可在煤矿突水事故发生后快速查明突水原因, 有效预防矿井水害。

## 关键词

矿井突水, 水源判别, 主成分分析法, Bayes判别

# Research on the Source of Water Inrush from Coal Roof in Qidong Mine Based on PCA-Bayes Comprehensive Discrimination Method

Taosheng Peng, Youbiao Hu, Qiding Ju, Taifeng Hu

School of Earth and Environment, Anhui University of Science and Technology, Huainan Anhui

Received: May 24<sup>th</sup>, 2022; accepted: Jun. 27<sup>th</sup>, 2022; published: Jul. 5<sup>th</sup>, 2022

## Abstract

Because water inrush accidents occurred frequently in Qidong Coal Mine's roof pressing support, it's the key to identify the source of water inrush accurately and rapidly to prevent water inrush from happening again. Based on the hydrogeological data collected in Qidong Coal Mine, according to the hydrochemical characteristics of aquifer,  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+ + \text{K}^+$ ,  $\text{HCO}_3^-$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$  were selected to establish a water inrush source discrimination model based on the combination of Principal Component Analysis and Bayes discrimination. Taking 28 water samples from Qidong Coal Mine as training samples and 11 water samples from Pan NO.2 Coal Mine as verification samples, the model is tested and applied. The Principal Component Analysis method is used to eliminate redundant ion variables in the training samples, and then the Bayes method is used to establish a model. The discriminant results show that the discriminant accuracy of Cenozoic loose aquifer (I) water samples is 100 %, the discriminant accuracy of coal measure sandstone water (II) water samples is 66.67 %, the discriminant accuracy of Taiyuan Formation limestone water (III) water samples is 100 %, and the comprehensive accuracy of the model is 90.91 %. Compared with the direct Bayes discrimination, the model reduces the calculation error caused by a large number of redundant original data and improves the calculation accuracy, makes the discrimination more accurate.

## Keywords

Mine Water Inrush, Identification of Water Source, Principal Component Analysis, Bayes Discriminant

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

煤炭是我国宝贵的能源资源，大规模的煤矿开采严重地影响了地下水赋存环境，地下水动力条件以及不同含水层之间的水力联系势必发生改变，从而导致煤矿区水文地质条件复杂化[1]。随着浅部煤炭资源的逐渐枯竭，煤炭开采正以每年 10~25 m 的速度快速向深部进行[2]，逐年增加的矿井突水事故严重影响了煤矿的安全生产，造成了重大的经济损失和人员伤亡。

近年来国内外学者从不同层次，采用各种方法识别矿井突水水源[3]。徐星[4]采取具有全局最优功能的遗传算法(GA)优化 BP 神经网络，从而建立具有仿真输出结果总误差最小、精度最高优点的突水水源的 GA-BP 神经网络；郭中安[5]结合主成分分析、遗传算法和极限学习机(Extreme Learning Machine, ELM)，通过 GA 优化 ELM，结合良庄煤矿 51101 工作面实测数据资料建立判别模型，有效地消除水化学离子指标间的相互影响，优化分类模型的权值和阈值，使矿井突水水源判别更为准确。张妹[6]结合 PCA 分析法和 Fisher 判别法建立矿井突水水源判别模型，并与 BP 神经网络判别模型进行对比，具有较高的可行性与准确性。曲兴玥等[7]用因子分析和距离判别法判别突水水源，提高了判别准确率。但以上判别方法都有一些不足，BP 神经网络较为繁琐，计算较为复杂；Fisher 判别方法只针对非线性问题，应用有一定局限性；因子分析与距离判别模型在计算中，因子得分容易失效。笔者针对祁东煤矿近年来压架突水事故，在主成分分析的基础上结合 Bayes 理论建立判别模型，该模型能高效且精准地判别突水水源，在此基础

上及时调查和处理含水层的充水通道和充水强度,可以有效防止突水事故的发生。该模型仅针对水化学信息的定量分析,因此,也可应用于其他类似矿井水源判别,具有广泛的适用性。

## 2. 研究区概况

### 2.1. 地层

#### 2.1.1. 矿井地层

祁东矿区地处淮北煤田的东南边缘,在地层区划上属于华北地层区鲁西地层分区徐宿小区,此区域的地层裸露出的较少,大部分是第四系冲积、洪积平原覆盖。该区域内发育的地层从老至新的顺序为青白口系、震旦系、寒武系、奥陶系、石炭系、二叠系、侏罗系、白垩系、上第三系和第四系。

##### 1) 石炭系上统太原组(C3t)

厚度 192.81 m,其中以石灰岩为主、占本组地层厚度的 40%左右。石灰岩有 8~14 层,一般 11~12 层。石灰岩编号自上而下为一灰~十二灰。其中第三层灰岩、第四层灰岩、第十二层灰岩比较厚,基本上处于稳定状态。顶部一灰厚度一般 2~3 m,顶部含泥质,富含动物化石及其碎屑,细晶-粗晶结构,是下部煤组对比的重要标志。本组含煤 6~8 层,总厚平均 3.58 m,属不稳定煤层。

##### 2) 二叠系下统山西组(P1s)

厚度 120~165 m,平均 135 m,以 125~150 m 为多。本组岩性由砂岩、粉砂岩、泥岩和煤层组成。含 10、11 两个煤组,含煤系数 0.80%,10 煤层上偶含薄煤线。10 煤层为本矿井局部可采煤层之一,11 煤层一般不可采且不稳定。与下伏地层整合接触。

##### 3) 二叠系下统下石盒子组(P1xs)

大部分为三角洲平原沉积,本组以铝质泥岩之底为下届,以 K3 砂岩之底为顶。地层厚度变化较大,揭露厚度为 205~245 m,平均 235 m,全区以 220~230 m 为主。含 4、6、7、8、9 五个煤层(组),含煤 10~13 层,煤层总厚 15.53 m。含煤系数 6.69%。本组为本矿井主要含煤段。岩性主要由砂岩、粉砂岩、泥岩、铝质泥岩和煤层组成。砂岩多在 63~9 煤层(组)间和 4 煤(组)上。

##### 4) 第三系(R)

第三系揭露有中新统、上新统,厚度约在 171.50 m 至 369.300 m 之间,平均 275.60 m,与下伏地层呈不整合接触。据钻孔岩性岩相组合特征,化石孢粉资料与有关地层剖面对比,第三系包括:渐新统(E3)、中新统(N1)、上新统(N2)。

##### 5) 第四系(Q)

厚度 70.55~101.60 m,平均 87.80 m,与下伏新近系呈假整合接触。

#### 2.1.2. 含煤地层

##### 1) 32 煤层

32 煤层在石盒子组下部,距离煤层 23 约 110 m,魏庙断层北部可作为采区,以南部分保留,煤层厚度在 0.31 m 至 4.11 m 之间,整体平均厚度约为 1.62 m,变异系数 53%,区内开采 80%,开采指数为 0.82,具有较为复杂的采区工作面,主要为 1~2 层泥岩或泥炭夹煤矸,煤层顶板以泥岩为主,砂岩、粉砂岩零星分布,是一个相对稳定的煤层,大部分地区均可开采。

##### 2) 61 煤层

位于下石盒子组的中部,位于 60 煤下大约 11 m 处是 6 煤组中发育较为完全的一层。此处区域可以提供大多数可开采的煤层,该煤层的厚度在 0 到 5.21 m 之间,平均厚度在 1.56 m,变异系数为 56%,可采指数为 0.77,可采面积率为 70%。泥岩夹矸石较少,煤层顶部有少量粉砂岩和砂岩,岩浆在四点穿透,

是该矿的重要组成部分，煤层稳定。

### 3) 71 煤层

为 7 煤组的上层，处于下石盒子中下部，63 煤一般约 30 m。该区大部分为可采区，厚度在 0 m 至 4.78 m 之间，平均 1.87 m，变异系数 57%，可采指数 0.83，可采面积 70%，碳结构复杂，多为 1~2 层夹矸；顶板以粘土和粉砂岩为主，其次为砂岩。71 煤层是一个相对稳定的煤炭网络，可以在大多数地区开采。

### 4) 82 煤层

位于下石盒子组下部，81 煤下 7~18 m，一般约 11 m，可开采区域面积相对较大，煤层厚度在 0 m 至 5.54 m 之间，平均 1.61 m，变异系数 49%，开采指数 0.95，面积可采率 90.5%，煤层顶板以砂岩为主，泥岩、粉砂岩较少。82 煤层相对于其他煤层较为稳定，可开采区域面积较大。

### 5) 9 煤层

靠近下石盒子组下部，82 煤层下部约 10 m 至 21 m 处，平均约 16 m，可开采区域面积相对较大，煤层厚度在 0 m 至 5.78 m 之间，平均 2.57 m，变异系数 57%，可采面积率为 70%，有 50 个岩浆岩断点，占穿过点的 25%，10 号煤层后第二个煤层破坏最为严重，以砂岩为主，粉砂岩和泥岩比重较小。

## 2.1.3. 含水层

### 1) 直接充水含水层

砂岩厚度处于 7.50 m 至 65 m 之间，大部分集中在 21 m，细粒砂岩占大部分。该层砂岩裂隙发育较不完全，一小部分发育较为完全，尤其是平坦部分，当 K3 砂岩处于风氧化带时，裂隙较发育。现场施工时，32 碳及下部砂岩进入矿井 24-25-10、30-8、j5-5 孔时，在 32 煤上、下砂岩发生漏水。根据矿井 25-26-7、27-12、j4-5 钻孔泵送数据： $q = 0.0039 \sim 0.00815 \text{ l/sm}$ ， $k = 0.002 \sim 0.1114 \text{ m/d}$ 。该含水层整体富水性较弱，但少部分区域有一定富水性。

主采煤层 71 层和 82 层的直接顶板主要为砂岩。地下含水层总厚度 11~65 m，平均 30 m。现场施工期间，30-6、30-3，28-5、24-12、29-3、29-6 孔有漏水现象。此处砂岩裂隙发育比较好，但是发育不全面且不均匀，各段存在较大的发育差距，富水性基本都没有其他段强。根据矿区已知钻孔，如 36-2、30-11 进行抽水实验测得该， $q$  最低为 0.0042 L/sm，最大为 0.0755 L/sm， $k$  最大为 0.0754 m/d，属于弱~中等含水层。因此，煤层间顶底板砂岩裂隙含水层为可采煤层的直接充水含水层。

### 2) 间接充水含水层

#### a) 第一含水层

底板埋深在 20 m 至 41.60 m 之间，一般约 31 m，含水砂总厚度 6.60~31.00 m，一般 15~20 m。通常在垂直深度约 20 m 处发育，厚度约 1~1.5 m，含化石或贝壳碎片，表面附近 0.50 m 为棕黄色。根据附近水井的抽水试验数据， $q = 0.5387 \text{ L/sm}$ ， $K$  最大为 2.9094 m/d，最小为 3.11 m/d，它是一个中等富水含水层。

#### b) 第二含水层

底板埋深 78.10~100.70 m，一般为 86~88 m，含水砂层厚度 5.45~39.30 m，一般 10~25 m，岩性以褐黄色为主，浅肉质、细砂、砂质，夹 2~3 层粘土或砂岩。砂层结构较为松散，水位的变化基本上与第一含水层的升降同步，滞后于第一含水层。

#### c) 第三含水层

底板埋深 159.80~218.50 m，一般为 185~200 m，含水层砂层厚度 19.20~99.20 m，一般 50~70 m。岩性主要为褐红色、粉砂、泥质砂，夹 3~5 层褐黄色。上部为 1~3 层透镜状灰岩，厚度 1~3 m，坚硬，局部分辨率较高，上部砂层厚度大，富水性强。下部砂层不发育。单层厚度小，质量不干净，泥浆质量增加，富水性差。

#### d) 第四含水层

底板埋深最小为 189.25 m, 最大为 453 m, 一般集中在 285 m、300 m、350 m 附近。含水砂层厚度集中在 0 至 59 m 之间。谷口冲洪积扇北部为粗粒相, 南部为细粒相。粗粒相砾石的粒径多数  $\geq 3$  mm, 一般夹粘土、砂质粘土 0~6 层, 夹层总厚 0~5 m。细粒相砾石粒径多数  $< 3$  mm, 一般夹粘土。据抽水试验资料, 水位标高最低为 16.77 m, 最高位 69.50 m,  $q$  最小为 0.0258 L/sm, 最大为 0.4318 L/sm,  $K$  最小为 0.079 m/d, 最大为 3.282 m/d, 含水层富水性比较低。矿化度最小为 1.458 g/L, 最大为 1.582 g/L, 全硬度最小为 31.52 德国度, 最大为 44.15 德国度, 水质为  $SO_4 \cdot Cl \cdot Na \cdot Ca \cdot Mg$  型。

#### e) 太原组石灰岩岩溶裂隙含水层(段)

太原组地层总厚度 192.81 m, 岩溶裂隙发育程度决定含水性强弱, 岩溶柱发育程度不一, 导致其含水性也有很大的差别。第一、二层石灰石厚度小, 质量纯; 第三层和第四层石灰岩较厚。1~4 号灰岩平均总厚度 39.27 m, 1~4 号灰岩岩溶柱发育, 水动力条件好, 含水率高。各断面距离仅几米, 石灰岩裂隙在平坦部位大, 随着向深部发展而逐渐减弱。富水性在裂隙发育时呈现较强性质, 反之富水性较弱。25-26-2、26-27-6、27-11、VII-3 和 J1-1 钻孔出现严重泄漏。对 25-26-2 井太原组 1~4 煤灰进行了抽水试验, 静水位标高 8.09~19.60 m,  $Q = 0.3798$  L/sm,  $K = 34.223$  m/d, 为中等富水含水层。5~14 煤灰埋深远离主煤层, 煤层发育差, 水动力条件相对较差。

但在受扰动影响的局部地段, 地下水管厚度较低, 岩性破碎, 降低了地下水导体的性能, 并且随着从断面采掘深度的增加, 矿井发生突水的情况非常严重, 因此可能导致 10 号煤夜间与灰分含量“对应”接触。太原市 1~2 号粉煤灰厚度较小, 难以形成大型地下水储运系统, 3~4 号灰岩为中厚灰岩, 岩溶裂隙发育, 储水疏水能力强。它是台州盆地最重要的地下水储存和转运系统, 是预防和控制水害的中心目标。

#### f) 奥灰水

区域厚度 500 多米, 本矿 26-27-6 孔揭露, 厚度 10.33 m。根据区内相关水文地质资料, 奥陶系灰岩中岩溶柱发育极不均匀。由于奥陶系灰岩的导水体距离开采煤层较远, 正常情况下对矿井无直接影响, 但是, 不排除在导水断层或波浪拉伸技术中的岩溶塌陷情况下发生水破裂的可能性。奥灰水对矿井生产的影响主要取决于奥陶系灰岩、太原灰岩与煤系砂岩水管之间是否存在水力联系。该水力连接的通道主要为引水或引水岩溶柱。

### 3. 理论方法

#### 3.1. 主成分分析法

主成分分析算法是经典的降维方法, 作为一种统计方法, 主成分分析旨在通过正交变换将一组潜在相关变量转换为线性不相关的新变量; 通过变换得到的新变量称为主成分, 它们能够在表达信息方面保持原始信息不变。基于 PCA 的数据处理在有效消除高维数据相关性、实现数据降维、简化数据结构等方面发挥了重要作用[8]。主成分分析的数学模型如下所示[9]。

原始数据矩阵  $X(X_1, X_2, \dots, X_p)$  的  $P$  变量形成一个线性组合, 表示为  $Y = AX$ , 即,

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

式中:  $Y_1, Y_2, \dots, Y_p$  表示各个变量的主成分值;  $X_1, X_2, \dots, X_p$  表示原始数据矩阵  $X$  的一个线性组合;  $a_{pp}$  表

示各个变量之间的协方差矩阵  $A$ 。

主成分分析法一般步骤如下：

1) 将原始数据变量先标准化，再计算各个变量之间的协方差矩阵  $A$ ；

2) 求出的协方差矩阵的特征向量可以排序为  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ ，对应的单位特征向量为  $T_1, T_2, \dots, T_p$ 。在转换矩阵  $A = T$  的情况下， $A$  的第  $i$  行代表第  $i$  个位置的特征值  $T_i$  总和；并且第  $i$  个主成分  $Y_i$  的方差也只是第  $i$  个位置的特征值  $\lambda_i$  总和；

3) 第  $m$  个主成分  $Y_m$  的方差贡献率为  $\xi_m = \lambda_m / \sum_{m=1}^s \lambda_m = 1$ 。若取  $n (n < s)$  个主成分，主成分累计贡献率为  $\xi_n = \sum_{m=1}^n \lambda_m / \sum_{m=1}^s \lambda_m$ 。

4) 主要成分数量的确定一般取决于累积方差贡献率。通常，累积方差贡献率达到至少 80% 的事实表明可以满足以下要求：提取的前  $m$  个主成分的样本信息包含大部分关于初级样本的信息。

### 3.2. Bayes 判别分析

贝叶斯判别有大量的统计理论予以支持，主要依据贝叶斯概率法则，其主要研究对象为多元分布的样本集，所得到的样本后验概率来源于多元正态分布的概率密度中包含的信息。从  $G$  个样本集中抽取了  $n$  个样本，这  $n$  个样本有  $p$  个变量，因此可以假设存在一个  $p$  维空间，这  $n$  个样本就可以看成空间中存在的  $n$  个离散的点。在判别过程中，划分错误难以避免，一旦划分出错便有产生损失量，如果将某一未知样本  $X = (x_1, x_2, \dots, x_p)$  划归为任意母体都会产生损失，但是未知样本被划归为  $A_g$  母体时，产生的损失量最小，那么将这个未知样本划归为  $A_g$  母体。

如果已知  $g$  个母体的概率分布为  $f_g(x)$ ，某一未知样本应该属于母体  $A_g$ ，设将这个样本错误的划归为母体  $A_h$  的概率是： $P\{h/g\}$ ，那么下式[10]成立：

$$P(h, g) = \int_{Rh} f_g(x) dx \quad (2)$$

式中： $P(h/g)$  表示样本错误划归的概率； $f_g(x)$  为  $g$  个母体的概率分布函数； $Rh$  表示求积分下限。

为了将样本被错误划归时所产生的损失量降到最低，Bayes 推导出划分空间  $\{R\}$  需要满足的条件为：

$$\{R_g\}: \sum_{g=1, g \neq h}^G L(h/g) q_g f_g(x) > \sum_{g=1, g \neq h}^G L(h/g) q_h f_h(x) \quad (3)$$

式中： $L(g/h)$  为将样本错误的划归为母体  $A_g$  时所产生的损失记为； $q_g f_g(x)$  为将样本错误的划归为母体  $A_g$  时的后验概率； $q_h f_h(x)$  为将样本错误的划归为母体  $A_g$  时的先验概率。

换言之，如果将未知样本划归为母体  $A_g$  的后验概率比未知样本被划归为其他母体的后验概率都要大，就将这个未知样本划归为母体  $A_g$ 。 $q_g f_g(x)$  最大就相当于后验概率最大。那么可以推导出在判别任意一个样本  $X = (x_1, x_2, \dots, x_p)$  属于某一母体时的判别函数， $G$  个判别函数表达式[11] [12]如下：

$$q_g f_g(x) = q_g (2\pi)^{-p/2} |\Sigma^{-1}|^{1/2} \exp \left[ -\frac{1}{2} (x - a_g)' \Sigma^{-1} (x - a_g) \right] \quad (4)$$

式中： $x = (x_1, x_2, \dots, x_p)'$ ；参数  $a_g$  表示  $A_g$  的均值； $\Sigma$  为一协方差矩阵； $g = 1, 2, \dots, G$ ；

依据 Bayes 准则，进一步计算可知，正态母体多类线性判别函数表达式如下：

$$Z(x) = b_{0g} + b_{1g} X_1 + \dots + b_{pg} X_p \quad (5)$$

式中： $X_p$  为样本集中的第  $p$  个指标的数值； $b_{pg}$  为判别系数。将最终结果  $Z_1(x), Z_2(x), \dots, Z_g(x)$  的数值大小进行比较，最大值所对应的母体就是待判样本  $X = (x_1, x_2, \dots, x_p)$  所属的母体。

### 3.3. PCA-Bayes 综合判别步骤

- 1) 对原始数据进行标准化处理, 排除数量级和量纲对结果造成的影响;
- 2) 计算各标量之间的协方差矩阵及相应特征向量与特征值;
- 3) 计算第  $k$  个主成分的方差贡献率 ( $k = 1, 2, \dots, n$ );
- 4) 按照累积方差贡献率  $> 80\%$  或特征值大于 1 的原则选取主成分;
- 5) 通过主成分得分系数矩阵, 由  $Y = AX$  线性组合得到主成分值;
- 6) 通过建立 Bayes 判别法, 将得出的主成分值代入线性判别函数;
- 7) 将最终结果的数值大小进行比较, 最大值所对应的母体就是待判样本所属的母体。

### 3.4. PCA-Bayes 综合判别影响因素

在主成分分析中, 我们首先应保证所提取的前几个主成分的累计贡献率达到一个较高的水平(即变量降维后的信息量须保持在一个较高水平上), 其次对这些被提取的主成分必须都能够给出符合实际背景和意义的解释(否则主成分将空有信息量而无实际含义)。

主成分的解释其含义一般多少带有点模糊性, 不像原始变量的含义那么清楚、确切, 这是变量降维过程中不得不付出的代价。因此, 提取的主成分个数  $m$  通常应明显小于原始变量个数  $p$  (除非  $p$  本身较小), 否则维数降低的优势可能抵不过主成分含义不如原始变量清楚的弊端。

在主成分分析和贝叶斯综合判别法识别突水水源类型时, 需要大量的基础数据进行训练, 从而提高模型判别的精度。

## 4. 突水水源判别模型的建立

### 4.1. 取样和测试

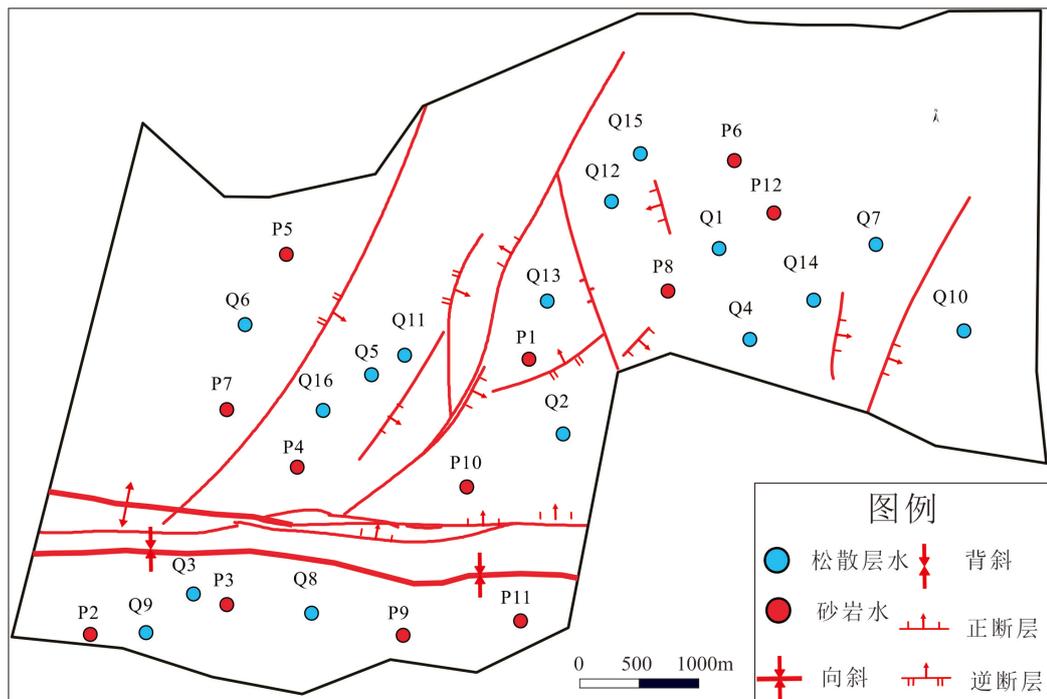


Figure 1. Location of sampling points

图 1. 取样点位置图

祁东煤矿水文地质条件复杂, 松散层突水事故频发, 主要充水含水层包括新生界水、煤系砂岩裂隙水、老空水等, 因此需要对这些影响矿井安全较大的含水层进行重点研究。于 2020 年 9 月在对祁东煤矿进行实地勘察选取其中 28 个水样(图 1), 其中新生界松散含水层(I)样本 12 个, 二叠系煤系砂岩裂隙含水层(II)样本 16 个, 取样点分布情况详见图 1。取样后主要对  $\text{Na}^+ + \text{K}^+$ 、 $\text{Ca}^{2+}$ 、 $\text{Mg}^{2+}$ 、 $\text{HCO}_3^-$ 、 $\text{Cl}^-$  和  $\text{SO}_4^{2-}$  六个参数进行测试, 取样时应携带便携式水质分析仪对 PH 和水温等参数进行测试, 所有水样均通过现场  $0.45 \mu\text{m}$  膜过滤, 用于阳离子分析的水样应用硝酸酸化至  $\text{PH} < 2$ , 采用电感耦合等离子体原子发射光谱法(ICP-AES)测定主要阳离子的浓度, 准确度为 3%。通过离子色谱法(DX-120, Dionex)测定氯化物和硫酸盐, 分析准确度小于 5%, 通过现场滴定法测定碳酸氢根离子浓度, 所有样品的离子电荷平衡小于 5%, 测试结果具体见表 1。

**Table 1.** Training samples of Qidong mine (unit: mg/L)

**表 1.** 祁东矿训练样本表(单位: mg/L)

编号	$\text{Ca}^{2+}$	$\text{Mg}^{2+}$	$\text{Na}^+ + \text{K}^+$	$\text{HCO}_3^-$	$\text{Cl}^-$	$\text{SO}_4^{2-}$	水样类型
1	23.01	14.99	341.07	535.8	179.04	83.55	I
2	19.94	12.57	299.71	564.51	115.22	32.52	I
3	7.18	9.67	784.37	1373.00	221.58	7.41	I
4	15.15	8.71	404.46	803.71	163.08	12.76	I
5	32.7	12.57	514.37	765.44	219.81	301.70	I
6	15.15	5.80	855.07	1911.2	242.85	11.52	I
7	68.11	30.93	408.32	525.10	230.44	412.42	I
8	27.92	19.83	307.60	526.24	189.67	67.50	I
9	10.37	13.54	967.36	2191.06	246.40	6.17	I
10	26.32	19.35	813.07	1863.36	248.17	34.57	I
11	35.09	17.41	287.16	583.64	189.67	9.88	I
12	15.15	13.54	619.83	1353.86	171.95	13.99	I
13	24.73	10.16	337.64	538.20	193.22	88.91	I
14	35.89	8.22	299.41	569.29	177.27	32.93	I
15	55.49	18.96	439.71	547.72	235.76	375.38	I
16	5.58	11.61	548.09	961.58	241.08	86.44	I
17	44.17	32.28	402.99	547.35	253.45	260.47	II
18	115.87	33.78	288.50	547.53	236.69	262.20	II
19	99.14	40.20	260.89	335.92	234.72	280.69	II
20	98.10	73.80	316.80	466.44	204.92	543.31	II
21	120.91	72.26	260.80	338.71	230.44	540.84	II
22	92.52	53.20	363.58	568.34	22.64	442.06	II
23	100.50	78.54	303.67	356.41	212.01	572.12	II

Continued

24	53.87	30.17	316.43	495.14	230.44	461.40	II
25	112.46	62.87	335.18	432.95	221.58	576.65	II
26	100.50	61.42	337.71	456.87	214.49	1524.00	II
27	109.27	29.99	389.23	552.55	239.31	433.83	II
28	98.10	49.82	365.61	485.57	242.85	484.04	II

#### 4.2. 训练样本主成分分析

首先采用 SPSS 26 对水样数据进行异常值判断, 判断出异常值之后采用使用平均值代替异常值的方法进行处理。求出训练样本 Person 相关系数矩阵, 通过表 2 可以看到: 各离子之间是有关联的, 例如  $\text{HCO}_3^-$  和  $\text{Na}^+ + \text{K}^+$  的关联度达到了 96.8%。这表明各个指标之间有重叠信息, 不能直接使用这 6 种离子进行水源判别, 否则会形成多余或重复的信息, 使其计算量增大, 也可能降低水源判别的精度, 造成错误判别。因此, 对样本进行主成分分析法处理是有必要的[13] [14] [15] [16]。

Table 2. Pearson correlation coefficient matrix

表 2. Pearson 相关系数矩阵

指标	$\text{Ca}^{2+}$	$\text{Mg}^{2+}$	$\text{Na}^+ + \text{K}^+$	$\text{HCO}_3^-$	$\text{Cl}^-$	$\text{SO}_4^{2-}$
$\text{Ca}^{2+}$	1					
$\text{Mg}^{2+}$	<b>0.862</b>	1				
$\text{Na}^+ + \text{K}^+$	-0.564	-0.461	1			
$\text{HCO}_3^-$	-0.599	-0.503	<b>0.968</b>	1		
$\text{Cl}^-$	0.078	0.008	0.275	0.2	1	
$\text{SO}_4^{2-}$	<b>0.718</b>	<b>0.751</b>	-0.387	-0.471	0.092	1

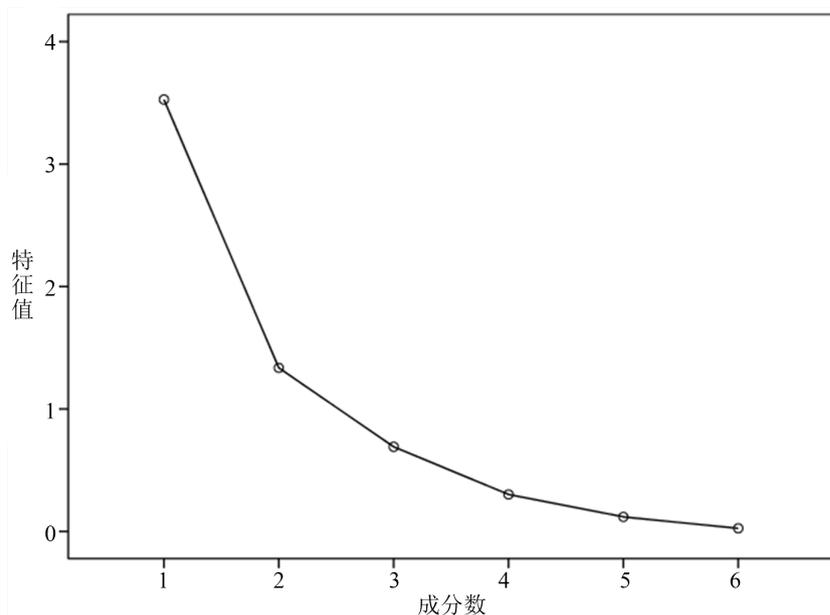
注: 黑体部分呈显著相关性。

对训练样本做主成分分析, 可以得到主成分解释方差率表(表 3)以及主成分分析碎石图(图 2), 通过表 3 可以得出训练样本前两个主成分累计贡献率约占总成分方差累计率的 81%, 可以对样本总体有效的描述。因此, 选前两个成分为主成分。

Table 3. Variance rate of principal component interpretation

表 3. 主成分解释方差率表

主成分因子	特征值	贡献率/%	累积贡献率/%
1	3.527	58.784	58.784
2	1.336	22.261	81.045
3	0.691	11.522	92.567
4	0.302	5.034	97.601
5	0.119	1.980	99.581
6	0.025	0.419	100



**Figure 2.** Composition of crushed stone  
**图 2.** 成分碎石图

通过主成分得分系数矩阵(见表 4)再结合公式(1)可以得出两个主成分  $F_1$  和  $F_2$  的线性函数模型, 见公式(6), 进而计算出主成分值。

$$\begin{cases} F_1 = 0.317X_1 + 0.319X_2 - 0.046X_3 - 0.084X_4 + 0.242X_5 + 0.325X_6 \\ F_2 = 0.069X_1 + 0.098X_2 + 0.424X_3 + 0.37X_4 + 0.549X_5 + 0.153X_6 \end{cases} \quad (6)$$

**Table 4.** Principal component score coefficient matrix  
**表 4.** 主成分得分系数矩阵

离子指标	$F_1$	$F_2$
$\text{Ca}^{2+}$	0.317	0.069
$\text{Mg}^{2+}$	0.319	0.098
$\text{Na}^+ + \text{K}^+$	-0.046	0.424
$\text{HCO}_3^-$	-0.084	0.37
$\text{Cl}^-$	0.242	0.549
$\text{SO}_4^{2-}$	0.325	0.153

## 5. 突水水源判别

### 5.1. 水源判别模型的建立

如不使用主成分分析进行 Bayes 判别建立判别模型, 由公式(1)可得:

$$\begin{cases} Y_1 = 0.086X_1 + 0.102X_2 + 0.042X_3 - 0.012X_4 + 0.064X_5 - 0.005X_6 - 13.955 \\ Y_2 = 0.217X_1 + 0.228X_2 + 0.037X_3 - 0.01X_4 + 0.072X_5 - 0.004X_6 - 20.217 \end{cases} \quad (7)$$

式中,  $Y_1$ ,  $Y_2$  分别为中新生界松散含水层和二叠系煤系砂岩裂隙含水层的判别函数值。明显数据较多,

容易出错且正确度不高。

如使用主成分分析, 用已经得出的两个主成分, 通过建立 Bayes 判别法, 根据公式(5)可得两个含水层水样的判别模型, 数学函数如公式(8)。

$$\begin{cases} Z_1 = 0.013F_1 + 0.014F_2 - 5.704 \\ Z_2 = 0.036F_1 + 0.015F_2 - 8.451 \end{cases} \quad (8)$$

式中,  $Z_1$ ,  $Z_2$  分别为中新生界松散含水层和二叠系煤系砂岩裂隙含水层的判别函数值。判别方法为哪一含水层的判别函数值大, 则此水样即为此含水层。

## 5.2. 水源判别模型效果验证

将 28 组训练样本数据代入到判别公式(5)中得出通过主成分分析后的 Bayes 判别结果与实际水源类型对比情况见表 5。

**Table 5.** Discrimination results  
**表 5.** 判别结果表

序号	$Z_1$	$Z_2$	实际结果	主成分 Bayes 判别结果	Bayes 判别结果
1	0.97808787	-0.80911686	I	I	I
2	-0.17050465	-2.7964362	I	I	I
3	6.63301848	2.78032833	I	I	I
4	1.71343798	-1.27455435	I	I	I
5	4.70845033	4.41935817	I	I	I
6	9.48632655	4.93944708	I	I	I
7	4.26140342	5.72165685	I	II	II
8	0.82498667	-0.91136931	I	I	I
9	11.2525896	6.2006415	I	I	I
10	9.40255046	5.34833772	I	I	I
11	0.60678042	-1.61102226	I	I	I
12	5.24266426	1.35832659	I	I	I
13	1.13940748	-0.5450265	I	I	I
14	0.57930066	-1.58629407	I	I	I
15	4.21296599	5.18609625	I	II	I
16	4.40684655	2.0713272	I	I	I
17	3.49135642	3.73570497	II	II	I
18	3.08343163	3.84965922	II	II	II
19	2.11770103	3.28556325	II	II	II
20	4.47917023	7.47091323	II	II	II
21	4.02632335	7.54672995	II	II	II
22	2.3842565	3.10596618	II	II	II

Continued

23	4.25788066	7.74583887	II	II	II
24	3.88364046	5.71211211	II	II	II
25	4.8458391	8.25885903	II	II	II
26	10.8432647	21.30154167	II	II	II
27	4.70830697	6.65129079	II	II	II
28	4.7190615	7.25151699	II	II	II

由表 5 可知,未使用主成分分析的 Bayes 判别结果中,I 类水样 16 个,判断正确 15 个,错误率 6.25%, II 类水样 12 个,判断正确 10 个,错误率 16.67%,综合错误率 10.71%;使用主成分分析的 Bayes 判别结果中 I 类水样 16 个,判断正确 14 个,错误率 12.50%, II 类水样 12 个,判断正确 12 个,错误率 0%,综合错误率 7.14%。PCA-Bayes 判别模型比传统 Bayes 判别模型正确率高,主成分可以减少冗余信息,方法高效,可用来实际判别突水水源问题。

### 5.3. 模型应用

利用已经建立的基于主成分分析的 Bayes 判别模型,以潘二矿新生界松散层水 I、煤系砂岩水 II、太原组灰岩水 III 三个类别突水水源的 11 个水样进行判别,判别结果见表 6。

Table 6. Results of judgment

表 6. 判别结果

编号	Ca <sup>2+</sup>	Mg <sup>2+</sup>	Na <sup>+</sup> + K <sup>+</sup>	HCO <sub>3</sub> <sup>-</sup>	Cl <sup>-</sup>	SO <sub>4</sub> <sup>2-</sup>	实际类别	PCA-Bayes 判别
1	56.12	18.68	89.65	482.36	8.69	6.68	I	I
2	45.78	17.5	106.53	481.33	12.36	12.89	I	I
3	46.3	20.78	88.78	463.56	11.05	8.36	I	I
4	61.42	34.42	738.51	263.68	934.3	386.25	I	I
5	37.88	20.73	825.33	389.98	967.36	331.29	I	I
6	2.5	2.63	232.06	524.35	51.25	36.42	II	II
7	7.32	4.21	708.56	908.78	584.67	8.06	II	III
8	7.07	3.07	707.68	786.38	621.85	0.35	II	II
9	19.31	5.12	1433.65	1432.12	758.63	210.37	III	III
10	110.32	2.33	1226.56	1388.59	733.65	148.35	III	III
11	6.89	3.26	1233.45	1756.64	943.14	2.89	III	III

Table 7. Judgment results of original index data

表 7. 原指标数据回判结果

	水样类型	I	II	III	总计
计数	I	5	0	0	5
	II	0	2	1	3
	III	0	0	3	3

Continued

准确率 (%)	I	100	0	0	100.0
	II	0	66.67	33.33	100.0
	III	0	0	100	100.0

从表 7 的判别结果可以得出, 在 5 个 I 类水样中, 5 个水样判别正确, 正确率为 100%; 3 个 II 类水样中, 2 个水样判别正确, 正确率为 66.67%; 3 个 III 类水样中, 3 个水样判断正确, 正确率为 100%; 结合三类水样, 综合判别正确率为 90.91%。结果表明本文的方法比单一 Bayes 判别更加准确, 极大的消除了样本之间的影响。

## 6. 结论

1) 结合主成分分析与 Bayes 判别法, 基于祁东矿地下水水样数据, 建立了突水水源判别模型。采用主成分分析法提取水化学指标的主成分因子, 用于压缩和对原始数据信息进行降维, 简化模型, 提高计算精度和效率, 避免信息冗余带来的误差。贝叶斯判别建立的判别函数模型简单, 只需将待测水样的指标浓度带入计算。将计算结果进行对比, 哪个数据最大即属于哪一类型, 因此可以较快地判别出突水水源类型。

2) 根据矿井突水水源的特点, 分别采用基于 PCA-Bayes 的突水水源判别模型和单独 Bayes 判别模型对比的方法, 以祁东煤矿的 28 组突水水样作为训练样本以及潘二矿的 11 组突水水样作为检验样本进行突水水源进行识别, 其中基于主成分分析的 Bayes 判别准确率分别为 92.86% 和 90.91%。结果表明, 采用主成分分析法对数据进行处理后, 矿井突水水源的识别精度比单纯采用贝叶斯判别法有较大提高。

3) 在本研究中, 当水样数据运用到识别模型时, 由于训练样本数量相对不足, 出现了一些误判。为了提高模型的预测精度, 在后续研究中, 应大量收集水化学数据, 建立矿井水化学综合数据库, 加强模型的训练, 提高识别模型的识别精度。

## 参考文献

- [1] 袁亮. 我国深部煤与瓦斯共采战略思考[J]. 煤炭学报, 2016, 41(1): 1-6.
- [2] Gui, H., Song, X. and Lin, M. (2017) Water-Inrush Mechanism Research Mining above Karst Confined Aquifer and Applications in North China Coalmines. *Arabian Journal of Geosciences*, **10**, Article No. 180. <https://doi.org/10.1007/s12517-017-2965-5>
- [3] Kalantari, H. and Ali Ghoreishi-Madiseh, S. (2019) Study of Mine Exhaust Heat Recovery System with Coupled Heat Exchangers. *Energy Procedia*, **158**, 3976-3981. <https://doi.org/10.1016/j.egypro.2019.01.844>
- [4] 徐星, 孙光中, 田坤云. GA-BP 神经网络在煤矿突水水源判别中的应用[J]. 煤炭技术, 2018, 37(10): 172-174.
- [5] 郭中安, 吴洪斌, 姜海滨, 李相通, 张海龙. 主成分分析与遗传算法优化 ELM 模型的矿井突水水源识别方法[J]. 中国科技论文, 2021, 16(9): 1010-1016+1022.
- [6] 张妹, 刘启蒙, 张宇通. 基于 PCA 水水源 Fisher 判别模型[J]. 煤炭技术, 2018, 37(3): 172-174.
- [7] 曲兴玥, 施龙青. 基于 Matlab 因子分析及距离判别模型的矿井突水水源识别[J]. 煤炭科学技术, 2018, 46(8):178-182
- [8] Farnham, I.M., Stetzenbach, K.J., Singh, A.K. and Johannesson, K.H. (2000) Deciphering Groundwater Flow Systems in Oasis Valley, Nevada, Using Trace Element Chemistry, Multivariate Statistics, and Geographical Information System. *Mathematical Geology*, **32**, 943-968. <https://doi.org/10.1023/A:1007522519268>
- [9] 胡友彪, 邢世平, 张淑莹. 基于可拓模型判别矿井突水水源[J]. 安徽理工大学学报(自然科学版), 2017, 37(6): 34-40.
- [10] Uddin, M.P., Mamun, M.A. and Hossain, M.A. (2019) Effective Feature Extraction through Segmentation-Based Folded-PCA for Hyperspectral Image Classification. *International Journal of Remote Sensing*, **40**, 7190-7220.

---

<https://doi.org/10.1080/01431161.2019.1601284>

- [11] 向晓蕊. 兴隆庄矿井水化学特征演化及识别模式研究[D]: [硕士学位论文]. 廊坊: 华北科技学院, 2016.
- [12] 徐星, 李垣志, 张文勇, 曾珠. MPSO-BP 模型在矿井突水水源判别中的应用[J]. 自然灾害学报, 2017, 26(5): 140-148.
- [13] 李小胜, 陈珍珍. 如何正确应用 SPSS 软件做主成分分析[J]. 统计研究, 2010, 27(8): 105-108.
- [14] 张好, 姚多喜, 鲁海峰, 朱宁宁, 薛凉. 主成分分析与 Bayes 判别法在突水水源判别中的应用[J]. 煤田地质与勘探, 2017, 45(5): 87-93.
- [15] 刘向远, 刘丹, 刘建. 灰色局势决策法在地下水水质评价中的应用[J]. 工业安全与环保, 2007, 33(1): 45-47.
- [16] 曹丽文, 吴圣林, 于宗仁, 王忠胜. 基于灰色局势决策理论的工程投标决策方法[J]. 华中科技大学学报(自然科学版). 2004, 32(7): 27-30.