

上海交通大学“深蓝计划”基金

项目申请书

项目名称 基于机器学习的印度洋洋壳地球化学
不均一特征识别及成因研究

项目类别 ☐ 重点 I 类 ☐ 重点 II 类 ☐ 面上项目

负责人 李金金

所在单位 上海交通大学 电院

合作者 余 星

所在单位 自然资源部第二海洋研究所

上海交通大学科研院

2021 年 06 月 01 日

填 报 说 明

一、项目重点支持海洋科学与技术的前瞻性创新性研究、发展规划和重大项目前期论证、“卡脖子”关键技术研发以及行业发展亟需的中小型海洋装备和科学仪器开发等。

二、项目名称：应确切反映研究内容，最多不超过 36 个汉字。

三、每人限主持 1 项，主持和参加项目总数不超过 2 项（含深蓝计划在研项目）

四、联系手机和电子邮件请填写常用联系方式。

五、申请书为 A4 幅面，双面打印，于左侧订书机简单装订成册。

基本信息

项目 负责 人	姓名	李金金	性别	女	出生 年月日	1985. 07. 14	所在单位	上海交通大学
	职称	研究员	学历	博士	手机	18202114296	Email 地 址	lijinjin@sjtu.edu.
	研究方向		人工智能与大数据计算					
项目 合作 者	姓名	余星	性别	男	出生 年月日	1981. 9. 30	所在院系	海底科学实验室
	职称	研究员	学历	博士	手机	13666660727	Email 地 址	yuxing@sio.org.cn
	研究方向		岩浆作用与大地构造					
关键词 (3-5)		印度洋, Dupal 异常, 地球化学, 机器学习, 大数据						
<h3>项目摘要</h3> <p>洋壳主要由洋中脊岩浆作用和板内岩浆作用形成, 不同岩浆作用形成的洋壳具有不同的地球化学特征, 同一种岩浆作用内部也具有地球化学组成上的分化。洋壳地球化学特征受控于地幔源区组成、熔融过程、岩浆分离结晶等直接过程, 而这些岩浆作用又受控于构造环境和地球动力学过程。因此了解洋壳的地球化学特征和异常分布, 有助于了解洋壳的成因以及地球动力演化的细节过程。印度洋具有全球最不均一的洋壳地球化学组成, 同时也是 Dupal 同位素异常分布的重点区域。目前由于研究区域复杂, 数据量大, 对印度洋不均一洋壳尚缺少整体的认识, 对其形成过程以及 Dupal 异常的成因仍存在争议。本项目拟利用印度洋区已有的洋壳地球化学数据(>10000 条, 每条参数>50 个), 融合监督学习和无监督学习算法, 全面分析地球化学指标相关性, 建立高精度、高效率的参数-区域异常的量化模型, 揭示印度洋在主量元素、微量元素以及同位素等多个维度上的不均一性特征, 了解不同维度不均一的相关性, 识别地球化学异常区, 包括 Dupal 异常的分布界线, 同时遴选异常的判别指标。结合区域上的地形、重磁、层析成像和构造演化等背景资料, 综合探讨印度洋不均一性洋壳的成因以及 Dupal 异常的形成机制。项目创新性地利用机器学习算法和大数据分析技术提出印度洋洋壳不同尺度不同维度的地球化学异常识别算法, 大大缩短了传统地球化学数据处理周期, 并使自动异常识别精度达到 90%以上。</p>								

一、立项依据

全球大洋基底即洋壳具有与陆壳截然不同的特征和形成机制。洋壳是地幔岩浆作用的产物，由地幔橄榄岩部分熔融形成岩浆，并经岩浆分异冷却而成。洋壳的地球化学组成受控于地幔源区组成、部分熔融作用及岩浆分异过程。由于这些受控因素的差异，洋壳表现出明显的地球化学不均一性。有关洋壳地球化学不均一性及其受控因素的研究一直是大洋固体圈层地球化学研究的重点(e.g. Doucet et al., 2020; Richter et al., 2020; Urann et al., 2020; Yang et al., 2020)。但以往的研究往往关注于局部样本，侧重于典型区域的对比，缺少全局视角分析。

随着研究工作的积累和大量海底基岩地球化学数据的发布，海底岩石地球化学数据库不断充实完善，如 EarthChem, GEOROC 等，已经为全局研究奠定了数据基础(如图 1)。另一方面，机器学习和大数据处理在科研领域已经取得了飞跃式的发展，如无监督学习算法（Unsupervised learning），可以对“无标记”的样本进行训练和预测，挖掘特征之间的相关性；迁移学习（Transfer learning）算法，在精度持平的条件下，可以降低对训练数据的规模，减少对数据的依赖性，克服数据稀缺的挑战；集成学习（Ensemble learning）算法，可以并行式处理单个训练器，提高模型泛化能力的同时，极大地提高海量数据的处理速度。这为开展基于海量数据的全局性问题研究提供了技术支撑。

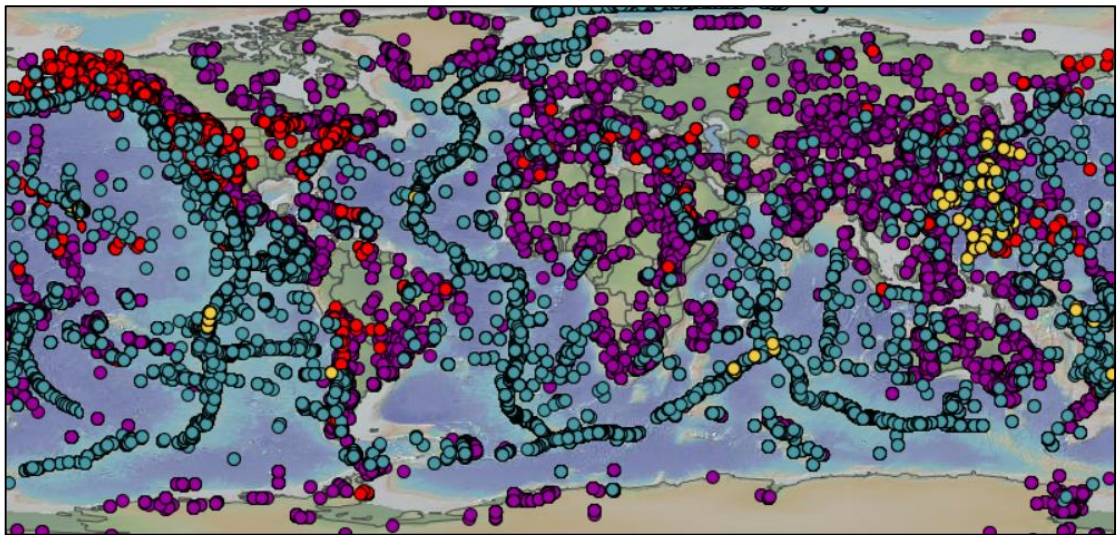


图 1. 全球岩石地球化学数据分布(玄武岩)

数据记录 220120 条（截至 2021 年 5 月 25 日）

(1) 地球化学异常的维度和尺度

目前已知的洋壳地球化学异常可以分为多个维度和尺度。

地球化学异常可以体现在主量元素、微量元素或同位素等多个方面,不同维度之间的相关性并不明确。放射性同位素,如 Sr-Nd-Pb 同位素,可以跟踪地幔源区;稀有气体同位素,如 He 同位素,可以反映原始地幔组分;微量元素中的相容元素和不相容元素在岩浆熔融和分异过程中表现出不同的行为,因此可以反映地幔源区熔融程度和分离结晶过程;主量元素是岩石的主要组成,且含量与岩浆的来源密切相关,体现源区熔融和岩浆演化过程,如衍生参数 Mg#, Na₈ 等。

地球化学异常的尺度可以从板块尺度到手标本尺度,各层次之间的内在联系仍十分模糊。全球最大尺度的地球化学异常是 Dupal 异常(Dupré and Allègre, 1983),其分布面积约 1 亿 km²。Dupal 异常具体指:在南半球赤道至 60°S 区域,如南大西洋和印度洋区,其玄武岩具有相对北大西洋和东太平洋更高的 ⁸⁷Sr/⁸⁶Sr(>0.7035),以及在相同 ²⁰⁶Pb/²⁰⁴Pb 比值条件下更高的 ²⁰⁷Pb/²⁰⁴Pb 和 ²⁰⁸Pb/²⁰⁴Pb 比值特征。目前已有的 Dupal 异常判别指标为 Δ7/4、Δ8/4 和 ΔSr(Hart, 1984)。Castillo(1988)、Wen(2006)和 White(2015)等认为 Dupal 异常的核心区有两个:印度洋-南大西洋区和中太平洋区,分别对应下地幔中的两个超大型地震波低速区(即 LLSVPs),同时对应全球两个最主要的活动热点聚集区。因此, Dupal 异常,这一地球化学异常的界线及成因的研究同时具有重要的地质和地球物理意义,有助于理解大尺度地幔不均一性、壳幔相互作用、板块运动规律、地球深部结构和动力学过程。

$$\Delta 7/4 = [(^{207}\text{Pb}/^{204}\text{Pb})_{\text{样品}} - 0.1084 \times (^{206}\text{Pb}/^{204}\text{Pb})_{\text{样品}} - 13.491] \times 100$$

$$\Delta 8/4 = [(^{208}\text{Pb}/^{204}\text{Pb})_{\text{样品}} - 1.209 \times (^{206}\text{Pb}/^{204}\text{Pb})_{\text{样品}} - 15.627] \times 100$$

$$\Delta \text{Sr} = [(^{87}\text{Sr}/^{86}\text{Sr})_{\text{样品}} - 0.7] \times 10^4$$

注:Δ7/4 和 Δ8/4 表示样品 ²⁰⁷Pb/²⁰⁴Pb、²⁰⁸Pb/²⁰⁴Pb 比值与北半球参考线(NHRL)的相对偏差,ΔSr 表示样品 ⁸⁷Sr/⁸⁶Sr 比值的绝对偏差

(2) 印度洋洋壳地球化学异常特征及研究现状

印度洋具有全球最不均一的洋壳地球化学组成,也是 Dupal 异常的最核心区域。目前关于印度洋洋壳不均一及 Dupal 异常的成因仍不清楚。其原因主要是缺少全局的认识,已有的研究多为点上的报道。

目前已发现的具有明显 Dupal 异常的板内火山作用包括马达加斯加海底高原、莫桑比克海底高原、厄加勒斯海底高原、凯尔盖朗海台、Broken 海脊、东经 90 度海岭、圣诞岛等 (Saunders et al., 1991; Mahoney et al., 1995; Weis, 2002; Taneja, 2016; Zhang et al., 2016; Jacques et al., 2019) (图 2)。

除了板内火山，印度洋洋中脊也广泛发育 Dupal 异常。西南印度洋中脊 Andrew Bain 转换断层(26 °E)以东洋脊段显示有 Dupal 异常(图 2 和图 3a)，且 Andrew Bain 转换断层至 Gallieni 转换断层间的洋脊段具有最高的 $\Delta 7/4$ 和 $\Delta 8/4$ ，分别可达+18.1 和+122.8 (余星等, 2020)。Yu and Dick (2020)认为西南印度洋中脊中段的 Dupal 异常属 EMII 型，与大陆裂解时大陆岩石圈拆沉进入软流圈有关。

东南印度洋中脊的 Dupal 异常边界在澳大利亚-南极错乱带，具体界线位于 127 °E 附近，界线以西显示 Dupal 异常($\Delta 7/4$ 和 $\Delta 8/4$ 分别可达+13.1 和+135.8)，以东洋脊段具有太平洋型 MORB 特征(图 2 和图 3b)。东南印度洋中脊上最强烈的 Dupal 异常信号位于 Amsterdam-St.Paul 海底高原一带(余星等, 2019a)。

西北印度洋中脊的 Dupal 异常界线仍不清楚。西北印度洋中脊(包括卡尔斯伯格脊和中印度洋脊)以及西北部亚丁湾的希贝洋脊(Sheba Ridge)和红海均显示有广义的 Dupal 异常特征($\Delta 8/4 > +10$)，但只有中印度洋脊存在典型 Dupal 异常($\Delta 8/4 > +60$)，其中罗德里格斯三联点处的异常最明显($\Delta 8/4$ 可达+71.4)。Nishio et al.(2007)通过 Li, Sr Nd 同位素分析，认为印度洋罗德里格斯三联点的 Dupal 异常主要是由于循环的大陆下地壳引起。从罗德里格斯三联点由南往北 Dupal 异常逐渐减弱，希贝洋脊(Sheba Ridge)和红海均不具有典型 Dupal 异常特征(图 3c)。

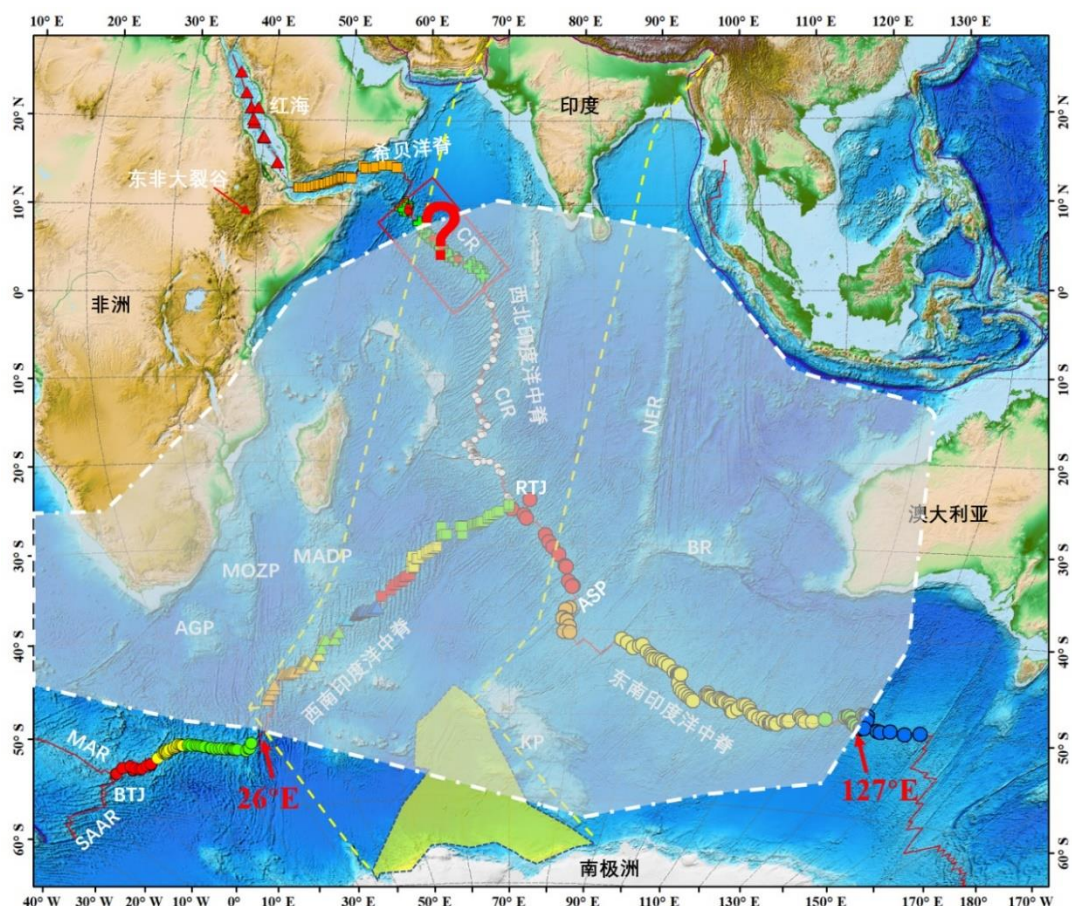


图 2. 印度洋 DUPAL 异常分布界线推测图 (余星, 李金金等, 2021, 未发表)

图中白色阴影区为基于 Hart(1984)绘制的 $\Delta 8/4 > 40$ 范围以及公开洋中脊玄武岩数据绘制的 Dupal 异常范围, 其中西南印度洋中脊及东南印度洋中脊的 Dupal 异常界线已比较明确, 而卡尔斯伯格脊的异常界线仍不清楚。卡尔斯伯格脊上红色圆点为已发表的同位素数据, 绿色十字为我国大洋调查航次样品。黄色虚线为印度大陆相对南极的移动轨迹。MAR—大西洋中脊, SAAR—南美洲-南极洲洋中脊, CR—卡尔斯伯格脊, CIR—中印度洋脊, BTJ—布维三联点, RTJ—罗德里格斯三联点。MADP—马达加斯加海底高原, MOZP—莫桑比克海底高原, AGP—厄加勒斯海底高原, KP—凯尔盖朗海底高原, ASP—Amsterdam-St.Paul 海底高原, BR—Broken 海脊, NER—东经 90 度海岭

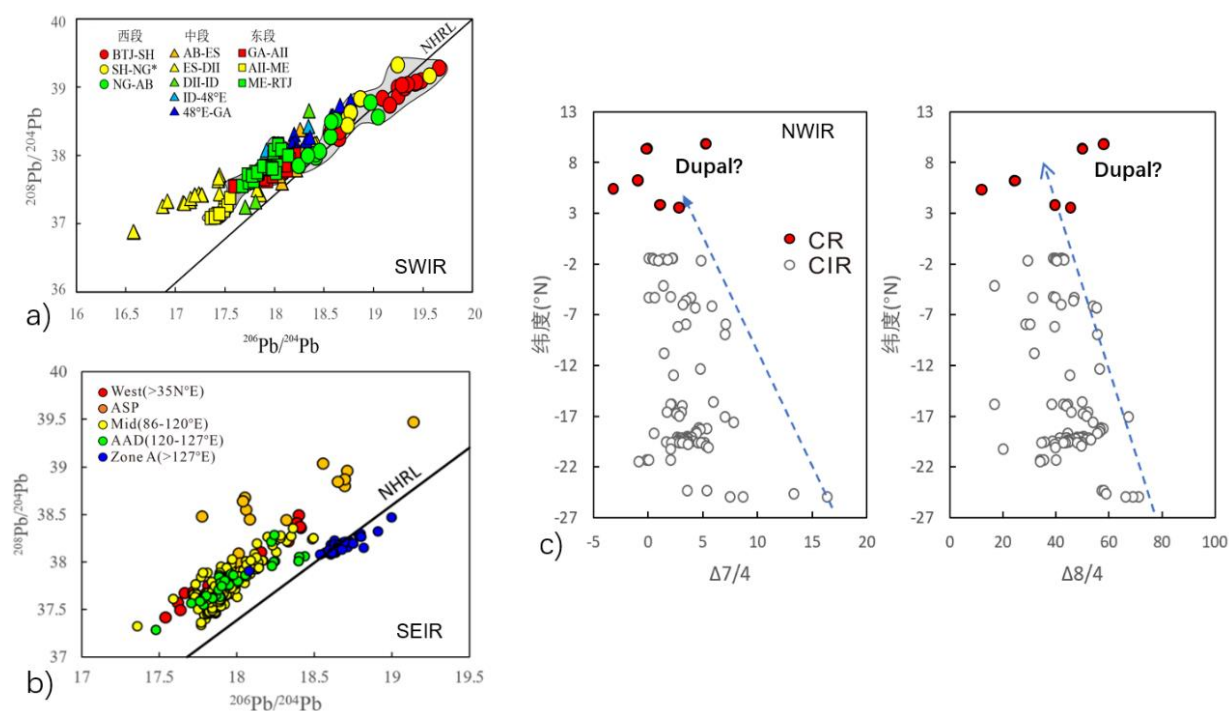


图 3. 印度洋三大洋脊的 Pb 同位素特征及 Dupal 异常指标（余星等，2019a,2019b,2020）

SWIR—西南印度洋中脊，SEIR—东南印度洋中脊，NWIR—西北印度洋中脊，
CR—卡尔斯伯格脊，CIR—中印度洋脊，ASP—Amsterdam-St.Paul 海底高原，
AAD—澳大利亚-南极错乱带

印度洋洋壳的地球化学组成极不均一，其不均一性的表现随区域而变化，但系统的变化规律仍不清楚。Dupal 异常在印度洋广泛分布，但其分布的界线仍存在争议，识别异常的地球化学指标也存在不确定性和片面性。对地球化学表象和客观规律的认识不足，限制了对印度洋洋壳地球化学不均一性成因的理解。

因此，亟待利用印度洋全局的洋壳地球化学数据，开展大数据分析和机器学习，全方位了解地球化学异常的特征及分布，明确多维度异常之间的关联性，理清 Dupal 异常的地球化学指标，圈定异常的界线，从而结合区域的地形、底质分布、重磁异常、层析成像等地质地球物理资料，结合区域构造演化，利用多图层空间分析和相关分析，揭示地球化学异常在地质及地球物理上的对应关系，讨论洋壳不均一及 Dupal 异常的成因。

二、研究内容和预期目标

1. 研究内容

(1) 印度洋基岩分布规律研究。

印度洋洋壳基岩的类型与洋壳地球化学不均一性密切相关。岩性与元素地球化学属于不同尺度的研究范畴，岩性控制了第一级别的地球化学组成，不同岩性可以有明显不同的地球化学特征。因此在研究洋壳地球化学不均一性，首先要求对比相同类型的岩石，比如全区的玄武岩对比，橄榄岩的对比。

基于印度洋区域海底地形、已有的沉积物分布、IODP 钻井数据、公开底质调查资料等，利用岩石地球化学数据库中的岩性记录，识别印度洋海底基岩的类型，圈定其范围或出露位置，掌握基岩的分布特征；叠合基岩空间分布和地形资料、离轴距离、离岸距离等参数，利用空间分析和相关分析，讨论基岩或底质的分布规律。

(2) 印度洋基岩地球化学分布特征研究。

对于同一种岩性，如沿洋中脊分布的玄武岩，不同区段表现出不同的地球化学特征。这些异常特征可以反映在主量元素、微量元素或同位素等不同维度，不同维度的异常代表了不同的成因过程，并且可能彼此关联。

选择洋壳的主体基岩类型，如玄武岩和橄榄岩，基于全岩或玻璃主微量元素丰度、同位素比值及其空间信息，利用聚类分析和判别分析，揭示印度洋洋中脊及洋盆基岩在主微量元素及同位素等多个维度的空间分布特征，识别异常区；利用不同维度异常区的空间相关关系，讨论印度洋洋壳的综合地球化学异常分布规律。

(3) 基于机器学习的洋壳地球化学异常指标遴选。

地球化学参数覆盖整个元素周期表，并包括同位素以及不同元素间的比值或其他衍生参数。通常所指的地球化学异常往往指代某一参数或某几个参数反映的异常值，而缺少对全局参数的系统性研究。随着机器学习算法和大数据技术的普及，利用增强采样、特征工程、模式识别手段进行各种参数及参数组合的模拟计算和机理分析，有望挖掘出更多有效的地球化学异常指标，为后续的异常判别提供帮助。

按不同维度或性质的地球化学参数，比如放射性同位素、稀有气体同位素、

稀土元素、微量元素、主量元素等，基于已知的反映不同地质过程的参数，如放射性同位素示踪地幔源区，主量元素衍生参数 $Mg\#$ 可指示岩浆演化， Na_8 和 La/Sm 可反映地幔源区熔融程度， $\Delta 7/4$ 和 $\Delta 8/4$ 反映 Dupal 异常等，利用分类算法，建立参数-区域异常的量化模型，识别出同类的地球化学异常指标，确定其识别标准，增进对异常的全面了解，从而帮助异常解释。

(4) 印度洋洋壳地球化学不均一性成因研究。

洋壳地球化学不均一的现象具有不同尺度多维的控制因素。岩性、蚀变过程、岩浆演化、部分熔融、地幔源区组成、形成时代等各种因素共同塑造了现今洋壳岩石的地球化学特征。而这些因素又与洋壳在时空背景下的演化过程密切相关，透过洋壳地球化学不均一性的研究可以探究地球系统相互作用的规律。

根据可靠的地球化学异常判断指标，明确印度洋各尺度各类地球化学异常的综合证据，如 Dupal 异常的综合判别指标，圈定各异常的空间分布，明确其表现尺度。结合地形、底质分布、重力磁力异常、层析成像等地质地球物理资料，考虑在时间维度上区域的构造演化过程，利用多图层空间分析和相关性分析，揭示地球化学异常在地质及地球物理上的对应关系，讨论地球化学异常的成因。

2. 预期目标

(1) 利用聚类分析和判别分析，揭示印度洋洋中脊及洋盆基岩在主微量元素及同位素等多个维度的空间分布特征，识别异常区；利用不同维度异常区的空间相关关系，揭示印度洋洋壳的综合地球化学异常分布规律，并明确异常的尺度。

(2) 明确印度洋各尺度各类地球化学异常的综合证据，结合地形、底质分布、重力磁力异常、层析成像等地质地球物理资料，探讨印度洋典型尺度地球化学异常的成因。

三、研究方案及技术路线

1. 研究思路

本研究拟采用岩石地球化学分析与大数据分析相结合的方法，希望通过大数据处理和机器学习等手段，从大量的岩石地球化学数据中，提取地球化学异常的指标及分布规律，利用多源地质、地球物理和地球化学数据的叠合空间分析，识

别地球化学异常对应的地质地球物理特征，从而探讨印度洋洋壳地球化学异常的成因（图 4）。

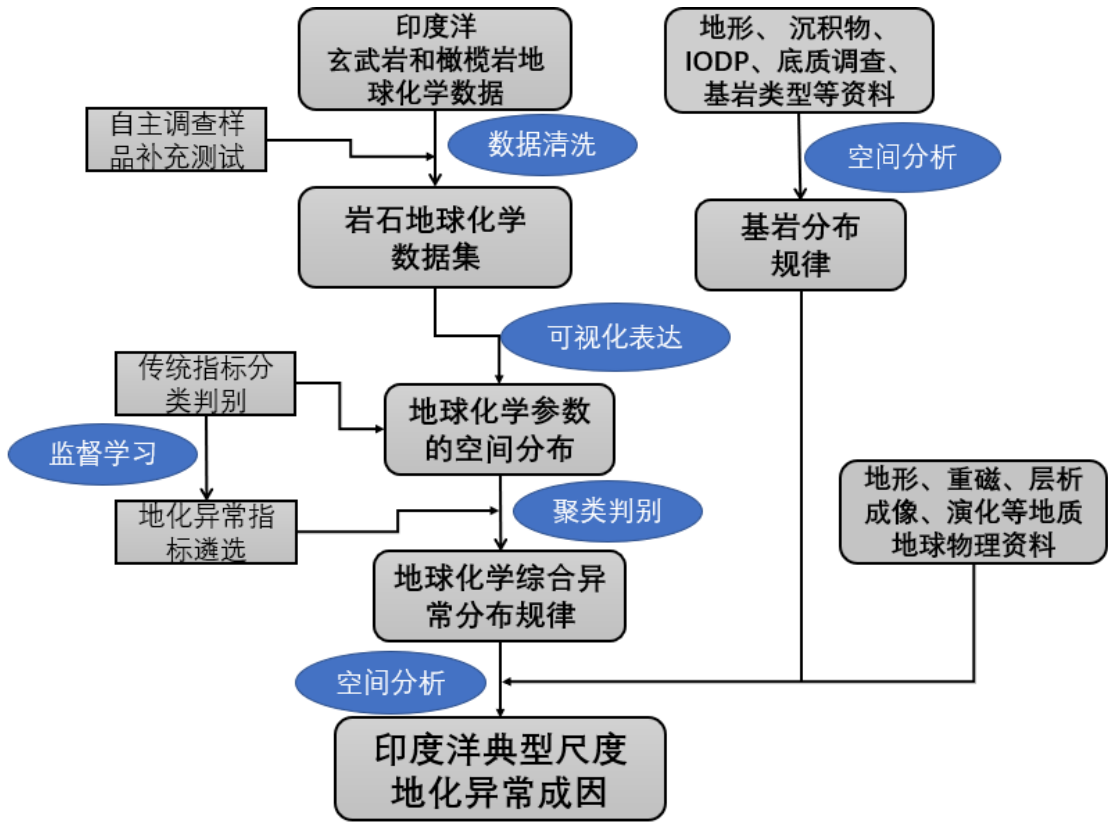


图 4. 项目研究思路流程图

2. 技术路线

本项目的技术路线如图 5 所示，分为 4 个模块：数据采集与清洗、机器学习模型建立与分析、地球化学可视化设计、地球化学成因判别与分析。涉及大数据、机器学习、地球化学等多学科领域问题，是集人工智能、化学、物理、地质学等多学科的一项交叉学科研究项目。本项目主要采用的研究技术有：多维相关性分析、增强采样、分类器、集成学习、特征工程、聚类分析、降维与度量学习等。

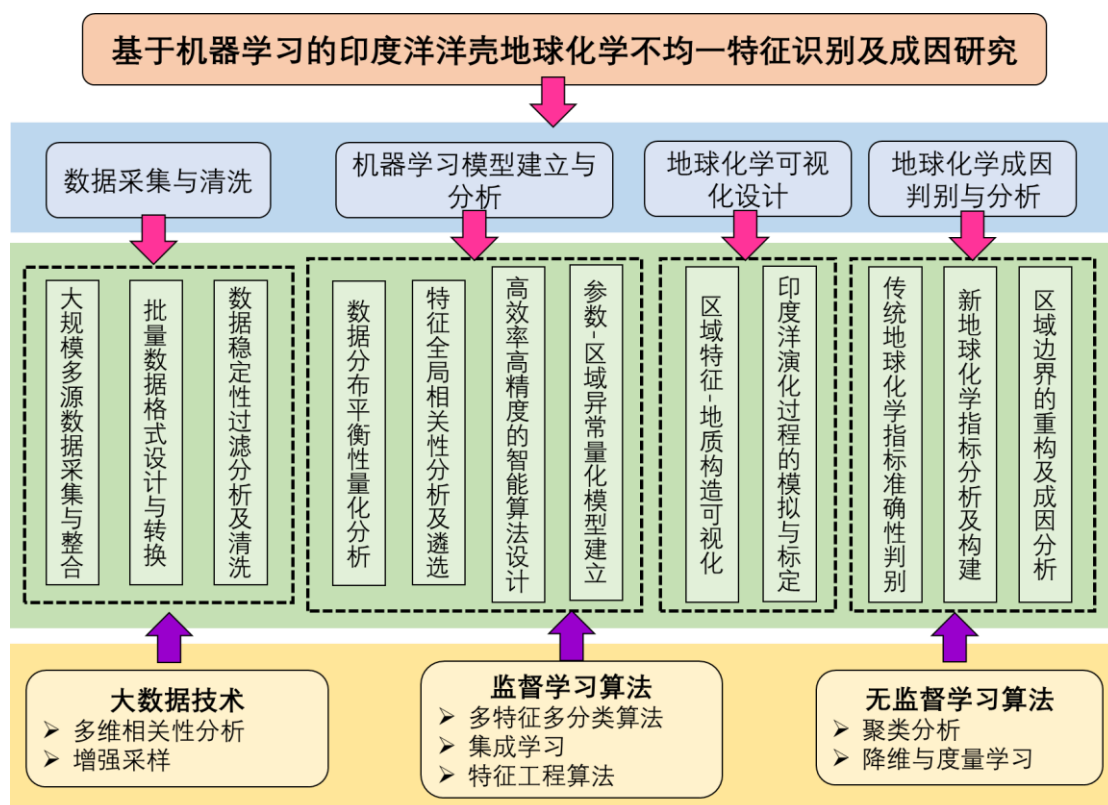


图 5. 项目技术路线图

(1) 数据采集与清洗

数据来源主要为网络公开数据库 (EarthChem、PetDB 及 GEOROC 等网站)。针对数据格式不统一、数据重叠、指标冗余等问题, 需要对数据进行整合, 并设计出高效可靠的易读、易存储的计算机数据格式。借助于传统的地球化学指标分析方法, 对数据的稳定性进行初步过滤性分析和清洗。最终建立一个全面、规范的印度洋洋壳地球化学指标数据库。

(2) 机器学习模型建立与分析

在大数据的基础上, 首先利用大数据技术对数据的平衡性进行鉴别与分析, 并采用增强采样 (下采样、过采样等) 的方法缓解数据不平衡的问题。为了全面考察印度洋洋壳地球化学指标对异常区域的成因, 结合地形、底质分布、重磁异常、层析成像、其他微量元素特征和区域构造动力学资料, 利用无监督算法 (聚类算法、相关性分析算法等) 分析特征的全局相关性, 利用特征工程 (末位淘汰制等策略) 提取出能够最大化表示原始数据的低维特征向量, 并分析其物理成因。根据提取出的特征, 结合监督学习算法 (分类器、集成学习等), 将 Dupal 异常

与关键地球化学指标联系起来，训练出高效、高精度的异常识别模型，从而可以快速地量化指标与异常区域之间的联系。

(3) 地球化学指标的遴选

传统上地球化学指标种，主量元素上可以通过 $Mg\#$ 识别出不同脊段其 MORB 经历的后期演化程度，通过 Na_8 得出不同脊段 MORB 所反应的源区熔融程度等，通过微量元素 (Sm/Yb , La/Yb) 可以识别出源区岩石矿物的物质组成，通过同位素特征 (Sr , Nd , Pb) 可以识别出其下伏地幔端元特征。通过机器学习和大数据分析，分析多源、高维的地球化学指标，基于传统地球化学指标，遴选出高相关性、可靠地的新地球化学指标，完善传统指标体系的不足。

(4) 地球化学可视化设计

根据上述的数据分析和识别模型，将地球化学指标与地质背景信息相结合得出直观的区域性特征(各脊段的地球化学边界，各地幔端元在印度洋区域的分布，Dupal 的边界等)，制作印度洋各类地球化学异常图，包括 Dupal 异常分布图，沿洋中脊的地球化学异常图等。

(5) 地球化学异常成因分析

基于遴选的地球化学异常判别指标，明确印度洋各尺度各类地球化学异常的分布，结合地形、底质分布、重力磁力异常、层析成像等地质地球物理资料，考虑印度洋洋壳的构造演化过程，对比地球化学异常与地质地球物理特征，开展空间分析，揭示印度洋典型尺度地球化学异常的成因。

3. 可行性分析

本项目融合了我国人工智能、智能算法、海洋科学、地质学等领域的一流科研团队，团队中多个成员此前已有紧密合作和科学探讨，积累了扎实的研究基础。从多个方面为本项目提供了强有力的可行性基础，具体表现为：

(1) 合作双方前期已展开良好的合作研究

合作双方团队在监督学习、无监督学习、大数据统计等领域具备优越的人才队伍，在海洋岩石数据采集、分析以及人工智能算法领域具备扎实的研究基础，提出了很多创新性的算法。上海交通大学李金金团队和自然资源部第二海洋研究所余星团队在本项目的前期关键技术问题上已经有了一定的研究基础。针对前期

的印度洋洋壳地球化学的小样本数据，团队设计了易读、易存储的数据格式，初步建立了一个小型印度洋洋壳数据平台（样本数接近 8000），结合多维相关性分析和聚类算法，考虑了地球化学指标之间的相关性和冗余程度，并根据异常区域给与了初步的成因分析。基于此，双方合作团队提出了一个基于人工智能的印度洋洋壳地球化学指标的分析 and 识别算法（An Artificial Intelligence-based Tool for the Analysis and Identification of Geochemical Indicators in the Indian Ocean crust, AI-GIIOC），该算法对于印度洋洋壳地球化学的指标分析是通用的，可以对不同规模的数据进行处理和分析。相关算法已经提交软件著作权。此外，上海交通大学李金金团队和自然资源部第二海洋研究所余星团队双方共同发表 SCI 科研论文一篇，Jinjin Li, Xing Yu, et al. Nano Energy (2021) (IF=16) (2021 年 6 月接收)。这也是本项目顺利实施的重要可行性保障和双方合作的研究基础。

（2）项目拥有优越的科研平台和创新环境

上海交通大学团队整合了计算机科学、人工智能、信息制造等多学科交叉的科研力量，建立了专业从事机器学习、深度学习算法研究平台，拥有全面的多学科人才队伍。团队具有 50 平方米，具备专业高性能超级计算机的无尘机房一间，在该机房中部署了 60 台计算服务器，约 1500 个 CPU 核心和 20 块 GPU 卡。此外还购买了上海交通大学 AI 计算平台，其计算能力为目前国内高校最强(图 6)。该平台深度学习张量计算能力可以达到 16PFLOPS；通过搭载 NVIDIA NVSwitch 创新技术，GPU 间带宽高达 2.4 TB/s；AI 计算平台采用可扩展架构，使得模型的复杂性和规模不受传统架构局限性的限制，从而可以应对众多复杂的人工智能挑战。团队雄厚的人才力量和超算平台，为本项目的顺利实施提供了技术平台支撑保障。

海洋二所团队依托的自然资源部海底科学重点实验室具有完善的实验分析平台，可以满足地球化学补充测试。实验室可完成样品的磨片、碎样、单矿物挑选、镜下鉴定、矿物化学分析和微量元素测试等工作。实验室中配备精密的岩石切割设备和研磨抛光设备，能够自动完成岩石薄片、探针片或包裹体薄片的加工。岩矿分析实验室还配备了各种显微镜，用于岩相观察和矿物鉴定。矿物显微结构分析和矿物成分分析可以在 JEOL JXA-8100 型电子探针上进行。X 射线荧光光谱仪(XRF)和带 CETAC 213 nm 激光器的 Elan DRC-e 电感耦合等离子体质谱仪

(ICP-MS)，可以满足全岩主量元素和微量元素的分析（图 7）。为本项目的顺利实施提供了强有力的实验条件。

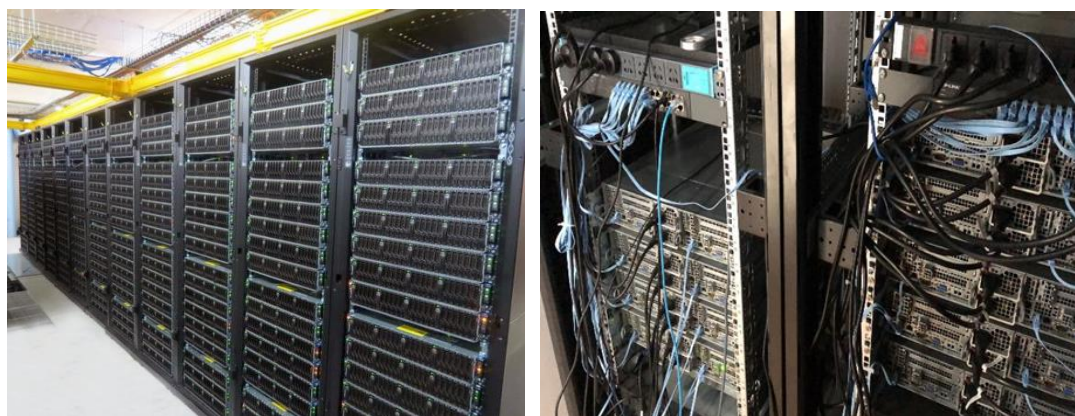


图 6. 上海交大团队拥有的超级服务器计算平台



图 7. 海洋二所团队拥有的地球化学分析平台

四、创新点和特色

(1) 利用地质学、地球化学、计算机科学、人工智能等多学科交叉的方法，首次结合机器学习算法和大数据技术，全面考察多种同位素、微量元素，融合实

验分析测试和数值模拟，建立高效、高精度的地球化学指标与地质背景之间的相关性，系统全面地分析地球化学参数对异常区域的影响，为解释印度洋洋壳地球化学指标分布及其成因提供指导。

(2) 基于大数据系统研究印度洋全区的地球化学异常特征，并从多尺度多维度揭示异常的表现形式、相应的地化指征以及对应的异常分布范围，同时综合其他地质与地球物理资料，从不同尺度上解释印度洋洋壳地球化学不均一的成因。研究成果将为认识印度洋岩浆作用和构造演化过程提供支撑。

五、年度计划

(1) 2021 年 7 月 - 2021 年 12 月

采集网络多源公开数据库 (PetDB、EarthChem 和 GEOROC)，整合数据格式，剔除冗余指标，设计出高效可靠的易读、易存储的计算机数据格式。建立规范全面的印度洋洋壳地球化学数据库，为后期进行基于机器学习的印度洋洋壳地球化学不均一特征识别及成因研究奠定数据库基础。

(1) 2022 年 01 月 - 2022 年 06 月

利用大数据技术对数据的平衡性进行鉴别与分析，并采用增强采样的方法缓解数据不平衡的问题。利用无监督算法 (聚类算法、相关性分析算法等) 分析特征的全局相关性，采用特征工程提取出能够最大化表示原始数据的低维特征向量，并分析其物理成因。进一步地，根据提取出的特征，结合监督学习算法，将 Dupal 异常与地球化学指标联系起来，训练出高效、高精度的异常识别模型，快速地量化指标与异常区域之间的联系。

(3) 2022 年 7 月 - 2022 年 12 月

根据上一步的数据分析和识别模型，将地球化学指标与地质背景信息相结合得出直观的区域性特征。通过地球化学所展现出的区域特征与地质构造背景相结合，得到对印度洋演化的深度解析，以及得出构造环境对岩浆过程，MORB 成分的控制机制。

(4) 2023 年 01 月 - 2023 年 06 月

对传统地球化学指标准确性进行判别，同时对表现出明显区域性特征的新地球化学指标分析其地球化学成因，对传统地球化学指标所划定到的区域边界或判

别标准进行校正。建立基于机器学习的印度洋洋壳地球化学不均一特征识别算法，结合多源地质与地球物理资料，探讨地球化学异常成因。

六、研究基础（相关的研究工作积累和已取得的工作成绩等）

上海交通大学项目负责人李金金研究员，国家千人计划青年特聘专家，“人工智能与微结构”实验室 PI。团队在人工智能交叉领域具备扎实的研究基础，提出了很多创新性的算法，发表 SCI 科研论文 100 余篇，相关科研成果发表在 Physics Reports (IF = 25.8)、Progress in Materials Science (IF = 38.1)、Advanced Science (IF = 15.6)、Nature communication (IF = 14.2)、Energy Storage Materials (IF = 17.1)、Nano Energy (IF = 17.2)、Briefings in Bioinformatics (IF = 10.2)、npj Computational Materials (IF = 11.8) 等顶级期刊上，团队授权/申请专利 40 余项，基于机器学习的软件著作权 10 余项。针对数据规模太大导致难以体系化研究的问题，团队巧妙利用“分而治之”的思想，将量子分块技术和神经网络(Neural network)结合，将分块子网络(Subnet)的贡献之和作为最终输出目标，该方法直接将指数级的时间复杂度降低为低线性标度，极大地提高了计算和训练效率。此外，团队将迁移学习(Transfer learning)应用小样本体系中，在精度持平条件下，数据量仅为从头训练(From scratch training)的十分之一，解决了因数据稀缺而导致精度低的问题(如图 8)。针对数据样本不均衡导致误差偏大的问题，团队提出首先采样无监督聚类算法(K-Means)对数据样本进行合理划分，再用监督学习算法支持向量机(Supporting vector machine)进行预测，降低了近 50% 的误差。

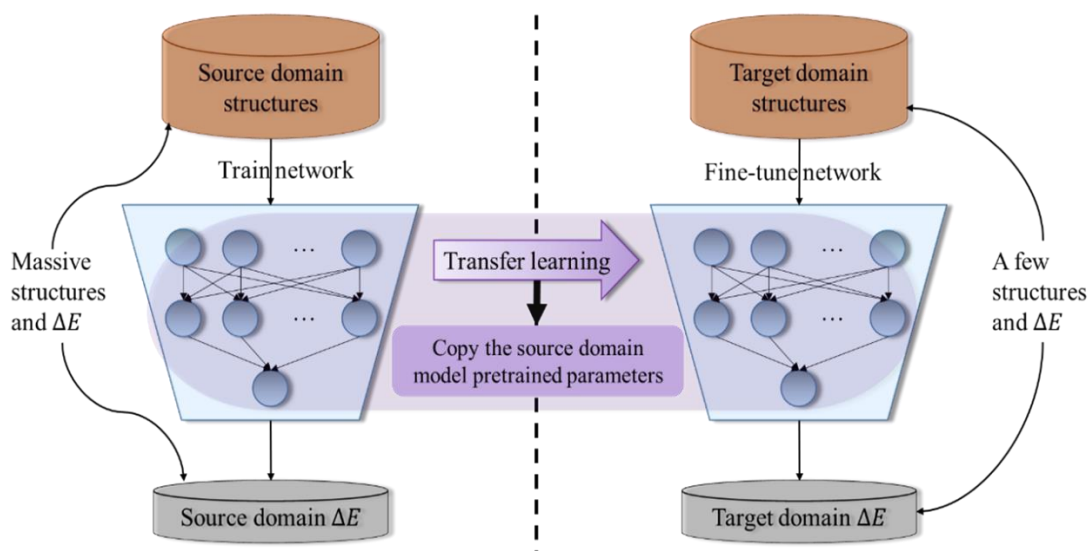


图 8. 上海交通大学课题组提出的深度神经网络和迁移学习模型示意图

自然资源部第二海洋研究所余星研究员团队，多年从事印度洋洋中脊的岩浆作用研究。从 2012 年开始，中国大洋 24、26、28、33、38、43、49 和 57 等 8 个航次前往西北印度洋区开展硫化物调查，其中 38 航次为“蛟龙”号载人深潜航次，团队作为下潜科学家参加了第 127 潜次的研究区海底调查，沿火山脊剖面采集了一系列 MORB 样品。通过调查研究，目前已积累了西南印度洋中脊和西北印度洋中脊大量的样品和数据资料，同时对洋中脊的海底岩石类型以及岩石地球化学特征有了较深入的认识。对洋中脊的地形地貌特征、构造发育情况、洋脊分段、扩张速率、岩性分布等都有了基本的了解。

另一方面，利用公开数据库资料，搜集了印度洋三大洋中脊的岩石地球化学数据以及地质地球物理资料，对三条洋中脊的作用域、演化历史以及地球化学特征有了初步的认识，研究成果总结发表于《地球物理学报》、《科学通报》、《地球科学》等刊物（余星等, 2019a, 2019b, 2020）。同时对于局部洋中脊的岩浆熔融异常以及地球化学异常的成因进行深入的探讨（Yu and Dick, 2020 EPSL; Zong et al., 2020）。通过前期的研究，已对印度洋西北印度洋脊、西南印度洋脊和东南印度洋脊的岩石分布、类型、地球化学特征，印度洋及周边各大火成岩省的时代、分布、演化特征等有了系统的认识，对印度洋区冈瓦纳大陆裂解的历史，各块体分离的时间，各洋脊发育顺序及持续时间均开展系统的梳理（图 9）。这为本项目的研究奠定了良好的基础。

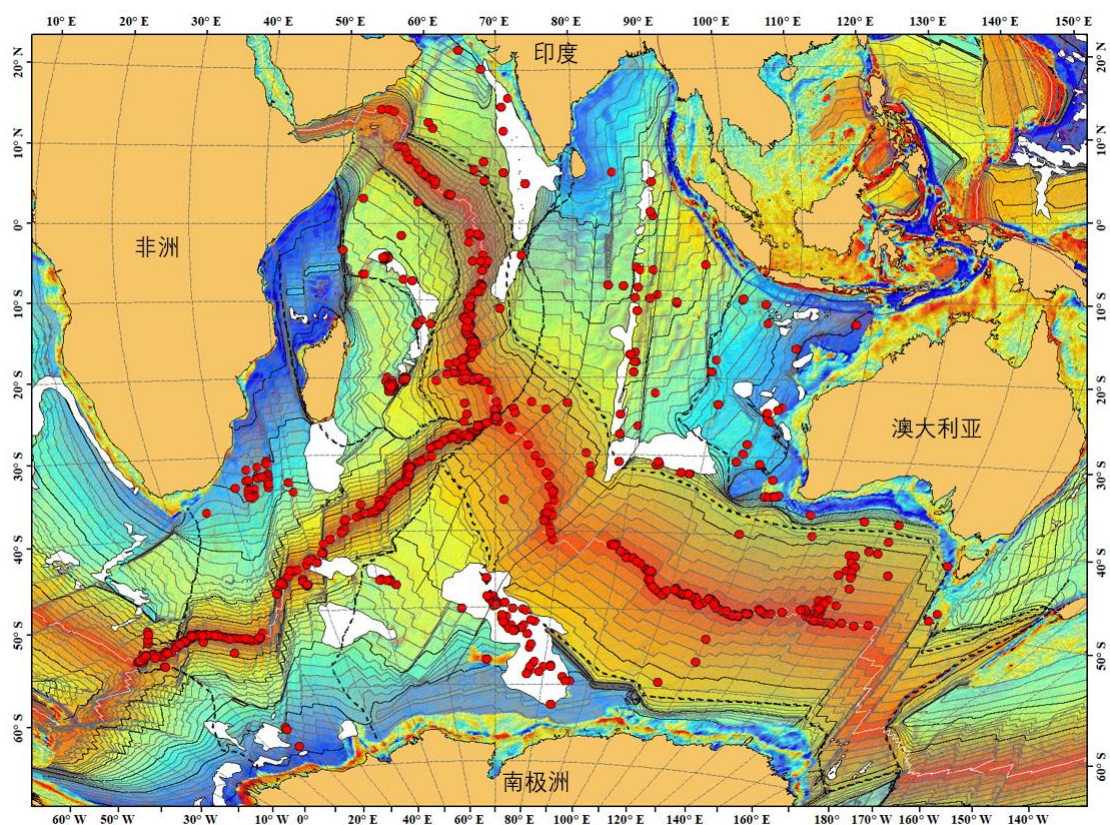


图 9. 印度洋岩石站位分布图

底图为自由空间重力异常图，黑色平行条带为洋壳的年龄等时线(5Ma 间隔)，
白色区域为海底高原，黑色虚线洋中脊作用域分区。

七、经费预算

科目名称	经费	计算依据及理由
设备费	0	/
图书资料	0	/
印刷费	8	项目预计发表2篇以上论文，申请2项以上专利。其中每篇论文需要提交版面费约2万元，每项专利提交申请费约1万元，共预算6万元。另额外需要2万的论文编辑费。印刷费共预算8万。

国外差旅费	0	/
国内差旅费	6	上海至北京、苏州等地往返机票平均按2000元，每人每次费用以3750元计。2年课题执行期每年按照业务调研8人次，预算为：差旅费=每次往返总费用×人次数×项目执行年限=3750×8×2=6万元。
市内交通费	2	用于上海市进行学术交流产生的市内交通费。1万元/年×2年=2万元。
耗材	0	/
会议费	0	/
维修费	3	对实验室和仪器设备产生的维修费，1.5万/年×2年=3万元。
人员费（发给校内专家）	0	/
劳务费（发给校外专家、学生等）	3	参加本项目的研究生助研津贴：5名硕士生每人每月300元，总额为：(300×5)元/月×10月/年×2年≈3万元。

其他（测试化验加工费、燃料动力费）	28	<p>(1) 材料表征和测试费： 根据《上海交通大学分析测试中心》和《上海交通大学 AEMD 加工公共平台》两个校内独立经济核算单位的价格进行预算。本项目中微加工的多个单项工艺需要在公共加工平台进行，主要包括：扫描电子显微镜：500 元/次×500 次=2.5 万元；透射电子显微镜：1000 元/次×25 次=2.5 万元；X 射线衍射：1000 元/次×25 次=2.5 万元；X 光电子能谱：1000 元/次×25 次=2.5 万元；Raman 光谱测试：1000 元/次×25 次=2.5 万元。共计 12.5 万。</p> <p>(2) 燃料动力费： 考虑到本课题需要对多层微结构进行微细加工测试，主要用于加工设备运行消耗的动能费用，包含水费 3.65 元/吨*2740 吨=1 万元；电费 0.65 元/度*50000 度/年*2 年=6.5 万元。共计 7.5 万元。</p> <p>(3) 高性能计算机使用机时费： 将租用“上海交通大学超算中心”、“天河二号”等计算测试中心的机组进行计算。平均每 24 小时计费 20 元。基于人工智能算法优化对结构进行计算，每 1 个结构需要在 24 核 CPU 下优化约 10 小时，400 个结构共需要 4000 小时。预算为 20 元/天×4000 小时=8 万元。 预算共计 12.5 万+7.5 万+8 万=28 万元。</p>
合 计	50	

八、项目组主要成员表

序号	姓名	性别	年龄	职称	所在单位	分工	参加月数	签名
1	李金金	女	37	研究员	上海交通大学	机器学习模型建立与分析	9 月/年	
2	余星	男	39	研究员	自然资源部第二海洋研究所	地化异常成因	9 月/年	
3	汪志龙	男	24	博士生	上海交通大学	大规模多元数据与整合	9 月/年	
4	韩彦强	男	28	博士生	上海交通大学	数据稳定性过滤分析及清洗	9 月/年	
5	林夕蓉	女	29	博士生	上海交通大学	数据分布平衡性量化分析	9 月/年	
6	蔡俊飞	男	23	硕士生	上海交通大学	参数-区域异常量化模型建立	9 月/年	
7	武思诚	男	23	硕士生	上海交通大学	区域特征-地址构造可视化	9 月/年	
8	张林	男	25	硕士生	上海交通大学	印度洋演化过程模拟与标定	9 月/年	
9	于泽灏	男	23	硕士生	上海交通大学	地球化学指标准确性判断	9 月/年	
10	唐洁群	女	24	硕士生	上海交通大学	区域边界重构及成因分析	9 月/年	
11	胡航	男	24	研究生	自然资源部第二海洋研究所	地球化学指标遴选	9 月/年	
12	党牛	男	24	研究生	自然资源部第二海洋研究所	地化数据可视化表达	9 月/年	

13	许绪成	男	26	研究生	自然资源部第二海洋研究所	数据处理与清洗	9 月/年	
----	-----	---	----	-----	--------------	---------	-------	--

九、申请者承诺

我保证申请书内容的真实性，如果获得基金资助，我将严格遵守上海交通大学“深蓝计划”基金的有关规定，认真开展研究工作。如果填报失实和违反规定，将承担全部责任。	
申请人签字：	合作申请人签字：
年 月 日	年 月 日

十、项目承担单位意见

单位负责人签章：	
单位公章：	
年 月 日	

十一、合作单位意见

合作单位负责人签章	合作单位负责人签章
单位公章：	单位公章：
年 月 日	年 月 日

十二、相关附件