基于人工智能算法的地浸采区产量预测

贾 皓[1,2],廖文胜[1,2],杜志明[1,2],王亚安[1,2],王立民[1,2]

(1.核工业北京化工冶金研究院,北京 101149;
2.中核矿业科技集团有限公司,北京 101149)

摘要:在地浸采铀生产后期,金属产量持续下降,其铀浓度也受浸采工艺影响而不断变化,抽液量与浓度的乘积不能准确反映金属产量。通过对某矿床 C10 采区生产数据进行归纳与分析,采用人工神经网络模型对该采区金属日产量进行预测,并使用多元线性回归模型与之比较。结果表明二者均具有良好的精确度和泛化能力,其中人工神经网络模型效果更优,对金属日产量预测平均误差仅为-0.36%,预测结果离群值较少且误差分布均衡。与简单计算金属日产量相比,使用人工神经网络模型预测采区产量,在精度上显著提高,能够为制定生产计划与增产提供理论指导。

关键词:地浸铀矿;产量预测;人工神经网络;多元回归

中图分类号:TP183;TD868 **文献标志码**:A **文章编号**:1000-8063(2023)03-0010-07

DOI:10.13426/j.cnki.yky.2023.02.07

随着地浸采铀工艺的日趋成熟,地浸勘探领域智能化逐渐完善,通过与大数据、人工智能算法的深度融合,大幅提高了勘探工作效率与质量[1];地浸与水冶过程也基本实现了实时监控与数据集成调度,但对于在生产过程中采集的数据仍然存在只收集不处理,或仅对数据直观判断的情况[2]。对于海量数据,未进行深层挖掘与分析建模,不仅无法根据已有的生产参数安排未来生产计划,也无法精准把握矿区的生产趋势[3-4]。

在铀矿采区开发的中后期,随着金属浸采率不断增加,其浸出液铀浓度与水量通常呈下降趋势,通过洗井、注液过滤、增强抽注循环等方式虽能够在短期内增加金属产量[5];但囿于资源总量,整体上金属产量仍呈降低趋势,导致开采成本显著增加。实际生产中的井场,在集控室与水冶车间对抽注液量、压力、金属浓度等参数的实时监控与采集,不仅能够对生产异常即时诊断,更有利于形成广泛全面的生产数据库。根据长周期下的生产参数建立数据模型,从而准确预测金属产量,可为采区未来生产制度的调整与矿区的整体经济性评价提供参考。

1 预测算法研究现状

1.1 智能算法应用领域

目前,铀矿采区金属日产量预测主要通过平均铀浓度与日抽液量乘积求取,误差较大。基于人工智能算法的采区产量预测在铀矿领域研究较少,而在石油天然气、煤炭、稀土等领域应用较为成熟。薛永超[6]等使用深度森林算法对油井产量做出预测,获得 92%左右的预测精度,该方法在传统的随机森林算法上做出了极大改进,但在精度上仍然存在优化空间。蔡光琪[7]等基于原煤剥离量等自变量提出了改进的切比雪夫神经网络,该方式较传统 BP 神经网络运算资源小且精度更高;但未考虑储量因素,具有一定局限性。马承杰[8]考虑到传统数值模拟方式的缺陷,从储层物性出发,使用长短期记忆神经网络与循环神经网络建立 LSTM 数据模型,该模型性能优越,能够更好地捕捉时序关系差异;但在长期预测上精度仍有局限性,需进一步优化输入参数。

1.2 地浸产量预测适用算法分析

当前人工智能算法主要用于解决分类和回归

收稿日期:2023-02-15

基金项目:中核集团集中研发项目(〔2021〕A100-3),总装和科工局项目(A90-1)

第一作者简介:贾皓(1996—),男,江苏沛县人,硕士,助理工程师,主要从事地浸采铀与人工举升工艺研究。

通信作者简介:廖文胜(1968—),男,江西赣州人,博士,正高级工程师,主要从事地浸采铀与化学方面研究。

问题；而本研究中的地浸产量预测问题，属于数据拟合回归范畴，因此诸如 SVM、决策树、KNN 等众多分类算法并不适用，应当从回归预测算法中选取合适算法进行。常见的回归算法有人工神经网络（ANN）、卷积神经网络、BP 神经网络、逻辑回归、多元线性回归等。

矿区内抽注液井数量多，现场针对单井能够获取的参数较为单一（仅有流量、金属浓度），数据维度过低，不利于高精度数值模型的建立。因此，笔者从数据集成角度考虑，以整个采区金属日产量为研究对象，增加注液压力、注氧压力、浸采率等输入参数，从资源总量与浸采工艺数据入手，建立金属产量预测模型。

本研究中可用的数据既有相关生产参数，也有相对应的每日金属量数据，可以采用有监督学习的算法。多元线性回归建模速度快、算力需求小、模型可解释性强，常被用于煤炭价格、地下水水位、煤矿井下瓦斯体积分数等具体指标数据的预测[9-11]，与本研究中的数据类型较为契合。人工神经网络属于黑箱模型，可以充分逼近各类非线性关系，还具有并行分布处理与自适应未知参数的特点，该方法在页岩油藏产量与地下水涌水量等地质参数预测方面都有成功运用的先例[12-13]。因此本研究拟采用多元线性回归和人工神经网络算法建立模型。

2　生产参数概况

以北方某矿床 C10 采区进行分析，该采区铀金属储量为 55.3 t，截至 2022 年 2 月浸采率约 48%。该采区共包含 21 口抽液井与 47 口注液井（目前关停抽液井 4 口），矿体平均厚度为 4.3 m（厚度 0～13.5 m），平均铀品位为 0.016%（品位 0.01%～0.054%），平均平米铀量为 1.33 kg/m^2。该采区自 2017 年 10 月开始投入生产，其金属日产量在 2018 年中旬达到高峰（约 35 kg/d），随后下降至约 10 kg/d。从井场集控室采集了对金属日产量存在相关性的变量（表 1）。

该采区信息化程度较高，可获取抽液、注液与储量等众多参数。针对单日金属量的预测，需要使用每日采集一次或单日采集多次取平均值的参数。因此，舍弃取样频率较低的参数，如浸出液 pH、离子浓度与电位等，以抽注液参数与金属浓度为主要研究变量。

表 1　采区生产参数汇总

Table 1　Summary of production parameters in mining area

变量	变化范围
瞬时抽液量（$Q_{p/in}$）/（m^3/h）	20.57～147.66
瞬时注液量（$Q_{i/in}$）/（m^3/h）	38.63～92.85
累计总抽液量（Q_t）/（10^4m^3）	4.07～169.89
日注液量（$Q_{i/d}$）/（m^3/d）	927～2 228
日抽液量（$Q_{p/d}$）/（m^3/d）	991～2 834
注液压力（P_i）/MPa	0.46～1.82
注氧压力（P_o）/MPa	1.28～1.92
平均浸采率（R_r）/%	0.07～39.04
剩余可采金属量（R_m）/t	xxx～xxx
金属平均浓度（ρ_a）/（mg/L）	1.07～20.60

金属平均浓度为该采区各抽孔每日采样浓度的加权平均值，考虑到每日采样时刻的差异与单日内浓度波动，本研究通过表 1 中的 10 个变量建立金属日产量预测模型，通过多元线性回归与人工神经网络方式实现产量预测。

3　铀产量预测模型构建

3.1　基于多元线性回归方法的预测模型

3.1.1　多元线性回归基本原理

多元回归分析是指通过对多个自变量与 1 个因变量的相关分析，建立预测模型进行预测的方法。当自变量与因变量之间存在线性关系时称为多元线性回归分析，公式为

$$f(x) = \omega_1 x_1 + \omega_2 x_2 + \cdots + \omega_n x_n + b。\quad (1)$$

式（1）形式简单，易于建模且可解释性强。ω 和 b 通过均方误差最小化来确定，该方式又称“最小二乘法”，即找到 1 条直线，使样本到直线上的欧氏距离之和最小，均方误差计算公式为

$$(\omega, b) = \arg_{\min} \sum_{i=1}^{m} (f(x_i) - y_i)^2。\quad (2)$$

3.1.2　多元线性回归模型搭建

以 10 个生产参数为自变量，金属日产量为因变量进行模型搭建。首先对所有自变量计算方差膨胀系数（variance inflation factor，VIF），VIF 是回归系数估计量的方差与假设自变量无线性相关时方差的比值，表示多元线性回归模型中的共线性程度。其计算方法见式（3），R^2 表示判定系数，是衡量模型拟合度的参量。

$$VIF = 1/(1 - R^2)。\quad (3)$$

VIF 值越接近于 1，多重共线性越轻；反之越重。通常，对于 *VIF*＞10 的自变量选择放弃；但考虑到本模型流量参数选取较多（存在一定的多重共线性），因此对 *VIF* 适当放宽，对于 *VIF*＞30 的自变量选择舍弃。各自变量的 *VIF* 值见表 2，平均浸采率与瞬时注液量参数的 *VIF*＞30，选择舍弃。

表 2 各生产参数 *VIF* 与标准化系数计算结果

Table 2 Calculation results of *VIF* and standardization coefficient of each production parameter

变量名称	*VIF*	标准化后系数
瞬时抽液量（$Q_{p/in}$）	2.330	−0.006
瞬时注液量（$Q_{i/in}$）	2.949×10^{5}	
累计总抽液量（Q_t）	10.529	0.004
日注液量（$Q_{i/d}$）	8.724	0.180
日抽液量（$Q_{p/d}$）	4.081	−0.089
注液压力（P_1）	1.494	−0.001
注氧压力（P_o）	1.808	−0.001
平均浸采率（R_r）	1.246×10^{11}	
剩余可采金属量（R_m）	24.761	−0.062
金属平均浓度（ρ_a）	3.515	0.927

完成变量预处理工作，即可进行多元线性回归建模，模型表达见式(4)，其中 $f(x)$ 表示该采区金属日产量。该模型由于自变量较多，判定系数（R^2）会随自变量数量的增加而增加，通常引入调整后的 R^2_{adjust} 来衡量模型效果。该参数能对添加的非显著变量给出调和，也就是说降低变量数量过多对模型拟合度的无效提升，该模型中 $R^2_{adjust}=0.988$，说明模型拟合效果良好。

$$f(x)=-0.005\times Q_{p/in}+8.147\times10^{-8}\times Q_t+0.007\times Q_{i/in}-0.4\times P_1-0.56\times P_0+0.003\times Q_{p/d}+1.942\times\rho_a-6.18\times10^{-5}\times R_m-14.341。\tag{4}$$

由于各变量数值在数量级上差异较大，故式中变量系数不能反映各自变量对于金属日产量的影响程度，需要对自变量系数进行标准化处理，得到标准化后的系数（表 2）。可以看出，金属平均浓度起决定性作用，与常规计算方法相对应；日抽液量与日注液量的影响程度较大，表明流场循环情况对于金属日产量也具有重要影响。

从多元线性回归模型效果（图 1）来看，在前期单日产量增长阶段，模型预测值与真实产量拟合良好；但在生产后期，预测值整体偏大且具有不稳定性。这是由于各自变量权重不变导致其某一部分适应性下降。单日金属产量预测值平均误差约为 1.56%；自 2017 年 11 月至 2020 年 12 月的总产量预测误差为 246.8 kg，相对误差为 1.24%，整体预测效果可靠。

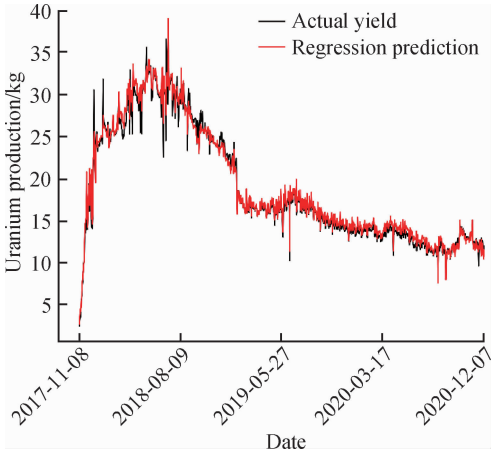


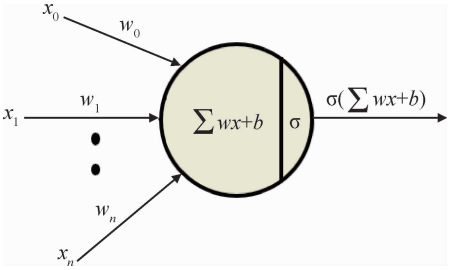
图 1 多元线性回归模型训练效果

Fig. 1 Training effect of multiple linear regression model

3.2 铀产量的人工神经网络预测模型

3.2.1 人工神经网络基本原理

近年来，有关神经网络在各学科领域的运用方兴未艾[14-16]。神经网络是由具有适应性的简单单元组成的广泛并行连接网络，能够模拟生物神经系统对外界情况做出交互反应，图 2 为单个神经元示意图。



ω—权值；b—偏置；σ—激活函数。

图 2 单个神经元计算原理

Fig. 2 Calculation principle of single neuron

根据激活函数选取的不同，人工神经网络既

可以解决分类问题，也可用来进行回归拟合，其模型结构见图 3。原始数据输入之前通常进行标准化处理以消除数据尺度间的差异，标准化计算公式为

$$\bar{x}_i = \frac{|x_i| - \min(|x_1|,\cdots,|x_n|)}{\max(|x_1|,\cdots,|x_n|) - \min(|x_1|,\cdots,|x_n|)} \quad i = 1,2,3,\cdots,n。 \tag{5}$$

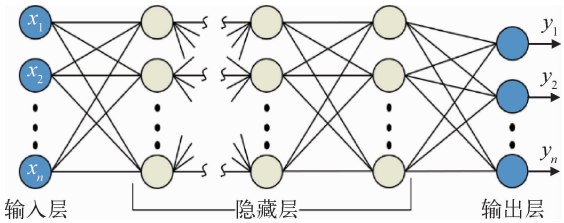


图 3　人工神经网络结构示意图

Fig. 3　Structure diagram of artificial neural network

多个隐藏层的设置能够识别出训练数据的突出特征；激活函数能够给输出层中经过线性计算的系统添加非线性特征。其中 ReLU 函数能够使神经网络模型更快速地收敛，避免过饱和现象，有效抵抗梯度消失问题，ReLU 函数的公式为

$$f(x) = \max(0,x)。 \tag{6}$$

在人工神经网络建立中，通常需要调节模型参数，性能最优的参数组合就是所求目标。贝叶斯优化算法基于现有的参数预测结果来建立替代函数（概率模型）[17]，通过替代函数得到目标函数最优解，可用来调节模型超参数。

3.2.2　人工神经网络模型搭建

搭建的人工神经网络产量预测模型，输入数据包含该采区 2017 年 11 月至 2020 年 12 月的生产数据，共 1 065 组。由于该模型不包含时间序列相关变量，部分日期无生产数据不影响模型精度。考虑到验证模型的泛化迁移能力，2021 年 1 月至 2022 年 2 月的生产数据不纳入模型训练与测试。

以表 1 中的 10 个自变量为数据输入层，在输入之前通过标准化处理以消除不同参数间的差异。考虑到样本的均匀覆盖，在近 3 年的 1 065 组生产数据中，按照 7∶3的比例随机选取训练集与测试集。使用训练集搭建模型，使用测试集对模型进行初步检验。

该模型结构除了输入层与输出层外，还包含 2 个隐藏层，通过贝叶斯优化算法调节得到最优模型参数。该模型第 1 个隐藏层包含 95 个隐藏神经元，第 2 个隐藏层包含 97 个隐藏神经元，最大迭代次数 196 次，初始学习率为 0.005。基于以上参数即可得到人工神经网络产量预测模型。

表 3　使用贝叶斯优化后的超参数取值

Table 3　The value of super parameters after Bayesian optimization

参数名称	调整范围	最优参数值
第 1 隐藏层节点数	15～100	95
第 2 隐藏层节点数	15～100	97
最大迭代次数	50～200	196
初始学习率	0.001～0.01	0.005

从基于 1 065 组训练测试数据的模型效果（图 4）看出，实际金属日产量与人工神经网络模型的预测值变化趋势基本相同；在前期到达产量峰值前，实际金属日产量跳动较大，而人工神经网络模型预测值变化较平缓。这可能是由于生产前期金属浓度增长较快，地下流场不稳定且变化较快，反应速率持续增加；也可能是因每日的金属浓度记录时间不同造成的。对于产量下降阶段，人工神经网络模型的预测值较为精准，单日金属产量平均误差仅有 0.88%；比多元线性回归模型更为精确。从整体来看，人工神经网络模型对这 3 年总产量的预测误差为 81.4 kg，相对误差为 0.41%；这也表明该模型预测精准，效果良好。

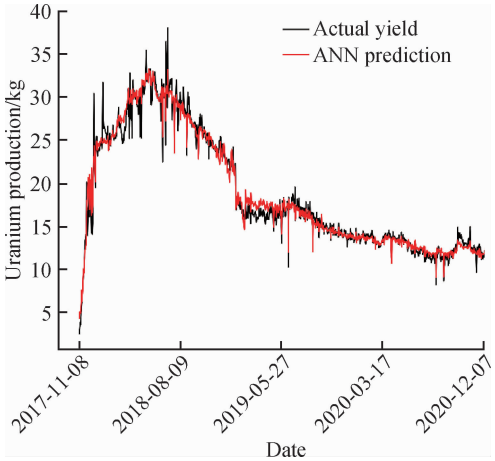


图 4　人工神经网络模型训练效果

Fig. 4　Training effect of artificial neural network model

4　实例运用与分析

基于以上 2 个预测模型，使用该矿床 C10 采

区的 2021 年 1 月至 2022 年 2 月的数据进行泛化能力检验。

4.1 多元线性回归模型检验

多元线性回归模型针对验证数据集的预测与误差分布见图 5。可以看出，在这段时间前期，金属日产量在 10～12 kg/d 波动；在生产中后期，金属日产量缓慢下降。多元线性回归模型的预测趋势与预期相符；但数据波动较为严重，预测值变化趋势为降低-增加-降低，预测结果均衡性较差，但相对误差大多在±5%之内。从期间金属总产量来看，整体预测较为可靠，相对误差为−1.68%。

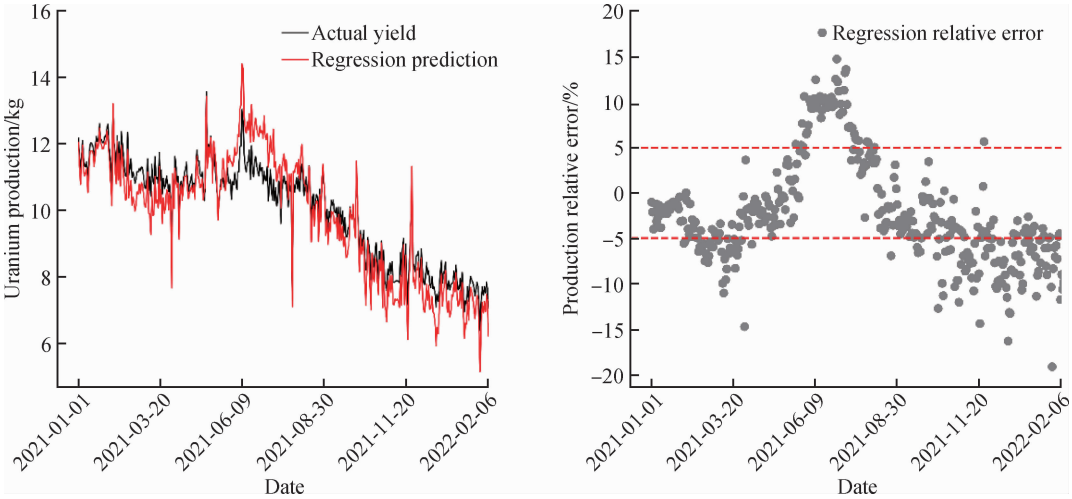


图 5 多元线性回归预测检验与误差分布

Fig. 5 Prediction test and error distribution of multiple linear regression

4.2 人工神经网络模型检验

人工神经网络的预测结果与误差分布见图 6。可以看出，该模型整体趋势拟合效果良好，对跳动较大的产量数据拟合效果一般；绝大多数预测值均分布在±5%以内，且无明显变化趋势。该模型泛化能力较多元线性回归模型有了很大提高，从预测期间金属总量来看，相对误差仅为−0.63%，说明该模型不仅能够准确预测金属日产量，也能精准确定某段时间的金属总产量。

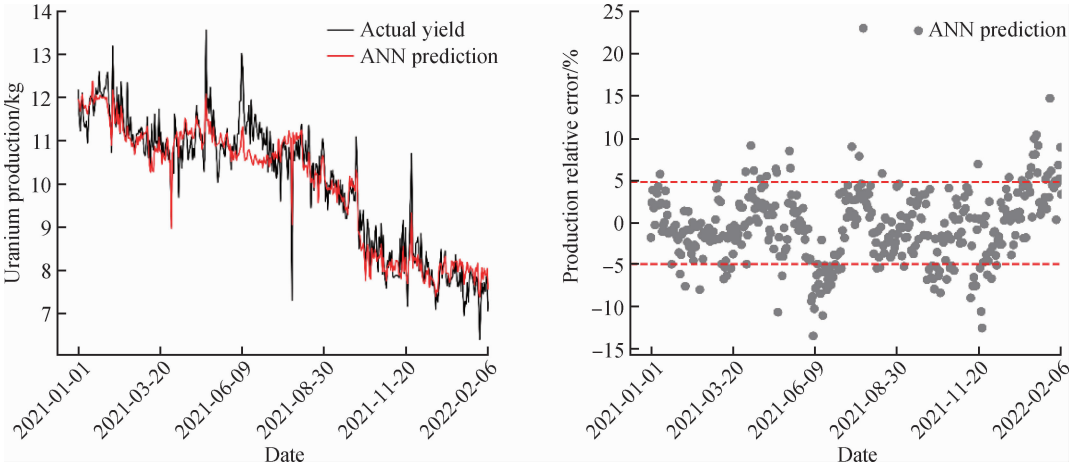


图 6 人工神经网络模型预测检验与误差分布

Fig. 6 Prediction test and error distribution of artificial neural network model

4.3　预测效果对比

3 种方式预测效果对比见表 4。可以看出，多元线性回归模型和人工神经网络模型的预测效果较简单金属日产量计算方法有了较大提高。通常使用金属浓度与当日抽液量相乘求解金属日产量，虽然计算误差大多分布在±5%内；但由于计算值普遍偏小，因此预测值并不均衡，在日产量与总产量的预测上表现一般。多元线性回归模型的性能介于计算值与人工神经网络模型之间，其预测误差较小；但存在一定比例的离群值（相对误差大于 10%）。人工神经网络模型表现优异，其金属日产量与总产量误差均小于 1%；虽然预测趋势较实际日产量更为平缓；但拟合效果最好，76.2%的预测值相对误差均在±5%以内，且离群值较少，说明该模型预测精准且泛化能力良好。

表 4　3 种预测方法的预测效果对比

Table 4　Comparison of prediction effects of three prediction methods

模型名称	日产量平均误差/%	总产量误差/%	误差±1%占比/%	误差±5%占比/%	误差大于 10%占比/%
多元线性回归	−2.20	−1.68	9.37	53.42	11.65
人工神经网络	−0.36	−0.63	18.48	76.20	2.53
简单计算	−3.90	−3.78	16.96	68.86	0.25

5　结论

针对地浸铀矿采区金属日产量预测问题，搭建了多元线性回归模型和人工神经网络模型，2 种模型较传统的金属日产量计算方法均有了较大提升。

1）通过对生产参数的分析，将变量精简后建立的多元线性回归模型能够对 C10 采区的金属产量做出较为精确的预测，预测结果准确可靠，具有一定的泛化能力。从多元线性回归预测模型公式可以直观看出各生产参数与金属日产量的相关性与影响权重，可为生产制度的调整与增产提供指导。

2）人工神经网络模型较多元线性模型在精度上进一步提高，对于近 13 个月采区预测的日产量平均误差仅为 −0.36%，总产量预测误差为 −0.63%，相对误差分布集中于±10%以内，模型预测效果优异。

参考文献：

[1]　张明林，刘洋，吴建勇，等. 中国铀矿地质勘查信息化建设现状及“十四五”发展思路[J]. 世界核地质科学，2021，38(3)：287-294.

[2]　袁志华，孙占学，周义朋，等. 绿色铀矿山研究进展[J]. 有色金属（冶炼部分），2021(11)：85-95.

[3]　陈军利，倪玉辉，于宝民. 绿色铀矿山建设标准构建思路[J]. 铀矿冶，2017，36(3)：155-158.

[4]　关键，侯江，邢妍. 数字地浸铀矿山顶层设计研究[J]. 铀矿冶，2021，40(2)：134-138.

[5]　廖文胜，王立民. 地浸采铀过程中抽注液量影响因素分析[J]. 铀矿冶，2022，41(2)：104-110.

[6]　薛永超，袁志乾，金青爽，等. 基于深度森林算法的油井产量预测[J]. 科学技术与工程，2022，22(11)：4327-4334.

[7]　蔡光琪，彭洪阁，周毅勇，等. 基于改进的切比雪夫神经网络的露天煤矿产量预测[J]. 中国矿业，2010，19(12)：88-89.

[8]　马承杰. 基于循环神经网络的油藏产量预测方法[J]. 内蒙古石油化工，2021，47(5)：108-112.

[9]　刘威，陈文昊，姜宇翔，等. 基于多元线性回归模型对煤炭问题的研究[J]. 智慧中国，2020，55(9)：83-84.

[10]　李啸. 基于多元线性回归对地下水水位的预测——以山西定襄县平原区为例[J]. 海河水利，2021(1)：19-21.

[11]　汝彦冬. 基于多元线性回归模型煤矿瓦斯体积分数预测[J]. 哈尔滨商业大学学报（自然科学版），2017，33(6)：735-737.

[12]　周玉辉，龚杰，盛广龙，等. 基于人工神经网络的页岩油藏产量预测定量表征模型[J/OL]. 大庆石油地质与开发：1-7[2023-06-01]. https：//doi.org/10.19597/J.ISSN.1000-3754.202208029.

[13]　何国富，张奇华，柳耀琦，等. 基于人工神经网络的地下水封洞库涌水量预测[J]. 水利与建筑工程学报，2022，20(5)：30-34＋74.

[14]　鲁中歧，肖文生，崔俊国，等. 基于参数优化神经网

络的海底油气管道腐蚀泄漏预测[J]. 科学技术与工程,2022,22(20):8673-8682.

[15] 盛茂,李根生,田守嶒,等. 人工智能在油气压裂增产中的研究现状与展望[J]. 钻采工艺,2022,45(4):1-8.

[16] 贾杨. 官地矿区采动地裂缝影响因素分析和发育规模预测[D]. 太原:太原理工大学,2020.

[17] XU Z, GUO Y, SALEH J H. Efficient hybrid Bayesian optimization algorithm with adaptive expected improvement acquisition function[J]. Engineering Optimization,2021,53(10):1786-1804.

Production Prediction of In-situ Leaching Mining Area Based on Artificial Intelligence Algorithm

JIA Hao[1,2], LIAO Wensheng[1,2], DU Zhiming[1,2], WANG Yaan[1,2], WANG Limin[1,2]

(1. Beijing Research Institute of Chemical Engineering and Metallurgy, CNNC, Beijing 101149, China;
2. China Nuclear Mining Science and Technology Corporation, CNNC, Beijing 101149, China)

Abstract: At the later stage of production in the uranium mining area, the metal production usually continues to decline, and the metal concentration also changes with the influence of the intensified leaching process. At this time, the production cannot be accurately calculated by the liquid pumping volume and the concentration. Based on the induction and analysis of production parameters, a multiple linear regression model and an artificial neural network model are established for the prediction of daily metal production in a single mining area. The results show that both of them have good accuracy and generalization ability, the effect of artificial neural network model is better, the average error of metal daily production prediction is only −0.36%, the outliers are less and the error distribution is balanced. The accuracy of this method is significantly improved compared with the simple calculation to obtain the daily metal production. The use of artificial neural network model to predict the production of the mining area can provide solutions and theoretical guidance for the adjustment and change of leaching process and production system.

Key words: in-situ leaching uranium; production prediction; ANN; multiple linear regression