



TAS-PF: 基于大数据概率场的 TAS 扩展图解

于秋野¹, 李绍竑², 王智岩¹

1. 吉林外国语大学 人工智能学院, 吉林 长春 130117;

2. 北京世纪好未来教育科技有限公司博物工作室, 北京 100085

摘 要: 大数据时代背景下, 地学数据规模持续增长. 以 TAS 为代表的传统图解面临困境: 一方面, 有限图幅内投点过多导致可读性降低, 无法呈现有效、直观的可视化效果; 另一方面, 原始数据陈旧的传统图解若引入新数据可能导致分类边界发生扰动, 从而降低判别分类结果的稳定性, 并难以兼容已有文献投图. 针对上述问题, 本文首先继承前期研究为 TAS 图解所做的扩展, 为经典图解中的各岩性标签构建基于空间位置的类别分区. 根据待分类数据投图位置与各类别分区空间关系进行判别, 并以数据表形式呈现分类结果, 从而弥补数据规模增大带来的投图可读性降低. 另外, 从 GEOROC 数据库中提取 24 万余条火成岩的主量元素数据, 将其在 TAS 图解上进行可视化, 并按岩性分类进行核密度分析. 基于分析结果在投图坐标范围内构建对应类别概率场, 基于待分类数据在各概率场中所处位置信息计算概率, 并对比不同岩性标签的概率结果. 基于概率场利用已知岩性标签数据判别待分类数据, 补充传统分类边界模式, 并提供更具有定量意义的判别结果.

关键词: TAS 图解; 火成岩; 岩性分类; 判别图; 大数据; 概率场

TAS-PF: Extended TAS diagram powered with probability field of big data

YU Qiu-ye¹, LI Shao-hong², WANG Zhi-yan¹

1. School of Artificial Intelligence, Jilin International Studies University, Changchun 130117, China;

2. TAL Education Group, Beijing 100085, China

Abstract: Under the background of big data, the continuous growth of geological data poses challenges to traditional discrimination diagrams represented by TAS: On one hand, the excessive data points within limited diagram space reduce readability and hinder effective visualization; On the other hand, the input of new data into traditional diagrams with outdated original data may lead to perturbation in classification boundaries, compromising the stability of discrimination results and compatibility with existing literature plots. To solve the above problems, this study first extends the previous research on TAS diagrams by constructing category partitions based on spatial positions for various lithology labels in classic diagrams. The discrimination is made then on the basis of spatial relationship between location of the data to be classified and category partitions, with results presented in data table to mitigate readability degradation caused by the increase of data volume. Besides, over 240 000 entries of major element data of igneous rocks are extracted from GEOROC database for TAS visualization and for kernel density analysis in terms of lithologic

收稿日期: 2024-02-01; 修回日期: 2024-03-04. 编辑: 张哲.

基金项目: 吉林省教育厅科学技术研究项目“机器学习方法扩展传统图解进行火成岩样品分类”(JJKH20210689KJ).

作者简介: 于秋野(1989—), 男, 博士, 主要从事地学数据分析与大数据技术交叉方向研究工作, 通信地址 吉林省长春市净月大街 3658 号, E-mail// yuqiuye@jisu.edu.cn

引用格式: 于秋野, 李绍竑, 王智岩. TAS-PF: 基于大数据概率场的 TAS 扩展图解[J]. 地质与资源, 2025, 34(2): 255-264.

Yu Q Y, Li S H, Wang Z Y. TAS-PF: Extended TAS diagram powered with probability field of big data[J]. Geology and Resources, 2025, 34(2): 255-264.

classification. The corresponding category probability field is constructed across the plotting coordinates based on the analysis results. The probability is calculated by the position of the data to be classified in each probability field, and the probability results of different lithologic labels are compared. Based on probability field, the known lithology label data are used to distinguish the data to be classified, supplement the traditional classification boundary model and form more quantitative discrimination results.

Key words: TAS diagram; igneous rock; lithologic classification; discrimination diagram; big data; probability field

0 引言

近年来,随着地球科学研究的时空范围和数据规模不断扩大,各类研究呈现出从定性到定量的趋势,基于火成岩地球化学成分数据的判别图解就是其中的典型场景^[1]。作为一种体现了地球科学中“相似类比”思想的定量化方法,判别图解被广泛应用于各种相关研究^[2]。例如,用于火成岩类别判断的 TAS 图解为众多相关研究提供了基础依据^[3-16]。然而,以 TAS 为代表的若干经典图解诞生年代久远^[17],且最初的原始数据规模有限,难以应对日益增长的数据规模。与其他图解类似,如果投图数据超过一定规模,TAS 图解上所呈现的只有一片稠密的数据点,基本丧失了直观的可读性。对此,有学者尝试采用全球尺度的岩石地球化学数据库,如 GEOROC (Geochemistry of Rocks of the Oceans and Continents, 意为海洋和大陆岩石地球化学) 数据库,进行数据挖掘,以评估和改进传统判别图解^[18],从特定类别岩石的划分方案^[19]开始以量化的视角审视经典图解。

向前兼容是“基于边界分类”所要面对的最大挑战,随着数据更新,传统分类边界将会发生显著变化。在这一背景下,有学者尝试应用概率方法^[20],以及通过数据密度方法在 TAS 图解中确定不同岩类的分布区域^[21],指出了 TAS 图解的不足^[22],提供了新的图解方案。然而这些研究主要着眼于分类边界的几何形态、特定岩石类别的数值范围以及具体区块划分方案等传统思考层面,所提出的新图解难以与传统图解兼容。近年来其他相关研究也未摆脱“基于边界分类”这一思路的影响^[23]。

数据叠覆是“基于边界分类”这一思想要面对的另一挑战。从地球化学的角度来看,成分相近甚至相同的物质,完全可以形成不同的矿物和岩石。具体到 TAS 图解中,就是多个不同岩石类别可能叠覆于某些

投图位置。针对这种情况,可采取的一个典型应对策略是提高数据维度,通过引入高维数据的分析方法来尽量避免不同类别标签之间的叠覆^[24]。另外,还可以对地球化学成分数据进行主成分特征提取,再利用支持向量机等方法进行全新的分类边界划分^[25]。但这类策略只是将“基于边界分类”的思想应用于更高的数据维度,并未解决本质问题^[26]。

针对上述问题,有研究人员尝试引入概率密度估计^[21],让数据进行自我描述^[22],但并未实现向前兼容。另外,相关研究也未能充分正视数据叠覆问题,而是普遍将标注岩石类别与经典 TAS 图解分类区域不匹配的样品推测为“分类错误”“录入错误”“测量错误”等,或是由于测量误差等因素导致。部分研究仍受到经典图解的思维束缚,在构建新图解的同时以旧图解为标准去考量数据质量,引入概率也只是为了辅助生成“分类边界”。

本文以经典 TAS 图解为基础:一方面利用前期研究的空间关系对比方法来兼容经典图解;另一方面针对数据叠覆问题,用 GEOROC 数据库中的火成岩数据生成分类概率场,评估待分类数据在图解位置上属于各分类标签的概率,将概率最高的分类标签作为判定结果。经上述步骤的扩展 TAS 图解最终将得到的判别结果整合生成成为一份数据结果。

本文将基于概率场 (Probabilistic Field) 扩展得到 TAS-PF 相关代码开源发布于 <https://github.com/GeoPyTool/TAS-PF>,本次研究所涉及的 GEOROC 数据库文件开源发布于 <https://github.com/GeoPyTool/GeoRocData>。

1 技术方案

1.1 保持向前兼容

以 TAS 图解为代表的地球化学成分数据判别图

解,既是地学发展的历史遗产,也是进一步发展所要面对的历史包袱。“基于边界分类”的基本思路是将平面划分成不同区域,各个区域之间有截然的边界,处于边界两边的两个待分类数据,无论彼此距离多近,都会被分成两个类别。以TAS图解为代表的这类经典判别图解,其原始数据在规模和范围两方面都可能因其时代因素而受到限制,其原始的分类线条设计也可能存在很多局限性^[22]。

对于经典TAS图解来说,在这一情境下,更新数据必然破坏向前兼容性。这首先是由于基于数据生成的分类边界必然将会随着数据的更新而发生变化,这种变化可能是边界形态从直线变为曲线,也可能是边界位置甚至类别分区整体移动^[18-19]。

在这一背景下,传统图解固然需要发展和改进,但对其进行直接修改也将带来难以估量的隐患:不同时代的研究将由于采取了不同的分类边界而出现不同的判别结论;采用不同分类边界的同时代的研究更将带来各种认知冲突。类似情况将导致相关研究共同语境的进一步碎片化。更新分类边界乃至分区的尝试虽然有探索意义,但更可能由于这种碎片化而引发歧义。为了保持向前兼容,并保持相关研究的后继可用性,经典TAS图解中的分类边界应当保留。

1.2 设计向后兼容

保留经典分类边界,而又避免停滞不前,可以平行于经典图解中原有的分类边界体系,构建一套另外的分类方案。保留经典分类的边界体系,可以接续过去文献中的判别结论,也能兼容同时代的其他版本图解。另一套分类方案,则可以降低开发难度,避免负担历史包袱,完全根据当前的地学领域数据现状来从头构建。

这种新方案的设计既需要考虑当前全球领域内的相关岩石地球化学数据状况,也需要考虑对未来数据更新的向后兼容。即在数据规模和形态发生巨大变化的未来,这套新的分类方案依然能被持续更新,并能提供相对稳定的分类参考依据。

借助全球数据设计全新的分类方案,可选路径包括K均值聚类、决策树、神经网络以及概率模型等。K均值聚类是一种无监督的机器学习方法,可以将数据分成K个类别,但其生成的分类结果依然是“基于边界分类”,并且其分类结果与现实中的岩石类别未必能

够建立对应关系。笔者前期研究中采用过的支持向量机方法也存在类似的问题。决策树是一种有监督功能的机器学习方法,但在过于复杂的场景下可能由于数据维度过高而发生拟合并导致误分类,并且可能受到细小数据变化带来的巨大扰动,难以保障判别结果的稳定性。神经网络适合各种复杂的非线性问题,但需要巨大数据规模,训练周期难以估计,且可能发生局部优化而难以全局优化。

使用概率模型可以在应对不确定性的同时兼顾可解释性,已有相关研究尝试利用高斯函数进行数据密度分析^[21-22]。本文拟使用高斯函数作为核函数,对概率密度函数进行核密度估算。

2 数据准备

2.1 数据源选择

前人研究^[18-19, 21-22]中多次使用了GEOROC数据库,并对原始数据进行了复杂的筛选,剔除了所有烧失量质量分数绝对值高于2.5%的数据,选取了所有的岩石类型,最终选择了约13.3万条数据。本文参考上述研究,选用GEOROC数据库2023年12月1日版本^[27],从中选择标注类型符合的数据共240 511条,并未对原始数据的数值进行人为筛选,这一方面是要考虑允许一定的冗余容错,另一方面也是避免过分带入主观判断,尽量保持数据的原始性。

在岩石类别方面,已有研究在试图挣脱TAS束缚重新构建图解的同时,又受到TAS类型判别的影响,倾向于按照TAS图解中的分类标签对数据库中岩石类别进行严格匹配。本次研究摒弃了这一限制,用于扩展的样本数据中的岩石类别标签与TAS图解中的经典分类标签完全并行,同步判断。

2.2 数据条目筛选

如表1、图1所示,GEOROC数据库中不同岩石类别的数据条目数差别巨大,数据条目数最多的有79 997条,最少的不足35条,中位数为964。这种情况下利用规模差异如此巨大的数据重新生成分类边界,必然带来极大偏差。根据中心极限定理,一维数据需要至少有30个观测值才能评估其高斯分布,二维数据则需要增加更多的观测值,但这个观测值的具体选取还要取决于数据特点和分析目标^[2]。因此可以将数据划分为3个规模层次:

表 1 GEOROC 数据库中岩石类别样品数统计表

Table 1 Statistics of rock category samples in the GEOROC database

原文名称	中文翻译	样品数	原文名称	中文翻译	样品数
Basalt	玄武岩	79997	Trachydacite	粗面英安岩	771
Andesite	安山岩	44560	Phonotephrite	响质碱玄岩	587
Rhyolite	流纹岩	22358	Leucitite	白榴岩	580
Andesite_basaltic	玄武质安山岩	16046	Comendite	钠闪碱流岩	575
Tholeiite	拉斑玄武岩	13306	Benmoreite	歪长粗面岩	569
Dacite	英安岩	12977	Pantellerite	碱流岩	510
Trachyte	粗面岩	7003	Ankaramite	富辉橄玄岩	478
Basanite	碧玄岩	5520	Carbonatite	碳酸岩	412
Trachyandesite	粗面安山岩	5455	Melilitite	黄长岩	357
Phonolite	响岩	3912	Foidite	副长岩	273
Komatiite	科马提岩	3680	Oceanite	大洋岩	164
Trachybasalt	粗面玄武岩	3424	Trachyphonolite	粗面响岩	121
Kimberlite	金伯利岩	3001	Trachydolerite	粗面辉绿岩	35
Picrite	苦橄岩	2687	Dolerite	辉绿岩(常规拼写)	26
Hawaiiite	夏威夷岩	2532	Ore	矿石	26
Nephelinite	霞石岩	1929	Trachybasanite	粗基碧玄岩	18
Rhyodacite	流纹英安岩	1725	Lamprophyre	煌斑岩	13
Tephrite	碱玄岩	1345	Diorite	闪长岩	9
Boninite	玻质古铜安山岩	1323	Granite	花岗岩	3
Latite	安粗岩	1186	Diabase	辉绿岩(北美拼写)	2
Shoshonite	钾质安粗岩	1138	Gabbro	辉长岩	2
Adakite	埃达克岩	1009	Peridotite	橄榄岩	2
Mugearite	钠质安粗岩	1003	Granodiorite	花岗闪长岩	1
Lamproite	钾镁煌斑岩	964			

岩石中文名称翻译据文献[21].

- 1) 规模不足 100 条的样品 11 类. 这部分数据的最大条目数仅为 35, 不适合进行数理统计, 也很难保证数据稳定性, 因此予以剔除.
- 2) 规模超过 100 条但不足中位数的样品 12 类. 这部分样品规模在 121~771 之间, 可以进行数理统计, 但其规模较小, 可能会随着未来的数据更新而有扰动.
- 3) 规模在中位数及以上的样品 24 类. 这部分数据规模充足, 即便面对后续更新也能保障有相对稳定

的整体分布形态.

综上所述, 本次研究剔除了不足以进行数理统计的 11 类岩石数据, 对后两种岩石类别数据进行了核密度估计以构建概率场, 仅推荐参考其中达到中位数及以上规模样本数据所生成的概率场.

2.3 数据库的生成

GEOROC 数据库本身的原始文件普遍为 CSV 文本格式, 这种格式实际上是使用逗号分隔文本而构建的表格. 其优势在于易于读写编辑, 但数据体积大, 且



图 1 GEOROC 中筛选出的火成岩数据条目分布图

Fig. 1 Distribution of igneous rock data entries selected in GEOROC

岩石名称中文翻译如表 1 (Chinese translation of rock names shown in Table 1)

读取过程中容易有字符编码等方面原因导致的各种问题。此前的研究^[18-19, 21-22]中仅利用 GEOROC 数据库的原始文件进行数据挖掘和信息提取,而未公开所整理好的数据集形态。

针对这一问题,本次研究基于 GEOROC 数据库 2023 年 12 月 1 日版本,整合生成了一份 SQLite 数据库。该数据库通过一种结构化方式来对 GEOROC 数据进行读取和更新,也便于使用 SQL 等查询语言来检索和操作数据。SQLite 数据库是轻量级数据库,可以被 Python 等编程语言原生访问,无需独立服务器,也不需要其他外置工具。

本文的数据和数据库文件都开源发布于 <https://github.com/GeoPyTool/GeoRocData>, 供下载使用。在保留原始条目基础上,对烧失量外的主量元素数据求和获得总和值,再用各项数据值除以对应总和值而实现归一化。所生成数据列放入新增表内,不影响原始数据。

3 概率场扩展 TAS 图解

本文所扩展的 TAS 图解由两部分构成:第一部分

是传统边界,基于经典 TAS 图解的传统分类边界坐标所生成的区域判别体系;第二部分是概率场,基于 GEOROC 数据库中所选岩石数据而生成。传统边界的保留保证了在此基础上扩展的 TAS 图解能向前兼容已有研究中所使用的 TAS 图解,其可视化形态与经典 TAS 图解中保持一致,其基本思想和数学构建构成在前期文献中^[24-26]已经说明,不再赘述。本节着重讲述概率场部分的实现过程。

3.1 分布函数选择

假设有一组二维数据 (x_i, y_i) ($i=1, \dots, N$), 其中 $(x_i, y_i) \in R^2$, N 是数据的数量。已有的研究^[21]在 X 和 Y 两个方向上分别使用高斯函数进行估计, 然后进行褶积(卷积)(convolution)来估计这组数据的概率密度函数 $f(x, y)$ 。

在 X 方向上,有高斯函数 $g_X(x)$:

$$g_X(x) = \frac{1}{\sqrt{2\pi} \sigma_X} \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right] \tag{1}$$

在 Y 方向上,有高斯函数 $g_Y(y)$:

$$g_Y(y) = \frac{1}{\sqrt{2\pi} \sigma_Y} \exp\left[-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}\right] \quad (2)$$

其中, (μ_X, μ_Y) 是均值, (σ_X, σ_Y) 是标准差。

然后, 通过卷积来计算二维的概率密度函数 $f(x, y)$:

$$f(x, y) = (g_X \cdot g_Y)(x, y) \quad (3)$$

使用卷积的方法(即在 X 和 Y 两个方向各自采用高斯函数然后进行卷积)在计算上通常比二维高斯分布的计算要简单。另外, 卷积本身是一种平滑操作, 可以减少数据的噪声和异常值的影响, 生成平滑的分类边界。

然而, 在地球化学成分数据之间往往存在相关性, 主量元素数据更是具有定和效应(constant sum effect, 指在一组数据中, 所有数据的和是一个固定的常数)。对于一组经过去除烧失量后重新计算的主量元素数据 x_j ($j=1, \dots, M$), 其中 $0 \leq x_j \leq 100\%$, 是成分 j 的质量分数, M 是氧化物的数量, 则所有成分的质量分数总和为 100%:

$$\sum_{j=1}^M x_j = 100\% \quad (4)$$

在 TAS 图解中, x 为 SiO_2 的质量分数(%), y 为 K_2O 与 Na_2O 两种氧化物的质量分数(%)之和, x 和 y 之间必然存在相关性。而上述卷积的方法假设了 x 和 y 是独立的, 而不能捕捉到 x 和 y 之间的相关性。因此, 不应简单地通过两个方向上各自的边缘分布(即在各自方向上的高斯函数)的卷积来估计整体的概率密度函数。

在 x 和 y 之间存在相关性的情况下, 使用二维的高斯分布函数可能会更合适, 因为二维的高斯分布函数考虑了 x 和 y 的协方差, 所以可以捕捉到 x 和 y 之间的相关性。因此, 本次研究选择使用高斯核函数来估计这组数据的概率密度函数 $f(x, y)$ 。

$$f(x, y) = \frac{\exp\left[-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right]\right]}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \quad (5)$$

需要注意的是, 这里得到的是概率密度函数, 而不是概率分布。理论上, 概率分布可以通过对概率密度函数进行积分得到, 这个概率分布描述的是二维数据

作为随机数据在二维空间内各个位置上出现的概率。对任意一点来说, 该组二维数据中有点出现于此的概率都为 0, 因此其概率分布函数显然无法直接用于图解。

3.2 生成概率场

在 TAS 图解中, 用于生成概率场的基础二维数据是已知且固定的, 所得到的概率密度函数可描述各类数据在二维空间中各个位置分布的密度, 然后将未分类二维数据按照坐标投入到当前二维空间, 计算其归属于各个类别的概率。一个合理的假设是, 概率密度最高的地方是其类别归属概率最高之处, 而概率密度最低的地方是其类别归属概率最低之处。

基于这个假设, 可以定义一个归一化的概率密度函数 $P(x, y)$, 如下:

$$P(x, y) = \frac{f(x, y)}{\max_{x, y} f(x, y)} \quad (6)$$

其中, $\max_{x, y} f(x, y)$ 是概率密度的最大值, 代表了该类数据最为集中的区域。这个归一化的概率密度函数 $P(x, y)$ 在二维空间中形成了一个状态空间, 每个位置都有一个对应的值, 表征该位置上的数据点属于该组二维数据的概率。至此实现了对概率场的构建。

值得注意的是, 高斯分布在整个实数域上都有定义, 但地球化学成分数据实际上只在正数域的一个有限范围内。这可能导致高斯核密度估计在负数域上产生一些不必要的估计值, 从而可能影响结果的准确性。此外, 如果数据在正数域上的分布存在偏斜, 那么由于高斯函数的对称性, 可能难以准确捕捉这种非对称的数据分布。然而, 尽管存在这些潜在问题, 高斯核密度估计仍然能够提供合理的结果。毕竟目标只是构建一个合理的定量化的概率场, 而通过二维高斯函数进行核密度估计已经足够满足这一需求。

3.3 数据可视化

上述过程所生成的概率场本身是难以直接观察的, 但基于投图空间位置对不同位置上的概率生成等高线, 则可以直观观察到对应类别概率场的基本形态(如图 2)。由于篇幅所限, 本文仅选择若干比较有代表性的概率场予以呈现。在实际使用中, 这些概率场不需要呈现在最终的 TAS 图解所生成的投图上, 而是以数据表的形式给出各待分类数据的判别结果和对

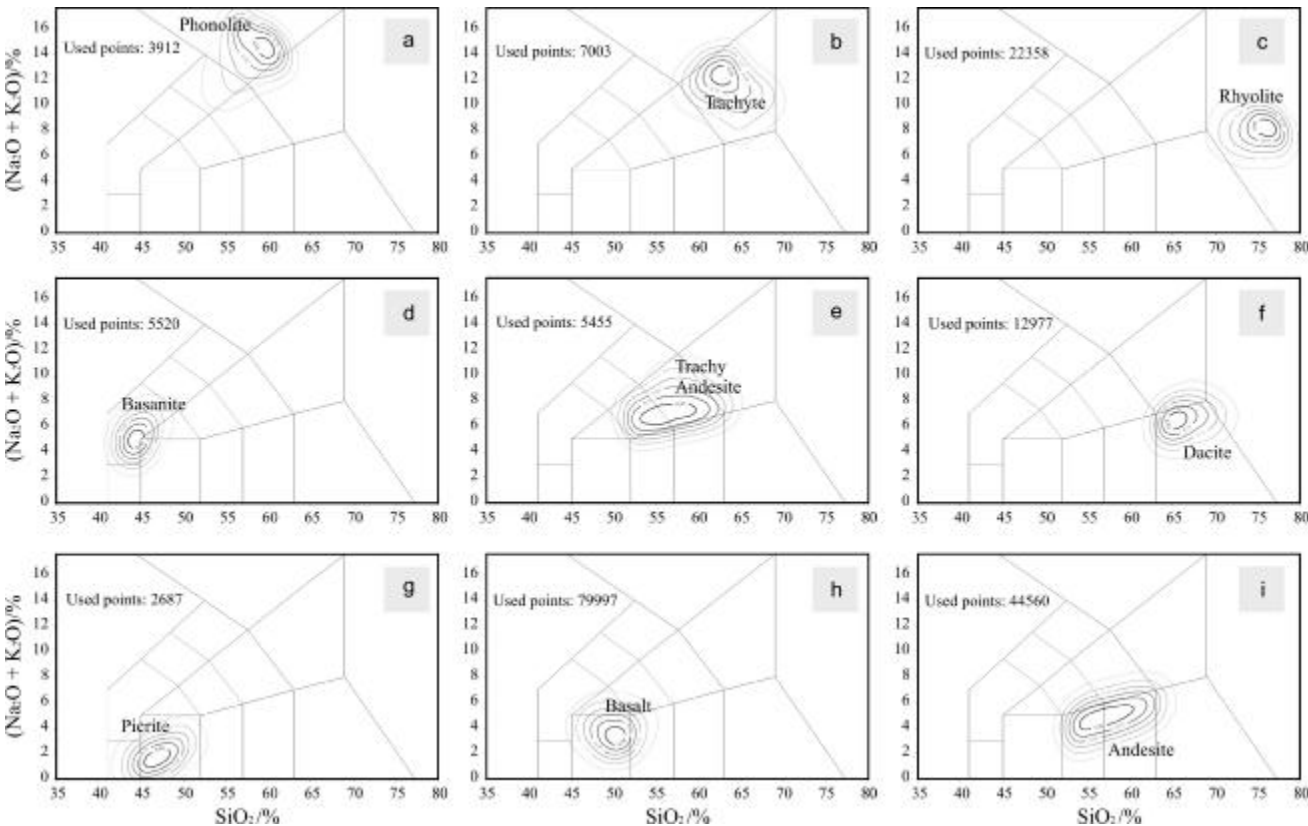


图 2 GEOROC 中部分岩石类型概率场等高线图

Fig. 2 Contour maps of probability field of selected rock types from GEOROC

a—响岩(phonolite); b—粗面岩(trachyte); c—流纹岩(rhyolite); d—碧玄岩(basanite); e—粗面安山岩(trachyandesite); f—英安岩(dacite); g—苦橄岩(picrite); h—玄武岩(basalt); i—安山岩(andesite)

应概率.

3.4 判别结果导出

将经典 TAS 图解的传统判别边界与上述过程所生成的概率场相结合,就得到了经过概率场扩展的 TAS 图解.如图 3 所示,该扩展图解包含了经典 TAS 图解原有分类边界,以及基于 GEOROC 数据库所选条目数据分布而构建的概率场,并且在图幅中显示有各概率场的岩石类别标签,以及总体选用数据规模.本次研究将上述数值过程以代码形式开源发布于 <https://github.com/GeoPyTool/TAS-PF>,下载源码或打包的安装文件即可以使用图形界面软件 TAS_Extended 来运行扩展图解.

利用 TAS_Extended 软件可如图 3 一样将任意待分类数据投影到扩展 TAS 图解上,并且可分组调整控制画面基础元素是否显示,所显示的投图效果可以用矢量图或者位图的格式来保存.当然,这种直观感知无法应对更大规模的数据,也不能直接提供有定量意

义的判别结果.对此,本研究在 TAS_Extended 软件上将概率场的判别结果和 TAS 经典边界的分类结果相结合,可以与原始数据一起直接导出为 CSV 或 Excel 格式的数据文件.如表 2 所示,表中内容为图 3 中部分投图点的判断结果与原始数据.

4 结论

本研究基于概率场的思想实现了对经典 TAS 图解的继承和发展,通过保留原有边界保持了向前兼容性,同时所选用的概率场设计路径也前瞻性地顾及了后续的数据补充.主要结论如下:

- 1) 继承经典分类边界使得扩展的 TAS 图解能兼容已有研究,避免碎片化;
- 2) 概率场的稳定性会随着后续数据更新和补充而日趋稳定,保障后续应用;
- 3) 基于概率场的扩展图解可以得到比传统图解更全面、更定量的描述.

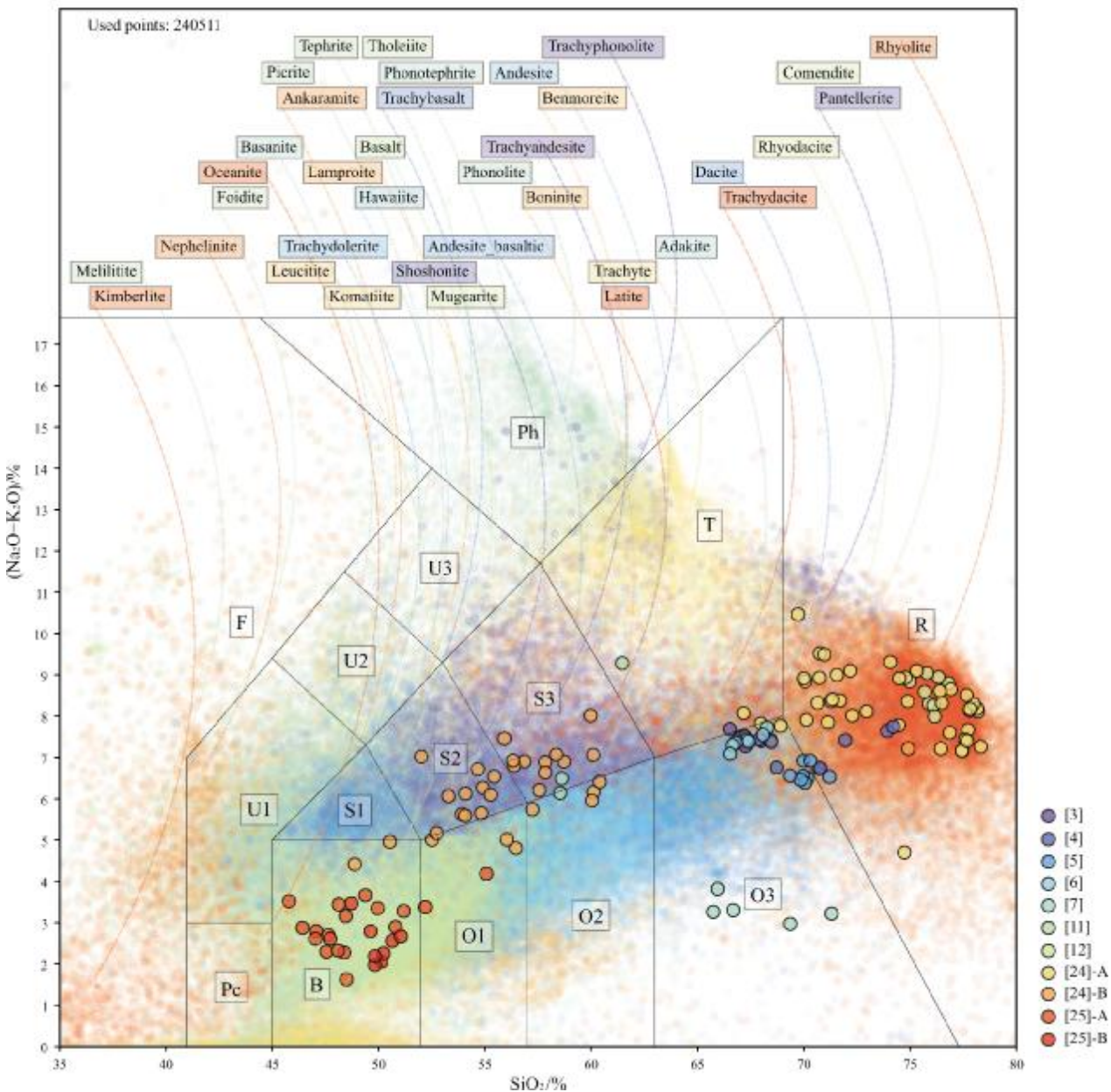


图 3 将文献数据投到概率场扩展的 TAS 图解

Fig. 3 Literature data plotted into the extended TAS diagram with probability field

经典 TAS 图解分类标签(classification label of typical TAS diagram): F—副长岩(foidite); Ph—响岩(phonolite); Pc—苦橄岩(picrite); U1—碱玄岩(橄榄石<10%) (tephrite), 碧玄岩(橄榄石>10%) (basanite); U2—响质碱玄岩(phonotephrite); U3—碱玄质响岩(tephriphonolite); B—玄武岩(basalt); S1—粗面玄武岩(trachybasalt); S2—玄武质粗安岩(basaltic trachyandesite); S3—粗面安山岩(trachyandesite); O1—玄武质安山岩(basaltic andesite); O2—安山岩(andesite); O3—英安岩(dacite); T—粗面岩(石英<20%) (trachyte); 粗面英安岩(石英>20%) (trachydacite); R—流纹岩(rhyolite); [x]—数据文献编号(reference number of data)

致谢:在此感谢并缅怀我的同学、好友、伙伴,李解博士,他曾多次批评我过去单纯以分类边界思想衡量地球化学成分数据的做法,十年前我们曾经一起装电

脑、学编程、写程序,这篇文章本是准备与他合作完成,奈何天妒英才,李解博士在 2023 年因病英年早逝,如今谨以此文来纪念.

表 2 将文献数据投到概率场扩展的 TAS 图解后生成
的部分判别结果展示表

Table 2 Discrimination results generated by the extended
TAS diagram with probability field

概率场 分类结果	概率	经典 TAS 分类结果	SiO ₂	K ₂ O	Na ₂ O	数据来源
Rhyolite	0.74	Rhyolite	77.7	4.42	3.1	文献[24]
Rhyolite	0.776	Rhyolite	77.58	4.86	3.5	文献[24]
Rhyolite	0.772	Rhyolite	77.55	4.94	3.49	文献[24]
Rhyolite	0.798	Rhyolite	77.38	5.68	2.86	文献[24]
Rhyolite	0.872	Rhyolite	77.21	4.96	3.46	文献[24]
Rhyolite	0.921	Rhyolite	77.12	4.25	3.67	文献[24]
Rhyolite	0.821	Rhyolite	77.06	5.52	3.23	文献[24]
Rhyolite	0.909	Rhyolite	76.98	5.18	2.51	文献[24]
Rhyolite	0.849	Rhyolite	76.86	3.26	4.16	文献[24]
Rhyolite	0.891	Rhyolite	76.32	5.1	3.79	文献[24]
Rhyolite	0.993	Rhyolite	75.9	4.29	4.27	文献[24]
Picrite	0.904	Basalt	48.17	0.72	1.88	文献[25]
Basalt	1	Basalt	50.56	0.77	2.44	文献[25]
Lamproite	0.775	Basalt	47.44	0.73	2.29	文献[25]
Basalt	0.888	Basalt	49.15	1.38	2.59	文献[25]
Basalt	0.734	Basalt	47.92	0.71	3.04	文献[25]
Tholeiite	0.884	Basalt	50	0.45	2.13	文献[25]
Basalt	0.821	Basalt	49.6	0.47	2.04	文献[25]
Tholeiite	0.971	Basalt	50.4	0.5	2.38	文献[25]
Tholeiite	1	Basalt	50.8	0.69	2.3	文献[25]
Basalt	0.83	Basalt	48.5	1.16	2.61	文献[25]
Basalt	0.943	Basalt	49.4	0.77	2.34	文献[25]

所用数据引自文献[3-7,11-12,24-25],考虑篇幅等因素仅展示投图中的
部分数据结果.含量单位:%(质量分数).

参考文献(References):

[1]刘宝珺,李廷栋.地质学的若干问题[J].地球科学进展,2001,16
(5):607-616.
Liu B J, Li T D. Some problems of geology[J]. Advances in Earth
Science, 2001, 16(5): 607-616.

[2]赵鹏大.定量地质学方法及应用[M].北京:高等教育出版社,2004.
Zhao P D. Quantitative geoscience: Methods and its applications[M].
Beijing: Higher Education Press, 2004.

[3]黄雄飞,莫宣学,喻学惠,等.西秦岭印支期高Sr/Y花岗岩类的成
因及动力学背景——以同仁地区舍哈力吉岩体为例[J].岩石学报,
2014,30(11):3255-3270.
Huang X F, Mo X X, Yu X H, et al. Origin and geodynamic settings
of the Indosinian high Sr/Y granitoids in the West Qinling: An example
from the Shehaliji pluton in Tongren area[J]. Acta Petrologica Sinica,

2014, 30(11): 3255-3270.

[4]罗明非,莫宣学,喻学惠,等.东昆仑香日德地区晚三叠世花岗岩
LA-ICP-MS 锆石 U-Pb 定年、岩石成因和构造意义[J].岩石学报,
2014,30(11):3229-3241.
Luo M F, Mo X X, Yu X H, et al. Zircon LA-ICP-MS U-Pb age
dating, petrogenesis and tectonic implications of the Late Triassic
granites from the Xiangride area, East Kunlun[J]. Acta Petrologica
Sinica, 2014, 30(11): 3229-3241.

[5]罗明非,莫宣学,喻学惠,等.东昆仑五龙沟晚二叠世花岗岩闪长岩
LA-ICP-MS 锆石 U-Pb 定年、岩石成因及意义[J].地学前缘,2015,
22(5):182-195.
Luo M F, Mo X H, Yu X H, et al. Zircon U-Pb geochronology,
petrogenesis and implication of the Later Permian granodiorite from the
Wulonggou area in East Kunlun, Qinhai Province[J]. Earth Science
Frontiers, 2015, 22(5): 182-195.

[6]田子龙,赵庆英,李子昊,等.内蒙古突泉县姜家屯侏罗纪花岗岩闪
长岩年代学及地球化学特征[J].地质与资源,2017,26(6):542-551.
Tian Z L, Zhao Q Y, Li Z H, et al. Geochronology and geochemistry
of the Jurassic Granodiorite in Jiangjiatun, Tuquan County of Inner
Mongolia[J]. Geology and Resources, 2017, 26(6): 542-551.

[7]张垚垚,张达,陈宣华,等.哀牢山中段镇沅金矿晚二叠世花岗岩斑
岩 U-Pb 年代学、岩石地球化学及其构造意义[J].岩石学报,2021,
37(6):1674-1690, doi: 10.18654/1000-0569/2021.06.03.
Zhang Y Y, Zhang D, Chen X H, et al. U-Pb chronology,
Lithogeochemistry and tectonic significance of Late Permian granite
porphyry in Zhenyuan gold deposit, the middle section of Ailaoshan
[J]. Acta Petrologica Sinica, 2021, 37(6): 1674-1690, doi: 10.18654
/1000-0569/2021.06.03.

[8]李利阳,冯欣,焦晓宇,等.大兴安岭乌奴耳地区白音高老组的厘
定及其地质意义[J].地质与资源,2024,33(2):135-142,151,
doi: 10.13686/j.cnki.dzyzy.2024.02.001.
Li L Y, Feng X, Jiao X Y, et al. Determination of the Baiyingaolao
Formation in Wunuer area, Daxinganling Mountains: Geological
implication [J]. Geology and Resources, 2024, 33(2): 135-142,
151, doi: 10.13686/j.cnki.dzyzy.2024.02.001.

[9]王文东,刘涛,周传芳,等.黑龙江富源沟晚三叠世中-基性火山岩
的厘定及地质意义[J].地质与资源,2023,32(6):670-680, doi:
10.13686/j.cnki.dzyzy.2023.06.003.
Wang W D, Liu T, Zhou C F, et al. Identification of the late Triassic
Fuyuangou intermediate-basic volcanic rocks in Heilongjiang province:
geological implication [J]. Geology and Resources, 2023, 32(6):
670-680, doi: 10.13686/j.cnki.dzyzy.2023.06.003.

[10]李晓海,张海华,李文博,等.大兴安岭南地区花岗岩 LA-ICP-MS
锆石 U-Pb 年龄、地球化学特征及其地质意义[J].地质与资源,
2023,32(6):655-663, doi: 10.13686/j.cnki.dzyzy.2023.06.001.
Li X H, Zhang H H, Li W B, et al. LA-ICP-MS zircon U-Pb dating
and geochemistry of the granites in Gannan area, Daxinganling

- Mountains: Geological implication[J]. *Geology and Resources*, 2023, 32(6): 655–663, doi: 10.13686/j.cnki.dzyz.2023.06.001.
- [11] 洪文武, 李松彬, 刘钧沅, 等. 大兴安岭北段卜奎沟岩体地球化学和锆石 U-Pb 年代学研究[J]. *矿产勘查*, 2023, 14(7): 1026–1033.
- Hong W W, Li S B, Liu J Y, et al. Geochemical and zircon U-Pb geochronology of Pukuigou granite, Northern Great Xing'an Rang[J]. *Mineral Exploration*, 2023, 14(7): 1026–1033.
- [12] 张善明, 胡雅璐, 王根厚, 等. 内蒙古东七—山花岗岩杂岩的形成演化及对成矿的贡献: 年代学及地球化学证据[J]. *岩石学报*, 2023, 39(6): 1791–1816, doi: 10.18654/1000-0569/2023.06.13.
- Zhang S M, Hu Y L, Wang G H, et al. Formation, evolution of the granitic complex and its contribution to mineralization in Dongqiyishan deposit, Inner Mongolia: Chronological and geochemical evidences [J]. *Acta Petrologica Sinica*, 2023, 39(6): 1791–1816, doi: 10.18654/1000-0569/2023.06.13.
- [13] 高子越, 鞠楠, 李经纬, 等. 吉林延边五凤地区早侏罗世粗面安山岩岩石成因[J]. *地质与资源*, 2024, 33(1): 12–23, 40, doi: 10.13686/j.cnki.dzyz.2024.01.002.
- Gao Z Y, Ju N, Li J W, et al. Petrogenesis of the Early Jurassic trachyandesite in Wufeng area of Yanbian, Jilin Province[J]. *Geology and Resources*, 2024, 33(1): 12–23, 40, doi: 10.13686/j.cnki.dzyz.2024.01.002.
- [14] 靳立杰, 刘晓, 刘伟, 等. 山东栖霞煌斑岩脉年代学、地球化学特征及其地质意义[J]. *地质与资源*, 2024, 33(2): 143–151, doi: 10.13686/j.cnki.dzyz.2024.02.002.
- Jin L J, Liu X, Liu W, et al. Geochronology and geochemistry of the lamprophyre dikes in Qixia area, Shandong Province: Geological implication[J]. *Geology and Resources*, 2024, 33(2): 143–151, doi: 10.13686/j.cnki.dzyz.2024.02.002.
- [15] 马维, 马婧轩, 洪阳百合, 等. 嫩江洋闭合时限: 来自早二叠世大乌苏岩体锆石年代学和地球化学的制约[J]. *地质与资源*, 2024, 33(3): 267–279, doi: 10.13686/j.cnki.dzyz.2024.03.001.
- Ma W, Ma J X, Hong Y B H, et al. Closure time of Nenjiang Ocean: Constrains from zircon geochronology and geochemistry of the Early Permian Dawusu monzogranite [J]. *Geology and Resources*, 2024, 33(3): 267–279, doi: 10.13686/j.cnki.dzyz.2024.03.001.
- [16] 张洪文, 平先权, 黄伟, 等. 黑龙江伊春地区早侏罗世 A 型花岗岩地球化学特征及构造环境探讨[J]. *地质与资源*, 2024, 33(3): 280–290, doi: 10.13686/j.cnki.dzyz.2024.03.002.
- Zhang H W, Ping X Q, Huang W, et al. Geochemical characteristics and tectonic environment of the Early Jurassic A-type granites in Yichun area, Heilongjiang Province [J]. *Geology and Resources*, 2024, 33(3): 280–290, doi: 10.13686/j.cnki.dzyz.2024.03.002.
- [17] Le Bas M J, Le Maitre R W, Streckeisen A, et al. A chemical classification of volcanic rocks based on the total alkali-silica diagram [J]. *Journal of Petrology*, 1986, 27: 745–750.
- [18] 王金荣, 陈万峰, 张旗, 等. N-MORB 和 E-MORB 数据挖掘——玄武岩判别图及洋中脊源区地幔性质的讨论[J]. *岩石学报*, 2017, 33(3): 993–1005.
- Wang J R, Chen W F, Zhang Q, et al. Preliminary research on data mining of N-MORB and E-MORB: Discussion on method of the basalt discrimination diagrams and the character of MORB's mantle source [J]. *Acta Petrologica Sinica*, 2017, 33(3): 993–1005.
- [19] 第鹏飞, 陈万峰, 张旗, 等. 全球 N-MORB 和 E-MORB 分类方案对比[J]. *岩石学报*, 2018, 34(2): 264–274.
- Di P F, Chen W F, Zhang Q, et al. Comparison of global N-MORB and E-MORB classification schemes [J]. *Acta Petrologica Sinica*, 2018, 34(2): 264–274.
- [20] 周永章, 陈烁, 张旗, 等. 大数据与数学地球科学研究进展——大数据与数学地球科学专题代序[J]. *岩石学报*, 2018, 34(2): 255–263.
- Zhou Y Z, Chen S, Zhang Q, et al. Advances and prospects of big data and mathematical geoscience[J]. *Acta Petrologica Sinica*, 2018, 34(2): 255–263.
- [21] 葛粲, 顾海欧, 汪方跃, 等. 基于数据密度确定分布区域的方法: 以 TAS 图解分析为例[J]. *地质科学*, 2018, 53(4): 1240–1253, doi: 10.12017/dzgx.2018.071.
- Ge C, Gu H O, Wang F Y, et al. Determination of distribution region based on data density: A case study of TAS diagram [J]. *Chinese Journal of Geology*, 2018, 53(4): 1240–1253, doi: 10.12017/dzgx.2018.071.
- [22] 张旗, 葛粲, 焦守涛, 等. 在大数据背景下看 TAS 分类的不足及可能的解决方案[J]. *地质通报*, 2019, 38(12): 1943–1954.
- Zhang Q, Ge C, Jiao S T, et al. The deficiencies and possible solutions of TAS classification in the context of big data[J]. *Geological Bulletin of China*, 2019, 38(12): 1943–1954.
- [23] 方思源, 董少春, 胡欢. 基于大数据的元素协变图自动绘制软件的设计与实现[J]. *高校地质学报*, 2021, 27(1): 73–79.
- Fang S Y, Dong S C, Hu H. Design and implementation of an element variation diagram automatic generation software based on big data[J]. *Geological Journal of China Universities*, 2021, 27(1): 73–79.
- [24] 于秋野. 基于高维数据的成矿信息分析——以内蒙古二道河矿床为例[D]. 北京: 中国地质大学, 2019.
- Yu Q Y. Analysis of ore-forming information from high-dimensional data: A case study of Erdaohe polymetallic deposit in Inner Mongolia [D]. Beijing: China University of Geosciences, 2019.
- [25] Yu Q Y, Bagas L, Yang P H, et al. GeoPyTool: A cross-platform software solution for common geological calculations and plots [J]. *Geoscience Frontiers*, 2019, 10(4): 1437–1447.
- [26] Yu Q Y, Zhang X M, Hu B J, et al. Separating volcanic rock groups: A novel method based on principal component analysis and a support vector machine[J]. *Arabian Journal of Geosciences*, 2021, 14(11): 967.
- [27] GEOROC Compilation; Rock types[EB/OL]. <https://doi.org/10.25625/2JETOAGEOROCdatabaseon>, 2023–12–01.