

X 射线荧光光谱结合判别分析识别铁矿石产地及品牌: 应用拓展

刘 曙¹, 张 博^{1, 2}, 闵 红¹, 安雅睿^{2*}, 朱志秀¹, 李 晨^{1*}

1. 上海海关工业品与原材料检测技术中心, 上海 200135
2. 上海理工大学理学院化学系, 上海 200093

摘 要 铁矿石是钢铁工业的重要原材料, 我国是铁矿石进口需求型国家, 是世界铁矿石消费第一大国。海关对进口铁矿石检验的主要目标是预防进口铁矿石中涉及安全、卫生、环保、欺诈等方面的风险。对进口铁矿石产地及品牌进行符合性验证, 可以快速筛选掺杂、掺假、以次充好, 支撑进口铁矿石的风险管理, 保障贸易便利化。在前期研究基础上进行应用拓展, 研究对象为澳大利亚、南非、巴西、哈萨克斯坦、印度 5 个国家、21 个品牌的 422 份进口铁矿石样品。考察了波长色散-X 射线荧光光谱无标样分析方法的准确度, 对于测量过程中未检出的元素含量, 选择了用检测限替代缺失值。对于测量过程中的异常值, 使用基于剩余方差的 F 检验进行异常值的剔除, 皮尔巴拉混合块、纽曼混合块铁矿、纽曼混合粉铁矿各有一组数据计算得出的 F 统计量大于 F 检验临界值($\alpha=0.01$), 因此将这 3 组数据剔除。采用逐步判别法筛选出 Fe, O, Si, Ca, Al, Mn, Ti, Mg, P, Na, Cr, K, Sr, S, Zn, V, Cu, Ba, Ni, Mo, Pb 共 21 个元素的含量作为产地识别模型的特征变量, 建立四维 Fisher 判别模型, 实现了对铁矿石产地的识别; 采用逐步判别法筛选出 Fe, O, Si, Ca, Al, Mn, Ti, Mg, P, Na, Cr, K, Sr, S, Zr, Zn, V, Cu, Ba, Cl, Ni, Mo 和 Pb 共 23 种元素含量作为品牌识别模型的特征变量, 建立二十维 Fisher 判别模型, 实现对 21 种品牌铁矿石的识别。考察了特征元素对分类识别模型的贡献, 并分析了误判品牌铁矿石的元素特征。总结出进口铁矿石产地及品牌判别分析模型的整体数据处理流程。

关键词 铁矿石; X 射线荧光光谱; 缺失值; 异常值; 判别分析

中图分类号: O657.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)01-0285-07

引 言

铁矿石是钢铁工业的重要原材料, 我国近 90% 的铁矿石依赖进口^[1], 2018 年进口量约 10.38 亿吨, 居世界第一位。澳大利亚、巴西、南非、哈萨克斯坦、印度是我国铁矿石最主要的进口国, 涉及国际大型矿业集团数十种品牌铁矿石。海关对进口铁矿石的检验包括放射性检验、外来夹杂物检测、固体废物属性鉴别、品质检验、有害元素监测等, 主要目标为预防进口铁矿石中涉及安全、卫生、环保、欺诈等方面的风险。原产地、品名等是铁矿石入境报关时的申报信息, 对进口铁矿石的产地及品牌进行符合性验证, 可以快速筛选掺杂、掺假、以次充好等现象, 支撑进口铁矿石的风险管理, 保障贸易便利化。

X 射线荧光光谱具有制样简单、无损分析、灵敏度高、

稳定性好等优点, 能够实现固体样品中主次元素的测定。判别分析是一种多变量统计分析方法, 当变量间相关系数较大时, 逐步判别分析能剔除不合适的变量, 实现有效的变量筛选。X 射线荧光光谱与判别分析相结合, 能够实现样品原产地及类别的识别, 如: 姬建飞^[2]利用波长色散-X 射线荧光光谱结合 Fisher 判别分析实现对 5 种火山岩(玄武岩、英安岩、流纹岩、粗面岩、安山岩)岩性的识别; Hondrogiannis^[3]利用波长色散-X 射线荧光光谱结合判别分析实现了对 4 种孜然产地(中国、印度、叙利亚、土耳其)的识别; Nganvongpanit^[4]利用手持式 X 射线荧光光谱结合判别分析实现了对 4 类物种(人类、大象、狗、海豚)骨骼的识别。课题组前期工作中^[5], 针对澳大利亚、南非、巴西 3 个国家、14 个品牌的 236 份进口铁矿石样品, 运用波长色散-X 射线荧光光谱无标样分析法结合逐步 Fisher 判别建立了包含 10 种有效变量的判别模型, 实现了对研究样品产地及品牌的识别。

收稿日期: 2019-12-03, 修订日期: 2020-04-27

基金项目: 国家重点研发计划项目(2018YFF0215400, 2017YFF0108905)资助

作者简介: 刘 曙, 1982 年生, 上海海关工业品与原材料检测技术中心高级工程师 e-mail: liu_shu@customs.gov.cn

* 通讯作者 e-mail: Li_chen@customs.gov.cn; anyarui@usst.edu.cn

本文是前期研究工作的应用拓展,研究对象拓展到澳大利亚、巴西、南非、哈萨克斯坦、印度 5 个国家的 21 种品牌铁矿石共 422 批进口铁矿石代表性样品,考察了波长色散-X 射线荧光光谱无标样分析方法的准确度,对比了未检出元素缺失值处理方式对模型准确度的影响,运用基于剩余方差的 F 检验进行了异常值剔除,考察了特征元素对分类识别模型的贡献,分析了误判品牌铁矿石的元素特征。总结出进口铁矿石产地及品牌判别分析模型的整体数据处理流程。

1 实验部分

1.1 样品收集

根据 GB/T 10322.1—2014《铁矿石取样和制样方法》,

从我国主要的铁矿石进口口岸采集并制备来自澳大利亚、巴西、南非、哈萨克斯坦、印度 5 个国家的 21 个品牌 422 批次进口铁矿石化学分析样品,样品容量大、种类丰富,基本包含了海关口岸日常检测中的铁矿石的主要类别。所述 21 个品牌铁矿石包括津布巴混合粉铁矿、巴西铁矿石精粉、皮尔巴拉混合块、纽曼混合块铁矿、国王粉铁矿、皮尔巴拉混合粉、澳大利亚球团矿、杨迪粉铁矿、哈杨粉铁矿、纽曼混合粉铁矿、南非铁矿石精粉、澳大利亚铁矿石精粉、弗特斯克混合粉、卡拉加斯铁矿石、哈萨克斯坦球团矿、哈萨克斯坦铁矿粉、昆巴标准矿粉、超特粉铁矿、麦克粉铁矿、昆巴标准块、印度球团矿。剔除异常值之后的样品共 419 批,信息如表 1 所示。

表 1 铁矿石样品信息

Table 1 The information of iron ore samples

中文名称	英文名称	国别	建模样品数量	测试样品数量	样品总数
津布巴混合粉铁矿	Jimblebar Blend Fine Ore	澳大利亚	9	3	12
巴西铁矿石精粉	Brazilian Iron Ore Concentrate	巴西	9	3	12
皮尔巴拉混合块	Pilbara Blend Lump	澳大利亚	27	9	36
纽曼混合块铁矿	Newman Blend Lump Ore	澳大利亚	26	8	34
国王粉铁矿	Kings Fines	澳大利亚	9	3	12
皮尔巴拉混合粉	Pilbara Blend Fines	澳大利亚	35	12	47
澳大利亚球团矿	Savage River Iron Ore Pellets	澳大利亚	9	3	12
杨迪粉铁矿	Yandi Fine Ore	澳大利亚	19	6	25
哈杨粉铁矿	HIY Fines	澳大利亚	15	6	21
纽曼混合粉铁矿	Newman Blend Fine Ore	澳大利亚	14	5	19
南非铁矿石精粉	South Africa Iron Ore Concentrate	南非	18	5	23
澳大利亚铁矿石精粉	Australian Iron Ore Concentrate	澳大利亚	18	5	23
弗特斯克混合粉	Fortescue Blend Fines	澳大利亚	24	7	31
卡拉加斯铁矿石	Carajas Iron Ore	巴西	13	4	17
哈萨克斯坦球团矿	Kazakhstan Iron Ore Pellet	哈萨克斯坦	9	3	12
哈萨克斯坦铁矿粉	Kazakhstan Iron Ore Fines	哈萨克斯坦	9	3	12
昆巴标准矿粉	Kumba Standard Fines	南非	15	4	19
超特粉铁矿	Super Special Fine	澳大利亚	9	3	12
麦克粉铁矿	MAC Fine Ore	澳大利亚	9	2	11
昆巴标准块	Kumba Standard Lump	南非	13	4	17
印度球团矿	Indian Iron Ore Pellet	印度	9	3	12

1.2 方法

将采集样品分装到玻璃广口瓶中于 105 °C 下烘干 4 h。采用压片机对烘干样品压片,压片前用乙醇清洗模具,使用聚乙烯环使粉末样品聚拢,压制样品在 30 t 压力下维持 30 ~ 60 s。检查压制样品表面均匀且无裂纹、脱落现象,测量前用洗耳球吹去样品表面浮粉。

使用德国布鲁克公司 S4 Pioneer 波长色散-X 射线荧光光谱仪中的无标样分析方法检测铁矿石中元素的含量。检测中使用铑靶光管、四个分析仪晶体(LiF200, XS-55, PET 和 Ge)、流气计数器(FC)、闪烁计数器(SC)等元件。

1.3 数据处理

1.3.1 整体数据处理流程

整体数据处理流程如图 1 所示,首先收集样品的 X 射线

荧光光谱(XRF)无标样分析数据,然后依次进行缺失值处理、异常数据剔除,对剩余数据划分训练集与测试集,采用逐步判别法提取特征变量,建立 Fisher 判别分析模型,通过建模样品验证、交叉验证、测试集样品验证评价模型的准确度。

1.3.2 异常数据剔除

分别对不同品牌铁矿石的测量数据做主成分分析,根据主成分贡献度选择合适的主成分,再根据式(1)和式(2)分别计算样品剩余方差、模型剩余方差,用样品剩余方差除以模型剩余方差得到 F 统计量[式(3)],再查询 $F_{0.01}$ 检验临界值表,通过比较判断该样品是否在 99% 置信区间内为异常值。

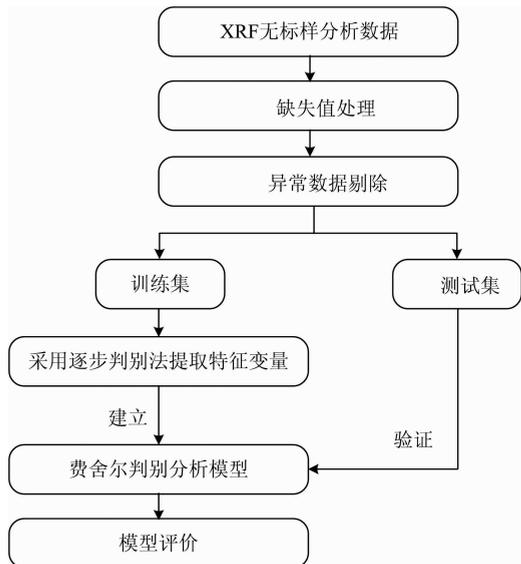


图 1 数据处理流程图
Fig. 1 Data processing flow chart

样品剩余方差公式

$$\hat{S}_i^2 = \frac{\hat{e}_i \hat{e}_i^T}{m - k} \quad (1)$$

模型剩余方差公式

$$S_0^2 = \frac{1}{n} \sum_i s_i^2 \quad (2)$$

F 统计量计算公式

$$F_i = \frac{\hat{S}_i^2}{S_0^2} \quad (3)$$

式中, \hat{e}_i 为样本第 i 行原始数据与其估计值的差 ($\hat{e}_i = x_i - \hat{x}_i$), x_i 为数据第 i 行的 26 个元素含量矩阵, \hat{x}_i 根据主成分分析计算, 其为得分矩阵 (提取主成分的得分) 与因子载荷矩阵的乘积, T 为矩阵转置符号, m 为项目数、 k 为主成分数、 n 为样品数。

1.3.3 逐步判别-费舍尔判别分析

逐步判别分析属于有监督的分类方式, 先对已知的样品进行分类来建立模型, 再对未知样品进行预测分类, 其判别函数的建立利用了方差分析的思想^[6]。判别时将一未知样品的变量代入判别式, 根据判别函数和质心处坐标函数, 计算每个样品坐标与质心的距离, 与哪个类别的质心最近, 该样品就判定为哪个类别。

在逐步判别分析中通过费舍尔分数 (F-score) 算法^[7] 进行变量评估和特征选择, F-score 值为组间均方与总平方和的比, F-score 值越大说明该变量在组间差异越大, 对判别的贡献度越大。具体描述如给定训练样本集 $X_k \in R_m$, $K = 1, 2, \dots, n$, 其中正类和负类的样本数分别为 n^+ 和 n^- , 则训练样本第 i 个特征的 F-score 值定义为 [式 (4)]

$$F\text{-score}_i = \frac{(X_i^+ - \bar{X}_i)^2 + (X_i^- - \bar{X}_i)^2}{\frac{1}{n^+ - 1} \sum_{k=1}^{n^+} (X_{k,i}^+ - \bar{X}_i^+)^2 + \frac{1}{n^- - 1} \sum_{k=1}^{n^-} (X_{k,i}^- - \bar{X}_i^-)^2} \quad (4)$$

式 (4) 中 \bar{X} , \bar{X}_i^+ 和 \bar{X}_i^- 分别为第 i 个特征在整个数据集上的平均值、在正类数据集上的平均值和在负类数据集上的平均值; $X_{k,i}^+$ 为第 i 个正类样本点的第 i 个特征的特征值; $X_{k,i}^-$ 为第 i 个负类样本点的第 i 个特征的特征值。逐步判别分析的使用可以使进行判别分析前首先剔除不重要的变量, 有利于数据分类的有效性, 选取的 F-score 值为系统默认数值, F-score 值大于 3.84 时变量进入判别模型, F-score 值小于 2.71 时变量从模型中剔除^[8]。

对于所建立的判别模型使用判别准确率 [式 (5)] 对模型进行评价。

$$\text{判别准确率} = \frac{N - E}{N} \times 100\% \quad (5)$$

式 (5) 中, N 为样品总数, E 为判别错误样品数量。

2 结果与讨论

2.1 分析方法的准确度

根据 GB/T 6379.1—2004《测量方法与结果的准确度 (正确度与精密度) 第 1 部分: 总则与定义》, 准确度是指测试结果与接受参照值之间的一致程度, 由正确度和精密度组成。正确度指由大量测试结果得到的平均数与接受参照值之间的一致程度, 精密度指在规定条件下独立测试结果间的一致程度。波长色散-X 射线荧光光谱无标样分析方法的基本思路是由仪器和软件制造商测定校准样品, 储存元素谱线强度和校准曲线, 然后将这些数据转到用户的 X 射线荧光分析系统中, 并用参考样品校正仪器的漂移, 考虑到实际样品与校准样品的基体差异, 无标样分析法只能是一种半定量方法, 准确度很难达到定量分析的要求。选取纽曼混合块铁矿、津布巴混合粉铁矿的化学分析样, 采用波长色散-X 射线荧光光谱无标样分析方法分别进行 3 次独立测量, 进一步考察了方法的精密度, 见表 2。结果表明 Fe, Si, Al, P, Mn 和 Ti 的相对标准偏差在 0.25%~2.43% 之间, S, Ca 和 Mg 三个元素含量的平均值在 0.010%~0.044% 之间, 相对标准偏差在 1.08%~9.93% 之间。除纽曼混合块铁矿中 S 和 Mg、津布巴混合粉铁矿中 Ca 外, 其余元素的相对标准偏差皆满

表 2 无标样分析方法测量结果

Table 2 Measurement results without standard analysis method

	纽曼混合块铁矿		津布巴混合粉铁矿	
	平均值 / %	相对标准偏差 / %	平均值 / %	相对标准偏差 / %
Fe	59.98	0.63	56.30	0.25
Si	2.32	1.74	3.11	1.22
Al	1.32	2.43	3.44	1.02
P	0.030	1.26	0.046	1.25
S	0.010	5.59	0.017	3.07
Mn	0.225	0.68	0.135	1.28
Ti	0.030	1.86	0.077	1.65
Ca	0.030	1.56	0.021	9.93
Mg	0.029	8.78	0.044	1.08

足 GB 27417—2017《合格评定化学分析方法确认和验证指南》对实验室内变异系数的要求,说明采用同一台仪器在固定的条件下检测,无标样分析方法的精密度基本能达到定量分析的要求。这也是 X 射线荧光光谱无标样分析方法检测结果结合判别分析能建立产地识别模型的原因。

2.2 缺失值处理

数据缺失是实验中不可避免的问题,数据缺失会使整个系统丢失一部分有价值的信息,增加了系统在使用过程中的不确定性,影响了最终输出的准确性。在大部分研究中,都将其作为无效数据删除,即使不人为删除,在诸如 SPSS, SAS, Stata 和 R 的很多统计软件的数据处理中,也会默认删除缺失值从而使整个数据处理过程能流畅进行^[9]。

针对采集的 422 个铁矿石样品,采用波长色散-X 射线荧光光谱无标样分析可以检测到 Fe, O, Si, Ca, Al, Mn, Tb, Ti, Mg, P, Na, Cr, K, Sr, S, Zr, Zn, V, Cu, Gd, Ba, Cl, Ni, Co, Mo 和 Pb 共 26 种元素的含量,其中 Na, Cr, K, Sr, S, Zr, Zn, V, Cu, Gd, Ba, Cl, Ni, Co, Mo 和 Pb 共 16 个元素含量存在未检出的情况。对于未检出的元素含量,对比了 4 种缺失值处理方式:(1)缺失值用 0 替代;(2)缺失值用检测限替代;(3)缺失值用检测限替代,并增加一组逻辑变量,元素检出标记为 1,未检出标记为 0;(4)删除存在缺失值的元素。4 种方式所建立的国家与品牌判别模型中建模验证准确率、交叉验证准确率、测试验证准确率如表 3 所示。结果表明,方式 2 与方式 3 所建立的判别模型准确率相差不大,考虑到增加逻辑变量会使数据处理过程复杂,选取用第 2 种缺失值处理方式。

表 3 四种缺失值处理方式比较

Table 3 Comparison of four missing value processing methods

		建模验证 准确率/%	交叉验证 准确率/%	测试验证 准确率/%
方式 1	国家判别模型	99.40	98.10	100
	品牌判别模型	95.60	92.80	92.10
方式 2	国家判别模型	99.10	98.40	100
	品牌判别模型	95.60	92.80	93.10
方式 3	国家判别模型	99.40	98.40	100
	品牌判别模型	96.00	92.20	93.10
方法 4	国家判别模型	94.70	93.50	87.10
	品牌判别模型	94.40	90.70	89.10

2.3 异常数据剔除

一批数据中有部分数据与其余数据相比明显不一致的称为异常值,或称离群值。实验过程中采集异常数据的可能来源包括:样品测试过程被污染、样品信息、检测结果采集错误等。异常数据会使样本均值与样本方差产生明显偏差,因此寻找合适的方法来发现和这些异常数据十分重要。本文使用 Pirouette 多元数据分析软件基于剩余方差的 F 检验进行异常值的剔除。皮尔巴拉混合块、纽曼混合块铁矿、纽曼混合粉铁矿各有一组数据计算得出的 F 统计量大于 $F_{0.01}$ 检验临界值,如表 4 所示,认为这 3 组数据为异常值,因此

将这 3 组数据剔除,剩余 419 组数据用于后续分析。

表 4 异常样品的样品剩余方差、模型剩余方差、 F 统计量、临界值、自由度

Table 4 Sample residual variance, model residual variance, F value, critical value and degree of freedom of the abnormal sample

	样品剩余方差	模型剩余方差	F 统计量	F 临界值	自由度
皮尔巴拉混合块	0.001 823 29	0.000 197 57	9.228 6	7.499	32
纽曼混合块铁矿	0.000 655 36	0.000 074 84	8.757 0	7.562	30
纽曼混合粉铁矿	0.003 226 24	0.000 314 76	10.249 9	8.683	15

2.4 特征变量筛选

分析来自我国主要铁矿石进口口岸的 21 个品牌共计 419 个铁矿石样品,建模过程中选取 318 个样品作为训练集,101 个样品作为验证集检验模型的准确性。训练样品及测试样品的选取如表 1 所示。铁矿石生产国家判别模型中,经过逐步判别分析筛选出 Ca, K, O, V, Mg, Sr, Na, Zn, Al, Ti, Ni, Pb, P, Cr, Cu, Mo, Mn, S, Ba, Fe 和 Si 共 21 个元素含量作为特征变量(F -score 值 > 3.84)保留在模型中,Zr, Tb, Cl, Gd 和 Co 等元素含量因未通过 F 检验(F -score 值 < 2.71)而从模型中剔除。铁矿石品牌的判别模型中,经过逐步判别分析筛选出 V, Ca, K, Al, O, Ti, Mn, Mo, Ni, P, Mg, Cu, Pb, Si, S, Na, Cl, Zn, Ba, Cr, Fe, Sr 和 Zr 共 23 个元素含量作为特征变量(F -score 值 > 3.84)保留在模型中,Gd 和 Co 因未通过 F 检验(F -score 值 < 2.71)而从模型中剔除。从地质成因上看,Ca, O, K 和 Na 等元素反映了海相沉积背景,V, Ti 和 Ni 等元素反映出岩浆活动、深部流体活动,筛选出的特征变量也说明地质成因差异是建立铁矿石产地和品牌判别分析模型的原因。

2.5 判别模型及评价

2.5.1 进口铁矿石生产国家判别模型

使用了 2.4 中选择的 21 个元素作为特征变量,建立澳大利亚、巴西、南非、哈萨克斯坦、印度产铁矿石的费舍尔判别模型,包括 4 个判别函数,5 个国家的组质心处的坐标可以通过计算得到。对于测试样品的预测,可以将该样品 21 个元素含量分别代入 4 个判别函数,分别计算 4 维坐标与 5 个国家组质心坐标的距离,最近距离对应的国家,即为该样品生产国家的预测结果。建模样品验证、交叉验证、测试样品验证的结果如表 5 所示,模型判别准确率分别为 99.1%, 98.4% 和 100%。使用前期工作^[5]建立的判别模型计算本实验数据,判别准确率分别为 96.3%, 94.4% 和 91.1%,可以看出增加特征变量的数量,再经过缺失值的处理与异常值的剔除,可以明显提高生产国家判别模型的准确率。不同国家铁矿石的类型、品位、成因以及分布情况存在一定的差异,如:澳大利亚铁 90% 矿石集中在皮尔巴拉地区^[10],大部分含磷低、埋藏浅、品位较高,铁含量一般在 56%~62% 左右;巴西铁矿主要由赤铁矿组成,具有高铁、中硅、低铝的特点;南非铁矿石主要分布在开普省北部赛申地区和德兰士瓦的西

部,多属于赤铁矿,品位高、杂质少,含有较高的钾、钠;哈萨克斯坦的铁矿属于富矿,其中约 60%为富矿和易选矿,铁精矿含量可达 65%左右;印度拥有丰富的铁矿资源且多为优质铁矿,主要为赤铁矿和磁铁矿,赤铁矿矿石铁品位均在 58%以上,磁铁矿矿石品位较低。图 2 为 2.4 节特征变量选择过程中 F-score 值排名前十元素含量平均值的条形图,从图 2 可以看出元素含量在不同国家间有较大的差异,因此可以利用不同国家铁矿石间元素含量的差异进行铁矿石产地国家的判别。

表 5 国家判别模型具体判别结果

Table 5 National discriminant model specific discriminant results

国家	建模样品验证/%	交叉验证	测试样品验证/%
澳大利亚	99.60	99.60	100
巴西	95.50	95.50	100
南非	100	100	100
哈萨克斯坦	100	88.90	100
印度	88.90	88.90	100
总计	99.10	98.40	100

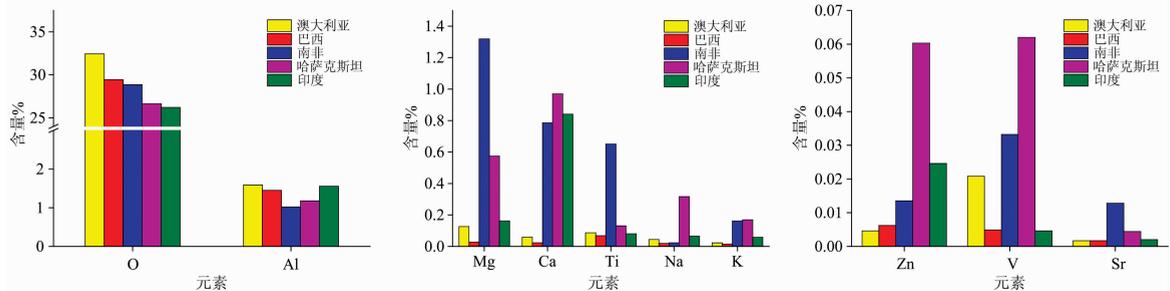


图 2 不同国家元素含量条形图

Fig. 2 Bar chart of elemental content in different countries

表 6 品牌判别模型具体判别结果

Table 6 Brand discriminant model specific discriminant results

样品名称	建模样品验证/%	交叉验证/%	测试样品验证/%	样品简称	建模样品验证/%	交叉验证/%	测试样品验证/%
津布巴混合粉铁矿	100	100	100	澳大利亚铁矿石精粉	100	94.40	100
巴西铁矿石精粉	77.80	66.70	100	弗特斯克混合粉	91.70	87.50	57.10
皮尔巴拉混合块	96.30	96.30	100	卡拉加斯铁矿石	100	100	100
纽曼混合块铁矿	92.30	88.50	100	哈萨克斯坦球团矿	100	100	100
国王粉铁矿	100	100	100	哈萨克斯坦粉铁矿	100	88.90	100
皮尔巴拉混合粉	100	94.30	100	昆巴标准粉	93.30	93.30	100
澳大利亚球团矿	100	100	100	超特粉铁矿	77.80	77.80	66.70
杨迪粉铁矿	94.70	94.70	83.30	麦克粉铁矿	100	100	100
哈杨粉铁矿	93.30	86.70	100	昆巴标准块	100	100	100
纽曼混合粉铁矿	100	100	100	印度球团矿	100	77.80	100
南非铁矿石精粉	100	100	100	总计	96.20	93.10	95.00

这些误判品牌铁矿石均来自澳大利亚皮尔巴拉克拉通的哈默斯利成矿省,成矿类型为受变质沉积改造型铁矿床^[14]。杨迪粉铁矿与哈杨粉铁矿均来自西澳皮尔巴拉地区的杨迪矿

2.5.2 进口铁矿石品牌判别模型

品牌判别模型使用 2.4 节选择的 23 个元素作为特征变量,建立了 21 种品牌铁矿石的费舍尔判别模型,包括 20 个判别函数。建模样品验证、交叉验证、测试样品验证的结果如表 6 所示,模型判别准确率分别为 96.2%, 93.1% 和 95.0%。使用前期工作^[5]建立的判别模型计算本次实验数据,模型判别准确率分别为 95.3%, 92.5% 和 91.1%,可以看出增加特征变量的数量,再经过缺失值的处理与异常值的剔除,可以提高品牌判别模型的准确率。铁矿石品牌判别模型对测试样品进行分类时,有 5 个样品判别错误,其中 1 个杨迪粉铁矿错分为哈杨粉铁矿、3 个弗特斯克混合粉错分为皮尔巴拉混合粉、1 个超特粉铁矿错分为弗特斯克混合粉。选择 2.4 节特征变量选择过程中 F-score 值排名前 10 的元素,作前 2 个主成分的散点图(如图 3 所示,横坐标为主成分 1,纵坐标为主成分 2)。从二维散点图中可以看出,分类错误的品牌与其被误判的品牌,在二维散点图中部分样品位置接近,甚至重叠,说明这些品牌中部分样品元素间含量比较接近,X 射线荧光光谱无标样分析方法所建立的判别模型对这些品牌铁矿石存在误判的可能。

山,分别由必和必拓与力拓公司开采,由于来自相同的矿山,矿床成因一致,元素含量差异很小;弗特斯克混合粉是澳大利亚 FMG 公司的产品,由位于汤姆普利斯的所罗门枢

纽生产。皮尔巴拉混合粉为澳大利亚力拓公司产品, 主要由布鲁克曼 2、布鲁克曼 4、霍普唐斯 4、汤姆普利斯、帕拉伯杜共 5 个地区的铁矿石混合而成, 其中布鲁克曼 2、布鲁克曼 4、汤姆普利斯矿区与弗特斯克混合粉产区所罗门枢纽十分接近; 超特粉铁矿与弗特斯克混合粉均为澳大利亚 FMG

公司产品, 通过调查了解到, 两种品牌铁矿石在元素含量方面差异比较接近。对于这几类品牌铁矿的识别, 存在误判的可能, 如需准确识别, 需进一步提高分析方法的准确度, 或借助其他分析技术或手段。

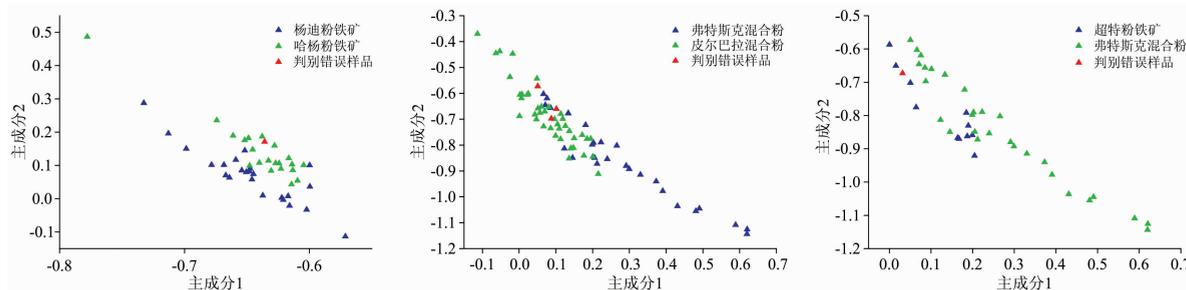


图 3 分类错误品牌与其被误判品牌散点图

Fig. 3 Classification error brand and its misjudged brand scatter plot

3 结 论

本文是前期研究的应用拓展, 研究对象为澳大利亚、巴西、南非、哈萨克斯坦、印度 5 个国家的 21 种品牌铁矿石共 422 批进口铁矿石代表性样品, 经过对比选择用检测限替代检测结果中的缺失值, 运用基于剩余方差的 F 检验剔除了 3 组异常数据, 利用逐步判别分析选择特征变量, 建立了铁矿石产地与品牌的费舍尔判别模型, 最后分析了特征元素对分类识别模型的贡献以及误判品牌铁矿石的元素特征。判别模

型的判别结果表明增加特征变量的数量, 再经过缺失值的处理与异常值的剔除, 可以提高模型判别准确率。在此基础上, 总结出进出口铁矿石产地及品牌判别分析模型的整体数据处理流程。形成固定的方法经验之后, 可以进一步增加铁矿石样品数量, 提高模型普适性, 实现对更多国家、品牌铁矿石的识别。使用 X 射线荧光无标样分析法测量元素的含量将大大缩短分析时间, 结合已建立的判别模型, 可以对铁矿石品牌及产地进行快速的溯源分类, 有利于保障贸易安全、提高通关效率。

References

- [1] ZHAO Hong-jun, CHEN Xiu-fa, HE Xue-zhou, et al(赵宏军, 陈秀法, 何学洲, 等). *Geology in China(中国地质)*, 2018, 45(5): 890.
- [2] JI Jian-fei, YUAN Sheng-bin, YANG Yi, et al(姬建飞, 袁胜斌, 杨毅, 等). *China Petroleum and Chemical Standard and Quality(中国石油和化工标准与质量)*, 2019, (3): 240.
- [3] Hondrogiannis E, Peterson K, Zapf C M, et al. *Food Chemistry*, 2012, 135(4): 2825.
- [4] Nganvongpanit K, Buddhachat K, Klinhom S, et al. *Forensic Science International*, 2016, 263: 101.
- [5] ZHANG Bo, MIN Hong, LIU Shu, et al(张博, 闵红, 刘曙, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2020, 40(8): 2640.
- [6] Granato D, Putnik P, Kovačević D. B, et al. *Comprehensive Reviews in Food Science and Food Safety*, 2018, 17(3): 663.
- [7] Sabzi S, Javadikia P, Rabani H, et al. *Measurement*, 2013, 46(9): 3333.
- [8] YAO Ting, FANG Kai-tai(尧庭, 方开泰). *Introduction to Multivariate Statistical Analysis(多元统计分析引论)*. Beijing: Science Press(北京: 科学出版社), 1982.
- [9] PAN Chuan-kuai, XIONG Wei, QI Chun-jie(潘传快, 熊巍, 祁春节). *Journal of Huazhong Agricultural University • Social Sciences Edition(华中农业大学学报 • 社会科学版)*, 2017, 127(1): 72.
- [10] HAN Ning, JIANG Si-hong, BAI Da-ming, et al(韩宁, 江思宏, 白大明, 等). *Geological Bulletin of China(地质通报)*, 2015, 34(6): 1086.
- [11] LÜ Li-na, WU Cen(吕立娜, 吴岑). *Coal Geology of China(中国煤炭地质)*, 2013, 25(12): 106.

X-Ray Fluorescence Spectroscopy Combined With Discriminant Analysis to Identify Imported Iron Ore Origin and Brand: Application Development

LIU Shu¹, ZHANG Bo^{1, 2}, MIN Hong¹, AN Ya-rui^{2*}, ZHU Zhi-xiu¹, LI Chen^{1*}

1. Technical Center for Industrial Product and Raw Material Inspection and Testing, Shanghai Customs, Shanghai 200135, China

2. Department of Chemistry, College of Science, University of Shanghai for Science and Technology, Shanghai 200093, China

Abstract Iron ore is an important raw material for the iron and steel industry. China is an iron ore import-demand country and the world's largest iron ore consumer. The main goal of the customs' inspection of imported iron ore is to prevent the risk of safety, health, environmental protection, fraud and other aspects of imported iron ore. The compliance verification of the origin and brand of imported iron ore can quickly screen the phenomena of adulteration, adulteration, and inferior charging, which support the risk management of imported iron ore and ensure trade facilitation. This article expands the application based on previous research. The research objects are 422 imported iron ore samples from 5 countries. In this paper, the accuracy of the non-standard sample analysis method of wavelength dispersive X-ray fluorescence spectrum is investigated. For the elements not detected in the measurement process, the detection limit was chosen to replace the missing values. For the outliers in the measurement process, F-test based on residual variance is used to eliminate the outliers. Each of the Pilbara Blend Lumps, Newman Blend Lumps, and Newman Blend Fines has one F statistic calculated from one set of data is greater than the F-test critical value ($\alpha=0.01$), so these three sets of data are eliminated. The contents of Fe, O, Si, Ca, Al, Mn, Ti, Mg, P, Na, Cr, K, Sr, S, Zn, V, Cu, Ba, Ni, Mo, and Pb are selected by the stepwise discriminant method as the characteristic variable of the original identification model, and a four-dimensional Fisher discriminant model is established to identify the origin of the iron ore. The contents of Fe, O, Si, Ca, Al, Mn, Ti, Mg, P, Na, Cr, K, Sr, S, Zr, Zn, V, Cu, Ba, Cl, Ni, Mo, and Pb are selected by the stepwise discrimination method as the feature variables of the brand recognition model, and a 20-dimensional Fisher discriminant model is established to realize the recognition of 21 brand iron ores. The contribution of characteristic elements to the classification and recognition model is investigated, and the element characteristics of misidentified brand iron ore are analyzed. On this basis, the paper summarizes the whole data processing flow of the discrimination analysis model of the origin and brand of imported iron ore.

Keywords Iron ore; X-ray fluorescence spectrum; Missing value; Outliers; Discriminant analysis

(Received Dec. 3, 2019; accepted Apr. 27, 2020)

* Corresponding authors