

基于数据挖掘的中国矿石储量预测模型的研究与应用

朱子健¹, 王向前²

¹安徽理工大学计算机科学与工程学院, 安徽 淮南

²安徽理工大学经济与管理学院, 安徽 淮南

收稿日期: 2026年7月2日; 录用日期: 2026年7月30日; 发布日期: 2026年8月6日

摘 要

随着中国矿产资源需求的持续增长, 矿产储量的科学预测对于资源管理和可持续发展至关重要。采用多种预测模型(如线性回归、多项式回归、Prophet、ARIMA及指数平滑法)对中国主要矿产的储量进行了预测分析。通过比较各模型在不同矿产资源上的表现, 发现大多数模型能够较好地反映矿产储量的变化趋势。对于储量变化平稳的矿产, 预测效果较为准确, 而对于波动较大的矿产, 误差有所增加。研究结果为矿业公司和政府提供了科学的决策支持, 促进了资源的合理开发和环境保护。随着技术的进步, 未来的储量预测模型有望进一步提升精度和适应性。

关键词

矿产资源, 储量预测, 模型比较, 时间序列分析, 资源管理

Research and Application of China's Mineral Reserves Prediction Model Based on Data Mining

Zijian Zhu¹, Xiangqian Wang²

¹College of Computer Science and Engineering, Anhui University of Science and Technology, Huainan Anhui

²School of Economics and Management, Anhui University of Science and Technology, Huainan Anhui

Received: July 2, 2026; accepted: July 30, 2026; published: August 6, 2026

Abstract

With the continuous growth of demand for mineral resources in China, the scientific prediction of mineral reserves is crucial for resource management and sustainable development. This study

adopts a variety of prediction models (such as linear regression, polynomial regression, Prophet, ARIMA, and exponential smoothing method) to conduct predictive analysis on the reserves of China's major minerals. By comparing the performance of each model on different mineral resources, it is found that most models can well reflect the changing trend of mineral reserves. For minerals with stable reserve changes, the prediction effect is relatively accurate; however, for minerals with large fluctuations, the prediction error increases. The research results provide scientific decision support for mining companies and the government, and promote the rational development of resources and environmental protection. With the advancement of technology, future reserve prediction models are expected to further improve their accuracy and adaptability.

Keywords

Mineral Resources, Reserve Prediction, Model Comparison, Time Series Analysis, Resource Management

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

矿产资源作为国家经济发展的重要物质基础,在工业化和现代化进程中扮演着举足轻重的角色。中国,作为世界上最大的矿产资源消费国之一,对矿产资源的需求日益增长,而传统的矿产资源评估方法已难以满足快速、准确评估的需求。

近年来,随着数据挖掘技术的不断发展,矿产资源预测领域得到了广泛关注,尤其是中国矿石储量的预测。通过对历史数据和现有地质数据的深入挖掘与分析,研究人员可以更准确地预测矿石储量和分布情况,从而提高资源勘探和开发的效率。数据挖掘技术通过整合多源数据,利用机器学习算法和统计模型,能够从大量地质、地球物理和地球化学数据中提取有价值的模式和信息,辅助决策。基于数据挖掘的矿石储量预测不仅可以显著提升资源评估的精度,还能为矿产资源的合理规划和开发提供科学依据。

在矿石储量预测方面,全球学者进行了大量研究。孙晓旭[1]提出了一种概率抽样的方法,它是在已知模型区大量资料的基础上(矿石量、品位等)来模拟未知区的资源量,应用蒙特卡洛法数学模型,可得出不同概率水平下的矿体资源量估计值。赵翠平[2]提出了动力费对矿石产量有着一定的影响。Dac Ngo Xuan 等人[3]使用地质结构、最优参数和计算工具方法构建了 3D 块体模型,用于表面和矿体分布的 3D 几何模型。Cu 和 Ag 储量是根据 3D 几何模型预测的。Chengkai Fan 等人[4]采用深度神经网络(DNN)以卡车运输信息和露天矿的天气条件作为训练数据来预测矿石产量。

随着数据挖掘技术的不断进步,国内学者在矿石储量预测领域的研究日趋深入。这些基于数据挖掘的预测模型,不仅提高了矿产资源评估的精度,还为未来矿产资源的开发和保护提供了重要的技术支持。

为了更清晰地了解中国矿石储量的变化,同时进行有效的矿石储量预测,本研究通过数据挖掘技术,对中国矿石储量进行分析与处理,构建矿石储量预测模型,以期提高矿产资源评估的效率和准确性。

2. 资料与方法

2.1. 研究问题

随着中国矿业资本市场的开放,矿业市场的竞争将会提速,矿业市场的成熟使其在储量估算方面的经验更加丰富[5]。本研究对于主要能源矿产储量和矿产查明资源储量的定义为:主要能源矿产储量通常

指的是已经探明并且技术上可开采的能源矿产的量, 包括煤炭、石油、天然气等。这些数据反映了在当前技术和经济条件下可以直接用于生产和消费的能源矿产数量。矿产查明资源储量这是一个更广泛的概念, 包括了所有已经探明的矿产资源, 不仅仅是能源矿产。这涵盖了能源矿产、金属矿产、非金属矿产以及水气矿产等。查明资源储量不仅包括技术上可开采的储量, 还可能包括一些目前技术上不可开采或者不经济的储量。研究基于数据挖掘技术, 探究预测中国矿石储量预测的方法。

2.2. 研究方法

分别采用线性回归、多项式回归、Prophet、自回归积分滑动平均模型(ARIMA)、指数平滑法、随机森林(random forest)、Logistic 回归、BP 神经网络、AdaBoost、Gradient Boosting (GBDT)、LightGBM、CatBoost 模型构建矿物储量预测模型, 采用 5 折交叉验证方法对模型性能进行评价。本文首先采用 RMSE、MAE 和 R^2 等指标对不同预测模型进行性能比较。在综合考虑预测精度、模型稳定性、时间序列适用性以及模型可解释性的基础上, 选择最适用于矿产储量预测任务的模型作为最终预测模型。

2.3. 数据源

本文利用中国主要矿产查明资源储量和中国主要能源矿产储量, 分析了中国主要能源矿产储量状况和涨幅程度。数据来源于 2013 年至 2023 年中国矿产资源报告, 共有 48 种资源, 数据包含 48 个中国主要矿产查明资源储量和中国主要能源矿产储量。这些数据参照国家标准《油气矿产资源储量分类》(GB/T 19492-2020)¹和《固体矿产资源储量分类》(GB/T 17766-2020)², 且主要来源于中华人民共和国自然资源部和中华人民共和国国家统计局, 未包括香港特别行政区、澳门特别行政区和台湾地区的统计数据。

2.4. 数据预处理

围绕 48 种矿石资源数据, 数据采用去除异常值(空值、负值、非数值的其他符号等)、最大 - 最小归一化等方法进行处理, 计算公式如下:

$$x^* = \frac{x - \min}{\max - \min} \quad (1)$$

其中 $\max$ 为样本数据的最大值, $\min$ 为样本数据的最小值。公式(1)这种方法有个缺陷就是当有新数据加入时, 可能导致 $\max$ 和 $\min$ 的变化, 需要重新定义。

数据处理过程中, 重要字段缺失情况常见。由于“十二五”期间, 能源矿产勘查取得重大进展, 煤层气、页岩气、锂矿、钠盐在 2015 年后被《中国矿产资源报告》³录入, 故之前年份的缺失值取其余年份的平均值。

2.5. 特征指标维度

在确定了预测算法之后, 为了确保模型的有效性和准确性, 我们需要构建一个统一的特征向量体系。从宏观视角出发, 该体系被划分为以下五个维度: 地质地理维度、勘查与调查维度、经济与社会维度、技术与环保维度以及生产与利用维度。这类数据采用独热编码、二进制编码以及赋予特殊数值编码进行处理, 如将“高”“中”“低”转为 3、2、1。这五个维度下各自包含的具体指标参见表 1。这些指标的选定和定义, 主要参考了《中国矿产资源报告》。

2.6. 特征选择

特征选择是指在构建模型之前从原始数据中选择最相关、最有信息价值的特征或变量的过程。正确

¹<https://openstd.samr.gov.cn/bzgk/std/newGbInfo?hcno=326DE28D2E98AC3A5DDF223A09C728EC&refer=outter>

²<https://www.dzxxzch.cn/zcjc/ch/202004/P020200423709369463784.pdf>

³<https://www.mmac.org.cn/UploadFile/UserWeb/laws/414.pdf>

的特征选择可以提高模型的性能, 降低过拟合的风险, 并加速训练过程。

本研究通过 2 个阶段进行特征过滤: 首先使用方差过滤, 过滤方差为 0 的特征; 然后使用互信息法来捕捉特征与目标变量的相关性, 过滤不相关的特征。其中, 互信息是用来捕捉每个特征与标签之间任意关系的过滤方法。计算公式如下:

$$I(X;Y)=\sum_{x\in X}\sum_{y\in Y}p(x,y)\lg\left(\frac{p(x,y)}{p(x)p(y)}\right) \quad (2)$$

其中, $I(X;Y)$ 表示 X 和 Y 之间的互信息, $p(x,y)$ 是联合分布, 即 X 与 Y 同时发生的概率, $p(x)$ 和 $p(y)$ 分别是 X 和 Y 的边缘分布。公式(2)返回“每个特征与目标之间的互信息量的估计”, 范围在 $[0, 1]$ 之间, 0 表示 2 个变量独立, 1 表示 2 个变量完全相关。

首先使用方差过滤, 舍弃方差为 0 的“副产品价值”等特征; 然后采用互信息法舍弃互信息量等于或接近 0 的“勘查成功率”“气候条件”“资源回收率”等 2 维特征。见表 1。

Table 1. Feature index dimension

表 1. 特征指标维度

维度	特征选择前特征	特征选择后特征
地质地理维度	矿床地质状况	矿床地质状况
	区域地质条件	区域地质条件
	矿区地理状况	矿区地理状况
	气候条件	\
勘查与调查维度	勘查投资比例	勘查投资比例
	勘查技术数据	勘查技术数据
	勘查成功率	\
	探明储量和资源量	探明储量和资源量
	地质勘探进展	地质勘探进展
经济与社会维度	投资强度	投资强度
	市场因素	市场因素
	矿产开采成本	矿产开采成本
	资源需求	资源需求
技术与环保维度	勘查和开采技术	勘查和开采技术
	环保指标	环保指标
	水文、环境条件	水文、环境条件
生产与利用维度	矿产产量与开采率	矿产产量与开采率
	资源回收率	\
	矿区利用率	矿区利用率
	储量衰减率	储量衰减率
	副产品价值	\

2.7. 模型构建与优化

数据处理后, 根据输入目标, 选择合适的算法进行建模, 并控制模型的精度, 最终达到预测矿物储量的目的。使用 python 常用的数据建模工具 Scikitlearn (0.24.2)、XGBoost (1.5.2)、LightGBM (4.1.0) 实现快速数据建模。本研究分别构建线性回归、多项式回归、Prophet、ARIMA、指数平滑法、Random Forest、Logistic 回归、BP 神经网络、AdaBoost、Gradient Boosting (GBDT)、LightGBM 及 CatBoost 等 12 种预测模型。以 RMSE、MAE 和 R^2 作为综合评价指标, 对各模型进行统一比较, 最终选择综合性能最优的 Gradient Boosting 模型作为最终预测模型, 并进一步进行参数优化与预测分析。

矿产储量预测属于典型的时间序列预测问题, 为避免未来数据泄漏, 本文按照时间先后顺序划分训练集、验证集和测试集, 而非随机划分数据集。具体划分如下: 训练集(Training Set): 2013~2018 年, 用于模型训练及参数学习; 验证集(Validation Set): 采用训练集内部时间序列交叉验证(Time Series Cross Validation), 不单独划分固定年份, 而是在 2013~2018 年的训练数据上采用滚动窗口(Rolling Window)方式进行 5 折时间序列交叉验证, 用于模型参数选择和模型比较; 测试集(Test Set): 2019~2023 年, 仅用于最终模型性能评估, 不参与模型训练和参数调整。

由于时间序列数据具有明显的时间依赖性, 普通随机交叉验证会导致未来信息泄漏, 因此本文采用时间序列交叉验证方式, 即每一折均以前一时间段作为训练数据, 以后一时间段作为验证数据。例如:

Fold1: 2013~2014 → 2015

Fold2: 2013~2015 → 2016

Fold3: 2013~2016 → 2017

Fold4: 2013~2017 → 2018

Fold5: 2013~2018 → 下一阶段预测

该方法能够更真实地模拟实际预测场景, 提高模型泛化能力。

对于线性回归、多项式回归、Random Forest、Gradient Boosting、AdaBoost、LightGBM、CatBoost、BP 神经网络等监督学习模型, 由于其本身不具备时间序列建模能力, 因此首先将时间序列转换为监督学习数据集。本文采用“滑动窗口(Sliding Window)”构造样本, 以连续 3 个历史年份(滑窗大小 = 3)作为输入特征, 对下一年的矿产储量进行预测, 即

$$(X_{t-3}, X_{t-2}, X_{t-1}) \rightarrow X_t \quad (3)$$

其中, X_t 表示第 t 年矿产储量, 滞后阶数(Lag)设定为 3。

所有监督学习模型均采用相同的时间序列特征构造方式, 以保证不同模型之间具有公平的比较基础。

对于 Prophet、ARIMA 以及指数平滑法等传统时间序列模型, 则直接利用原始时间序列进行建模, 无需构造滑动窗口特征。其中, ARIMA 模型依据训练集的平稳性检验及 AIC 准则确定模型参数; Prophet 模型采用默认趋势项与自动节假日处理策略; 指数平滑法根据最小化训练误差自动估计平滑参数。

模型训练完成后, 采用 RMSE、MAE 和 R^2 三个指标对各模型进行统一评价, 并最终选择综合性能最优的 Gradient Boosting 模型开展矿产储量预测分析。

3. 实验结果

3.1. 模型评价指标

本文选用均方根误差(RMSE)、平均绝对误差(MAE)和 R-Squared 特征参数值作为评价模型的拟合效果。

均方根误差(RMSE)用于衡量模型预测值与真实值之间的差距。RMSE 会放大的误差值, 因此对极端误差较为敏感。RMSE 的值越小, 说明预测值与真实值的差距越小, 模型的预测精度更高。

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^n (Y_i - f(x_i))^2} \quad (4)$$

平均绝对误差(MAE)表示预测值与实际值的平均绝对误差。相比于 RMSE, MAE 对较大的误差不敏感, 因此提供了一种更加“平均”的误差衡量方式。MAE 越小, 模型的预测效果越好。

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (5)$$

R-Squared 用来衡量模型对数据的拟合程度。 R^2 越接近 1, 说明模型能够更好地解释数据的变异性。较高的 R^2 值意味着模型拟合度较高, 能够较好地描述数据趋势。

$$R^2_{\text{adjusted}} = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (6)$$

3.2. 结果分析

Table 2. Comparison of evaluation metrics for algorithm models
表 2. 算法模型评价指标对比

评价指标 模型	RMSE	MAE	R^2
RandomForest	534.25242	197.458421	0.996841
AdaBoost	425.332096	162.611458	0.997997
GradientBoosting	422.524489	160.24839	0.998024
LightGBM	1374.142615	525.146216	0.979097
CatBoost	458.594626	172.727677	0.997672
BP_NeuralNetwork	888.192988	265.559014	0.991267
线性回归	208.086903	194.653781	0.997457
多项式回归	1769.365763	1550.535185	0.994268
Prophet	215.182306	197.665584	0.995126
ARIMA	193.876268	175.593515	0.995254
指数平滑法	233.634535	218.395112	0.991417

表 2 给出了不同预测模型在矿产储量预测任务上的性能比较结果。本文采用 RMSE、MAE 和 R^2 三个指标作为模型评价标准。其中, RMSE 和 MAE 越小表示预测误差越低, R^2 越接近 1 说明模型拟合能力越强。由表 2 可知, Gradient Boosting 模型取得了最佳综合性能, 其 RMSE、MAE 分别为 422.52 和 160.25, R^2 达到 0.998024, 在三项评价指标中均处于最优或接近最优水平。说明 GBDT 能够更加准确地学习矿产储量变化过程中的非线性关系, 具有较高的预测精度和较好的泛化能力。

AdaBoost 模型整体性能与 GBDT 较为接近, 但 RMSE 和 MAE 均略高于 GBDT, 说明其预测稳定性稍逊。CatBoost 和 Random Forest 同样取得了较高的拟合精度, 但误差指标仍高于 GBDT, 表明其在复杂样本上的预测能力略有不足。

相比之下, LightGBM 和 BP 神经网络预测误差较大, 其中 LightGBM 的 RMSE 达到 1374.14, 说明

其在当前样本规模下存在一定的过拟合或欠拟合现象;BP 神经网络由于训练样本数量有限,模型参数较多,泛化能力受到一定影响,因此预测效果不够理想。

传统统计模型中,线性回归、Prophet、ARIMA 及指数平滑法虽然能够较好地反映总体变化趋势,但对于矿产储量中复杂的非线性特征刻画能力有限,因此整体预测性能低于 GBDT 模型。其中,ARIMA 能够较好地描述时间序列的线性变化规律,但难以充分挖掘多维特征之间的复杂关联关系,因此综合性能不及 Gradient Boosting 模型。

综合 RMSE、MAE 及 R^2 三个评价指标,Gradient Boosting 模型具有最优的预测精度和稳定性。因此,本文选择 Gradient Boosting 作为最终预测模型,并基于该模型开展后续矿产储量预测分析。

基于 Gradient Boosting 模型得到的预测结果如表 3 所示。总体来看,中国主要矿产资源储量未来几年仍呈现以平稳增长为主、局部资源差异化发展的趋势,尽管中国正处于能源结构转型的过程中[6],煤炭仍然是主要的能源来源之一。预测结果与实际需求保持一致,反映了煤炭作为基础能源的重要性。

对于石油和天然气,模型同样显示了缓慢增长的趋势。尤其是在天然气方面,由于政策推动和环保需求的增加,天然气的需求逐渐上升,这与预测结果吻合。预测显示天然气储量将继续增加,但增幅相对平稳,表明储量的增加与市场需求密切相关。

Table 3. Prediction results of Gradient Boosting model
表 3. Gradient Boosting 模型预测结果

矿产名称	2019	2020	2021	2022	2023
煤炭(亿吨)	17564.63291	17550.27165	17673.8568	17746.12327	17811.86642
石油(亿吨)	35.93615	36.05028	36.12882	36.17852	36.20997
天然气(亿立方米)	58189.87348	58225.50292	58227.993	58228.23493	58228.25844
煤层气(亿立方米)	3036.105088	3041.068614	3038.652057	3039.828589	3039.255779
页岩气(亿立方米)	2038.204642	2122.136897	2064.391885	2104.120192	2076.787296
铁矿(亿吨)	853.2073617	853.52001	853.61616	853.64570	853.6547885
锰矿(亿吨)	17.98955167	17.89270955	17.83768763	17.80642631	17.78866484
铬铁矿(万吨)	1189.780237	1189.328682	1189.270254	1189.262693	1189.261715
钒矿(万吨)	6633.597445	6672.856274	6694.174533	6705.750736	6712.036825
钛矿(亿吨)	8.297803886	8.318220084	8.329245964	8.335200551	8.338416359
铜矿(万吨)	12009.0683	12391.81751	12650.83903	12826.12913	12944.75488
铅矿(万吨)	9396.236171	9526.088272	9619.802087	9687.435026	9736.245486
锌矿(万吨)	18787.0829	18790.85179	18791.30398	18791.35823	18791.36474
铝土矿(亿吨)	52.1563264	52.21307764	52.28203394	52.31236844	52.32571287
镍矿(万吨)	1237.939512	1273.836299	1299.577248	1318.035625	1331.2718
钴矿(万吨)	69.98104674	70.10701449	70.15494691	70.17318585	70.18012601
钨矿(万吨)	1082.805245	1085.87282	1086.710365	1086.939041	1087.001476
锡矿(万吨)	453.461785	453.515239	453.5223506	453.5232968	453.5234226
钼矿(万吨)	3039.058161	3044.058806	3046.452189	3047.597697	3048.145954
铋矿(万吨)	334.6331303	340.7374259	346.0965122	350.8013638	354.9318495
金矿(吨)	13818.91479	13892.49801	13922.49272	13934.71946	13939.70345

续表

银矿(万吨)	33.82124343	34.45510953	34.89603038	35.20273743	35.41608459
铂族金属(吨)	401.6773376	401.6901887	401.6904325	401.6904372	401.6904373
锑矿(万吨)	5641.194858	5641.189627	5641.189846	5641.189837	5641.189837
锂矿(万吨)	1116.39917	1121.176248	1122.111545	1122.294665	1122.330518
菱镁矿(亿吨)	30.96954239	30.93908304	30.92373721	30.91600577	30.91211057
萤石(亿吨)	2.648237114	2.689044087	2.71032822	2.721429613	2.727219887
耐火粘土(亿吨)	26.67449284	26.8630277	26.98372807	27.06100066	27.11047072
硫铁矿(亿吨)	64.07735645	64.56098016	64.77807813	64.87553309	64.91928047
磷矿(亿吨)	252.8044047	252.792244	252.7827615	252.7753674	252.7696017
钾盐(亿吨)	10.20448543	10.18649494	10.19377053	10.19082819	10.19201811
硼矿(万吨)	7841.666735	7843.011981	7843.367049	7843.460766	7843.485503
钠盐(亿吨)	14231.27125	14237.10675	14233.58478	14235.71044	14234.42752
芒硝(亿吨)	1173.053356	1173.057281	1173.057466	1173.057475	1173.057475
重晶石(亿吨)	3.784142945	3.810792568	3.823909741	3.830366126	3.833544015
水泥用灰岩(亿吨)	1486.129001	1532.525375	1572.567479	1607.125571	1636.950719
玻璃硅质原料(亿吨)	103.2945763	110.2500171	117.0024271	123.5577329	129.921688
石膏(亿吨)	843.4956992	841.3232403	841.576495	841.5469718	841.5504134
高岭土(亿吨)	34.99587011	35.00171859	35.00267216	35.00282764	35.00285299
膨润土(亿吨)	29.81964381	29.78979554	29.78344798	29.78209811	29.78181104
硅藻土(亿吨)	5.111203052	5.111130685	5.111135038	5.111134776	5.111134792
饰面花岗岩(亿立方米)	55.81389041	57.06954199	57.85243508	58.34056538	58.64491243
饰面大理岩(亿立方米)	18.68801401	19.48848921	20.19416172	20.81625933	21.36468005
金刚石(千克)	3126.599893	3126.599893	3126.599893	3126.599893	3126.599893
晶质石墨(亿吨)	4.987578803	5.532441058	6.013148742	6.437255709	6.811426384
石棉(万吨)	9235.227188	9233.224061	9233.056613	9233.042615	9233.041445
滑石(亿吨)	2.878730855	2.878569782	2.878549339	2.878546745	2.878546415
硅灰石(亿吨)	2.411755626	2.436881783	2.442066954	2.443136994	2.443357813

而锰矿、铬铁矿和钛矿等矿产, Gradient Boosting 模型预测结果显示储量变化较小, 几乎维持平稳状态。这表明这些矿产的开采量已经接近其资源的上限, 未来储量的增长将面临更多挑战。

此外, 一些稀有矿产(如锂矿、锑矿)的预测储量增长较为明显。随着全球对新能源汽车和高科技产品的需求上升[7][8], 锂矿等稀有矿产的储量预计会继续增长。这一趋势反映了全球产业结构变化对矿产资源需求的影响。

从宏观背景来看, 矿产资源储量变化不仅受资源禀赋影响, 同时受到国家产业政策、市场需求及技术进步等多方面因素共同作用。近年来, 《中华人民共和国矿产资源法》⁴《“十四五”省级矿产资源总体规划环境影响评价技术要点》⁵等政策持续推动战略性矿产资源保障能力建设, 加大深部找矿和资源综合利用力度, 为煤炭、天然气、铜矿、锂矿等重点矿产资源增长提供了政策支持。同时, 在新能源汽车、储能、新材料及高端制造产业快速发展的带动下, 锂、镍、钴、铜等战略矿产市场需求持续增加, 推动资

⁴<https://nrp.jinan.gov.cn/attach/0/074d3a1d141b4a35afe4ac7cf934aab4.pdf>

⁵<https://www.mee.gov.cn/xxgk/xxgk06/202112/W020211208637869863252.pdf>

源勘查投入不断增长。另一方面, 随着智能勘探、三维地质建模、大数据找矿及人工智能预测等技术的发展, 矿产资源探测精度不断提高, 也促进了资源储量的持续增长。因此, 本文预测结果不仅反映了历史数据演化规律, 也在一定程度上映射了产业政策、市场需求及技术进步共同作用下我国矿产资源发展的总体趋势。

4. 中国矿石储量预测模型的应用场景

4.1. 国家资源管理与战略矿产保障

由表 3 可知, 煤炭储量由 2019 年的 17,564 亿吨增长至 2023 年的 17,811 亿吨, 累计增长 247 亿吨, 增长率约 1.41%; 天然气储量由 58,189 亿立方米增长至 58,228 亿立方米, 整体保持稳定增长趋势。

这一预测结果表明, 在“双碳”目标背景下, 虽然我国能源结构正逐步向清洁能源转型, 但煤炭仍将在未来较长时期承担能源安全“压舱石”作用。同时, 天然气储量持续增长符合《国家发展改革委 国家能源局关于印发〈“十四五”现代能源体系规划〉的通知》⁶中提高天然气消费比重的发展方向, 为天然气管网建设、储气库布局及能源结构优化提供了数据支撑。因此, 政府可依据预测结果合理制定矿产资源勘查规划、战略储备计划及能源安全保障措施, 提高资源配置效率。

4.2. 新能源产业链资源保障

近年来新能源汽车、动力电池及储能产业快速发展, 对锂、镍、钴等战略性矿产需求持续增长。预测结果显示, 锂矿储量由 1116 万吨增长至 1122 万吨, 增长约 6 万吨; 镍矿由 1238 万吨增长至 1331 万吨, 增长约 93 万吨; 钴矿由 69.98 万吨增长至 70.18 万吨, 整体保持稳定增长。预测结果说明, 我国新能源关键矿产仍具有一定资源保障能力, 但增长幅度有限。未来仍需加强海外资源布局、资源循环利用以及盐湖提锂、深部找矿等关键技术研发, 以缓解资源供需矛盾。

4.3. 风险管理与应急准备

矿业活动存在许多不可预测的风险, 包括自然灾害、技术故障、市场波动等^[9]。通过矿石储量预测模型, 矿业公司可以提早识别潜在风险, 调整资源开采计划, 制定应急响应策略。如果预测显示某些矿区储量将迅速枯竭, 公司可以提前调配资源, 启动其他矿区的开采, 避免因储量不足导致的生产中断。此外, 储量预测模型还可以用于灾后恢复评估, 帮助企业在矿区发生灾害后合理分配资源, 确保恢复工作顺利进行。

4.4. 市场预测与价格波动分析

矿产资源的价格受到全球供需关系的影响。通过对矿石储量的预测, 矿业公司能够准确把握市场供需变化, 预测矿产资源价格的未来走势。如果储量较大的矿产在未来几年内会逐渐减少, 这可能会导致价格上涨。公司可以利用这一信息提前调整开采计划, 确保产销准备充分, 抓住价格上涨带来的市场机会。反之, 如果某些矿产的储量长期稳定且开采量过大, 可能导致市场过剩, 价格下降, 这时企业则应适当减少开采量, 避免过度生产。

4.5. 环境保护与可持续发展

矿石开采活动对环境有一定的影响, 尤其是大规模的露天开采, 会破坏生态环境, 导致水土流失、空气污染等问题^[10]。通过矿石储量的精准预测, 矿业公司和政府能够提前了解资源的开采周期, 合理安

⁶https://www.gov.cn/zhengce/zhengceku/2022-03/23/content_5680759.htm

排开采计划, 避免过度开发, 保护环境。例如, 预测模型可以预测某个矿区在未来几年内储量将枯竭, 并且该区域的生态环境脆弱。在这种情况下, 企业和政府可以提前停止开采, 并启动环境恢复工程, 减少对生态系统的负面影响。

4.6. 技术研发与创新

矿石储量预测模型在学术研究和技术创新方面也有重要的作用。通过对历史储量数据的分析, 研究人员可以深入了解矿产资源的开采规律, 推动新技术的发展和应用。例如, 基于预测模型, 学者们可以探索更高效、更环保的矿产资源开采技术, 提高资源回收率, 减少资源浪费。随着人工智能和大数据技术的不断发展, 越来越多的研究者尝试将深度学习和机器学习等先进技术应用用于矿产储量预测, 进一步提高预测的准确性和可靠性, 推动矿业领域的技术创新。

5. 结论、局限性与展望

5.1. 结论

本研究基于数据挖掘技术, 结合 Gradient Boosting 模型和其他预测模型, 对中国矿产资源储量进行了预测分析。通过对中国主要矿产资源储量的历史数据进行深入分析, 本文不仅提高了矿产资源评估的精度, 还为资源的合理开发、环境保护及政策制定提供了科学依据。

研究结果表明, Gradient Boosting 模型在预测大多数矿产储量的变化趋势时具有较好的表现。特别是在煤炭、天然气等基础能源矿产的储量预测上, Gradient Boosting 模型能够较为准确地反映出储量增长的稳定性。然而, 对于某些变化较为复杂的矿产, 如锂矿和稀有金属矿产, 虽然 Gradient Boosting 模型提供了合理的趋势预测, 但仍有一定的误差。这表明, 在处理具有较大波动性的矿产储量时, 可能需要结合更多的外部数据(如政策变化、市场需求等)来进一步优化预测模型。

5.2. 局限性分析

尽管 Gradient Boosting 模型在整体预测任务中取得了较高的预测精度, 但仍存在一定局限性。

(1) 时间序列样本数量有限。

本文采用 2013~2023 年《中国矿产资源报告》数据, 仅包含 11 个年度样本。对于机器学习模型而言, 样本规模相对较小, 容易导致模型学习到局部规律, 而难以充分刻画长期变化趋势。所以, 对于部分波动性较大的矿产, 其预测结果仍存在一定误差。

(2) 部分矿产具有明显的非平稳性。

例如页岩气、煤层气、锂矿等矿产, 其储量变化受到国家政策调整、勘探技术突破以及市场需求变化等因素影响较大, 具有明显的阶段性增长特征。Gradient Boosting 主要依据历史数据进行拟合, 难以准确预测由政策或重大技术突破引起的突变, 因此预测结果可能存在一定偏差。

(3) 外部影响因素未完全纳入模型。

本文主要采用矿产资源报告中的历史统计数据建模, 尚未引入国际矿产品价格、GDP 增长率、矿产投资规模、碳减排政策、国际贸易环境以及新能源产业发展速度等宏观变量。这些因素均可能对矿产储量变化产生重要影响, 因此模型对长期预测能力仍存在一定限制。

(4) 模型可解释性不足。

Gradient Boosting 属于集成学习模型, 虽然预测精度较高, 但难以直观解释各影响因素对预测结果的具体贡献, 不利于资源管理部门开展决策分析。未来可结合 SHAP (SHapley Additive Explanations)、Permutation Importance 等模型解释方法, 对关键影响因素进行定量分析, 提高模型透明性。

5.3. 展望

随着大数据技术和机器学习算法的持续发展, 矿产资源的预测模型将进一步得到完善。未来的研究可以考虑引入更多的非线性模型和外部数据源, 如气候变化、全球经济波动、矿产政策等, 这将有助于提高矿产储量预测的准确性。同时, 随着人工智能技术的应用, 结合深度学习等技术进行更复杂的模型构建, 或许能为矿产资源的预测提供更多的可能性。

在模型构建方面, 本文采用的 Gradient Boosting 模型能够较好地学习矿产储量变化规律, 但对于受政策调整、技术突破及市场波动影响较大的矿产资源, 其预测能力仍有进一步提升空间。未来可探索 ARIMA、Prophet 等传统时间序列模型与长短期记忆网络(LSTM)、Transformer 等深度学习模型的融合方法, 充分发挥不同模型的优势, 提高对复杂非线性特征和长期依赖关系的建模能力。同时, 可结合 SHAP 等模型解释方法, 对影响矿产储量变化的关键因素进行量化分析, 提高模型的可解释性和决策支持能力。

在应用层面, 未来矿产储量预测应进一步服务于国家资源安全战略和矿业高质量发展需求。结合新能源、新材料、高端装备制造等战略性新兴产业的发展趋势, 构建不同政策情景和市场需求下的矿产储量预测模型, 为战略性矿产资源勘查布局、资源储备、矿山开发规划及绿色矿山建设提供更加科学的决策依据。此外, 随着数字矿山、智能矿山及数字孪生技术的不断发展, 矿产资源预测将逐步实现由静态预测向动态监测、实时预警和智能决策转变, 为我国矿产资源的安全保障和可持续开发提供更加有力的技术支撑。

参考文献

- [1] 孙晓旭. 蒙特卡洛法在内蒙太平地地区金矿储量预测中的应用[J]. 世界有色金属, 2018(3): 143-144.
- [2] 赵翠平, 赵岩, 张继梅. 根据动力费对矿石产量进行回归预测——北洛河铁矿为例[C]//山东金属学会. 鲁冀晋琼粤川辽七省金属(冶金)学会第十九届矿山学术交流会论文集(管理及综合卷). 2012: 222-225.
- [3] Ngo Xuan, D., Phan Trong, T., Trinh Hai, S., Bui Vinh, H., Nguyen Quoc, D. and Van-Hao, D. (2024) 3D Models for Hydrothermal Copper Ore Bodies at Sin Quyen Deposit, North Vietnam: A Case Report for Ore Reserves and Prediction of Hidden Mineral Resource Potential. *Helvion*, **10**, e33017. <https://doi.org/10.1016/j.helivon.2024.e33017>
- [4] Fan, C., Zhang, N., Jiang, B. and Liu, W.V. (2024) Using Deep Neural Networks Coupled with Principal Component Analysis for Ore Production Forecasting at Open-Pit Mines. *Journal of Rock Mechanics and Geotechnical Engineering*, **16**, 727-740. <https://doi.org/10.1016/j.jrmge.2023.06.005>
- [5] 翟建波, 刘育明, 孙学森. 关于矿石储量估算的探讨和应用[J]. 中国矿山工程, 2022, 51(2): 1-7.
- [6] 余志强. 碳中和目标下能源和经济转型中的能源安全因素[J]. 财讯, 2023(7): 12-15.
- [7] 董波, 田庆华, 许志鹏, 等. 新能源战略金属镍钴锂资源清洁提取研究进展[J]. 材料导报, 2023, 37(22): 123-137.
- [8] 中国铝业盈测调整 2026 年盈利预期达 138 亿[J]. 铝加工, 2024(2): 10.
- [9] 黄中杰, 邓仁健, 周赛军, 等. 矿业活动对镉矿区土壤的重金属污染特征及生态风险影响[J]. 土木与环境工程学报(中英文), 2020, 42(4): 194-202.
- [10] 陈迎辉. 矿山开采对地质环境的影响[J]. 城市建设理论研究(电子版), 2014(8). <http://dx.chinadot.cn/10.3969/j.issn.2095-2104.2014.08.2313>