

基于机器学习的闪锌矿微量元素特征 在铅锌矿床类型识别中的应用

董赛娜¹, 王 达¹, 马国桃², 魏守才³, 代克刚⁴, 张晓暄¹, 徐大兴¹

1. 中国地质大学 地球科学与资源学院, 北京 100083;

2. 中国地质调查局成都地质调查中心 矿产资源室, 成都 610081;

3. 西藏自治区地质矿产勘查开发局 第二地质大队, 拉萨 850000;

4. 西藏自治区地质矿产勘查开发局 第五地质大队, 拉萨 850000

[摘要] 闪锌矿微量元素特征是识别铅锌矿床成因类型的重要指标。拟通过机器学习方法识别出判断铅锌矿床成因类型的关键控制元素, 建立基于闪锌矿微量元素特征的广义铅锌矿成因类型判别图。系统收集了密西西比河谷型、火山块状硫化物型、喷流-沉积型、矽卡岩型 4 种成因类型铅锌矿床中 3 700 条闪锌矿的 12 种微量元素数据 (Cd, Mn, Ag, Cu, Pb, Sn, Ga, In, Sb, Co, Ge 和 Fe), 使用支持向量机和随机森林 2 个机器学习分类模型对其进行分类, 对这些特征元素重要性进行排序。基于闪锌矿微量元素特征, 利用大数据和机器学习方法构建的铅锌矿床分类模型可以有效地区分不同成因类型的铅锌矿床, 支持向量机和随机森林分类模型在测试集上的准确率分别为 98.5% 和 96.9%。同时, 通过主成分分析对 12 种元素特征进行统计分析和降维可视化, 并结合随机森林模型特征元素重要性排序, 识别出能区分铅锌矿床 4 种成因类型的关键化学元素。结果显示, 闪锌矿的 12 种微量元素中, Mn, Ge, In, Co, Sb 和 Ga 这 6 种元素用于区分铅锌矿床成因类型效果显著。新构建了 4 种成因类型铅锌矿床闪锌矿 $\ln(\text{Mn}/\text{Cd})-\ln(\text{Ge}/\text{Cd})-\ln(\text{Co}/\text{Cd})$ 三元二维图, 以及 $\ln(\text{Mn})-\ln(\text{Sb})$ 、 $\ln(\text{Co})-\ln(\text{Ga})$ 和 $\ln(\text{Mn})-\ln(\text{In}/\text{Ge})$ 二元二维图, 可用来有效区分密西西比河谷型、火山块状硫化物型、喷流-沉积型、矽卡岩型 4 种类型铅锌矿床。

[关键词] 闪锌矿; 微量元素; 机器学习; 大数据分析; 铅锌矿床

[分类号] P618.13

[文献标志码] A

[收稿日期] 2023-09-21。

[基金项目] 北京市科学技术协会 2023—2025 青年人才托举工程 (BYESS2023411); 地质工程与矿产资源国家重点实验室开放研究课题 (GPMR202308); 中国地质大学 (北京) 新教师基本科研能力提升项目 (2-9-2020-010)。

[第一作者] 董赛娜 (1998—), 女, 硕士研究生, 研究方向: 地质工程, E-mail: dsn_cugb@163.com。

[通信作者] 王达 (1990—), 男, 副教授, 博士生导师, 研究方向: 金属稳定同位素在矿床学及找矿勘查中的应用, E-mail: WangDa900909@cugb.edu.cn。

[引用格式] 董赛娜, 王达, 马国桃, 等, 2024. 基于机器学习的闪锌矿微量元素特征在铅锌矿床类型识别中的应用 [J]. 成都理工大学学报 (自然科学版), 51(4): 614-629.

Dong S N, Wang D, Ma G T, *et al.*, 2024. Application of machine learning to predict types of Pb-Zn deposits by using trace elemental characteristics of sphalerite [J]. Journal of Chengdu University of Technology (Science & Technology Edition), 51(4): 614-629.

Application of machine learning to predict types of Pb-Zn deposits by using trace elemental characteristics of sphalerite

DONG Saina¹, WANG Da¹, MA Guotao², WEI Shoucai³, DAI Kegang⁴,
ZHANG Xiaoxuan¹, Xu Daxing¹

1. School of Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China;

2. Chengdu Center, China Geological Survey (Geosciences Innovation Center of Southwest China),
Chengdu 610081, China;

3. The Second Geological Brigade, Bureau of Geology and Mineral Exploration and Development of
Tibet Autonomous Region, Lhasa 850000, China;

4. The Fifth Geological Brigade, Bureau of Geology and Mineral Exploration and Development of
Tibet Autonomous Region, Lhasa 850000, China

Abstract: The trace elemental characteristics of sphalerite are crucial indicators for identifying the genetic types of Pb-Zn deposits. In this study, we identify the key control elements to distinguish between different genetic types of Pb-Zn deposits by using machine learning, and establish a generalized map of their identification based on the characteristics of their trace elements of sphalerite. To this end, we collected data on 12 trace elements (Cd, Mn, Ag, Cu, Pb, Sn, Ga, In, Sb, Co, Ge, and Fe; 3 700 samples) in sphalerite from four genetic types of Pb-Zn deposits: Mississippi Valley, Volcanogenic Massive Sulfide, Sedimentary Exhalative, and Skarn deposits. We applied two machine learning-based models of classification—the support vector machine and random forest—to classify the elemental data, rank the importance of these characteristic elements, and identify the key controlling elements to distinguish between the genetic types of Pb-Zn deposits. The use of big data in conjunction with the machine learning-based techniques enabled the accurate identification of the different genetic types of Pb-Zn deposits based on the characteristics of the trace elements of sphalerite. The support vector machine and random forest-based models of classification achieved accuracies of 98.5% and 96.9%, respectively, on the test dataset. Moreover, we subjected the 12-dimensional elemental features to principal component analysis. This, in conjunction with statistical analysis, visualized dimension reduction, and the ranking of importance of the elements according to their features based on the random forest model, enabled us to identify six key elements (Mn, Ge, In, Co, Sb, and Ga) that can be used to distinguish among the four genetic types of Pb-Zn deposits within the 12 trace elements of sphalerite. Following this, we constructed a 2D ternary diagram of $\ln(\text{Mn}/\text{Cd})$ - $\ln(\text{Ge}/\text{Cd})$ - $\ln(\text{Co}/\text{Cd})$, and 2D binary diagrams of $\ln(\text{Mn})$ - $\ln(\text{Sb})$, $\ln(\text{Co})$ - $\ln(\text{Ga})$, and $\ln(\text{Mn})$ - $\ln(\text{In}/\text{Ge})$ for sphalerite to distinguish among the four genetic types of Pb-Zn deposits.

Key words: sphalerite; trace elements; machine learning; big data analysis; Pb-Zn deposits

铅锌是十分重要的矿产资源,在电气、机械、冶金、医药、核和石油工业等领域有着广泛的应用,对促进国民经济发展和提高人民生活品质有重大作用,各个国家也越来越重视铅锌资源的勘探与开发。在勘探过程中,区分铅锌矿床类型是必不可少的环节,对于建立找矿模型和指导找矿具有重要意义,多元找矿信息的提取对指导深部找矿越来越重要(张弘等,2021;罗杰等,2022;

郭广慧等,2023)。

在铅锌矿床中,矿物类型多样,其中闪锌矿是最具代表性的硫化物矿物之一。在成矿过程中,成矿温度(Wark and Watson,2006)、金属来源、氧逸度(Keith et al.,2014)等因素都会影响闪锌矿微量元素含量(如 Ga, In, Ge, Cd, Fe, Mn 和 Co 等),因此闪锌矿中微量元素特征可以间接地反映出铅锌矿床的成矿信息(Reich et al.,2013;金露

英等,2015)。许多学者依据检测出来的元素含量或者比值,建立了基于闪锌矿成分判断矿床成因类型的判别方法。例如,胡鹏等(2014)对马元铅锌矿床闪锌矿进行高精度的激光剥蚀-电感耦合等离子体质谱仪(LA-ICP-MS)原位微量元素分析,低含量的 Fe、Mn 和 Ga/Ge 比值特征,均指示了矿床成矿流体温度不高,推断出马元铅锌矿床成矿作用与中低温盆地流体有关;Wen et al. (2016)统计了全中国范围内 9 个不同成因的铅锌矿床的闪锌矿 Cd 含量、 $\delta^{114/110}\text{Cd}$ 值变化范围以及 Zn/Cd 比值,并将这些矿床分为 3 大类(高温系统、低温系统、喷流系统),首次提出了 Cd 的分布与成矿流体环境有关(温度、pH 值和盐度),得出不同成矿系统 Cd 含量和 Zn/Cd 比值不同;张茂富等(2016)对会泽矿区闪锌矿进行元素分析,将闪锌矿中 Zn/Cd 比值作为地质温度计进行测温(Zn/Cd > 600: 指示高温; 100 < Zn/Cd < 600: 指示中温; Zn/Cd < 100: 指示低温);康凯等(2020)利用 LA-ICP-MS 原位微量元素分析技术测试花牛山铅锌矿床闪锌矿的微量元素组成,推测出花牛山铅锌矿床成矿作用与岩浆热液有关;陈钧渝等(2023)利用 LA-ICP-MS 对中天山狼牙泉铅锌矿床中 3 个闪锌矿样本进行了原位微量元素分析,推断狼牙泉铅锌矿床属于高温矽卡岩型铅锌矿床。以上研究均表明不同成因类型的铅锌矿床中闪锌矿具有明显不同的微量元素特征。

近年来,随着测试技术的发展,测试的速度和结果的精度均不断提高,地质数据呈现出爆炸式增长。同时,随着科学技术的不断发展,这些数据的丰富度、可信度和生产速度都处在一个较高的水平,使得地质大数据呈现出大量性、高速性和多样性的特征(路英川等,2021)。在大数据和智能化技术的推动下,一些学者尝试使用机器学习的方法对矿物的地球化学特征进行研究和分析,并取得了良好的研究成果。例如,Li et al. (2023)使用随机森林(RF)和 t-分布随机近邻嵌入(t-SNE)2 种机器学习方法对收集到的 5 个不同成因的铅锌矿的闪锌矿数据进行分析,结果表明,除火山块状硫化物型(VMS)矿床外,喷流-沉积型(SEDEX)、密西西比河谷型(MVT)、矽卡岩型和浅成热液型 4 种类型的铅锌矿床均可以被区分出来,证明闪锌矿微量元素数据可以应用于铅锌矿床的成因分类;Zou et al. (2022)通过对全球典型

斑岩型铜矿床中成矿与不成矿岩浆岩的微量元素进行系统分析,使用 RF 和深度神经网络(DNN)2 种新型的机器学习方法来获得成矿潜力指数,模型的特征重要性分析表明, Eu/Eu^* , $\text{Eu}/\text{Eu}^*/\text{Y}$, Ce/Nd , Ce/Ce^* , Dy, Hf 和 Ti 是判断成矿潜力指数的重要参数,通过成矿指数缩小找矿勘查范围,从而提高了找矿效率;洪双等(2021)收集国内外 7 388 个磁铁矿地球化学数据,采用 RF 算法对矿床成因类型进行分类,结果表明,磁铁矿中的主微量元素能有效区分矿床成因类型。上述研究案例说明,RF、t-SNE 和 DNN 机器学习方法可用来对矿床进行成因分类,但 RF 和 DNN 等高维机器学习方法对分类结果可视化有限,t-SNE 方法通过降维将分类结果可视化,虽然弥补了高维分类方法中对分类结果进行可视化局限性或与原始地球化学数据集建立联系,但降维可视化后的结果是由新的变量组合或新的复杂表达式组合而成,这种结果的可视化不利于其后续地质工作的开展,因此需要形成一种广义的判别图来对其后续研究提供价值。

基于此,本文以不同成因类型铅锌矿床中的闪锌矿为研究对象,在系统收集闪锌矿 LA-ICP-MS 元素含量数据的基础上,使用支持向量机(SVM)和 RF 对数据进行分析,并结合元素主成分分析结果,新建立了基于闪锌矿微量元素特征的广义铅锌矿成因类型判别图。

1 闪锌矿元素地球化学数据集

本文系统收集了已发表在国内外期刊上的 3 772 条闪锌矿 LA-ICP-MS 元素数据。数据收集后,对数据进行检查,发现一些矿床成因存在争议,会影响后续判断,所以剔除了 72 条相关数据。整理形成的闪锌矿地球化学大数据集(表 1)来源于全球 89 个铅锌矿床(图 1),包含澳大利亚的 Mt. Isa 铅锌矿床、伊朗的安戈帕斯(Angouran)铅锌矿床等典型的大型-超大型矿床,涉及 MVT 型、VMS 型、SEDEX 型和矽卡岩型 4 个常见的铅锌矿成因类型;每条闪锌矿数据包含 12 种元素: Cd, Mn, Ag, Cu, Pb, Sn, Ga, In, Sb, Co, Ge 和 Fe。

通过箱线图对不同成因类型铅锌矿床中闪锌矿微量元素含量分布特征进行可视化(图 2),发现微量元素特征可以用于区分不同类型的铅锌矿床。具体如下:(1)与岩浆热液无关(MVT 型、

SEDEX 型)和与岩浆热液有关(矽卡岩型、VMS 型)的矿床中,Mn 元素含量差别明显,在与岩浆热液有关的矿床中相对比较富集 Mn;(2)In 元素在 MVT 型矿床中含量较低,明显区别于其他 3

类铅锌矿床;(3)在矽卡岩型矿床中,Sb 元素相对亏损,而 Co 元素相对富集,与其他 3 类矿床差异显著;(4)MVT 型矿床中 Ge 元素含量相对富集,与其他 3 类矿床重叠度较低。

表 1 不同成因类型铅锌矿床中闪锌矿微量元素数据量
Table 1 Amounts of trace elements in sphalerite from different genetic deposits

成因类型	数据量	矿床名称及参考文献
MVT 型	1424	勐兴(刘锦康等,2020);纳雍枝(Wei et al.,2021);火烧云(袁鑫等,2022);湘西花垣(魏含涛等,2021);马元(胡鹏等,2014);Angouran(Zhuang et al.,2019);西邑(Xiao and Li,2019);Alvia mine(Pašava et al.,2014);North German Basin(Knorsch et al.,2020);大梁子(Yuan et al.,2018);会泽,勐兴(Zhong et al.,2021);李子坪,富隆厂(Liu et al.,2022);牛角塘(Ye et al.,2011);金顶(Zhong et al.,2021);毛坪(Wei et al.,2021);富乐(任涛等,2019);Tres Marias TM,Kapp Mineral,Sinkholmen(Cook et al.,2009);Lisheen(Frenzel et al.,2020);清水塘(Liu et al.,2022);毛坪,天桥,垭都,筲箕湾,青山,杉树林(Yang et al.,2022)
SEDEX 型	1181	Hilton,Mount Isa(Cave et al.,2020);Mofjället,Bleikvassli,Mount Isa,Broken Hill(Lockington et al.,2014);Haerdaban(Jiang et al.,2021);Mount Novit(Cave et al.,2022);Bentailou(Cugerone et al.,2021);Bobija(Maurer et al.,2019)
VMS 型	595	老厂(Ye et al.,2011);María Teresa(Torró et al.,2022);东风南山(Lu et al.,2019);Zinkgruvan(Frenzel et al.,2020);Vorta DMV,Eskay Creek,Sauda Sa,Zinkgruvan,Marketorp,Kaveltorp(Cook et al.,2009);Çayeli paleohydrothermal chimney(Revan et al.,2014);别鲁乌图(Li et al.,2021);Sulitjelma,Roros(Lockington et al.,2014);Falun(Kampmann et al.,2018);Kaiei Vent Field(Wang et al.,2018);Banskhapa,Jangaldehy,Biskhan,Bhuyari(Mishra et al.,2021)
矽卡岩型	500	核桃坪,芦子园,都龙(Ye et al.,2011);丁家山(Xing et al.,2021);云南(Zhao et al.,2021);庙山(邢波等,2016);Morococha(Benites et al.,2021);Majdanpek,Ocna de Fier,Baita Bihor,Varlea Seacă,Lefevre,Konnerudkollen,Kamioka(Cook et al.,2009);Baita Bihor(Frenzel et al.,2020);Yeonhwa,Gagok,Uljin,Sagok,Janggung,hinyemi,Dongnam,Wondong,Seojin,Sangdong(Lee et al.,2019);姚家岭(刘政,2019);峰岩(Xing et al.,2022);白音诺尔,开原(马腾瀚,2020);姚家岭(Xiong et al.,2022)

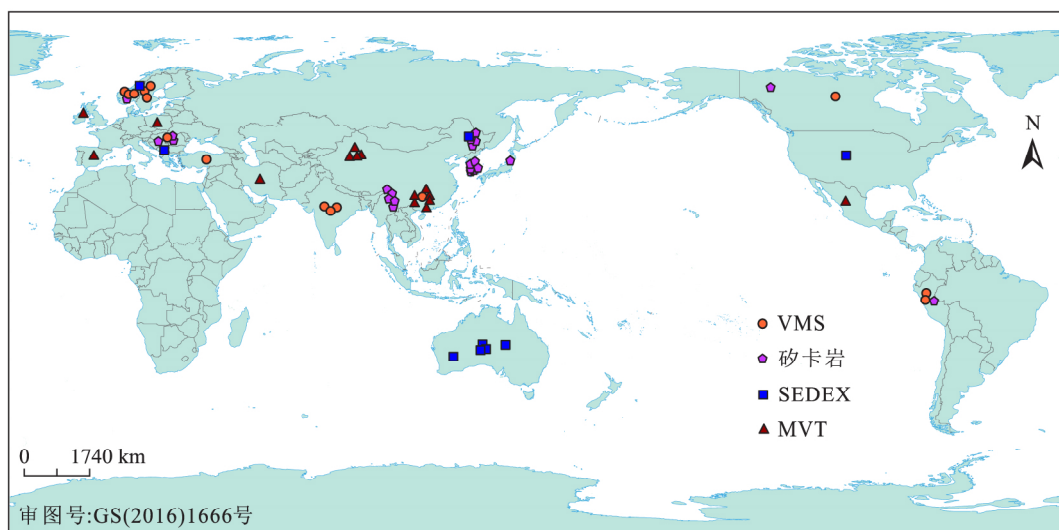


图 1 闪锌矿数据集矿床位置分布图(数据来源同表 1)
Fig. 1 Distribution of locations of ore deposits used to form the sphalerite dataset

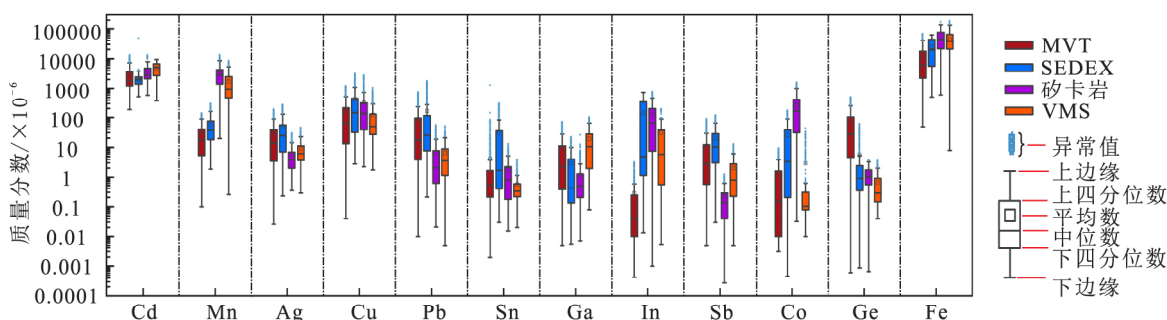


图 2 4 类不同成因类型铅锌矿床闪锌矿的 12 种微量元素箱线图(不含 3 倍四分位数范围值的数据)
Fig. 2 Box plots of the 12 trace elements of sphalerite from deposits of its four genetic types

2 方法与流程

在收集完数据后,利用机器学习方法对数据进行进一步地分析(图 3)。先对数据进行预处理,用 K 近邻算法(KNN)进行插补;然后使用闪锌矿微量元素含量数据来训练 SVM 和 RF 分类模型,初步评估分类模型、优化模型(调参),直到得到合理模型为止;最后,通过主成分分析(PCA)和 RF 特征重要性排序,找出矿床控制关键元素,用这些关键元素构建一个新的二维铅锌矿床的分类图。

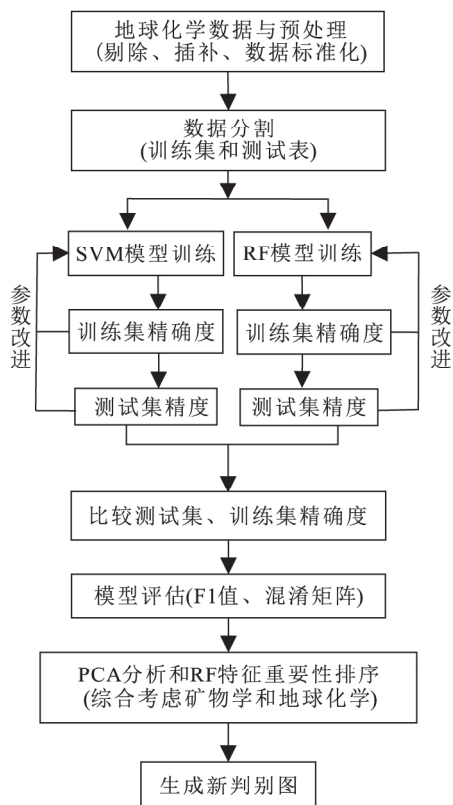


图 3 分析流程图
Fig. 3 Flowchart of analysis for this study

2.1 机器学习方法

(1)支持向量机(SVM)作为一种有标签的监督学习算法,其基本思想是在样本空间上找到一个最优的超平面,这个超平面到各个样本的距离最大,使得模型的泛化能力最强,分类结果最稳定(Cortes and Vapnik,1995)。SVM 在处理线性不可分的情况时,需要使用核函数(多项式函数和径向基核函数等)将样本投射到更高维的空间,从而将他们转化为高维空间上的线性可分的问题。本文用到的是带有径向基核函数的 SVM,此模型包含 2 个参数,参数 C 和参数 gamma。C 是平衡模型准确率与模型复杂度之间的一个参数,较小的 C 使超平面更加平滑,C 较高时可以提高其模型的准确率,但模型的复杂度也会增高;gamma 主要是定义了单个样本对整个分类超平面的影响,较小时,单个样本对超平面的影响就比较小,较大时,单个样本对整个分类的影响比较大,更容易被选择为支持向量,或者说有更多样本用来充当支持向量。SVM 作为常见的机器学习方法,已应用于矿物远景圈定、矿物地球化学和岩性判断,且都具有良好的分类能力和泛化能力(Petrelli and Perugini,2016;Geranian et al.,2016;张振杰等,2021;Nathwani et al.,2022)。

(2)随机森林(RF)是基于决策树提出来的方法,是一种分类效率高、参数易调整的集成 bagging 方法。决策树是 RF 的基础块,一般是二叉树结构,每个非叶节点按照一定的规则分成 2 个子节点,直到分裂成叶节点为止(Yang et al.,2017)。RF 的优点是不需要数据集服从某种分布,且是由多个决策树集成,通过对每个决策树进行投票,选出最佳的决策树来作为最终的结果,从而消除了单个决策树精度低和过拟合现象(Breiman,1996)。RF 作为高效机器学习方法,

已被应用到矿床勘探、靶区圈定(Harris et al., 2015; Zhang et al., 2016; Nathwani et al., 2022)和地球化学分析等领域(Gregory et al., 2019; Wang et al., 2020; 洪双等, 2021; Wang et al., 2021)。

(3)主成分分析(PCA)是一种无监督学习的线性降维方法,是通过计算协方差矩阵的特征向量,用特征向量的值来重建新的维度空间,可将高维空间样本线性映射到低维空间上,并尽可能保留多的信息,可以用来可视化高维多特征变量数据和突出变量之间的关系(Wold et al., 1987; Geladi et al., 1989; Nathwani et al., 2022)。该方法已被众多地质工作者用来判断矿物或矿石元素之间的关系,解析元素之间的转换和辅助判断矿床类型(Frenzel et al., 2016; Yuan et al., 2018; Zhou et al., 2022; Hu et al., 2022)。

本文中模型的训练、模型参数的调整和模型性能的评估均在 Python 3.9.7 环境中进行,用到的 Python 3.9.7 包有:Pandas 1.3.4 和 Numpy 1.20.3(用于对数据操作和分析)、Matplotlib 3.4.3(用来可视化数据和分别类结果)、Scikit-learn 1.2.0(用来学习数据并构建分类模型);PCA 和三元图的制作在 Origin 2023b 学习版软件中进行。

2.2 数据预处理

由于本文数据来源广泛,涵盖了全球不同地区不同成因类型铅锌矿床的闪锌矿地球化学数据,因其研究目的、用途、侧重点不同,导致某些元素检测值缺失。因此,在建立模型之前,需要对数据集进行预处理。本文对闪锌矿数据集做了如下预处理操作:

(1)收集的数据来源于国内外不同的论文,由于元素的检测限不同或未给出检测限值,检测限值无法统一。因此,本文将低于检测限的数据视为缺失值。为减少过多的人为干预,需要对数据集中的元素进行筛选,剔除了缺失值大于 40% 的元素,保留 12 种微量元素(Cd, Mn, Ag, Cu, Pb, Sn, Ga, In, Sb, Co, Ge 和 Fe)。

(2)为了减少极值对分类产生的影响,根据每种矿床类型闪锌矿元素数据箱线图(图 2),消除了落在 3 倍四分位数范围之外的高度可疑离群值。

(3)将数据按照矿床成因类型划分,对缺失的元素值用 scikit-learn 的 KNNImputer 函数进行

补充。KNNImputer 原理可以理解为它是通过距离测量来识别数据集中空间相似或者相近的 k 个样本,使用这 k 个相似的平均值来估计缺失点的值(Troyanskaya et al., 2001)。

(4)为使数据符合正态分布,首先需要对数据进行对数变换;其次,由于各个元素含量范围差别较大,为避免含量值高的元素对分类起支配作用,因此需要对数据进行对数和归一化处理。

2.3 数据分割

经过筛选和预处理后,将 3 700 条符合标准的数据集(包含 12 种闪锌矿微量元素: Cd, Mn, Ag, Cu, Pb, Sn, Ga, In, Sb, Co, Ge 和 Fe),按照每一类矿床的原始比例细分为训练集(80%)和测试集(20%),训练集用于训练分类模型(SVM, RF),测试集用来评估模型的性能。

2.4 参数调整

SVM 和 RF 模型除受样本规模和数据质量影响外,其模型的参数也对模型的分类结果起着关键作用。本文为避免模型出现过拟合现象和欠拟合现象,对模型的参数进行了调整。SVM 含有 2 个参数,分别为参数 C 和参数 gamma,将其 C 的范围设置在 $2^{-5} \sim 2^8$, gamma 的范围设置在 $2^{-10} \sim 2^3$ 。通过网格搜索法(图 4)和 10 折交叉验证法(Wang et al., 2021),最后确定模型参数 C 为 1, gamma 为 0.25。

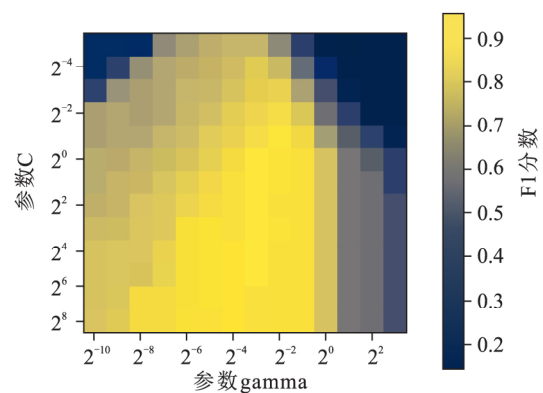


图 4 SVM 网格搜索超参数 C 和 gamma 模型的 F1 得分

Fig. 4 F1 score of the SVM model with the hyperparameters of grid search C and gamma

RF 的参数有很多,主要调整了树的最大生长深度(max_depth)、子树的数量(n_estimators)和最大特征数(max_features),RF 调参的具体步骤如图 5-A 所示。一般而言,RF 树的深度越深,

其模型的精度就越高,但模型的计算复杂度会增高且会出现过拟合现象(Nadi and Moradi, 2019),因此,本文将最大生长深度设置为 8。将子树的数量范围设置为 1~200,步长设置为 10,采用 10 折交叉验证方法,并绘制出得分曲线如

图 5-B 所示,得到子树的数量为 31 时,其模型在训练集上的准确率为 98.8%。同理,在固定 RF 模型子树数量为 31 时,将最大特征数设置在 1~12(有 12 个特征元素),确定其模型参数最大特征数为 3。

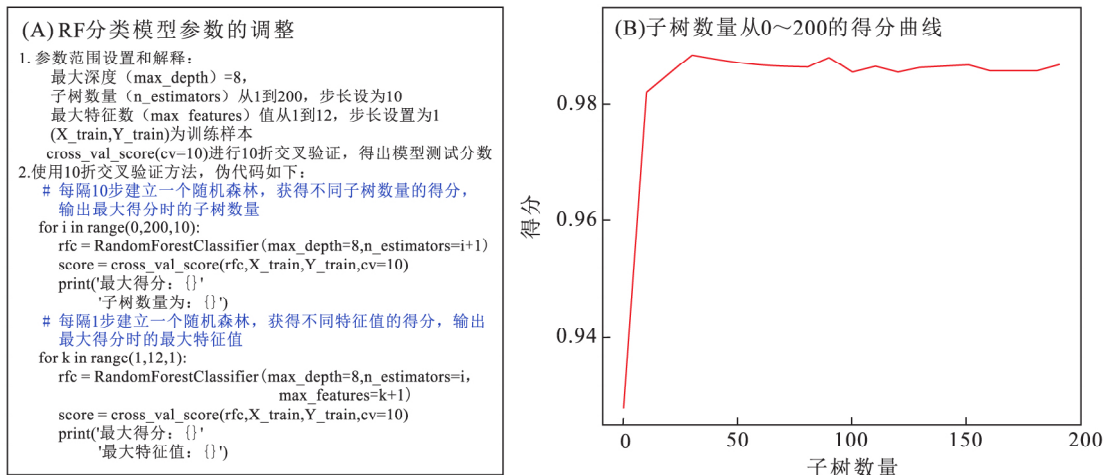


图 5 RF 分类模型超参数调整
Fig. 5 Tuning the hyper-parameters of the RF classification model

2.5 模型评估

模型评估使用准确率、精确率、召回率和 F1 分数值(图 6)。准确率是预测正确的结果占总样本的百分比;精确率是针对预测结果而言的,它是指在所有被预测为正的样本中实际为正样本的概率;召回率是针对原样本而言的,它是指在真实值中正样本被预测为正样本的概率;F1 分数是精确率和召回率的调和平均值。独立计算每一类的 F1 分数,然后所有类取平均,F1 分数值越高代表模型拟合得越好,反之亦然。

预测值和真实值,斜主对角线的值代表预测正确的样本个数。

3 结果与讨论

3.1 高维空间分类器

学习曲线描述了模型随着训练样本数量的增加,模型准确率的变化情况。图 7 是 SVM 和 RF 这 2 个模型的学习曲线。其中,从 SVM 的学习曲线可以看出(图 7-A),其训练集和 10 折交叉验证的得分随着样本数量的增加而变化,当样本数量达到 2 500 时,其模型的准确率逐渐稳定在 0.98 附近;RF 模型的学习曲线显示(图 7-B),训练集和 10 折交叉验证的得分在训练样本数量达到 2 500 时,其模型的准确率逐渐收敛于 0.98~0.99。

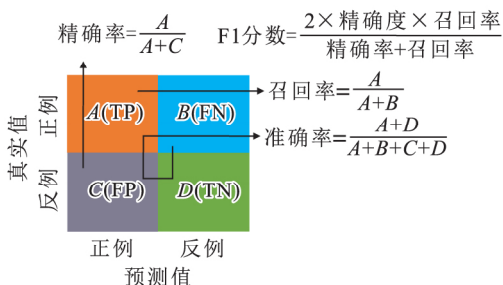


图 6 四个模型评估指标的示意图
Fig. 6 Simple representation of the evaluation metrics for the four models

另一个评估模型方法是混淆矩阵,混淆矩阵是机器学习中常用来表示分类模型预测结果的矩阵,以 n 行 n 列组成矩阵的形式来记录数据集真实值和模型预测值,其中 x 轴和 y 轴分别代表着

为进一步对这 2 种模型进行评估,本文将随机使用 20% 的数据并分别执行 50 次,来测试 SVM 和 RF 模型的综合性能,表 2 包括了每类矿床的精确度、召回率和 F1 分数。MVT 型、SEDEX 型、矽卡岩型和 VMS 型在其 SVM 分类模型的 F1 分数上的得分分别为 0.97±0.00, 0.98±0.01, 0.99±0.01 和 0.99±0.01;这 4 种矿床在 RF 的 F1 得分情况分别为 0.98±0.01, 0.97±0.01, 0.98±0.01 和 0.98±0.01。

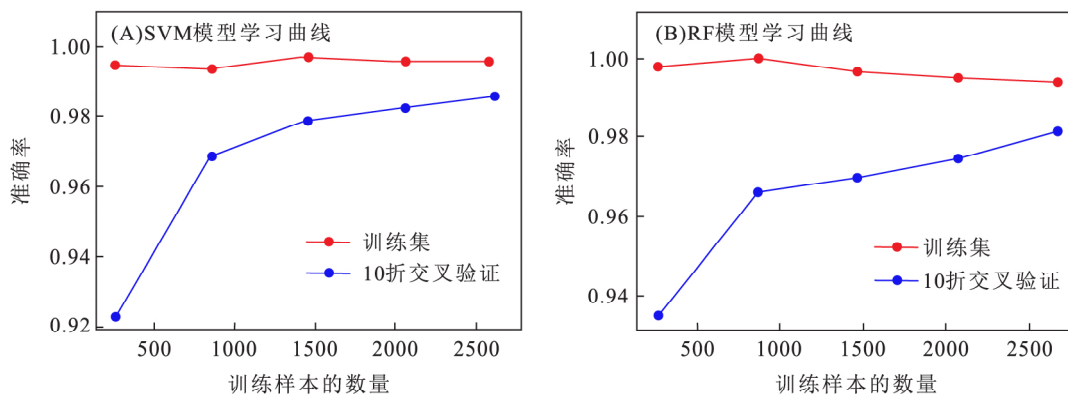


图 7 两个模型的学习曲线
Fig. 7 Learning curves of two models

表 2 分类报告
Table 2 Results of classification

分类模型	SVM				RF		
矿床类型	样本数量	精确度	召回率	F1 分数	精确度	召回率	F1 分数
MVT	284.44±10.85	0.98±0.01	0.99±0.01	0.99±0.00	0.97±0.01	0.98±0.01	0.98±0.01
SEDEX	236.06±12.17	0.99±0.01	0.97±0.01	0.98±0.01	0.97±0.01	0.97±0.01	0.97±0.01
矽卡岩型	100.70±8.42	0.99±0.01	0.98±0.01	0.99±0.01	0.98±0.01	0.97±0.02	0.98±0.01
VMS	118.80±8.48	0.98±0.01	0.99±0.01	0.99±0.01	0.97±0.01	0.98±0.01	0.98±0.01

注:执行 50 次的分类报告。

图 8 基于 SVM 和 RF 分类模型的混淆矩阵。SVM 的混淆矩阵显示(图 8-A),SEDEX 类型矿床有 1 个样本被误分为 MV T 型;MV T 类型矿床的准确率相对较低,7 个样本被误分为 SEDEX 型;VMS 型矿床有 2 个样本被误分为矽卡岩型矿床;矽卡岩型类型的矿床分类准确率较高,有 1 个

样本被错误分类为 VMS 型矿床。相比之下,基于 RF 的混淆矩阵中(图 8-B),SEDEX 型矿床有 3 个样本被误分成 MV T 型;MV T 型矿床在 RF 分类模型的分 类准确率也是最低,有 10 个样本被错误分类;VMS 矿床类型中有 5 个样本被错误分类;矽卡岩型矿床类型有 3 个样本被错误分类。

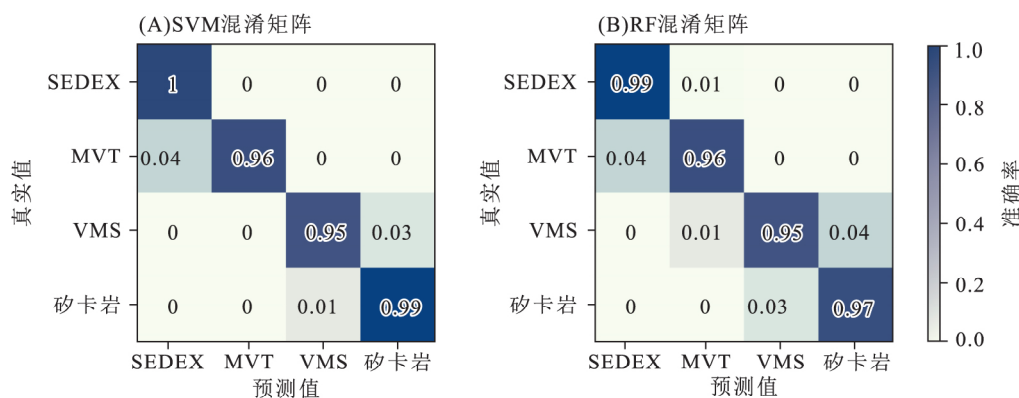


图 8 SVM 和 RF 分类模型在验证集的混淆矩阵
Fig. 8 Confusion matrices of the SVM and RF models on the validation set

由于 SVM 和 RF 机器学习算法是一种黑盒操作,无法可视化其模型工作机制背后的原理。然而,在 RF 分类模型中,存在一种特征重要性排序的方法,以此了解每种元素特征对分类模型的

影响。为了区分 12 种闪锌矿微量元素中每个特征元素对铅锌矿床分类的影响程度,我们使用上述训练好的 RF 分类模型对闪锌矿元素特征重要性进行排序。如图 9 所示,在这些元素中,

Mn(0.185),In(0.177),Co(0.124),Ge(0.105),Fe(0.091)的重要性指数排名靠前,表明它们在区分4种不同成因类型铅锌矿床中起着较大的作用,这意味着Mn,In,Ge,Co和Fe这些特征元素可以用来有效地区分不同成因类型的铅锌矿床。

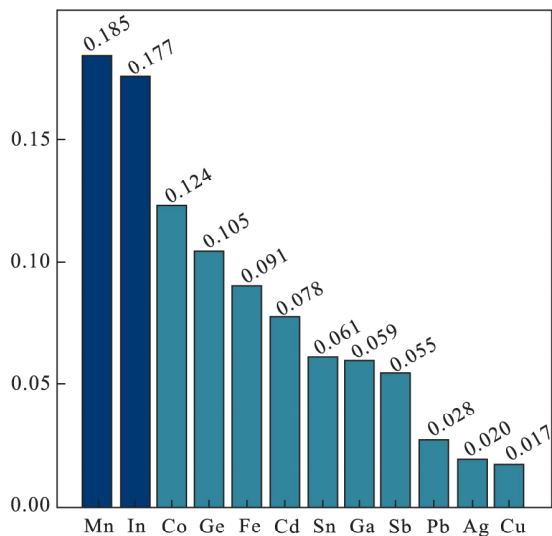


图 9 元素特征重要性排序

Fig. 9 Ranking of the feature importance of the elements

PCA 分析结果:PCA 的目标是通过正交变换,将原始数据转换为的一组线性无关的新变量,这些新变量称为主成分。每个主成分都是原始变量的线性组合,且彼此正交(即不相关)。将12种闪锌矿元素数据使用PCA进行降维,其结果如图10和图11所示,PC1~PC4(第一主成分PC1:具有最大的方差,保留数据最大的信息量;第二主成分PC2:与第一主成分正交,且具有次大的方差,保留次大的信息量;第三主成分PC3:在三维数据中,PC3是第三个主成分,代表第三大的方差方向;第四主成分PC4:当处理更高维度的数据时,如四维或更高,PC4代表第四大的方差方向)保留了原始数据的66.33%。其中,PC1和PC2保留了46.3%,虽然其保留的信息低于50%,但仍可以从这个模型得到数据集的一些特征。图10-A显示4种矿床类型在原点附近出现了同类之间不同程度的聚集,4种铅锌矿床类型主要在PC1方向存在差异,图10-B的元素载荷图中显示Mn和Ge元素与PC1方向夹角最小且长度最长,这表明Mn和Ge元素对PC1贡献大,对4种矿床

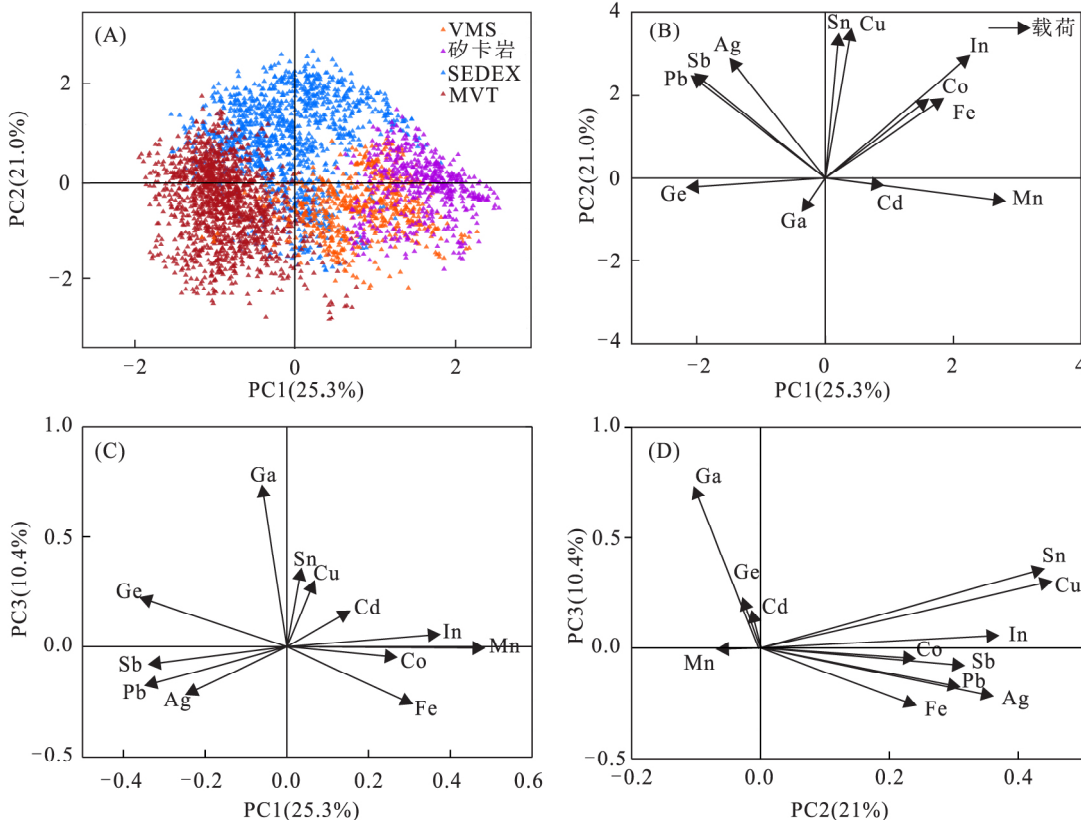


图 10 PCA 分析结果

Fig. 10 Results of PCA

(A)在PC1-PC2可视化结果图;(B)元素在PC1-PC2载荷图;(C)元素在PC1-PC3载荷图;(D)元素在PC2-PC3载荷图

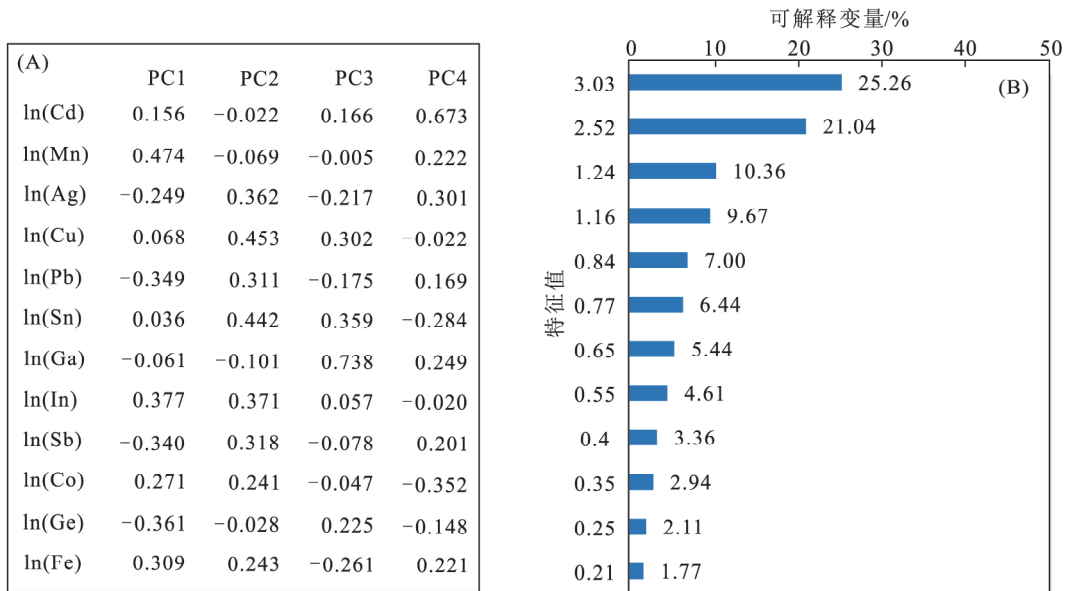


图 11 PCA 分析结果值
 Fig. 11 Results of PCA analysis
 (A)主成分(PC1~PC4)的载荷(特征向量);(B)特征值和可解释变量

类型的分类贡献大,这一现象和 RF 元素重要性的结果显示一致。载荷图中(ln 和 Co)、(Sn 和 Cu)、(Pb、Ag 和 Sb)3 组元素方向一致,这说明元素含量在 PC1 和 PC2 这 2 个维度里相关性强,对铅锌矿分类影响极其相似。同时,图 12 计算出 12 种闪锌矿元素中两两元素的相关系数,可以看出上述 3 组相关性系数均在 0.5 附近,再次证明了这 3 组元素在铅锌矿床分类中行为相似。

3.2 铅锌矿床类型判别新图解

如前文所述,SVM 和 RF 分类模型具有良好的性能,然而这种基于多变量的分类模型很难进行可视化。虽然传统的二维判别图会丢失一些信息,但是却更直观和更方便,因此二维判别图通常是学者们比较青睐的选择。先前大量的研究已表明,矿物的微量元素关系图可以用来判别矿床成因类型。例如,张乾(1987)利用闪锌矿的ln(Ga)-

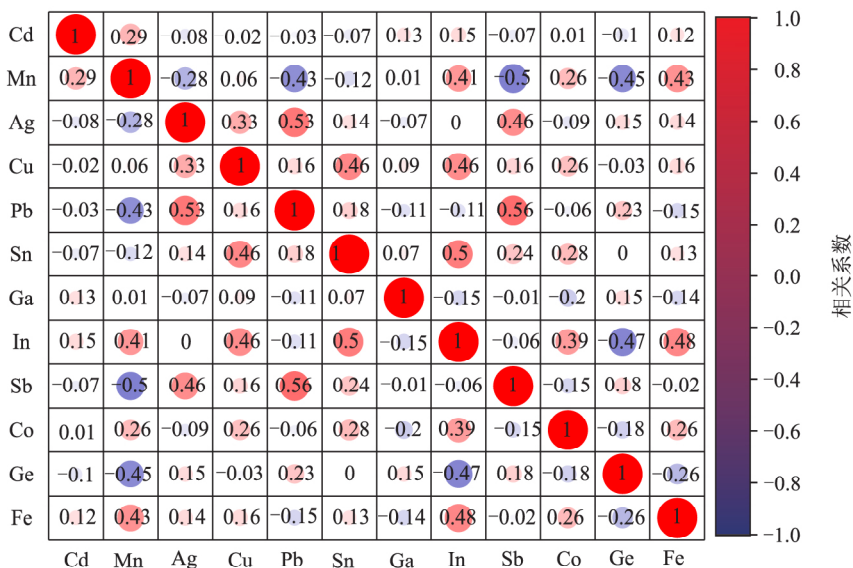


图 12 闪锌矿微量元素之间的相关系数
 Fig. 12 Correlation coefficient of the trace elements of sphalerite
 (系数值越大圈越大颜色越深)

ln(In)图和 Zn/Cd-Se/Te-Ga/In 二维三角图来判别不同成因类型的铅锌矿床。ln(Ga)-ln(In)图显示岩浆热液型矿床和沉积改造型矿床分别位于 Ga/In=1 的两侧, Zn/Cd-Se/Te-Ga/In 二维三角图则显示岩浆热液型铅锌矿、火山岩型铅锌矿、沉积改造型铅锌矿和沉积变质混合岩化型铅锌矿 4 类矿床均有独立的区域。说明利用闪锌矿微量元素图解可以将不同成因类型的铅锌矿区分开, 奠定了用闪锌矿元素特征来判断铅锌矿床成因类型的基础。Sun et al. (2022) 使用决策树和 RF 对收集到的 7 种不同成因类型矿床中 2 200 条闪锌矿微量元素数据进行分类, 得出的决策树和 RF 分类精度分别达到了 91.2% 和 95.4%, 最后结合 RF 模型对影响矿床类型分类重要性进行了排序, 认为 Mn、Co 和 Ge 元素对铅锌矿床分类影响较大。Hu et al. (2022) 利用 PCA 和 t-SNE 对收集到的 876 个全球高温成矿系统原生高钛磁铁矿数据集进行分析, 识别磁铁矿成因, 结果显示铁氧化物-磷灰石型磁铁矿与来自高温热液成因的铁氧化物铜金型和斑岩型元素特征相同, 表明铁氧化物-磷灰石型磁铁矿极大可能是热液成因, 并提出了 $\lg(\text{Al}) + \lg(\text{Ti}) + \lg(\text{V}) - \lg(\text{Mn}) / [\lg(\text{Co}) + \lg(\text{Mg})]$ 高钛磁铁矿新的判别图, 判别图在 SVM 中的准确率达到了 97%。

本文在机器学习的基础上, 试图构建一个比较完善的二维图, 以便后续使用闪锌矿微量元素对铅锌矿床成因类型进行初步判别。综合考虑 RF 分类模型中特征元素重要性排序、PCA 元素载荷值、元素存在的普遍性和其在数据集中缺失比例等指标, 本文构建了 $\ln(\text{Mn}/\text{Cd}) - \ln(\text{Ge}/\text{Cd}) - \ln(\text{Co}/\text{Cd})$ 二维三元图, 选择相关元素的理由如下。

(1) 秉承方便简洁的宗旨。总共选取了 4 种元素来进行区分, 既节省操作时间, 同时也节省测试元素含量的费用。这 4 种元素在闪锌矿中都是普遍存在的, 其缺失值均在 25% 以下 (Cd, Mn, Ge 和 Co 的缺失率分别为 2.23%, 5.73%, 22.16% 和 24.76%), 这样能保留较多原始数据本身的特征, 减少一些人为干预, 使得此判别图更具有普适性。

(2) Cd, Mn, Ge, Co 元素在元素特征重要性排序中位列前 4, 从图 10 也可以看出 Mn 和 Ge 与 PC1 轴呈锐角, 表明其对 PC1 的贡献大, 可以

较好地将矿床类型在 PC1 方向上进行区分。

(3) 尽管张乾(1987)利用 In 元素和 Ga 元素以及 Ga/In 比值来联合判断铅锌矿成因类型, 但 Ga 元素在 4 种类型矿床闪锌矿中的质量分数重合度高, 同时, Ga 元素在 PCA 元素载荷图中对 PC1 和 PC2 的贡献值很小, 几乎为零; In 元素除了在 MVT 型矿床闪锌矿中质量分数比较低外, 在其他 3 类矿床闪锌矿中的重合度很高。因此, 这 2 种元素 (Ga, In) 无法用来有效区分这 4 种铅锌矿床类型。

(4) Cd 元素质量分数在 4 种类型矿床的闪锌矿中都很高, 同时具有很高的重合度, 将 Mn、Ge 和 Co 这 3 种元素与 Cd 元素进行比值是为了将数据尽可能地收敛, 从而使数据点落在三角形的内部。

图 13 是本文构建的 4 种成因类型铅锌矿床中闪锌矿的微量元素特征三元判别图, 如图所示, 与岩浆热液有关的矿床 (矽卡岩型、VMS 型) 和非岩浆热液型矿床 (MVT 型、SEDEX 型) 区别明显, 矽卡岩型矿床主要集中在三角形上部分, 有部分样本数据与 VMS 型矿床重合; SEDEX 型矿床和 MVT 型矿床主要集中在三角形的中部, SEDEX 型矿床数据较为分散, 部分样本与 MVT 型矿床出现了重叠。

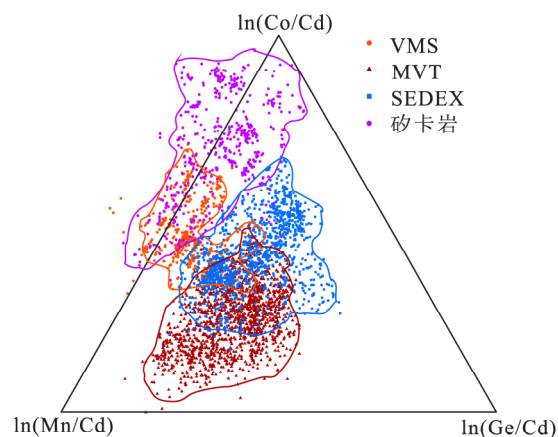


图 13 四种成因类型铅锌矿床闪锌矿 $\ln(\text{Mn}/\text{Cd}) - \ln(\text{Ge}/\text{Cd}) - \ln(\text{Co}/\text{Cd})$ 三元判别图

Fig. 13 Ternary discrimination plot of $\ln(\text{Mn}/\text{Cd}) - \ln(\text{Ge}/\text{Cd}) - \ln(\text{Co}/\text{Cd})$ for four genetic types of Pb-Zn deposits in sphalerite

由于二维空间的局限性, 上述三元图的相关数据点重叠是不可避免的, 为进一步区分重叠部分较高的矿床类型, 在上述三元图基础上, 尝试建

立 2 种铅锌矿床类型的分类图解。如图 14-A~C 所示, $\ln(\text{Mn}/\text{Cd})-\ln(\text{Ge}/\text{Cd})-\ln(\text{Co}/\text{Cd})$ 可以有效地区分 MVT 型与矽卡岩型、SEDEX 型与矽卡岩型、VMS 型与 MVT 型矿床。但是, VMS 型与矽卡岩型、MVT 型与 SEDEX 型、VMS 型与 SEDEX 型重叠部分较高, 无法很好地区分。因此, 需要建立其他元素判别图, 进一步区分这 3 组矿床。根据 RF 元素重要性排序以及图 2 箱线图可知, Co 元素在矽卡岩型矿床相对富集, 在 VMS 型矿床相对亏损, Ga 元素在矽卡岩型矿床相对亏损, 在 VMS 型矿床相对富集。因此, $\ln(\text{Co})-\ln(\text{Ga})$ 二维图(图 14-D)可以有效区分 VMS 型和矽卡岩型矿床。图 2 箱线图显示 MVT 型和

SEDEX 型矿床的 12 种微量元素富集状况大致相同, 虽然 In 元素和 Ge 元素箱体部分无重叠, 但这 2 种元素在这 2 种矿床的质量分数分布较为分散, 其质量分数不能很好区分 MVT 型和 SEDEX 型矿床, 本文采用 $\ln(\text{In}/\text{Ge})$ 扩大元素质量分数之间的差距。同时, Mn 元素在 RF 元素重要性排在首位, 说明 Mn 元素在矿床分类的作用显著, 用 $\ln(\text{Mn})$ 元素作为 x 轴, 构建 $\ln(\text{Mn})-\ln(\text{In}/\text{Ge})$ 判别图来区分这 2 种类型的矿床(图 14-E)。Mn 元素在 SEDEX 型矿床中亏损, 在 VMS 型矿床中富集; 相反, Sb 元素在 SEDEX 型矿床中富集, 在 VMS 型矿床中亏损。基于此, $\ln(\text{Mn})-\ln(\text{Sb})$ 二维图可以用来区分 VMS 型和 SEDEX 型 2 类矿床(图 14-F)。

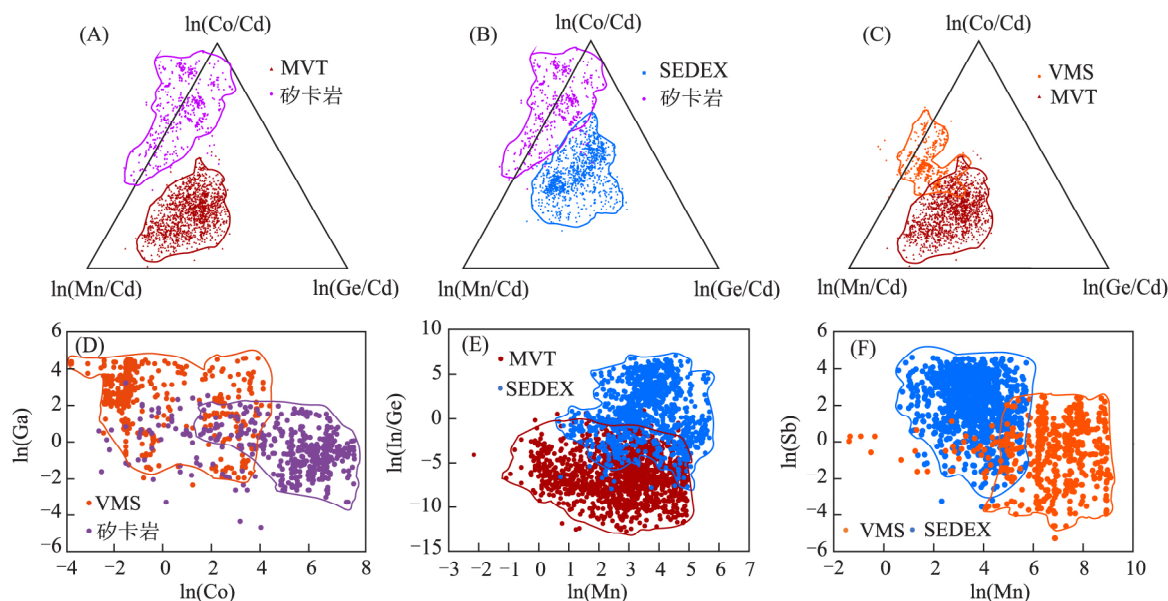


图 14 基于闪锌矿微量元素的铅锌矿床成因类型判别图

Fig. 14 Classification diagrams of genetic types of Pb-Zn deposits based on their sphalerite trace elements (A) MVT 型和矽卡岩型矿床三元判别图; (B) SEDEX 型和矽卡岩型矿床三元判别图; (C) VMS 型和 MVT 型矿床三元判别图; (D) VMS 型和矽卡岩型矿床二元判别图; (E) MVT 型和 SEDEX 型矿床二元判别图; (F) VMS 型和 SEDEX 型矿床二元判别图

4 结论

a. 本研究通过系统收集全球闪锌矿元素数据 3 700 条, 分别使用 SVM 和 RF 2 种机器学习方法构建了 4 种成因类型铅锌矿床的成因分类模型。结果表明, 闪锌矿作为勘查指示矿物, 其化学成分能够用来区分不同成因类型的铅锌矿。

b. SVM 和 RF 分类模型具有良好的性能, 测试集上的准确率分别为 98.5% 和 96.9%。

c. 从主成分分析的结果和 RF 特征重要性排

序上可以看出, Mn, Ge, Co 是区分不同成因铅锌矿床的关键元素。

d. 本文构建了 $\ln(\text{Mn}/\text{Cd})-\ln(\text{Ge}/\text{Cd})-\ln(\text{Co}/\text{Cd})$ 二维三元图来判断铅锌矿床的成因类型, 并在此基础上进一步新构建了 $\ln(\text{Co})-\ln(\text{Ga})$ 、 $\ln(\text{Mn})-\ln(\text{In}/\text{Ge})$ 和 $\ln(\text{Mn})-\ln(\text{Sb})$ 判别图对二维三元图中重叠较大的矿床类型进行详细区分, 综合上述判别图可以有效区分 SEDEX 型、MVT 型、VMS 型、矽卡岩型 4 种成因类型的铅锌矿床。

[参 考 文 献]

- 陈钧渝,沈鸿杰,颜伟裕,2023.中天山狼牙泉铅锌矿床闪锌矿 LA-ICP-MS 微量元素特征对矿床成因的指示意义[J].矿产勘查,14(3):377-391.
- Chen J Y, Shen H J, Yan W Y, 2023. LA-ICP-MS trace element geochemistry of sphalerite: Metallogenic constraints on the Langyaquan Pb-Zn deposit in the Middle Tianshan metallogenic belt[J]. Mineral Exploration, 14(3): 377-391. (in Chinese)
- 郭广慧,钟世华,李三忠,等,2023.运用机器学习和锆石微量元素构建花岗岩成矿潜力判别图解:以东昆仑祁漫塔格为例[J].西北地质,56(6):57-70.
- Guo G H, Zhong S H, Li S Z, *et al.*, 2023. Constructing discrimination diagrams for granite mineralization potential by using machine learning and zircon trace elements: Example from the Qimantagh, East Kunlun [J]. Northwestern Geology, 56(6): 57-70. (in Chinese)
- 洪双,左仁广,胡浩,等,2021.磁铁矿元素地球化学大数据构建及其在矿床成因分类中的应用[J].地学前缘,28(3):87-96.
- Hong S, Zuo R G, Hu H, *et al.*, 2021. Magnetite geochemical big data: Dataset construction and application in genetic classification of ore deposits[J]. Earth Science Frontiers, 28(3): 87-96. (in Chinese)
- 胡鹏,吴越,张长青,等,2014.扬子板块北缘马元铅锌矿床闪锌矿 LA-ICP-MS 微量元素特征与指示意义[J].矿物学报,34(4):461-468.
- Hu P, Wu Y, Zhang C Q, *et al.*, 2014. Trace and minor elements in sphalerite from the Mayuan [J]. Acta Mineralogica Sinica, 34(4): 461-468. (in Chinese)
- 金露英,秦克章,李光明,等,2015.大兴安岭北段岔路口斑岩 Mo-热液脉状 Zn-Pb 成矿系统硫化物微量元素的分布、起源及其勘探指示[J].岩石学报,31(8):2417-2434.
- Jin L Y, Qin K Z, Li G M, *et al.*, 2015. Trace element distribution in sulfides from the Chalukou porphyry Mo-vein-type Zn-Pb system, northern Great Xing'an Range, China: Implications for metal source and ore exploration [J]. Acta Petrologica Sinica, 31(8): 2417-2434. (in Chinese)
- 康凯,杜泽忠,于晓飞,等,2020.甘肃花牛山铅锌矿床闪锌矿 LA-ICP-MS 微量元素组成及其地质意义[J].吉林大学学报(地球科学版),50(5):1418-1432.
- Kang K, Du Z Z, Yu X F, *et al.*, 2020. LA-ICP-MS trace element analysis of sphalerite in Huaniushan Pb-Zn deposit in Gansu Province and its geological significance[J]. Journal of Jilin University (Earth Science Edition), 50(5): 1418-1432. (in Chinese)
- 刘锦康,邓明国,李余华,等,2020.滇西勐兴铅锌矿床闪锌矿 ICP-MS 微量元素特征及其地质意义[J].矿物岩石地球化学通报,39(1):133-143.
- Liu J K, Deng M G, Li Y H, *et al.*, 2020. The ICP-MS trace element characteristics of sphalerites from the Mengxing Pb-Zn deposit, Western Yunnan and their geological significances [J]. Bulletin of Mineralogy, Petrology and Geochemistry, 39(1): 133-143. (in Chinese)
- 刘政,2019.安徽铜陵姚家岭锌金多金属矿床成矿作用研究[D].合肥:合肥工业大学档案馆.
- Liu Z, 2019. The Petrogenesis and Metallogenesis of the Yaojialing Polymetallic Zn-Au Deposit in Tongling City, Anhui Province [D]. Hefei: The Archive of Hefei University of Technology. (in Chinese)
- 路英川,李鹏,王浩,等,2021.大数据时代矿床学研究发展状况综述[J].世界核地质科学,38(3):295-310.
- Lu Y C, Li P, Wang H, *et al.*, 2021. A review on the development of mineral deposit science in the era of big data [J]. World Nuclear Geoscience, 38(3): 295-310. (in Chinese)
- 罗杰,周仲礼,邹天一,等,2022.基于 PSO-CNN 的深部找矿预测模型构建[J].成都理工大学学报(自然科学版),49(6):697-708.
- Luo J, Zhou Z L, Zou T Y, *et al.*, 2022. Construction of deep prospecting prediction model based on PSO-CNN [J]. Journal of Chengdu University of Technology (Science & Technology Edition), 49(6): 697-708. (in Chinese)
- 马腾瀚,2020.内蒙古白音诺尔铅锌矿床石榴子石和闪锌矿成因矿物学研究[D].北京:中国地质大学档案馆.
- Ma T H, 2020. Genetic Mineralogy of Garnet and Sphalerite in Baiyinnuo'er Lead-zinc Deposit, Inner Mongoli [D]. Beijing: The Archive of China University of Geosciences. (in Chinese)
- 任涛,周家喜,王蝶,等,2019.滇东北富乐铅锌矿床微量元素和 S-Pb 同位素地球化学研究[J].岩石学报,35(11):3493-3505.
- Ren T, Zhou J X, Wang D, *et al.*, 2019. Trace elemental and S-Pb isotopic geochemistry of the Fule Pb-Zn deposit, NE Yunnan Province [J]. Acta Petrologica Sinica, 35(11): 3493-3505. (in Chinese)
- 隗含涛,邵拥军,叶周,等,2021.湘西花垣铅锌矿田闪锌矿痕量元素地球化学特征[J].成都理工大学学报(自然科学版),48(2):142-153.
- Wei H T, Shao Y J, Ye Z, *et al.*, 2021. Geochemical characteristics of trace elements of sphalerite from Huayuan Pb-Zn ore field, western Hunan, China [J]. Jour-

- nal of Chengdu University of Technology (Science & Technology Edition), 48(2):142-153. (in Chinese)
- 邢波,郑伟,欧阳志侠,等,2016. 粤西庙山铜多金属矿床硫化物原位微区分析及S同位素对矿床成因的制约[J]. 地质学报, 90(5):971-986.
- Xing B, Zheng W, Ouyang Z X, *et al.*, 2016. Sulfide micro-analysis and S isotope of the Miaoshan Cu polymetallic deposit in Western Guangdong Province, and its constraints on the ore genesis[J]. *Acta Geologica Sinica*, 90(5):971-986. (in Chinese)
- 袁鑫,吴越,段登飞,等,2022. 新疆火烧云超大型铅锌矿床闪锌矿微量(稀散)元素组成特征与指示意义[J]. 地质与勘探, 58(3):545-560.
- Yuan X, Wu Y, Duan D F, *et al.*, 2022. The trace (dispersed) elements in sphalerite from Huoshaoyun giant lead-zinc deposit, Xinjiang and their geological implications[J]. *Geology and Exploration*, 58(3):545-560. (in Chinese)
- 张茂富,周宗桂,熊索菲,等,2016. 云南会泽铅锌矿床闪锌矿化学成分特征及其指示意义[J]. 岩石矿物学杂志, 35(1):111-123.
- Zhang M F, Zhou Z G, Xiong S F, *et al.*, 2016. A typomorphic study of sphalerite from the Huize lead-zinc deposit, Yunnan Province [J]. *Acta Petrologica et Mineralogica*, 35(1):111-123. (in Chinese)
- 张弘,高卿楠,郭东旭,2021. 花岗伟晶岩型锂矿热红外反射光谱特征及锂元素定量反演研究[J]. 矿物岩石, 41(1):25-31.
- Zhang H, Gao Q N, Guo X D, 2021. Thermal infrared reflectance spectral characteristics of granitic pegmatite type lithium ore and quantitative inversion of lithium element [J]. *Mineralogy and Petrology*, 41(1):25-31. (in Chinese)
- 张乾,1987. 利用方铅矿、闪锌矿的微量元素图解法区分铅锌矿床的成因类型[J]. 地质地球化学, (9):64-66.
- Zhang Q, 1987. Distinguish the exposure types of lead-zinc deposits using the trace element graphical method of galena and sphalerite[J]. *Geology and Geochemistry*, (9):64-66. (in Chinese)
- 张振杰,成秋明,杨玠,等,2021. 机器学习与成矿预测:以闽西南铁多金属矿预测为例[J]. 地学前缘, 28(3):221-235.
- Zhang Z J, Cheng Q M, Yang J, *et al.*, 2021. Machine learning for mineral prospectivity: A case study of iron-polymetallic mineral prospectivity in southwestern Fujian[J]. *Earth Science Frontiers*, 28(3):221-235. (in Chinese)
- Benites D, Torró L, Vallance J, *et al.*, 2021. Distribution of indium, germanium, gallium and other minor and trace elements in polymetallic ores from a porphyry system: The Morococha district, Peru[J]. *Ore Geology Reviews*, 136:104236.
- Breiman L, 1996. Bagging predictors[J]. *Machine Learning*, 24:123-140.
- Cave B, Lilly R, Barovich K, 2020. Textural and geochemical analysis of chalcopyrite, galena and sphalerite across the Mount Isa Cu to Pb-Zn transition: Implications for a zoned Cu-Pb-Zn system[J]. *Ore Geology Reviews*, 124:103647.
- Cave B, Perkins W, Lilly R, 2022. Linking uplift and mineralisation at the Mount Novit Zn-Pb-Ag deposit, northern Australia: Evidence from geology, U-Pb geochronology and sphalerite geochemistry[J]. *Geoscience Frontiers*, 13(2):101347.
- Cook N J, Ciobanu C L, Pring A, *et al.*, 2009. Trace and minor elements in sphalerite: A LA-ICPMS study [J]. *Geochimica et Cosmochimica Acta*, 73(16):4761-4791.
- Cortes C, Vapnik V, 1995. Support-vector networks[J]. *Machine Learning*, 20:273-297.
- Cugerone A, Cenki-Tok B, Muñoz M, *et al.*, 2021. Behavior of critical metals in metamorphosed Pb-Zn ore deposits: Example from the Pyrenean Axial Zone[J]. *Mineralium Deposita*, 56:685-705.
- Frenzel M, Cook N J, Ciobanu C L, *et al.*, 2020. Halogens in hydrothermal sphalerite record origin of ore-forming fluids[J]. *Geology*, 48(8):766-770.
- Frenzel M, Hirsch T, Gutzmer J, 2016. Gallium, germanium, indium, and other trace and minor elements in sphalerite as a function of deposit type—A meta-analysis[J]. *Ore Geology Reviews*, 76:52-78.
- Geladi P, Isaksson H, Lindqvist L, *et al.*, 1989. Principal component analysis of multivariate images[J]. *Chemometrics and Intelligent Laboratory Systems*, 5(3):209-220.
- Geranian H, Tabatabaei S H, Asadi H H, *et al.*, 2016. Application of discriminant analysis and support vector machine in mapping gold potential areas for further drilling in the Sari-Gunay gold deposit, NW Iran [J]. *Natural Resources Research*, 25:145-159.
- Gregory D D, Cracknell M J, Large R R, *et al.*, 2019. Distinguishing ore deposit type and barren sedimentary pyrite using laser ablation-inductively coupled plasma-mass spectrometry trace element data and statistical analysis of large data sets [J]. *Economic Geology*, 114(4):771-786.
- Harris J R, Grunsky E, Behnia P, *et al.*, 2015. Data-and

- knowledge-driven mineral prospectivity maps for Canada's North[J]. *Ore Geology Reviews*, 71: 788–803.
- Hu B, Zeng L P, Liao W, *et al.*, 2022. The origin and discrimination of high-Ti magnetite in magmatic-hydrothermal systems: Insight from machine learning analysis [J]. *Economic Geology*, 117(7): 1613–1627.
- Jiang Z, Zhang Z, Duan S, *et al.*, 2021. Genesis of the sediment-hosted Haerdaban Zn-Pb deposit, Western Tianshan, NW China: Constraints from textural, compositional and sulfur isotope variations of sulfides [J]. *Ore Geology Reviews*, 139: 104527.
- Kampmann T C, Jansson N F, Stephens M B, *et al.*, 2018. Syn-tectonic sulphide remobilization and trace element redistribution at the Falun pyritic Zn-Pb-Cu-(Au-Ag) sulphide deposit, Bergslagen, Sweden [J]. *Ore Geology Reviews*, 96: 48–71.
- Keith M, Haase K M, Schwarz-Schampera U, *et al.*, 2014. Effects of temperature, sulfur, and oxygen fugacity on the composition of sphalerite from submarine hydrothermal vents [J]. *Geology*, 42(8): 699–702.
- Knorsch M, Nadoll P, Klemd R, 2020. Trace elements and textures of hydrothermal sphalerite and pyrite in Upper Permian (Zechstein) carbonates of the North German Basin [J]. *Journal of Geochemical Exploration*, 209: 106416.
- Lee J H, Yoo B C, Yang Y S, *et al.*, 2019. Sphalerite geochemistry of the Zn-Pb orebodies in the Taebaeksan metallogenic province, Korea [J]. *Ore Geology Reviews*, 107: 1046–1067.
- Li W, Zhang F, Hu C, *et al.*, 2021. Ore genesis and tectonic setting of the Bieluwutu Cu-Pb-Zn volcanogenic massive sulfide deposit in Xing'an-Mongolia orogenic belt, China [J]. *Ore Geology Reviews*, 130: 103951.
- Li X M, Zhang Y X, Li Z K, *et al.*, 2023. Discrimination of Pb-Zn deposit types using sphalerite geochemistry: New insights from machine learning algorithm [J]. *Geoscience Frontiers*, 14(4): 101580.
- Liu S, Zhang Y, Ai G, *et al.*, 2022. LA-ICP-MS trace element geochemistry of sphalerite: Metallogenic constraints on the Qingshuitang Pb-Zn deposit in the Qinhang ore belt, South China [J]. *Ore Geology Reviews*, 141: 104659.
- Lockington J A, Cook N J, Ciobanu C L, 2014. Trace and minor elements in sphalerite from metamorphosed sulphide deposits [J]. *Mineralogy and Petrology*, 108: 873–890.
- Lu S, Ren Y, Yang Q, *et al.*, 2019. Ore genesis for stratiform ore bodies of the Dongfengnanshan copper polymetallic deposit in the Yanbian area, NE China: Constraints from LA-ICP-MS in situ trace elements and sulfide S-Pb isotopes [J]. *Acta Geologica Sinica*, 93(5): 1591–1606.
- Maurer M, Prelević D, Mertz-Kraus R, *et al.*, 2019. Genesis and metallogenic setting of the polymetallic barite-sulphide deposit, Bobija, Western Serbia [J]. *International Journal of Earth Sciences*, 108: 1725–1740.
- Mishra B P, Pati P, Dora M L, *et al.*, 2021. Trace-element systematics and isotopic characteristics of sphalerite-pyrite from volcanogenic massive sulfide deposits of Betul belt, central Indian tectonic zone: Insight of ore genesis to exploration [J]. *Ore Geology Reviews*, 134: 104149.
- Nadi A, Moradi H, 2019. Increasing the views and reducing the depth in random forest [J]. *Expert Systems with Applications*, 138: 112801.
- Nathwani C L, Wilkinson J J, Fry G, *et al.*, 2022. Machine learning for geochemical exploration: Classifying metallogenic fertility in arc magmas and insights into porphyry copper deposit formation [J]. *Mineralium Deposita*, 57(7): 1143–1166.
- Pašava J, Tornos F, Chrastný V, 2014. Zinc and sulfur isotope variation in sphalerite from carbonate-hosted zinc deposits, Cantabria, Spain [J]. *Mineralium Deposita*, 49: 797–807.
- Petrelli M, Perugini D, 2016. Solving petrological problems through machine learning: The study case of tectonic discrimination using geochemical and isotopic data [J]. *Contributions to Mineralogy and Petrology*, 171: 1–15.
- Reich M, Deditius A, Chryssoulis S, *et al.*, 2013. Pyrite as a record of hydrothermal fluid evolution in a porphyry copper system: A SIMS/EMPA trace element study [J]. *Geochimica et Cosmochimica Acta*, 104: 42–62.
- Revan M K, Genc Y, Maslennikov V V, *et al.*, 2014. Mineralogy and trace-element geochemistry of sulfide minerals in hydrothermal chimneys from the Upper-Cretaceous VMS deposits of the eastern Pontide orogenic belt (NE Turkey) [J]. *Ore Geology Reviews*, 63: 129–149.
- Sun G, Zeng Q, Zhou J X, 2022. Machine learning coupled with mineral geochemistry reveals the origin of ore deposits [J]. *Ore Geology Reviews*, 142: 104753.
- Torró L, Benites D, Vallance J, *et al.*, 2022. Trace element geochemistry of sphalerite and chalcopyrite in arc-hosted VMS deposits [J]. *Journal of Geochemical Exploration*, 232: 106882.

- Troyanskaya O, Cantor M, Sherlock G, *et al.*, 2001. Missing value estimation methods for DNA microarrays [J]. *Bioinformatics*, 17(6): 520–525.
- Wang C, Pan Y, Chen J, *et al.*, 2020. Indicator element selection and geochemical anomaly mapping using recursive feature elimination and random forest methods in the Jingdezhen region of Jiangxi Province, South China [J]. *Applied Geochemistry*, 122: 104760.
- Wang Y, Han X, Petersen S, *et al.*, 2018. Trace metal distribution in sulfide minerals from ultramafic-hosted hydrothermal systems: Examples from the Kairei Vent Field, Central Indian Ridge [J]. *Minerals*, 8(11): 526.
- Wang Y, Qiu K, Müller A, *et al.*, 2021. Machine learning prediction of quartz forming-environments [J]. *Journal of Geophysical Research: Solid Earth*, 126(8): e2021JB021925.
- Wark D A, Watson E B, 2006. TitaniQ: A titanium-in-quartz geothermometer [J]. *Contributions to Mineralogy and Petrology*, 152(6): 743–754.
- Wei C, Ye L, Hu Y, *et al.*, 2021. LA-ICP-MS analyses of trace elements in base metal sulfides from carbonate-hosted Zn-Pb deposits, South China: A case study of the Maoping deposit [J]. *Ore Geology Reviews*, 130: 103945.
- Wen H, Zhu C, Zhang Y, *et al.*, 2016. Zn/Cd ratios and cadmium isotope evidence for the classification of lead-zinc deposits [J]. *Scientific Reports*, 6(1): 25273.
- Wold S, Geladi P, Esbensen K, *et al.*, 1987. Multi-way principal components-and PLS-analysis [J]. *Journal of Chemometrics*, 1(1): 41–56.
- Xiao C, Li G, 2019. Geological, sulfur isotopic, and mineral trace element constraints on the genesis of the Xiyi Pb-Zn deposit, Baoshan Block, SW China [J]. *Journal of Asian Earth Sciences*, 186: 104056.
- Xing B, Mao J, Xiao X, *et al.*, 2022. Metallogenic discrimination by sphalerite trace element geochemistry: An example from the Fengyan Zn-Pb deposit in central Fujian, SE China [J]. *Ore Geology Reviews*, 141: 104651.
- Xing B, Mao J, Xiao X, *et al.*, 2021. Genetic discrimination of the Dingjiashan Pb-Zn deposit, SE China, based on sphalerite chemistry [J]. *Ore Geology Reviews*, 135: 104212.
- Xiong Y, Zhou T, Fan Y, *et al.*, 2022. Enrichment mechanisms and occurrence regularity of critical minerals resources in the Yaojialing Zn skarn polymetallic deposit, Tongling district, eastern China [J]. *Ore Geology Reviews*, 144: 104822.
- Yang L, Liu S, Tsoka S, *et al.*, 2017. A regression tree approach using mathematical programming [J]. *Expert Systems with Applications*, 78: 347–357.
- Yang Q, Zhang X, Ulrich T, *et al.*, 2022. Trace element compositions of sulfides from Pb-Zn deposits in the northeast Yunnan and northwest Guizhou Provinces, SW China: Insights from LA-ICP-MS analyses of sphalerite and pyrite [J]. *Ore Geology Reviews*, 141: 104639.
- Ye L, Cook N J, Ciobanu C L, *et al.*, 2011. Trace and minor elements in sphalerite from base metal deposits in South China: A LA-ICPMS study [J]. *Ore Geology Reviews*, 39(4): 188–217.
- Yuan B, Zhang C, Yu H, *et al.*, 2018. Element enrichment characteristics: Insights from element geochemistry of sphalerite in Daliangzi Pb-Zn deposit, Sichuan, Southwest China [J]. *Journal of Geochemical Exploration*, 186: 187–201.
- Zhang Z, Zuo R, Xiong Y, 2016. A comparative study of fuzzy weights of evidence and random forests for mapping mineral prospectivity for skarn-type Fe deposits in the southwestern Fujian metallogenic belt, China [J]. *Science China Earth Sciences*, 59: 556–572.
- Zhao Y, Chen S, Tian H, *et al.*, 2021. Trace element and S isotope characterization of sulfides from skarn Cu ore in the Laochang Sn-Cu deposit, Gejiu district, Yunnan, China: Implications for the ore-forming process [J]. *Ore Geology Reviews*, 134: 104155.
- Zhong R, Deng Y, Li W, *et al.*, 2021. Revealing the multi-stage ore-forming history of a mineral deposit using pyrite geochemistry and machine learning-based data interpretation [J]. *Ore Geology Reviews*, 133: 104079.
- Zhou C, Yang Z, Sun H, *et al.*, 2022. LA-ICP-MS trace element analysis of sphalerite and pyrite from the Beishan Pb-Zn ore district, south China: Implications for ore genesis [J]. *Ore Geology Reviews*, 150: 105128.
- Zhuang L, Song Y, Liu Y, *et al.*, 2019. Major and trace elements and sulfur isotopes in two stages of sphalerite from the world-class Angouran Zn-Pb deposit, Iran: Implications for mineralization conditions and type [J]. *Ore Geology Reviews*, 109: 184–200.
- Zou S, Chen X, Brzozowski M J, *et al.*, 2022. Application of machine learning to characterizing magma fertility in porphyry Cu deposits [J]. *Journal of Geophysical Research: Solid Earth*, 127(8): e2022JB024584.