文章编号：1004—5589（2025）01—0131—11

朴泰圣，赵庆英，范国宇，等. 基于蜂群优化的单类支持向量机在多元地球化学异常识别中的应用：以内蒙古阿木辉伊勒特地区为例［J］. 世界地质，2025，44（1）：131-141. DOI：10.3969/j.issn.1004-5589.2025.01.010.

PIAO T S，ZHAO Q Y，FAN G Y，et al. Application of one-class support vector machine based on bee colony optimization in multivariate geochemical anomaly identification：a case study of Amuhuiyilete region in Inner Mongolia［J］. World Geology，2025，44（1）：131-141. DOI：10.3969/j.issn.1004-5589.2025.01.010.

基于蜂群优化的单类支持向量机在多元地球化学异常识别中的应用：以内蒙古阿木辉伊勒特地区为例

朴泰圣，赵庆英*，范国宇，赵科宇，张晟硕

吉林大学 地球科学学院，长春 130061

摘要：勘查地球化学是快速圈定区域找矿远景区最有效的方法之一。这种方法虽然能够快速圈定地球化学找矿远景区，却忽略了地球化学背景的空间变化性，存在遗漏弱缓地球化学异常。为在复杂地质环境下识别多元地球化学异常，笔者选择单类支持向量机模型进行研究。该方法可以在无需对数据分布做出任何假设的情况下进行高维异常检测。以阿木辉伊勒特地区为例，在1∶5万区域地质调查成果基础上，用Surfer软件对研究区水系沉积物数据中的11种地球化学元素数据进行网格化处理，以研究区已知矿点的空间位置为基础，生成网格化“地真”数据，统计检测每一种地球化学元素与已知矿点之间的空间关联性，把元素浓集作用与已知矿点存在显著关联性的元素作为找矿指示元素。在研究区共选出3种指示元素，将3种指示元素的网格化数据作为单类支持向量机的输入数据，进行多元地球化学异常识别研究。用试错法和人工蜂群优化算法对模型进行优化，获得2种模型的输出结果，结合“地真”数据，绘制试错法和人工蜂群优化算法优化后模型的受试者工作特征曲线（ROC），并计算相应的曲线下面积（AUC）值。结果显示，试错法优化的模型AUC值为0.879 6，而人工蜂群算法优化的模型AUC值为0.897 8。同时，2种模型识别的异常网格数量占比分别为27.14%和23.65%。表明在异常检测任务中，人工蜂群算法优化的模型性能略优于试错法优化的模型。基于人工蜂群算法优化的单类支持向量机能够更加有效地识别异常数据点，提升整体模型的准确性。

关键词：单类支持向量机；人工蜂群优化；地球化学异常；受试者工作特征曲线；曲线下面积；约登指数；阿木辉伊勒特地区；内蒙古

中图分类号：**TP**751　　**文献标识码**：**A**　　**DOI**：10.3969/j.issn.1004-5589.2025.01.010

Application of one-class support vector machine based on bee colony optimization in multivariate geochemical anomaly identification：a case study of Amuhuiyilete region in Inner Mongolia

PIAO Taisheng，ZHAO Qingying*，FAN Guoyu，ZHAO Keyu，ZHANG Shengshuo

College of Earth Sciences，Jilin University，Changchun 130061，China

收稿日期/Received：2024-08-30；**修订日期/Revised**：2024-11-22；**出版日期/Published**：2025-02-25

基金项目：国家自然科学基金项目（42172324）

第一作者：朴泰圣（1999—），男，硕士研究生，主要从事数字地质学研究工作。E-mail：piaots22@mails.jlu.edu.cn

*通信作者：赵庆英（1966—），女，教授，主要从事数字地质学研究工作。E-mail：zhaoqy@jlu.edu.cn

Abstract: The exploration geochemical method is one of the most effective methods to quickly delineate regional prospective areas. Although this method can quickly delineate geochemical prospective areas, it often ignores the spatial variability of geochemical backgrounds, potentially missing weak geochemical anomaly. In order to identify multivariate geochemical anomaly in complex geological environments, the authors select a one-class support vector machine (OCSVM) model for this study. The model allows for high-dimensional anomaly detection without making any assumptions about data distribution. Taking the Amuhuiyilete region as an example, based on the 1∶50 000 regional geological survey results, the data of 11 geochemical elements from stream sediments in the study area were gridded using Surfer software. The gridded "true" data were generated based on the spatial locations of known mining points in the study area. The spatial correlation between each geochemical element and the known mining points was statistically analyzed, and elements with significant correlation to the known mining points and concentrated elemental distribution were identified as prospecting indicator elements. In the study area, three indicator elements were selected. The gridded data of these three indicator elements were used as input data for OCSVM to conduct multivariate geochemical anomaly identification research. The models were optimized using both the trial-and-test method and the artificial bee colony (ABC) optimization algorithm. The output results of both models were obtained and combined with the "true" data, receiver operating characteristic (ROC) curves were then plotted for the models optimized by the trial-and-test method and the ABC optimization algorithm, and corresponding area under the curve (AUC) values were calculated. The results show that the AUC value of the model optimized by the trial-and-test method is 0.879 6, while the AUC value of the model optimized by the ABC algorithm is 0.897 8. At the same time, the proportion of anomalous grids identified by the two models is 27.14% and 23.65%, respectively. This indicates that, in the anomaly detection task, the model optimized by the ABC algorithm performs slightly better than the model optimized by the trial-and-test method. The OCSVM optimized by the ABC algorithm is more effective in identifying anomalous data points, and improving the overall model accuracy.

Keywords: one-class support vector machine; artificial bee colony optimization; geochemical anomaly; receiver operating characteristic curve; area under the curve; Youden index; Amuhuiyilete region; Inner Mongolia

0 引言

异常识别是通过发现数据集中偏离预期行为的模式来解决问题的一种方法。在地质学领域，异常往往是指与周边显著不同的地质结构、构造现象或其组合。这些地质异常通过地球物理、地球化学和遥感影像等手段表现出来。对这些异常的深入研究与正确解读，对矿产资源的预测以及某一区域成矿规律的总结具有至关重要的意义。在矿物勘探中，地球化学异常被定义为给定区域背景地球化学模式的变化[1]。但在勘探过程中，无论是异常检测方法还是监督分类方法都无法取得满意的结果。这是由于异常检测方法本身的局限性以及监督分类方法无法建立极端类别不平衡地球化学勘探数据的分类模型[2]。如何突破异常检测方法和监督分类方法的局限性，建立高性能地球化学异常探测模型是一个值得研究的科学问题。

近年来，机器学习和深度学习异常识别方法较常用的传统异常检测技术获得更令人满意的高维地球化学异常检测结果，因为机器学习和深度学习技术不需要地球化学数据遵循任何已知的理论分布。目前已应用于多元化探异常检测的无监督机器学习算法主要有连续受限玻尔兹曼机[3]、支持向量机[4]、高斯混合模型[5]、距离异常因素[6]、隔离森林[7]、无监督最近邻算法[8]和字典学习算法[9]。把机器学习中的异常识别算法应用于地球化学异常识别，是实现复杂地质环境中弱缓地球化学异常识别的可行途径之一。

单类支持向量机模型是支持向量机的一种拓展[10]，是数据挖掘中一种经典的无监督异常值检测方法。它可以在对高维数据进行建模时产生有用的结果，而无需对内在数据的分布进行任何假设。该算法的异常识别原理是，在给定的置信水平下，寻找输入空间的一个最小区域（输入空间的子集）

作为观测数据高维统计分布的支持集，多元异常值是从高维统计分布中抽取但位于支持集之外的样本。该模型已成功应用于新颖性和异常值检测[11-13]，并在对高维数据建模时获得有用的结果。

人工蜂群（artificial bee colony，ABC）算法是一种群体智能优化算法，可以解决各种复杂组合的优化问题。同时具备与蝙蝠算法[14]相似的全局搜索能力和局部搜索能力，因此可以与机器学习算法结合解决大规模的优化问题。单类支持向量机需要在地球化学数据建模中定义一组初始化参数。该模型中，使用默认或未经优化的超参数将导致模型性能下降。因此，针对这些超参数的调优至关重要。笔者采用人工蜂群算法对模型的超参数进行优化。在蜂群优化单类支持向量机模型中，搜索空间是一个由 σ 和 υ 组成的二维空间。蜂群的搜索范围定义为（$0<\sigma\leqslant 1$）和（$0<\upsilon<1$）。蜂群优化的迭代搜索过程最大化的适应度值是单类支持向量机模型的 AUC 值，迭代搜索过程从搜索空间内的 N 个随机位置开始。在每次迭代中，每个蜜蜂占据的空间位置的 2 个坐标被用作初始化单类支持向量机模型的 σ 和 υ 值，模型在数据 $\{x_1, x_2, \cdots, x_n\}$ 上进行训练。每个单元格的异常评分根据训练好的单类支持向量机模型计算。最后，单类支持向量机模型的最大 AUC 值位置被选为当前最优位置。

以内蒙古阿木辉伊勒特地区为例，在1∶5万区域地质调查成果基础上，分析整理了研究区水系沉积物数据和矿产数据。使用 Surfer 软件对 11 种观测数据进行网格化处理，以研究区已知矿点的空间位置为基础，生成网格化“地真”数据，统计检测每种地球化学元素与已知矿点之间的空间关联性，把元素浓集作用与已知矿点存在显著关联的元素作为找矿指示元素。共选出 3 种指示元素，将其网格化数据作为单类支持向量机的输入数据，进行多元地球化学异常识别研究。用试错法和 ABC 优化算法对模型优化，获得 2 个优化方法的模型输出结果，与“地真”数据结合，绘制试错法和人工蜂群算法优化后模型的 ROC 曲线，并计算相应的 AUC 值。比较 2 种模型识别多元地球化学异常的优化效果，人工蜂群算法的优化效果更好。

1　地质概况

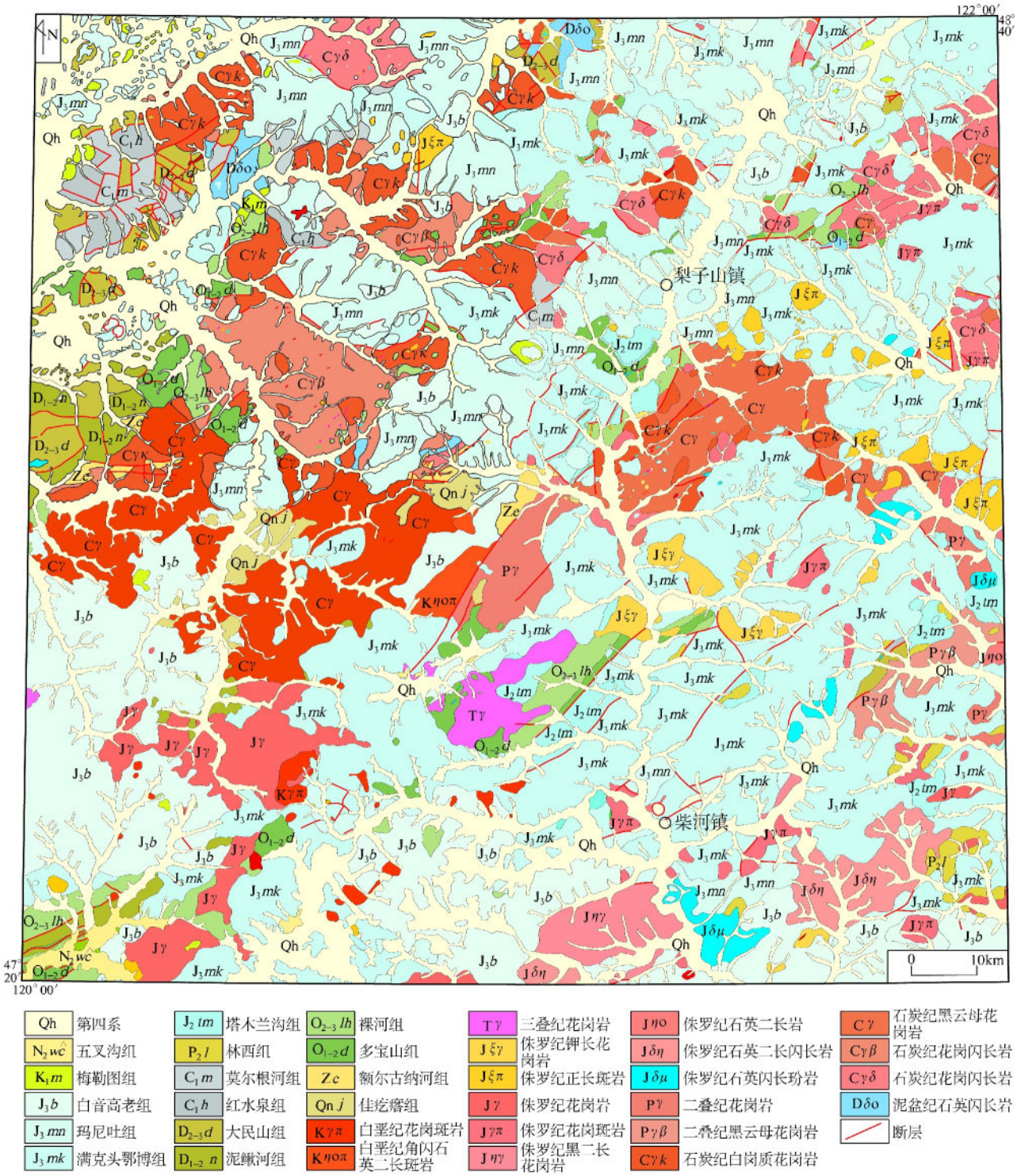
阿木辉伊勒特地区位于内蒙古东部呼伦贝尔市，大地构造位置属于内蒙古高原东部大兴安岭主峰的中段，前中生代时期处于华北板块与西伯利亚板块之间的中亚—蒙古造山带的东部—东乌珠穆沁旗早华力西造山带上；中生代以后，进入滨太平洋构造域—大兴安岭中生代火山活动带范畴。区内出露的地层由老至新为新元古界青白口系、中生界侏罗系及新生界。区内古生代属于东乌珠穆沁旗—扎兰屯火山型被动陆缘，中生代转化为大兴安岭岩浆岩带，形成了双层结构。其基底建造为陆缘碎屑岩-碳酸盐-中酸性火山岩建造组成，区内的基底地层主要表现为新元古界青白口系佳疙疸组（Qn*j*）及震旦系额尔古纳河组（Z*e*）。盖层建造为基性-中性-酸性火山岩建造，在区内形成的地层有塔木兰沟组（J_2tm）、满克头鄂博组（J_3mk）、玛尼吐组（J_3mn）、白音高老组（J_3b）（图 1）。

该矿区属于博克图—朝不楞钨、铁、锌、铅成矿带[15]，成矿带内的主要矿床有塔尔气铁矿（矽卡岩型）、绰源八十公里铁-铅锌多金属矿床等。区内的矿（床）点有塔尔气铁矿，大南沟铅、锌、银、金矿点，成矿围岩为二叠纪花岗岩，属于中-低温热液型；小南沟钼矿点，成矿围岩为二叠纪花岗岩，属于斑岩型钼矿；苏格河金、银、锌多金属矿点，成矿围岩为中生代塔木兰沟组玄武岩，成矿类型属于火山岩型金银矿；中心沟金铜矿点，成矿围岩为石炭纪花岗岩，属于浅成中-低温热液（石英脉型）金铜矿。

2　水系沉积物地球化学测量

内蒙古自治区第六地质矿产开发院①近年来在阿木辉伊勒特地区完成了1∶5万的水系沉积物地球化学测量（图 2）。研究区水系沉积物地球化学测量符合文献［16］地球化学水系沉积物采样技术要求，水系沉积物测量面积为 1 382.98 km^2，采样点密度为 8 点/km^2。对全测区样品分析结果进行统计，发现 Au、Ag、As、Sb、Cu、Pb、Zn、W、Sn、Bi、Mo 等 11 种矿化元素。使用 Surfer 软件对

① 吴国学，陈跃军，关继东，等. 内蒙古自治区呼伦贝尔市阿木辉伊勒特等四幅 1∶5 万区域矿产地质调查报告. 呼伦贝尔：内蒙古自治区第六地质矿产开发院，2012.



据文献①－③修改。

图1 阿木辉伊勒特地区地质简图

Fig. 1 Simplified geological map of Amuhuiyilete region

① 窦奇生，姜盛庭，秦鼐，等. 大黑沟幅1∶20万地质报告. 哈尔滨：黑龙江省地质局大兴安岭区域地质测量大队第十一分队，1958.

② 朱洪森，齐方云，董启贤，等. 一二五公里幅1∶20万区域地质调查报告. 呼和浩特：内蒙古自治区地质矿产局，1990.

③ 周沛然，马家骏，王莹，等. 塔尔其幅、绰尔幅1∶20万区域地质调查报告. 齐齐哈尔：黑龙江省区域地质调查第二队一分队，1981.

上述收集到的 11 种地球化学元素的 1∶5 万水系沉积物采样数据进行网格化处理，用反距离权重插值法生成 128×80 的网格数据[17]，当权重为 1 时会抑制以插值网格为中心的局部异常，权重为 3 时会加剧以插值网格为中心的局部异常，因此在本次研究中，将权重设为 2，搜索半径设为 0.02，生成一共包含 5 个已知矿床（矿化）点的 9 840 个约 0.14 km^2的网格。

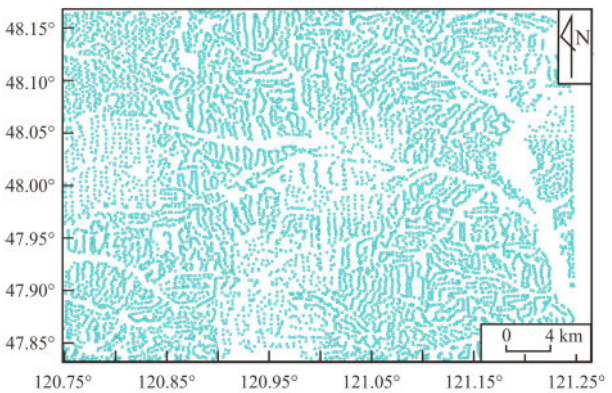


图 2　水系沉积物采样点位图

Fig. 2　Sample location of stream sediment survey

基于各地球化学指标的插值数据，可以通过其 AUC 值来衡量该地球化学指标与研究区已知矿点是否存在显著空间关联[4]，通过 S_{AUC}（即 AUC 的标准差）和 Z_{AUC}统计量进一步检验。AUC 取值在 0～1 之间，当某一元素 AUC 值接近于 0.5 时，该地球化学元素与已知矿点不存在显著空间关系；当某一元素 AUC 值接近于 1 时，说明该地球化学元素与已知矿点空间关系密切。Z_{AUC}是一个正太分布统计量，用于统计检验 AUC 值是否与 0.5 之间存在显著差异。如果计算出的 Z_{AUC}在 $\alpha=0.1$ 水平上超过临界值 1.64，则可以认为 AUC 值与 0.5 之间存在显著差异。即元素与已知矿点之间存在显著的空间相关性。根据表 1 的分析结果，可以选择 Ag、Mo 和 Zn 作为找矿指示元素。

3　异常识别方法

3.1　单类支持向量机

单类支持向量机在地球化学勘探中可用于将多元地球化学异常与背景分离。假设研究区域的多元

表 1　11 种地球化学元素的 AUC、S_{AUC}、Z_{AUC}

Table 1　AUC, S_{AUC} and Z_{AUC} of 11 geochemical elements

元素	AUC	S_{AUC}	Z_{AUC}
Au	0.547 6	0.132 4	0.359 1
Ag	0.823 1	0.114 6	2.818 0
As	0.464 7	0.125 7	−0.280 6
Bi	0.547 6	0.132 4	0.359 1
Cu	0.676 6	0.132 9	1.328 5
Mo	0.713 1	0.130 5	1.642 6
Pb	0.553 2	0.132 7	0.401 0
Sb	0.415 8	0.119 5	−0.704 1
Sn	0.439 8	0.122 7	−0.490 7
W	0.486 4	0.127 9	−0.106 3
Zn	0.729 3	0.129 1	1.777 0

地球化学数据满足未知的高维概率分布；并进一步假设 $X=\{x_1, x_2, \cdots, x_n\}$，$x_i \in \mathbf{R}^d$是一个训练样本集，其中 $n \in \mathbf{N}$ 是从未知高维概率分布中抽取的训练样本数量。那么，可以使用单类支持向量机算法来估计一个输入空间的子集 Γ，使得从未知高维概率分布中抽取的样本位于子集 Γ 之外的概率等于预定义的实值 $\varepsilon \in [0, 1]$[10]。参数 ε 表示位于子集 Γ 之外的训练样本的最大比例。它直接影响单类支持向量机算法在多元异常检测中的性能[13]。子集 Γ 可用于表示多元地球化学异常识别中的背景，并且位于子集 Γ 之外的测试样本被识别为异常样本。

给定参数 $\upsilon=1-\varepsilon$，以控制至少需要位于输入空间子集 Γ 中的训练样本的最小比例，即至少有 υ 个训练样本被分类为背景样本；参数 υ 还控制位于输入空间中允许落在子集 Γ 之外的训练样本的最大比例，即最多有 $(1-\upsilon)$ 个训练样本被识别为异常样本。参数 υ 的值可以通过经验预定义，或根据研究区域的矿产勘探水平采用试错法确定。

根据 Schölkopf 等[10]，单类支持向量机算法通过定义子集 Γ 的边界来估计未知高维概率分布的支持，即至少有 υ 个训练样本来自高维概率分布。在多元异常检测中，使用决策函数 $f(x)$ 来判断样本 x 是否为正常样本，即判断样本是否属于子集 Γ[11-13]。在多元地球化学异常识别中，可以定义相同形式的决策函数 $f(x)$ 来判断样本 x 是否为背景样本，记作 $f: \mathbf{R}^d \rightarrow \mathbf{R}$，使其满足：

$$f(x) > 0, \text{ if } x \in \Gamma; \quad f(x) < 0, otherwise \tag{1}$$

这个函数的含义是，当$f(x)$的值大于0时，样本x被视为背景样本；否则x被视为异常样本。

在支持向量机中，可能的函数空间$f(x)$被限制在一个再生核希尔伯特空间中，核函数为K：$\mathbf{R}^{\mathbf{d}} \rightarrow \mathbf{R}$。该核通过特征映射$\Phi$：$\mathbf{R}^{\mathbf{d}} \rightarrow H$引入所谓的特征空间H。可以使用$\langle , \rangle_H$来表示特征空间H中的点积。常用的核函数是高斯核[11-13]，定义为：

$$K(x,y)\langle \Phi(x), \Phi(y) \rangle_H = \exp\left(\frac{- \| x - y \|^2}{2\sigma^2}\right) \tag{2}$$

式（2）中$\| \cdot \|^2$是定义在$\mathbf{R}^{\mathbf{d}}$上的l_2范数。该核依赖于一个参数σ，与核的传播度相关。根据公式（2），可以有$K(x, x) = 1$。因此，所有地球化学样本都被映射到以空间H的原点为中心的单位半径超球面上。训练单类支持向量机模型的核心是定义分离超平面$W = \{h \in H \mid h, w_H - b = 0\}$，使得超平面到原点的距离（由$\| w \|_H$表示）最大化。参数$w$和$b$是以下优化问题的解。

$$Minimize\left[\frac{1}{2} \| w \|_H^2\right] + \left[\frac{1}{(1-v)m}\sum_{i=1}^{m} \xi_i - b\right] \tag{3}$$

约束条件为：$\langle w, \Phi(x_i) \rangle_H \geqslant b - \xi_i$，式（3）中，$\xi_i \leqslant 0$，$i = 1, 2, \cdots, m$，$\xi_i$为表示样本$x_i$相关损失的松弛变量（非零$\xi_i$允许一些异常样本）。与此问题相关的拉格朗日乘数$\alpha$可以确定$w$和$b$。使用Loqo[18]算法计算出对偶变量后，决策函数变为：

$$f(x) = \sum_{i=1}^{m} \alpha_i K(x_i, x) - b \tag{4}$$

式（4）中：$b = \sum_{i=1}^{m} \alpha_i K(x_i, x_j)$，$j \in [1, 2, \cdots, m]$，$\alpha_i (i = 1, 2, \cdots, m)$为拉格朗日不定乘数。

根据公式（4），计算地球化学样本群体中每个样本x的决策函数值。如果样本x的决策函数值为正，则样本x是背景样本；否则，它是异常样本。然而，在多元地球化学异常识别中，对异常样本比背景样本更感兴趣。因此，通过将$f(x)$的值乘以-1来修改决策函数值。结果是，修改后的决策函数值越大，样本x是异常的概率越高。因此，如果地球化学样本的决策函数值大于零，则可以将其识别为多元地球化学异常。在研究区域中存在已知矿床的情况下，可以通过选择一个阈值来确定最佳阈值，该阈值介于决策函数值的最大值和最小值之间，并且可以最大化识别出的地球化学异常与已知矿床之间的空间关联。约登指数[19-21]可用于表达地球化学异常与已知矿床之间的空间关联，并选择与最大约登指数对应的值作为最佳阈值。

3.2 蜂群优化的单类支持向量机

蜂群优化算法是一种受蜜蜂觅食行为启发的优化算法，通过模拟蜜蜂使用花蜜源检测食物和避开障碍。它将蜂群中的个体映射到D维问题空间中的可行解，并利用蜜蜂寻找花蜜的过程来模拟优化搜索过程。解决问题的适应度值用于评估蜜蜂的位置，利用优胜劣汰的进化过程模拟迭代搜索过程，以替代较差的可行解。蜂群优化算法动态控制局部搜索和全局搜索之间的转换，以避免算法陷入局部最优，在解决大规模目标优化问题上具有良好的全局收敛性和优异的性能。

在ABC算法中，人工蜂群包含3类蜜蜂：雇佣蜂、观察蜂和侦察蜂。负责在食物源（解）附近搜索新的食物源（解）称为雇佣蜂，根据雇佣蜂共享的信息（解的质量），选择一个食物源进行搜索的称为观察蜂。在当前食物源（解）失效或没有进展时，寻找新的食物源（解）称为侦查蜂。

ABC算法的主要步骤如下：

（1）初始化；

（2）重复：

（a）将雇佣蜂放置在内存中的食物上；

（b）将观察蜂放置在记忆中的食物源上；

（c）派遣侦查蜂前往搜索区域寻找新的食物来源；

（3）直到满足要求。

根据Dervis[22]的研究，在ABC算法中，食物源的位置代表优化问题的可能解，食物源的花蜜量对应于相关解的质量（适应度）。雇佣蜂或观察蜂的数量等于种群中解的数量。第一步，ABC算法生成N解（食物源位置）的随机分布初始种群P，

其中 N 表示种群大小。每个解（食物源）x_i（$i = 1, 2, \cdots, N$）都是一个 D 维向量。这里，D 是优化参数的数量。在初始化后，蜂群的食物源位置（即问题的解决方案）会经过雇佣蜂、观察蜂和侦查蜂的搜索过程，这一过程以循环的形式重复进行，依次为 C_1，C_2，…，$C_{最大限度}$。雇佣蜂或观察蜂概率性地对其记忆中的食物源位置（解决方案）进行修改，以寻找新的食物源，并测试新食物源（新解决方案）的花蜜量（适应度值）。对于真正的蜜蜂，新食物源的生产是基于对一个地区食物源的比较过程，具体取决于蜜蜂通过视觉收集的信息。在我们的模型中，新食物源位置的产生也是基于食物源位置的比较过程。然而，在模型中，人工蜜蜂没有使用任何信息进行比较。他们随机选择一个食物源位置，并对记忆中存在的位置进行修改。如果新来源的花蜜量高于前一来源的花蜜量，蜜蜂就会记住新位置并忘记旧位置。否则它会保留前一个的位置。所有雇佣蜂完成搜索过程后，将食物源的花蜜信息及其位置信息与信息交流区的观察蜜蜂分享。观察蜂评估所有受雇蜜蜂获取的花蜜信息，并以与其花蜜量相关的概率选择食物源，与雇佣蜂的情况一样，观察蜂对记忆中的位置进行修改，并检查候选源的花蜜量。如果花蜜比前一个候选源更高，蜜蜂就会记住新的位置并忘记旧的位置。

观察蜂根据与该食物源相关的概率值 $p_i(prob[i])$ 选择食物源，该概率通过以下表达式计算：

$$prob[i] = \frac{fitness[i]}{\sum_{n=1}^{N} fitness[n]} \tag{5}$$

式（5）中：$fitness[i]$ 是解决方案，由其雇佣蜂评估适应度值，它与位置 i 的食物源的花蜜量成正比，N 是食物源的数量，等于雇佣蜂的数量。通过这种方式，雇佣蜂与观察蜂交换信息。

在人工蜂群算法优化单类支持向量机模型中，搜索空间是一个由 σ 和 υ 组成的二维空间。蜂群的搜索范围定义为（$0 < \sigma \leqslant 1$）和（$0 < \upsilon < 1$）。人工蜂群优化的迭代搜索过程最大化的适应度值是单类支持向量机模型的 AUC 值。

在人工蜂群优化算法中，适应度值通常是目标函数值的一个转换。因为很多优化算法都是最小化问题，为了兼容最小化问题，通常使用负的目标函数值作为适应度值，这样目标函数值越大，适应度值越小。

具体来说，适应度值的计算公式是：

$$fitness = -\mathrm{AUC} \tag{6}$$

这是因为我们的目标是最大化 AUC 值，而 ABC 算法是一个最小化优化算法。负号的使用将最大化问题转换为最小化问题。

迭代搜索过程从搜索空间内的 N 个随机位置开始。在每次迭代中，每个蜜蜂占据的空间位置的 2 个坐标被用作初始化单类支持向量机模型的 σ 和 υ 值，然后模型在数据 $\{x_1, x_2, \cdots, x_n\}$ 上进行训练。每个单元格的异常评分根据训练好的单类支持向量机模型计算。最后，单类支持向量机模型的最大 AUC 值的位置被选为当前最优位置。

4　结果与讨论

4.1　案例分析

选择中国内蒙古阿木辉伊勒特地区作为案例研究区，因为该地区地质环境复杂，并且几年前已经完成水系沉积物地球化学采样调查。水系沉积物中的 11 种地球化学元素的浓度数据在第 2 节已完成网格化处理，选取 Ag、Mo 和 Zn 三种化学元素作为找矿指示元素。由于这三种化学元素指标与研究区已知矿点（床）具有显著的空间关联性，把这些化学指标的网格化数据作为 OCSVM 模型的输入数据，用于提取研究区的地球化学异常。在研究过程中，试错法和人工蜂群优化算法用于优化 OCSVM 模型参数，使模型能够更好地识别异常。同时，比较试错法和人工蜂群优化算法优化后的模型在多元地球化学异常识别中的性能。

4.2　模型的参数选择

在异常识别工作中，单类支持向量机的识别效果受参数 σ 和 υ 的取值影响，在确定参数的单类支持向量机模型上用选定的指示元素作为输入数据进行训练，然后计算每个晶胞的异常分数。根据所有单元的异常得分，通过约登指数确定的最佳阈值提取地质异常[19-21]，所提取的地质异常通常在空间上与已知矿藏密切相关。在本研究中，单类支持向量机或使用试错法确定参数，或通过蜂群算法来进

行优化。经实验确定模型的参数 υ，其他参数可以根据文献［4］，使用默认值进行初始化（“kernel”用“rbf”初始化，“σ”用“$1/d$”初始化，“d”表示指示元素的数量）。本研究中参数 υ 的取值分别定义为 0.50、0.55、0.60、0.65、0.70、0.75、0.80、0.85、0.90 和 0.95，σ 默认为 0.33。然后计算不同 υ 值下的 AUC 值，绘制 AUC 随参数 υ 值变化的图（图 3a）。

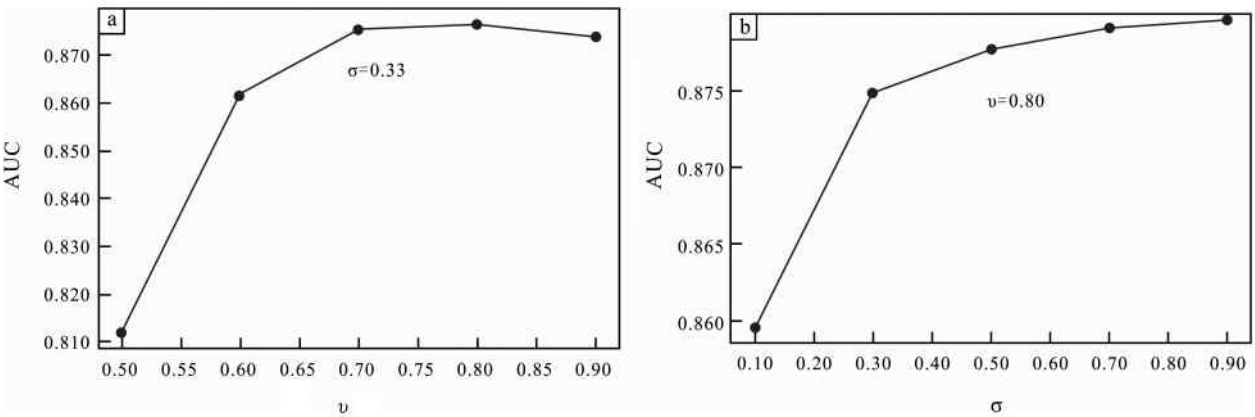


图 3 单类支持向量机的 AUC 随 υ 值（a）和随 σ 值（b）变化曲线图

Fig. 3 Variation curves of AUC value of OCSVM model changing with υ (a) and σ (b)

由图 3a 可以看出，模型的 AUC 在 υ = 0.80 达到最大值，因此，选择 υ = 0.80 来进行模型的“最佳”υ 值，然后设置 σ = 0.10、0.30、0.50、0.70 和 0.90，使用之前确定好“最佳”υ 值，计算每个 σ 值下的模型 AUC 值，置于图 3b 中。根据图 3a 和图 3b 可以看出，υ 和 σ 的“最佳”值分别为 0.80 和 0.90，相应的最大 AUC 值为 0.879 6。

在使用人工蜂群算法优化模型参数时，根据 3.2 节中 ABC 使用了 3 个控制参数：食物源的数量（等于雇佣蜂或观察蜂的数量）、限制值和最大迭代次数。

通过经验和交流，将最大迭代次数和限制值均设定为 50 次，之后在 20、30、40 和 50 的蜂群数量下分别对单类支持向量机模型的 2 个关键参数 σ 和 υ 在给定的范围找寻最优解，将不同参数下的模型 AUC 值作为蜂群优化的最大适应度值，绘制随着迭代次数不同蜂群数量下 AUC 值的变化曲线。选择 AUC 值最大的参数组合用于单类支持向量机模型的训练（图 4），可以看出，经过 50 次迭代后，不同蜂群数量下模型的 AUC 值最终都稳定在一个值上，且在 40 的蜂群数量下，模型的 AUC 值最大。因此，在本研究中选择 40 蜂群得出的参数组合（σ = 0.81，υ = 0.06）用于单类支持向量机的训练。

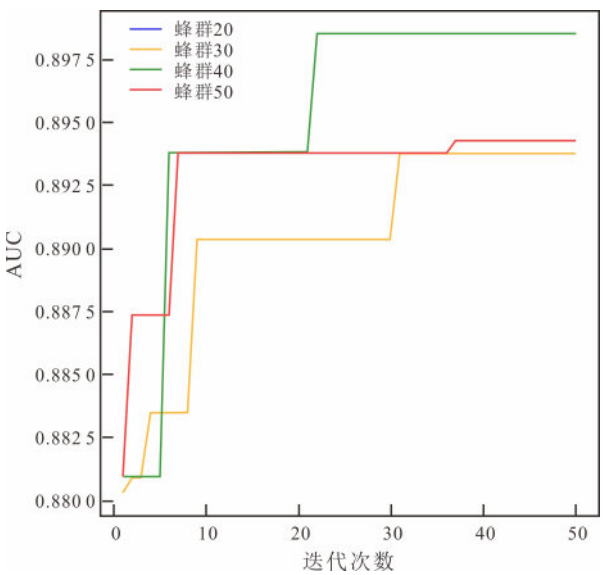


图 4 AUC 在不同蜂群下的变化曲线图

Fig. 4 Variation curve of AUC under different bee colonies

4.3 异常识别结果

为了评价多元地球化学异常识别结果的有效性，必须确定含矿床网格点（即真阳性样本）和非矿床网格点（即真阴性样本）来计算成本、收益和约登指数，进行 ROC 曲线分析[19-21]。

应用约登指数[19-21]来量化已识别的多元地球

化学异常与已知矿床之间的空间关联。基于模型决策函数值，使用均匀分布在异常指数数据的最大值和最小值之间的多个阈值（本文中使用的值为 1 000）对异常指数数据进行分类。将网格点划分为多元地球化学异常和背景；针对每个阈值，计算含有已知矿藏的异常网格点的个数、不含已知矿藏的异常网格点的个数、含有已知矿藏的背景网格点的个数、以及存在已知矿藏的背景网格点的个数。对不含已知矿藏的背景网格点进行计数并用于计算成本、收益[20]和约登指数。最终选择相对最大约登指数的最佳阈值来描绘多元地球化学异常，以此最优阈值识别多元地球化学异常热点图与已知矿床点具有最大程度的空间关联性。故所提取的元素浓度异常具有更好的找矿指示意义。

使用在 4.1 中确定的参数，将网格化处理的指示元素数据分别输入到试错法单类支持向量机模型和 ABC 优化单类支持向量机模型中，得到相应的决策函数值，根据最大约登指数确定最佳阈值，并绘制约登指数随阈值变化的曲线图（图 5、6），同时列于表 2 中。

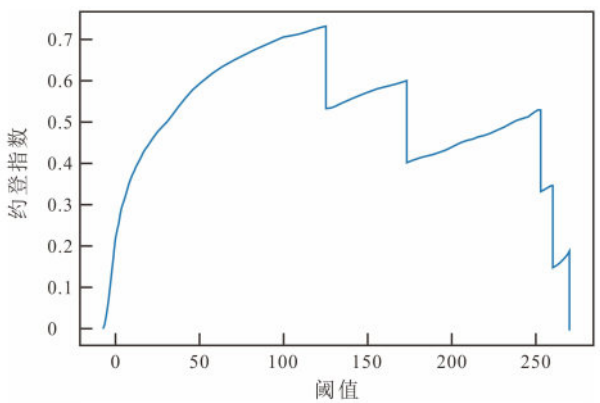


图 5　试错法约登指数随阈值变化曲线

Fig. 5　Variation curve of Youden Index with threshold using trial-and-error method

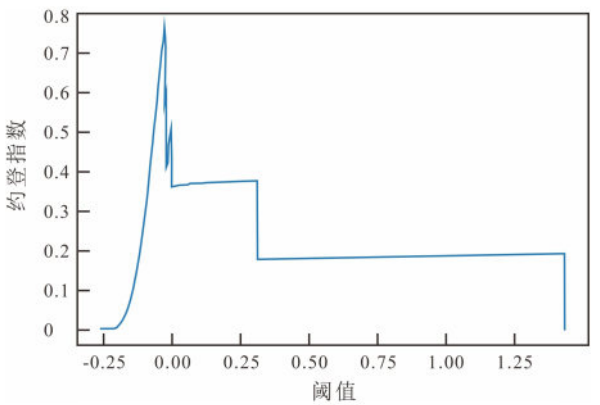


图 6　ABC 优化算法约登指数随阈值变化曲线

Fig. 6　Variation curve of Youden Index with threshold using ABC optimized algorithm

表 2　两种方法决策函数值的最大约登指数及最佳阈值

Table 2　Maximum Youden index and optimal threshold of decision function values for two methods

	约登指数	最佳阈值
试错法	0.728 967	125.363 3
ABC 优化法	0.763 925	−0.024 9

如表 2 所示，得到两种方法的最佳阈值后，使用 Surfer 软件将两个方法所得的决策函数值绘制成等值线图，作为矿产前景图，并将已知矿点投在图上（图 7、8）。

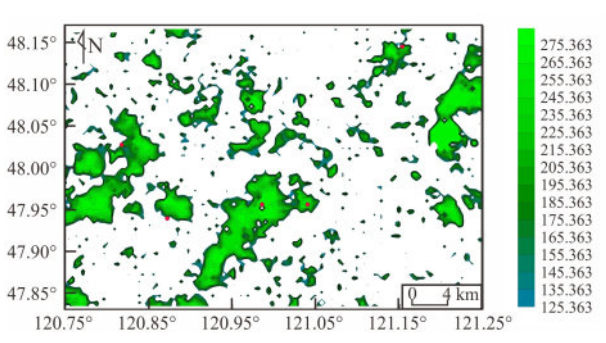


图 7　试错法多元地球化学异常识别得分等值线图

Fig. 7　Trial-and-test method for contour map of multivariate geochemical anomaly identification scores

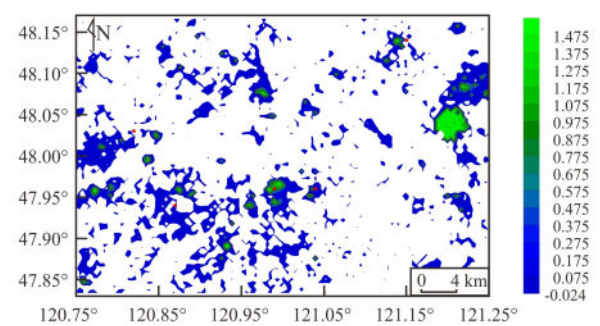


图 8　ABC 优化法多元地球化学异常识别得分等值线图

Fig. 8　ABC optimized algorithm for contour map of multivariate geochemical anomaly identification scores

4.4 讨论

本文将AUC值、Z_{AUC}值、异常网格数量占比以及异常网格中含矿点比例作为单类支持向量机模型的多元地球化学异常识别效果的量化指标。绘制ROC曲线（图9），并计算试错法和人工蜂群算法优化后单类支持向量机模型的AUC值，同时计算相应的Z_{AUC}值，异常网格数量占比和异常网格含矿点数量均由程序运行所得（表3）。

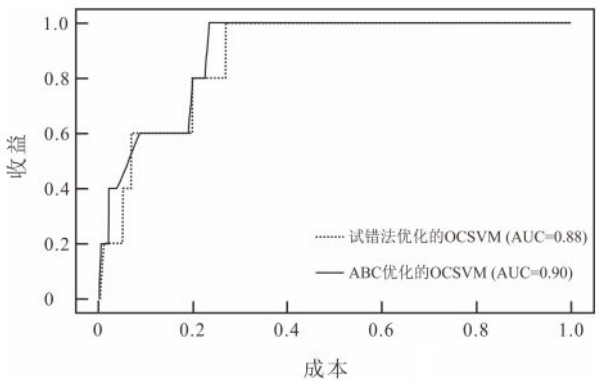


图9 优化后OCSVM的ROC曲线图

Fig. 9 ROC curves of optimized OCSVMs

表3 两种模型的Z_{AUC}、识别的异常网格占比和异常网格中的含矿比例

Table 3 Z_{AUC}, proportion of identified anomalous grids, and mineral content ratio within anomalous grids for two models

模型	AUC	Z_{AUC}	异常网格占比/%	矿点占比异常网格/%
试错法-OCSVM	0.879 6	3.812 7	27.14	100
ABC-OCSVM	0.897 8	4.268 6	23.65	100

从表3和图9中可以看出：①试错法和ABC优化算法优化后的模型Z_{AUC}值均大于置信水平为0.05时的临界值1.96。因此，使用上述两种方法确定参数的单类支持向量机模型的多元地球化学异常识别结果与研究区已知矿点之间存在显著的空间关联性；②经过人工蜂群优化后的单类支持向量机模型的AUC值大于使用试错法的“最佳”参数的单类支持向量机模型，表明人工蜂群优化后的单类支持向量机的多元异常识别效果更优；③在矿点占比相同的情况下，经过人工蜂群优化的模型识别出来的异常网格数量更少，说明模型对异常网格的识别更加精准，人工蜂群优化算法有效提高了模型的精确度，达到实验前对人工蜂群优化模型的期望效果。

5 结论

（1）无论是试错法优化的模型，还是人工蜂群优化的模型，基于两种优化方法的单类支持向量机提取的多元地球化学异常与研究区已知矿点存在的显著空间关联性，表明单类支持向量机是一种可行的多元地球化学异常识别方法。

（2）在异常检测任务中，试错法优化的模型AUC值为0.879 6，而人工蜂群优化的模型AUC值为0.897 8。同时，基于人工蜂群优化的单类支持向量机能够更加有效地识别异常数据点，提升整体模型的准确性。人工蜂群优化方法比试错法具有更为明显的优势。因此，蜂群优化单类支持向量机在多元地球化学异常识别中展现出极大的潜力和实用价值，是一种高性能的找矿预测方法。

参考文献：

[1] CARRANZA E J M. Geochemical anomaly and mineral prospectivity mapping in GIS. 11 [M]. Oxford: Elsevier Science & Technology, 2008: 368.

[2] CHEN Y L, SUI Y H, SHAYILAN A. Constructing a high-performance self-training model based on support vector classifiers to detect gold mineralization-related geochemical anomaly for gold exploration targeting [J]. Ore Geology Reviews, 2023, 153: 105265.

[3] CHEN Y L, LU L J, LI X B. Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly [J]. Journal of Geochemical Exploration, 2014, 140: 56-63.

[4] CHEN Y L, WU W. Application of one-class support vector machine to quickly identify multivariate anomaly from geochemical exploration data [J]. Geochemistry Exploration Environment Analysis, 2017, 17: 231-238.

[5] CHEN Y L, WU W. Separation of geochemical anomaly from the sample data of unknown distribution population using Gaussian mixture model [J]. Computers & Geosciences, 2019, 125: 9-18.

[6] CHEN Y L, SUN G S, ZHAO Q Y. Detection of multivariate geochemical anomaly associated with gold deposits by using distance anomaly factors [J]. Journal of Geochemical Exploration, 2021, 221: 106704.

[7] CHEN Y L, WANG S C, ZHAO Q Y, et al. Detection of multivariate geochemical anomaly using the bat-optimized isolation forest and bat-optimized elliptic envelope models [J]. Journal of Earth Science, 2021, 32 (2): 415–426.

[8] CHEN Y L, ZHAO Q Y, LU L J. Combining the outputs of various k-nearest neighbor anomaly detectors to form a robust ensemble model for high-dimensional geochemical anomaly detection [J]. Journal of Geochemical Exploration, 2021, 231 (Suppl. C): 106785.

[9] CHEN Y L, SHAYILAN A. Dictionary learning for multivariate geochemical anomaly detection for mineral exploration targeting [J]. Journal of Geochemical Exploration, 2022, 235: 106958.

[10] SCHÖLKOPF B, PLATT J C, SMOLA A J, et al. Estimating the support of a high-dimensional distribution [J]. Neural Computation, 2001, 13 (7): 1443–1471.

[11] HAYTON P, SCHÖLKOPF B, TARASSENKO L, et al. Support vector novelty detection applied to jet engine vibration spectra [J]. Advances in Neural Information Processing Systems, 2001, 13: 946–952.

[12] DAVY M, GODSILL S. Detection of abrupt spectral changes using support vector machines: an application to audio signal segmentation [C] // DAVY M, GODSILL S. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002). Orlando, Florida, USA: IEEE, 2002: 1313–1316.

[13] LECOMTE S, LENGELLE R, RICHARD C, et al. Abnormal events detection using unsupervised one-class SVM: application to audio surveillance and evaluation [C] // LECOMTE S, LENGELLE R, RICHARD C, et al. Proceedings of 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS 2011). Klagenfurt, Austria: IEEE, 2011: 124–129.

[14] CHEN Y L, WU W, ZHAO Q Y. A bat-optimized one-class support vector machine for mineral prospectivity mapping [J]. Minerals, 2019, 9 (5): 317.

[15] 陈郑辉，朱裕生，王保良，等. 内蒙古主要成矿区带及其矿产资源潜力分析 [J]. 西部资源，2005，4：4–9.
CHEN Z H, ZHU Y S, WANG B L, et al. Analysis of major metallogenic belts and their mineral resource potential in Inner Mongolia [J]. Western Resources, 2005, 4: 4–9.

[16] 陈国光，张华，叶家瑜，等. 地球化学普查规范 (1∶50 000)：DZ/T 0011—2015 [S]. 中华人民共和国国土资源部，中国地质调查局南京地质调查中心，2015：1–39.
CHEN G G, ZHANG H, YE J Y, et al. Geochemical survey specification (1∶50 000): DZ/T 0011–2015 [S]. Nanjing Geological Survey Center of China Geological Survey, Ministry of Natural Resources of the People's Republic of China, 2015: 1–39.

[17] CHEN Y L, AN A J. Application of ant colony algorithm to geochemical anomaly detection [J]. Journal of Geochemical Exploration, 2016, 164: 75–85.

[18] VANDERBEI, ROBERT J. LOQO: an interior point code for quadratic programming [J]. Optimization Methods and Software, 1999, 11/12 (1/4): 451–484.

[19] CHEN Y L. Mineral potential mapping with a restricted Boltzmann machine [J]. Ore Geology Reviews, 2015, 71: 749–760.

[20] CHEN Y L, WU W. A prospecting cost-benefit strategy for mineral potential mapping based on ROC curve analysis (Article) [J]. Ore Geology Reviews, 2016, 74: 26–38.

[21] CHEN Y L, WU W. Mineral prospectivity mapping using an extreme learning machine regression [J]. Ore Geology Reviews, 2017, 80: 200–213.

[22] DERVIS K, BAHRIYE B. A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm [J]. Journal of Global Optimization, 2007, 39 (3): 459–471.