

第一次实验

1 实验要求

实验2 对11个辉石的样品的分析得到其中几种主要氧化物的百分含量如表所示。进行主成分分析。

表1 氧化物成分表

| 样号 | SiO2(z_1) | FeO(z_2) | MgO(z_3) | CaO(z_4) |
|----|-----------|----------|----------|----------|
| 1 | 2.11 | -1.33 | 2.33 | -1.19 |
| 2 | -0.35 | 0.63 | 0.15 | -1.04 |
| 3 | -1.45 | 2.07 | -1.14 | -1.04 |
| 4 | 1.24 | -1.49 | 0.31 | 1.40 |
| 5 | -0.74 | 0.48 | -1.40 | 1.41 |
| 6 | 0.56 | -1.17 | 0.22 | 1.22 |
| 7 | -1.10 | 0.08 | -0.73 | 0.93 |
| 8 | -0.30 | -0.14 | 0.21 | 0.01 |
| 9 | -0.40 | 0.44 | -0.77 | -0.23 |
| 10 | 0.63 | -0.25 | 0.97 | -0.72 |
| 11 | -0.18 | 0.67 | -0.15 | -0.75 |

2 实验步骤

首先写出表1数据的矩阵形式，记为X。

$$X = \begin{pmatrix} 57.75 & 3.65 & 36.50 & 0.23 \\ 49.97 & 28.64 & 15.75 & 1.44 \\ 46.48 & 47.21 & 3.53 & 1.45 \\ 54.99 & 1.59 & 17.27 & 21.45 \\ 48.72 & 26.85 & 1.07 & 21.47 \\ 52.85 & 5.71 & 16.48 & 19.94 \\ 47.58 & 21.73 & 7.42 & 17.60 \\ 50.11 & 18.90 & 16.33 & 9.99 \\ 49.79 & 26.33 & 6.99 & 8.10 \\ 53.06 & 17.52 & 23.61 & 4.08 \\ 50.51 & 29.21 & 12.89 & 3.86 \end{pmatrix}$$

(1)

2.1 数据规范化

数据规范化会使每个变量的能量规范化为1，即将每个变量的均值为0，方差为1。这一步有助于消除不同量纲和量级的影响，使得各变量在PCA分析中具有同等的重要性。具体方法为式2。

$$z_{ij} = \frac{x_{ij} - \overline{x_j}}{\sqrt{s_j}}$$

(2)

其中，z为数据规范后的元素，xj为j列的均值，sj为j列的元素个数。经过式2处理后得到以下数据：

$$Z = \begin{pmatrix} 2.0133 & -1.2674 & 2.2228 & -1.1336 \\ -0.3328 & 0.5938 & 0.1406 & -0.9927 \\ -1.3852 & 1.9769 & -1.0857 & -0.9915 \\ 1.1810 & -1.4208 & 0.2931 & 1.3375 \\ -0.7097 & 0.4605 & -1.3325 & 1.3398 \\ 0.5357 & -1.1140 & 0.2138 & 1.1617 \\ -1.0535 & 0.0792 & -0.6953 & 0.8892 \\ -0.2906 & -0.1316 & 0.1988 & 0.0030 \\ -0.3871 & 0.4218 & -0.7385 & -0.2171 \\ 0.5990 & -0.2344 & 0.9293 & -0.6853 \\ -0.1700 & 0.6362 & -0.1464 & -0.7109 \end{pmatrix} \quad (3)$$

具体代码为：

```
% data为式1提到的数据
data_normalized = zscore(data);
```

2.2 计算协方差矩阵（相关系数矩阵R）

对协方差矩阵进行特征值分解，得到特征值和特征向量。特征值代表各主成分的方差大小，特征向量则表示主成分的方向。

其中，协方差矩阵计算方式为：

$$\text{Cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \quad (4)$$

其中， X_i 和 Y_i 为样本值即每一列的变量值。在本实验中，每两列都需要计算一个协方差。最后，根据协方差我们可以得到一个相关系数：

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} \quad (5)$$

其中， σ_x 和 σ_y 为每列变量的标准差。

计算协方差矩阵具体代码为：

```
% 计算协方差矩阵
cov_matrix = cov(data_normalized);
% disp('协方差矩阵: ');
% disp(cov_matrix);
```

得到的协方差矩阵为：

$$Z' = \begin{pmatrix} 1.0000 & -0.8546 & 0.8881 & -0.0471 \\ -0.8546 & 1.0000 & -0.6852 & -0.4004 \\ 0.8881 & -0.6852 & 1.0000 & -0.3718 \\ -0.0471 & -0.4004 & -0.3718 & 1.0000 \end{pmatrix} \quad (6)$$

相关系数矩阵在下一步给出。

2.3 相关系数矩阵谱分解

对协方差矩阵进行特征值分解后可以得到特征值和特征向量。特征值代表各主成分的方差大小，特征向量则表示主成分的方向。应该直接从协方差矩阵中得出。

具体代码为：

```
% 相关系数矩阵谱分解
[eigenvectors, eigenvalues_matrix] = eig(cov_matrix);
eigenvalues = diag(eigenvalues_matrix);
[eigenvalues, idx] = sort(eigenvalues, 'descend'); %输出时进行了排序，但排序后无法知道原来的值。排序信息保存在idx，为4 3 2 1。即第一个是原来的第四个
eigenvectors = eigenvectors(:, idx);

% disp('特征值: ');
% disp(eigenvalues);
% disp('特征向量: ');
% disp(eigenvectors);
```

得到的特征向量矩阵为：

$$U = \begin{pmatrix} 0.6052 & 0.0256 & 0.7868 & -0.1181 \\ -0.5567 & 0.3674 & 0.3148 & -0.6753 \\ 0.5690 & 0.3160 & -0.5296 & -0.5440 \\ -0.0106 & -0.8744 & -0.0361 & -0.4838 \end{pmatrix} \quad (7)$$

得到的特征值为：

$$\lambda = (2.6219 \quad 1.3040 \quad 0.0625 \quad 0.0117) \quad (8)$$

实际上，根据特征值我们已经得出了主要成分，根据排序数组idx，结果为

1. $CaO(z_4)$ 2. $MgO(z_3)$ 3. $FeO(z_2)$ 4. $SiO_2(z_1)$

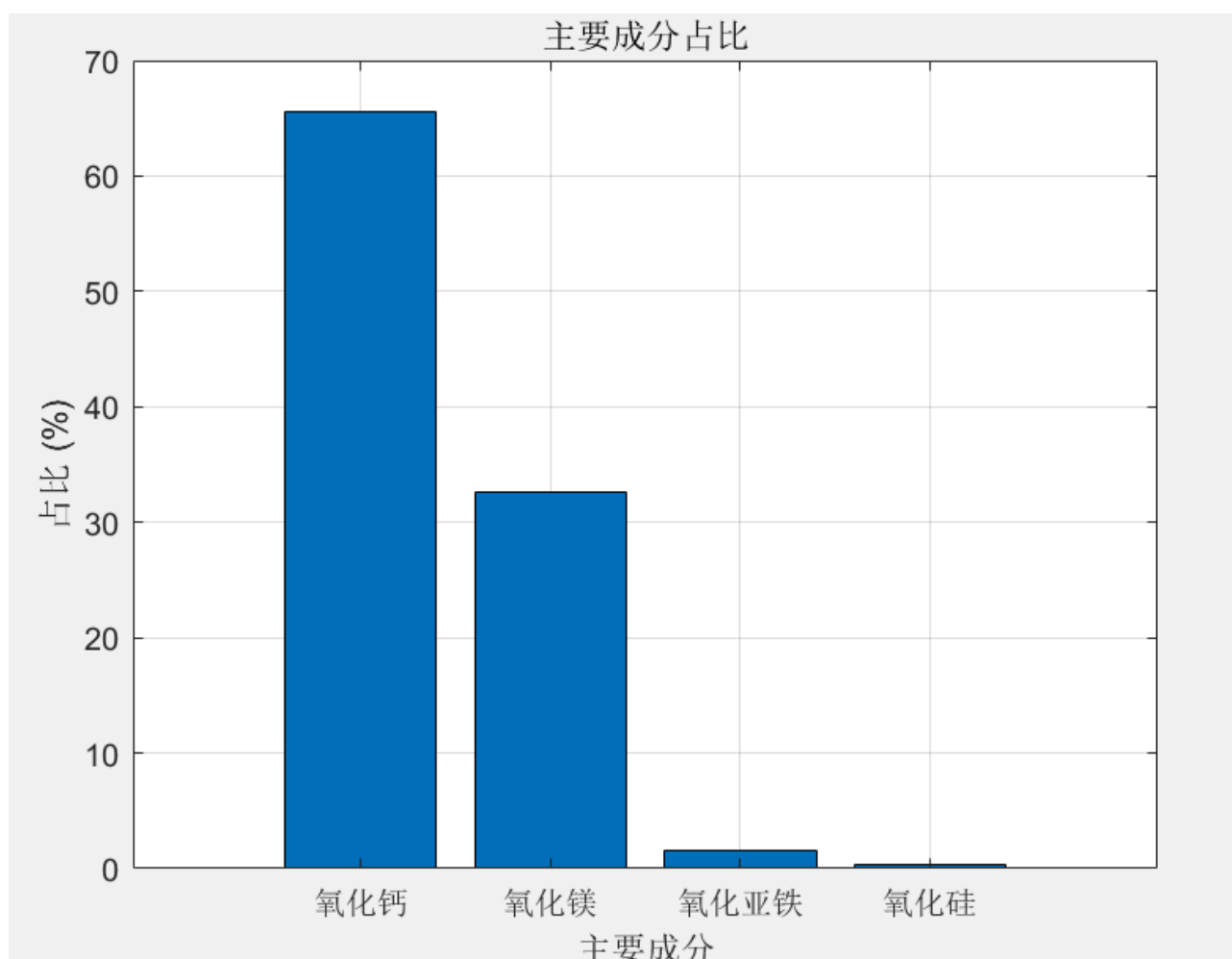


图1 成分占比

根据以上操作，我们根据11个样品的数据，得到了可信的主要成分占比。

