

任梦姣, 孙记红, 王晓宇, 等. 海洋地质数据挖掘系统研发及在海山年龄预测中的应用[J]. 海洋地质前沿, 2023, 39(2): 79-85.

REN Mengjiao, SUN Jihong, WANG Xiaoyu, et al. Marine geological data mining system development and its application in seamount age prediction[J]. Marine Geology Frontiers, 2023, 39(2): 79-85.

# 海洋地质数据挖掘系统研发及在海山年龄预测中的应用

任梦姣<sup>1</sup>, 孙记红<sup>2\*</sup>, 王晓宇<sup>1</sup>, 容伊霖<sup>1</sup>, 白永良<sup>1</sup>

(1 中国石油大学(华东)海洋与空间信息学院, 青岛, 266580; 2 中国地质调查局青岛海洋地质研究所, 青岛, 266237)

**摘要:** 利用数据挖掘技术分析海洋地质调查数据, 以获取其中隐藏信息, 对推进海洋地质数据的科学高效利用具有重要意义。在模块化设计原则下, 利用 Python 语言开发海洋地质数据挖掘相关的核心功能, 利用 WinForm 搭建人机交互界面, 并通过参数交互的方式实现了界面和后台功能间的互动。基于综合地质地球物理资料, 利用软件预测了太平洋海山年龄。预测结果精度高于利用传统克里金插值方法所得结果的精度。应用结果表明, 所开发软件的数据预处理、指标分析、综合评价等功能具有很好的实用性。

**关键词:** 海洋地质; 数据挖掘; 系统研发; 海山年龄

中图分类号: P628.4 文献标识码: A DOI: 10.16028/j.1009-2722.2022.251

## 0 引言

随着海洋地质调查数据采集效率的不断提升, 数据获取速度和周期得到明显改善。许多国家已经积累了大量多来源、多类型、多时段、多维度的海洋地质调查数据<sup>[1,2]</sup>。利用不同地球物理手段对同一区域进行探测时, 可以采集到不同类型的调查数据, 每类数据从不同角度反应了海洋地质信息<sup>[3,4]</sup>。充分利用多类型调查数据进行综合分析, 有利于准确揭示海洋地质、海底资源的规律和特征。目前, 对多源海洋地质调查数据进行协同分析的程度较弱。如何深入挖掘海量调查数据中的潜在信息, 发现海洋地质、海底资源的特征与规律, 是当前面临的一个关键问题。数据挖掘是借助统计分析、机器

学习、专家系统和模式识别等方法, 从海量数据中提取、整理、分析数据, 并发现数据中潜在规律的技术手段。因此, 利用数据挖掘技术分析海洋地质调查数据, 以获取其中隐藏的信息, 对推进海洋地质数据的科学高效利用具有重要意义。

目前, 随着数据库、人工智能和数理统计技术的不断进步, 数据挖掘技术得到了长足发展, 被广泛应用于商业金融、社会网络、公共服务<sup>[5-7]</sup>和地球科学等诸多领域<sup>[8]</sup>。现有海洋地质调查数据处理与分析系统多是针对单一来源或单一类型数据而设计的, 数据综合分析能力较弱, 导致大量数据无法得到有效利用<sup>[9]</sup>。近期有多位学者研究了利用多源调查数据挖掘潜在地质规律的方法和技术<sup>[10-12]</sup>, 但是多处于理论研究阶段, 尚无成熟的实用系统。因此, 有必要基于目前成熟的数据分析、挖掘等方法技术体系, 开发一套适用于海洋地质数据分析和挖掘的软件系统, 以提高海洋地质领域的的数据利用能力。

本文综合多种适用于海洋地质数据的挖掘分析方法, 充分利用计算机、数据挖掘、可视化、统计学等技术手段, 研发海洋地质数据挖掘系统。通过将该系统应用于太平洋海山年龄预测, 验证其实用性。

收稿日期: 2022-09-05

资助项目: 中国地质调查局项目(DD20190214, DD20221711); 国家自然科学基金面上项目(42176068); 山东省自然科学基金面上项目(ZR2020MD065)

作者简介: 任梦姣(1997—), 女, 在读硕士, 主要从事智能海洋地质方面的研究工作. E-mail: renmail1105@163.com

\* 通讯作者: 孙记红(1984—), 男, 硕士, 高级工程师, 主要从事海洋地质大数据及智能化应用方面的研究工作. E-mail: sjihong@mail.cgs.gov.cn

# 1 海洋地质数据挖掘方法

海洋地质调查数据包括空间位置数据、野外综合调查数据、样品实验分析数据、资料解释数据和调查成果数据等<sup>[13-14]</sup>(图1)。海洋地质调查数据类型多样,具有空间性强、尺度多、体量大、数据维度高等特点,导致预处理过程复杂、综合分析涉及计算量大等问题。为了解决上述问题并获得有效的海洋地质规律,需要对海洋地质数据进行数据预处理、指标分析以及综合评价等操作。

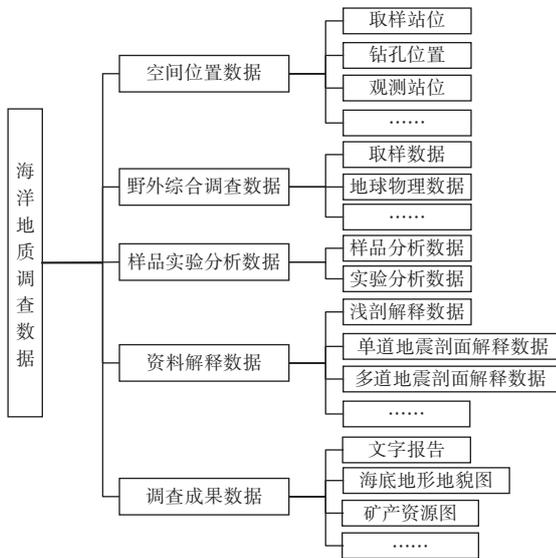


图1 海洋地质调查数据情况

Fig.1 The classification and composition of marine geological survey data

## 1.1 数据预处理

海洋地质调查资料存在数据量纲不同、采样间隔不等、隐藏畸变数据等问题。因此,在进行数据综合分析之前,需要进行预处理。当多源数据存在差异化量纲时,数据分布范围差别大。同时,纳入同一数据分析方法后,会导致不同数据的权重出现明显差异。为此,利用数据归一化的方法,采取线性变换的方式消除不同量纲带来的影响。当地质调查工作没能覆盖到重要研究区域时,通常采用数据插值的方法推测空白区域的调查数据。地学领域中常用的数据插值方法为克里金方法<sup>[15]</sup>。此外,通过频谱分析,能够有效筛选出畸变数据、发现数据的整体分布趋势和局部特征。

## 1.2 指标分析

不同类型海洋地质调查数据与研究目标间的

关联程度不同,且不同调查数据之间可能存在相关性。因此,需要明确不同变量/指标的权重,对相关性强的指标体系还需进行降维处理,提升综合评价的效率和精度。平均影响值(Mean Impact Value, MIV)是用于表征输入变量对输出结果影响程度的一个有效参数<sup>[16]</sup>。因子分析和主成分分析是常用的指标相关性分析方法。

MIV 计算的具体过程为:①选择1个输入参数,在原值基础上分别增加、减少10%,构成2组新样本;②基于2组样本,通过预测模型的模拟计算,得到2组预测结果;③将2组预测结果的差值按样本数求平均,得到该参数的MIV值。重复上述步骤,求出各相关参数的MIV值。MIV值的正负号代表参数对预测的影响方向,绝对值大小代表对预测结果的影响权重<sup>[17-19]</sup>。

因子分析能够揭示数据指标之间的关联性,进而提取主导指标、简化输入数据集<sup>[20]</sup>,得到的代表性指标可以有效保留原始指标的主要信息。通过线性变换的方式,主成分分析可以将海洋地质调查数据中的多个指标转化为一组在任何维度下都能保证线性无关的数据表示<sup>[21-22]</sup>,继而达到提取海洋地质调查数据中主要指标的目标。

## 1.3 综合评价

综合评价就是基于筛选出来的代表性指标集合,利用统计分析、机器学习等方法定量表征或者预测研究目标特征的综合分析方法。针对评价目标的特征,需要选择合适的数学方法。常用的海洋地质数据综合评价方法有灰色关联法、聚类分析、回归分析和神经网络法等。

误差后向传播(Back Propagation, BP)神经网络是按照误差逆向传播算法训练的多层前馈神经网络,其主要特点为信号前向传递、误差反向传播<sup>[23]</sup>。训练BP神经网络的步骤主要包括前向计算、误差反向传播及判断迭代是否结束。前向计算指输入数据通过权、阈值计算得到预测结果。误差反向传播的过程包括确定误差下降方向和下降速度,进而确定网络参数,以降低预测值与真实值之间的偏差。当偏差满足条件时,迭代结束,模型训练完成。

因BP神经网络无法区分极值点与最值点,最终结果大概率趋于局部极小。而应用遗传算法(Genetic Algorithm, GA)优化BP神经网络参数,则可以较好地避免陷入局部最优的问题,优化后的方法成为GA-BP方法<sup>[24]</sup>。遗传算法是模拟自然界遗

传机制和生物进化论而设计的一种并行随机搜索的最优化方法, 具有高效启发式搜索、并行计算等特点<sup>[25]</sup>。遗传算法主要包括种群初始化、适应度计算、选择、交叉与变异等计算。

## 2 海洋地质数据挖掘系统设计与实现

海洋地质数据挖掘系统的性能应满足实用性、方便性、兼容性和可扩展性的要求。本文将以此为

准则, 设计并实现该系统。

### 2.1 系统结构设计

本文的海洋地质数据挖掘系统以模块化设计为准则, 将各功能进行模块化封装, 各模块间松散耦合。每一模块均按照面向对象的思想进行设计, 结构明确、独立性强, 满足系统更新和扩展的需要。系统结构包括数据处理层、平台功能层和业务应用层 3 部分(图 2)。

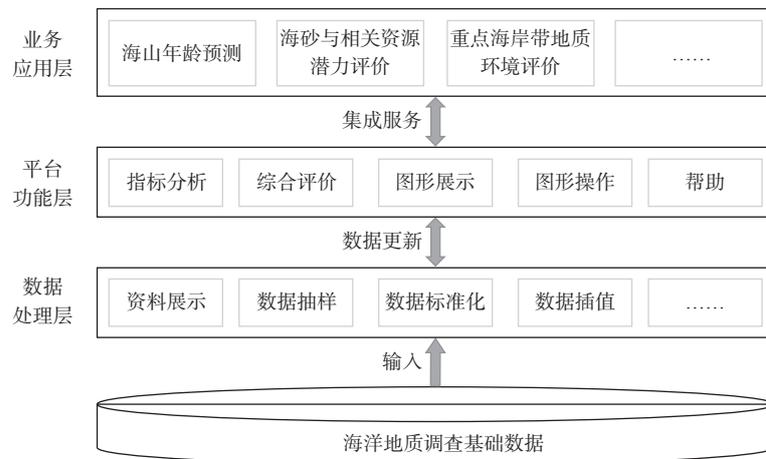


图 2 海洋地质数据挖掘系统总体架构

Fig.2 Structural tiers of the marine geological data mining system

### 2.2 系统功能设计

海洋地质数据挖掘系统包括 6 个功能模块: 数据预处理、指标分析、综合评价、图形展示、图形操作和帮助模块。主要模块包含的具体功能如下:

(1) 数据预处理模块包括数据清洗、数据抽样、数据标准化、数据离散化、数据平滑和数据插值等功能;

(2) 数据指标分析模块包括 MIV 分析、因子分析和主成分分析等功能;

(3) 数据综合评价模块包括聚类分析、灰色关联分析、随机森林、支持向量机和神经网络分析等功能;

(4) 数据图形展示模块可以利用统计图表展示数据, 具体图形形式包括散点图、直方图、折线图和饼状图等。

(5) 图形操作模块包括图形放大、图形缩小、图形漫游、图形清除、图形保存等功能。

(6) 帮助模块提供软件操作说明及版权归属说明。

### 2.3 系统开发

为了提高系统开发效率, 本文以权威的 Python 开源库为基础, 开发各功能的核心算法; 利用 Win-Form 设计人机交互界面; 最终利用参数交互的方式, 关联起软件界面与核心代码, 形成海洋地质数据挖掘系统。因此, 系统开发流程包括后台功能实现、软件界面搭建与参数交互 3 部分。

#### (1) 后台功能实现

利用 Python 标准库和第三方库, 有助于数据挖掘算法的快速实现。本文采用的第三方库主要有 NumPy、Scikit-learn 和 Matplotlib。NumPy 库主要用于数组、矩阵的基本运算, 能够进行高性能的数据处理。系统中的数据预处理、指标分析和综合评价模块采用的 Python 的 Scikit-learn 库。Scikit-learn 库是专门为数据挖掘开发的第三方库, 涵盖了当前常见的大部分机器学习算法。系统中的图形展示模块采用 Matplotlib 库。Matplotlib 库支持交互式和非交互式绘图, 可以绘制曲线图、条形图、饼状图等多种图形, 并能够将图形保存成多种栅格与矢量

格式。

PyInstaller 是 Python 的另一个第三方库,能够读取 Python 脚本,分析代码以发现脚本执行所需的所有模块和库。利用 PyInstaller 将相关 Python 程序打包成一个标准的独立可执行文件,这些文件可在未安装 Python 环境的计算机上运行。

### (2) 软件界面搭建

海洋地质数据挖掘系统采用 WinForm 框架构建系统界面。WinForm 框架可以进行 Windows 风格窗体设计,整体风格简约、美观,界面搭建的效率高。它可以采用插件化的开发模式,能够集成自主开发的模块和插件,便于系统更新和使用。WinForm 框架还提供了软件系统需具备的常见功能,并能够提供已封装好的、有关数据访问的基类,便于提高开发效率。

### (3) 参数交互

系统后台功能开发与软件界面搭建完成后,需要将 WinForm 框架与基于 Python 语言编写的后台功能进行连接,实现两者间的参数交互,最终完成系统的开发工作。

WinForm 框架获取到用户输入的必要参数并收到开始运算指令后,启动对应的后台可执行文件,并利用 Process.StartInfo.Arguments 将参数传递给已经启动的进程。当后台运算结束时,将运算结果

按照用户设定目录保存至相应文件,将计算结果以图表形式绘制到 WinForm 界面内,并传回运算结束的指令,由 WinForm 向用户给出运算完毕的提示。至此,基于图形界面,可以按照人机交互的方式,进行海洋地质数据分析、挖掘与成图等工作。

## 3 太平洋海山年龄预测实例

海山年龄不仅是研究大洋构造演化的一个重要参数<sup>[26]</sup>,而且对推进海洋资源的探寻、开发和利用具有重要意义<sup>[27]</sup>。太平洋板块与岩浆补给源的相对运动形成了一系列海山链、海山群<sup>[28]</sup>(图3)。受到岩石样本采集成本的限制,太平洋内大部分海山年龄仍是未知的。因此,本文以太平洋海山年龄预测来验证本文所开发系统的实用性。

### 3.1 指标筛选与预处理

许多研究已经证实重力异常<sup>[29]</sup>、磁力异常<sup>[30-31]</sup>、洋壳年龄<sup>[32]</sup>、沉积层厚度<sup>[33]</sup>、海山高度<sup>[34]</sup>和大洋热流<sup>[35]</sup>等地质地球物理参数与海山年龄相关。因此,选取上述地球物理观测数据作为海山年龄预测的评价指标。除了与海山年龄相关的地球物理资料外,还搜集了太平洋海山部分测年数据。具体数据来源见表1。

表1 用于太平洋海山年龄预测的基础数据源

Table 1 Data sources for the Pacific seamount age prediction

| 数据名称   | 分辨率/弧分 | 发布者  | 备注          |
|--------|--------|--|-------------|
| 空间重力异常 | 1      | SANDWELL等 <sup>[36]</sup>                  | 卫星测高数据为基础   |
| 磁力异常   | 2      | MAUS等 <sup>[37]</sup>                      | 4 km高度处的磁异常 |
| 洋壳年龄   | 2      | MÜLLER等 <sup>[38]</sup>                    |             |
| 沉积层厚度  | 5      | STRAUME等 <sup>[39]</sup>                   |             |
| 水深与地形  | 1      | AMANTE等 <sup>[40]</sup>                    |             |
| 海底热流   | 站位数据   | 国际热流委员会                                    |             |
| 海山年龄   | 点位数据   | V CLOUARD AND A BONNEVILLE <sup>[41]</sup> | 前人成果的汇编     |

利用海底地形等值线圈定海山范围。对于每类地球物理数据而言,取海山范围内数据点的平均值,作为对应评价指标的具体数值。当目标海山范围内缺少海底热流数据时,则根据周边站位的热流数据插值得到目标海山的热流值。海山高程即为海山范围内最高点与基础趋势面之间的高程差。

### 3.2 指标分析

在预测海山年龄前,为了排除弱相关变量的干

扰,采用 MIV 方法对潜在的输入指标进行分析。6 类指标的 MIV 值如表 2 所示, MIV 绝对值由大到小的顺序为洋壳年龄、重力异常、海山高度、沉积层厚度、磁异常和海底热流。海底热流的 MIV 绝对值仅为 0.21 Ma,其主要原因可能是多数海山的热流值是由海山周边实测热流值插值得到的,数据精度较差。因此,在综合评价中,不再将海底热流作为一个评价指标。

表 2 与海山年龄相关地球物理数据的 MIV 绝对值  
Table 2 The absolute mean impact values of the seamount-age related geophysical observables

| 地球物理观测值 | 数据类型 | 处理方法 | MIV /Ma |
|---------|------|------|---------|
| 洋壳年龄    | 网格   | 提取   | 9.23    |
| 重力异常    | 网格   | 提取   | 3.50    |
| 海山高度    | 网格   | 提取   | 1.42    |
| 沉积层厚度   | 网格   | 提取   | 1.08    |
| 磁异常     | 网格   | 提取   | 0.82    |
| 海底热流    | 地点   | 内插   | 0.21    |

### 3.3 海山年龄预测

预测海山年龄的总体思路是, 先训练 GA-BP 预测模型, 再利用模型进行实际预测。随机地将已知海山年龄的数据分成 2 组(图 3), 一组作为训练集, 用于预测模型的建立及参数选择; 剩余部分作为测试集, 用于预测效果的评价。将筛选出的地质地球物理指标输入 GA-BP 模型, 由输入输出指标的个数初步确定 BP 神经网络结构; 经反复训练后, 优选第 1 层含 12 个节点、第 2 层含 10 个节点、学习率为 0.01 的 GA-BP 神经网络。最后, 将训练好的模型应用于测试集, 预测结果如图 4 所示。

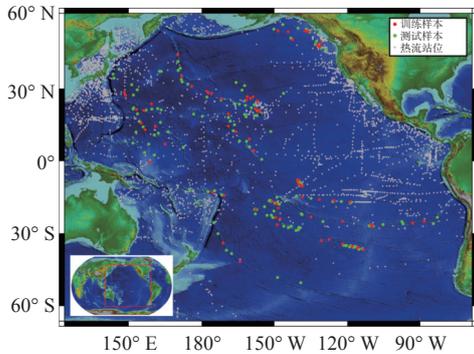


图 3 太平洋地形<sup>[40]</sup>、海山与热流站位分布<sup>[42]</sup>  
Fig.3 Topography of the Pacific<sup>[40]</sup> and the seamounts-heat flows distribution<sup>[42]</sup>

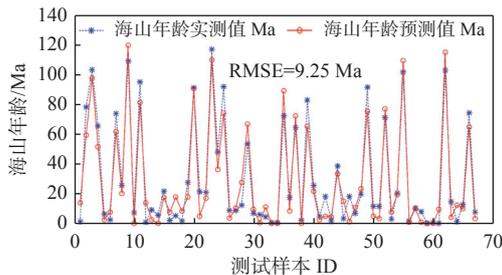


图 4 基于 GA-BP 模型的海山年龄实测值与预测值对比  
Fig.4 Comparison between the measured and predicted seamount ages of individual seamount

### 3.4 精度评价

当以测试集的实测年龄为基准时, 预测结果的均方根误差(Root Mean Squared Error, RMSE)和决定系数(coefficient of determination,  $R^2$ )分别为 9.69 Ma 和 0.93。图 4、5 展示了预测年龄与实测年龄之间的对比关系。图 5 表明, 基于 GA-BP 模型的海山年龄预测值与实测值之间呈现出较好的线性关系。为了更好地说明应用 GA-BP 模型的预测效果, 从均方根误差和决定系数角度, 将 GA-BP 神经网络模型的预测结果与利用传统克里金空间插值方法所得结果、仅用 BP 模型的预测结果进行对比(表 3)。无论从 RMSE 还是从  $R^2$  角度评价, 插值方法和 BP 模型的计算精度均不如 GA-BP 神经网络方法的预测精度。

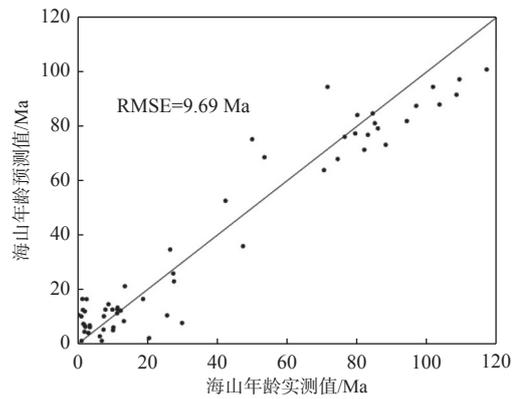


图 5 基于测试集的 GA-BP 海山年龄预测值与实测值散点图

Fig.5 Scatter plot of GA-BP's prediction and measured seamount ages based on test set

表 3 克里金插值算法、BP 和 GA-BP 算法预测精度对比  
Table 3 Comparison in prediction accuracy among the Kriging, BP, and GA-BP algorithms

| 预测模型    | RMSE/Ma | $R^2$ |
|---------|---------|-------|
| 克里金     | 27.54   | 0.60  |
| BP模型    | 20.35   | 0.74  |
| GA-BP模型 | 9.69    | 0.93  |

由于各海山链和海山群具有差异化的构造背景和形成过程, 导致不同海山链和海山群具有不同的地质地球物理场表现特征<sup>[43-45]</sup>。本文所搜集已知年龄的海山分散于不同海山链/海山群, 致使预测模型只学习了太平洋海山年龄分布的整体规律, 而忽视了不同海山链差异化的特征, 最终导致海山年龄的预测精度仍然有限。但是限于已知年龄样品的数量, 无法分区域学习规律、预测年龄。

## 4 结论

在数据挖掘理论的指导下,按照系统开发流程,利用 Python 语言开发系统核心运算功能,利用 WinForm 搭建人机交互界面,并通过参数交互的方式实现了人机交互界面与后台功能的互动,完成了海洋地质数据挖掘系统的开发工作。

以太平洋区域为例,通过指标筛选、数据预处理、MIV 分析、GA-BP 神经网络模型训练和实际年龄预测等步骤,得到了测试集上的海山年龄预测结果。精度评价结果表明,相比于仅用已知年龄插值未知数据的方式,综合考虑与海山年龄相关的已有多源数据集,可以有效提升预测精度,也验证了本文所开发系统的实用性。

### 参考文献:

- [1] LI W L, GAO S Y, HAN C H, et al. A brief analysis on data mining for deep-sea mineral resources based on big data[J]. *Procedia Computer Science*, 2019, 154: 699-705.
- [2] 杜艳玲. 混合云存储环境下海洋大数据的布局及迁移算法研究[D]. 上海: 上海海洋大学, 2014.
- [3] JIANG K, JIANG X. Technological development of ocean mineral resources exploitation[J]. *Nonferrous Metals Engineering*, 2011, 1(1): 3-8.
- [4] HE G, DENG X, YANG S. Geological characteristics of polymetallic nodules in the central Indian Ocean and comparison with those from CC Zone in the eastern Pacific[J]. *Marine Geology Quaternary Geology*, 2011, 31(2): 21-30.
- [5] HAND D J. Principles of data mining[J]. *Drug safety*, 2007, 30(7): 621-622.
- [6] ZONGXI Y, JINRONG T, PING Z. Earth science research in US Geological Survey under the big data revolution[J]. *Geological Bulletin of China*, 2013, 32(9): 1337-1343.
- [7] DUNHAM M H. Data mining: Introductory and advanced topics[M]. Pearson Education India, 2006.
- [8] SHARMA P, SHARMA S. Past Present & Future of Data Mining[J]. *IITM Journal of Information Technology*, 2018: 42.
- [9] LI C, YAN W, XIAO K, et al. Analysis and application mode of geological big data[J]. *Journal of Geology*, 2015, 39(3): 352-357.
- [10] LI J. Predictive modelling using random forest and its hybrid methods with geostatistical techniques in marine environmental geosciences[M]//The Proceedings of the Eleventh Australasian Data Mining Conference, 2013.
- [11] OBELCZ J, XU K, BENTLEY S J, et al. Sneaky submarine landslides, and how to quantify them: a case study from the Mississippi River Delta front contrasting geophysical and machine learning techniques[C]//AGU Fall Meeting Abstracts, 2017: NH44B-03.
- [12] SHENG J, SUN J, BAI Y, et al. Evaluation of hydrocarbon potential using fuzzy AHP-based grey relational analysis: a case study in the Laoshan Uplift, South Yellow Sea, China[J]. *Journal of Geophysics Engineering*, 2020, 17(1): 189-202.
- [13] 刘兰法. 海洋天然气水合物三维地质建模研究[D]. 青岛: 中国石油大学(华东), 2014.
- [14] 李连伟, 许明明, 刘展, 等. 天然气水合物数据挖掘服务组件的设计与实现[J]. *计算机应用与软件*, 2016, 33(10): 32-36.
- [15] 郑文棠. 岭澳核电三期高边坡三维地形可视化[C]//岩石力学与工程的创新和实践: 第十一次全国岩石力学与工程学术大会, 2010: 5.
- [16] DOMBI G W, NANDI P, SAXE J M, et al. Prediction of rib fracture injury outcome by an artificial neural network[J]. *Journal of Trauma*, 1995, 39(5): 915-921.
- [17] JIANG J L, SU X, ZHANG H, et al. A novel approach to active compounds identification based on support vector regression model and mean impact value[J]. *Chemical Biology & Drug Design*, 2013, 81(5): 650-657.
- [18] LI H, ZHONG Z, LI L, et al. A cascaded QSAR model for efficient prediction of overall power conversion efficiency of all-organic dye-sensitized solar cells[J]. *Journal of Computational Chemistry*, 2015, 36(14): 1036-1046.
- [19] ZHENG J, LAN Q, ZHANG X, et al. Prediction of MRI RF exposure for implantable plate devices using artificial neural network[J]. *IEEE Transactions on Electromagnetic Compatibility*, 2020, 62(3): 673-681.
- [20] RUMMEL R J. Applied factor analysis[M]. Northwestern University Press, 1988.
- [21] 郭晓伟. 基于深度学习的车型识别算法与DSP平台实现[D]. 长沙: 国防科学技术大学, 2016.
- [22] 李静. 煤矿震动波CT反演探测技术的优化与应用[D]. 徐州: 中国矿业大学, 2017.
- [23] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. *Nature*, 1986, 323(6088): 533-536.
- [24] DING S, SU C, YU J. An optimizing BP neural network algorithm based on genetic algorithm[J]. *Artificial Intelligence Review*, 2011, 36(2): 153-162.
- [25] BOOKER L B, GOLDBERG D E, HOLLAND J H. Classifier systems and genetic algorithms[J]. *Artificial Intelligence*, 1989, 40(1-3): 235-282.
- [26] TARDUNO J A, GEE J. Large-scale motion between Pacific and Atlantic hotspots[J]. *Nature*, 1995, 378(6556): 477-480.
- [27] 朱本铎, 吕文超, 张伙带. 中西太平洋海山年龄及其地质意义[J]. *矿床地质*, 2014: 1151-1152.
- [28] TARDUNO J A, DUNCAN R A, SCHOLL D W, et al. The Emperor Seamounts: southward motion of the Hawaiian Hotspot Plume in earth's mantle[J]. *Science*, 2003, 301(5636): 1064.
- [29] WATTS A B, SANDWELL D T, SMITH W H F, et al. Global gravity, bathymetry, and the distribution of submarine volcanism through space and time[J]. *Journal of Geophysical Research: Solid Earth*, 2006, 111: B08408.

- [30] FAGGION O, PINNA E, SAVELLI C, et al. Geomagnetism and age study of Tyrrhenian seamounts[J]. *Geophysical Journal International*, 1995, 123(3): 915-930.
- [31] CANDE S C, KENT D V. Ultrahigh resolution marine magnetic anomaly profiles: A record of continuous paleointensity variations?[J]. *Journal of Geophysical Research Solid Earth*, 1992, 97(B11): 15075-15083.
- [32] SETON M, MULLER R D, ZAHIROVIC S, et al. A Global data set of present-day oceanic crustal age and seafloor spreading parameters[J]. *Geochemistry, Geophysics, Geosystems*, 2020, 21(10): e2020GC009214.
- [33] VOLKER D, GEERSEN J, CONTRERAS-REYES E, et al. Sedimentary fill of the Chile Trench (32–46°S): volumetric distribution and causal factors[J]. *Journal of the Geological Society*, 2013, 170(5): 723.
- [34] CLAGUE D A, DALRYMPLE G B. Cretaceous K - Ar ages of volcanic rocks from the Musicians Seamounts and the Hawaiian Ridge[J]. *Geophysical Research Letters*, 2013, 2(7): 305-308.
- [35] HARRIS R N, FISHER A T, CHAPMAN D S. Fluid flow through seamounts and implications for global mass fluxes[J]. *Geology*, 2004, 32(8): 725-728.
- [36] SANDWELL D T, MUELLER R D, SMITH W H F, et al. New global marine gravity model from CryoSat-2 and Jason-1 reveals buried tectonic structure[J]. *Science*, 2014, 346(6205): 65-7.
- [37] MAUS S, BARCKHAUSEN U, BERKENBOSCH H, et al. EMAG2: a 2-arc min resolution earth magnetic anomaly grid compiled from satellite, airborne, and marine magnetic measurements[J]. *Geochemistry Geophysics Geosystems*, 2009, 10(8): Q08005.
- [38] MULLER R D. Age, spreading rates, and spreading asymmetry of the world's ocean crust[J]. *Geochemistry Geophysics Geosystems*, 2008, 9(4): Q04006.
- [39] STRAUME E O, GAINA C, MEDVEDEV S, et al. GlobSed: updated total sediment thickness in the World's Oceans[J]. *Geochemistry Geophysics Geosystems*, 2019, 20(4): 1756-1772.
- [40] AMANTE C. ETOPO1 1 Arc-Minute Global Relief Model: procedures, data sources and analysis[EB/OL]. <http://www.ngdc.noaa.gov/mgg/global/global.html>, 2009.
- [41] CLOUARD V, BONNEVILLE A. Ages of seamounts, islands, and plateaus on the Pacific plate[J]. *Geological Society of America Bulletin*, 2005: 388.
- [42] YESSON C, CLARK M R, TAYLOR M L, et al. The global distribution of seamounts based on 30 arc seconds bathymetry data[J]. *Deep Sea Research Part I Oceanographic Research Papers*, 2011, 58(4): 442-453.
- [43] LEE H B, OH J K, PARK C K, et al. Geophysical and sedimentological characteristics of Lomilik Seamount, West Pacific[J]. *Ocean & Polar Research*, 2004, 26(2): 207-218.
- [44] CARBOTTE S M, DIXON J M, FARRAR E, et al. Geological and geophysical characteristics of the Tuzo Wilson Seamounts: implications for plate geometry in the vicinity of the Pacific - North America - Explorer triple junction[J]. *Canadian Journal of Earth Sciences*, 1989, 26(11): 2365-2384.
- [45] KIM C H. Magnetic characteristics of TA19-1 and TA19-2 Seamounts in the Lau Basin, the South Western Pacific[J]. *Economic and Environmental Geology*, 2014, 47(4): 395-404.

## Marine geological data mining system development and its application in seamount age prediction

REN Mengjiao<sup>1</sup>, SUN Jihong<sup>2\*</sup>, WANG Xiaoyu<sup>1</sup>, RONG Yilin<sup>1</sup>, BAI Yongliang<sup>1</sup>

(1 College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China;

2 Qingdao Institute of Marine Geology, China Geological Survey, Qingdao 266237, China)

**Abstract:** Using data mining technology to find hidden information in the marine exploration data is important for increasing the marine data using efficiently. The core functions of marine geological data mining system (MG-DMS) was developed in the Python, the graphical user interface (GUI) was designed in the WinForm, and finally the interaction between GUI and the core functions was realized via parameter transformation and transference. Based on the geological and geophysical data, the ages of seamount in the Pacific are predicted using the MG-DMS, and the predicted ages are more precise than those of conventional Kriging method. Our case study result indicates that the functions of data preprocessing, index analysis, and overall evaluation with the MGDMS are suitable for marine geophysical data mining.

**Key words:** marine geology; data mining; system development; seamount age