

# 基于多任务学习多目标优化的 稀土元素组分含量与浓度多维度软测量

张水平, 张奇涵, 王碧<sup>†</sup>

(江西理工大学 信息工程学院, 江西 赣州 341000)

**摘要:** 稀土混合萃取溶液中各元素组分含量的在线软测量是优化连续萃取生产过程、确保产品高纯化的前提。现有软测量方法可独立求解单个稀土元素组分含量, 但忽略了多元素组分含量间或组分含量与其它相关因素(如浓度)间的共性。本文为探索多稀土元素组分含量间及组分含量与浓度间的共性, 将多任务学习方法用于稀土元素组分含量软测量中。首先, 构建多任务深度神经网络, 提高模型的泛化能力和鲁棒性。其次, 提出基于多目标优化算法的稀土多元素组分含量预测方法, 通过搜索Pareto最优以提升各任务的预测精度。经多组对比实验表明, 该方法在多元素组分含量或多元素组分含量与浓度同时训练时性能最佳, 能满足稀土元素组分含量在线检测的精确性和实时性。

**关键词:** 稀土萃取; 组分含量; 多任务学习; 多目标优化; 机器学习; 深度学习; 帕累托

**引用格式:** 张水平, 张奇涵, 王碧. 基于多任务学习多目标优化的稀土元素组分含量与浓度多维度软测量. 控制理论与应用, 2024, 41(3): 454 – 467

DOI: 10.7641/CTA.2023.20871

## Content and concentration of rare earth element components based on multi-task learning multi-objective optimization multidimensional soft measurement

ZHANG Shui-ping, ZHANG Qi-han, WANG Bi<sup>†</sup>

(School of Information Engineering, Jiangxi University of Science and Technology, Ganzhou Jiangxi 341000, China)

**Abstract:** Online soft measurement of the component content of each element in a mixed rare earth extraction solution is a prerequisite for optimizing the continuous extraction production process and ensuring high purity of the product. Existing soft measurement methods can solve for individual rare earth element fractions independently, but ignore the commonality between multi-element fractions or between fractions and other relevant factors (e.g. concentration). A multi-task learning approach is used to explore the commonality between the component content of multiple rare earth elements and between the component content and concentration in soft measurements of rare earth elements. Firstly, a multi-task deep neural network is constructed to improve the generalization ability and robustness of the model. Secondly, a multi-objective optimization algorithm is proposed to improve the prediction accuracy of each task by searching the Pareto optimum. After several sets of comparison experimental results, it is shown that the method has the best performance when the multi-element component content or multi-element component content and concentration are trained at the same time, which can meet the accuracy and real-time performance of online detection of rare earth elemental component content.

**Key words:** rare earth extraction; component content; multi-task learning; multi-objective optimization; machine learning; deep learning; Pareto

**Citation:** ZHANG Shuiping, ZHANG Qihan, WANG Bi. Content and concentration of rare earth element components based on multi-task learning multi-objective optimization multidimensional soft measurement. *Control Theory & Applications*, 2024, 41(3): 454 – 467

收稿日期: 2022–10–08; 录用日期: 2023–04–20.

<sup>†</sup>通信作者. E-mail: wangbi@jxust.edu.cn; Tel.: +86 18146699752.

本文责任编辑: 周平.

江西理工大学博士科研启动基金项目(2022205200100595), 国家自然科学基金委员会项目(72261018), 江西省教育厅青年项目(GJJ2200868)资助。Supported by the Jiangxi University of Science and Technology Doctoral Research Start-up Fund (2022205200100595), the National Natural Science Foundation of China (72261018) and the Science and Technology Research Project (GJJ2200868) from Jiangxi Education Department.

## 1 引言

稀土因其特有的光、电、磁等性质<sup>[1]</sup>, 成为永磁材料、催化材料、储氢材料等材料必须的关键元素, 而这些高性能的稀土材料在冶金、化工、航空航天、国防科技等高新技术领域起核心作用, 受到国内外科学界的广泛关注<sup>[2-5]</sup>. 我国是稀土储量大国, 稀土的分离工业规模和产量雄踞世界首位, 然而稀土分离工业自动化生产过程中获取组分含量值的方式还基本停留在离线分析的低水平阶段. 稀土联动萃取分离工艺流程<sup>[6]</sup>主要依靠在连续的生产过程中对监测点处的稀土各元素组分含量进行在线检测, 通过检测数据来优化控制料液、有机相、洗涤液等物料流量, 以此保证分离产品的高纯化. 而目前常用的在线检测装置价格高, 结构复杂, 维护保养困难, 因此, 探寻具备成本低、维护保养简单、连续检测迅速等优点的稀土元素组分含量软测量<sup>[7]</sup>方法为提升稀土分离工业自动化水平的关键.

目前, 国内有学者使用基于传统机器视觉的机器学习方法<sup>[8-10]</sup>对稀土元素组分含量进行软测量, 虽然该类方法能够独立求解各元素组分含量, 并达到萃取过程所需精度, 但需要人工筛选并提取稀土混合萃取溶液图像特征, 特征提取过程繁琐, 相较于深度学习方法<sup>[11-12]</sup>具有局限性<sup>[13]</sup>; 且该类方法未将混合萃取溶液图像中各元素组分含量之间或组分含量与其他相关因素之间联合进行训练及预测, 忽略了多个元素组分含量之间或组分含量与其他相关因素之间在稀土混合萃取溶液图像中颜色特征之间的共性, 泛化能力不足. 深度学习模型能够自动提取图像抽象表征, 通过卷积、池化等操作<sup>[14]</sup>让所提取表征更具泛化能力, 而深度学习中的多任务学习方法<sup>[15]</sup>可以将多个任务联合训练, 旨在利用相关任务的训练信号中包含的特定特征来改进泛化能力. 与单任务学习相比(每个任务由其单独的网络进行求解), 多任务学习具有以下优势:

- 1) 多个特定任务共享一个共享层网络, 大大减小内存的占用;
- 2) 避免重复计算共享层网络中的特征, 提升模型训练速度;
- 3) 若相关任务共享互补信息, 则具有提高模型整体性能的潜力.

从多任务学习方法的角度进行分类, 可以分为基于多任务模型结构和基于多任务优化策略两种多任务学习方法<sup>[15-16]</sup>. 其中, 将多任务模型结构按照处理多图像级分类任务的角度归类, 可分为基于硬参数共享和基于软参数共享两种模型结构: 在硬参数共享中, 模型参数可分为共享层参数和特定任务层参数, 将由一个编码器构成的共享层参数分支到各特定任务层

首层, 实现参数共享, 如Kokkinos<sup>[17]</sup>提出的Ubernet和Teichmann等<sup>[18]</sup>提出的Multinet; 在软参数共享中, 每个特定任务被分配各自的参数集, 并由特征共享机制对各特定任务参数集进行交互处理, 如Ruder等<sup>[19]</sup>提出的水闸网络(slucice networks)和Liu等<sup>[20]</sup>提出的MT-AN网络(multi-task attention network). 基于多任务优化策略的多任务学习主流方法有Chen等<sup>[21]</sup>提出的梯度归一化(gradient normalization), 旨在平衡多任务网络训练的梯度和速率, 鼓励网络以平等的速度学习所有任务, 但该方法可能在不同任务的量级不一时导致其中某些任务不能取得良好的解; Kendall等<sup>[22]</sup>提出的同方差不确定性(uncertainty weighting), 使用噪声参数 $\sigma$ 作用在损失函数上, 用于平衡各特定任务间的权重; 上述方法试图通过一些启发式设置损失中的特定权重来解决不同任务间性能相互冲突的问题, 不同的是, Désidéri<sup>[23]</sup>将多任务学习视为一个多目标优化问题, 提出了一种多重梯度下降算法(multiple gradient descent algorithm)来解决这个问题, 通过寻找不同任务间的Pareto平稳点, 使任何特定任务的损失值都在不增加其他损失值的前提下减少; Sener等<sup>[24]</sup>将在多重梯度下降算法的研究基础上, 针对多重梯度下降算法需要对每个任务的梯度进行显示计算比较耗时的问题, 提出多目标损失上界, 在所有任务中利用基于上界的多重梯度下降算法(multiple gradient descent algorithm-upper bound)找到各任务间的Pareto最优解, 通过无需明确特定任务梯度的单个反向计算过程, 使训练迭代时间大大减少, 并取得了与原方法相近甚至更优的测试性能. 基于多任务模型结构和基于多任务优化策略两种多任务学习方法可以交互使用, 共同提升模型整体性能.

国内也有学者对多目标优化问题进行研究, 张波等<sup>[25]</sup>在基于多重梯度下降和Frank-Wolfe两个算法上将多目标优化问题分解为多个两目标优化问题, 对梯度下降的方向进行梯度更新和线性搜索. 周晓君等<sup>[26]</sup>将多重梯度下降算法用于多车牌识别中, 提出基于多目标优化多任务学习的端到端车牌识别方法; 赵佳铭等<sup>[27]</sup>将多重梯度下降算法用于泡沫浮选过程工况识别中, 取得了良好的预测效果, 但周、赵并没有对训练过程的迭代时间进行统计与比较, 无法得知不同算法在不同任务训练迭代过程中的时间性能.

本文受到Sener等的启发, 针对稀土萃取在连续的生产过程中对监测点处的稀土各元素组分含量进行在线检测的需求, 探索多元素组分含量之间及组分含量与浓度之间的共性, 将稀土多个元素组分含量或组分含量与浓度同时训练及预测, 并将多个元素组分含量或组分含量与浓度的预测任务转换为一个多任务学习问题, 在基于多图像级分类任务的硬参数共享结构上构建多任务学习模型以提升模型泛化力; 针对多

个任务间可能存在竞争和冲突的问题,将稀土多元素组分含量或组分含量与浓度预测任务视为多目标优化问题,提出基于优化上界的多重梯度下降算法(multiple gradient descent algorithm based on optimal upper bound),找到多个任务间的Pareto最优解,在节省训练时间的同时提升模型预测精度.如图1所示.

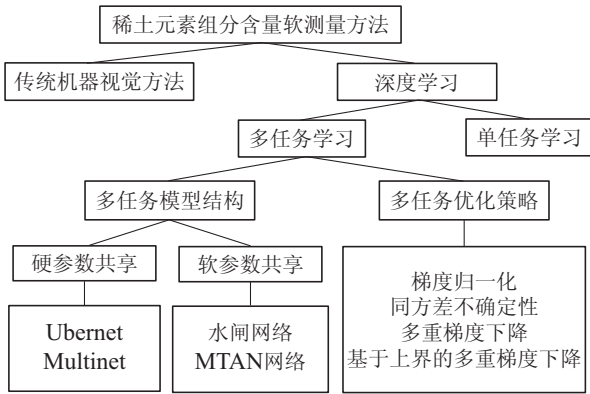


图1 稀土组分含量软测量方法

Fig. 1 Soft measurement method for content of rare earth element components

接下来,本文将在第2节介绍建立在稀土联动萃取分离流程下的稀土多元素组分含量与浓度预测多任务学习模型;第3节详细描述稀土多元素组分含量与浓度预测多目标优化算法;第4节将本文方法与单任务学习、多重梯度下降和其他主流多任务优化方法从以下两个角度在多个预测任务中呈现的性能进行多组对比实验:1)验证集的损失函数值变化、训练过程单次迭代平均时长;2)10个测试集样本预测值与真实值的误差评价指标值;最后,在第5节对本文进行总结,并对未来工作进行展望.

## 2 稀土多元素组分含量与浓度预测问题建模

### 2.1 稀土联动萃取分离流程

稀土联动萃取分离流程中为降低萃取流程所需物料(有机相、洗涤液等)的消耗,使用纵横相连的方式将萃取分离流程中相关分离单元关联,每个单元间彼此提供物料.这种关联方式使稀土分离整体流程中各物料仅需一个输入口,即可设置同时产出所需萃取元素纯产品的各出口.如La/Ce/Pr/Nd这4个稀土元素的联动萃取分离过程(图2),共有6个分离单元:(LaCePr)/(CePrNd), (LaCe)/(CePr), (CePr)/(PrNd), La/Ce, Ce/Pr, Pr/Nd, 6个单元相互纵横相连,通过(LaCePr)/(CePrNd), La/Ce和 Pr/Nd 这3个单元分别输入需萃取分离的混合稀土料液、有机相和洗涤液,便可从各元素出口获取4种元素萃取后的纯溶液.

为便于优化整个萃取流程的进料过程,需通过各监测点对萃取液中各元素组分含量的检测值计算出

3个进料单元所输入的各料液流量.而由于部分稀土元素(如Pr, Nd)具有独特的离子颜色特征,其萃取溶液具备对光线呈一定反射性和折射性的特点,透光性较好,能在可见光区域呈现颜色特征,满足基于机器视觉技术对稀土元素组分含量软测量的光学成像条件.因此,可在稀土联动萃取分离流程中的待测稀土元素组分含量处设置检测点,采集稀土混合萃取溶液图像,用于模型训练与预测.将稀土元素组分含量软测量数据通过理论公式计算,便可得到各分离单元所需的理论最小萃取量、最小洗涤量及分离单元横纵间联动衔接所需的最优化流量,从而优化控制萃取流程,以此保证分离产品的高纯化.

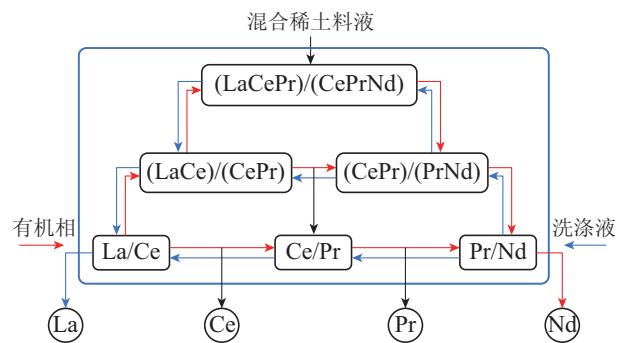


图2 La/Ce/Pr/Nd四元素联动萃取分离流程<sup>[6]</sup>

Fig. 2 La/Ce/Pr/Nd four-element linkage extraction separation process<sup>[6]</sup>

### 2.2 基于多任务学习的稀土多元素组分含量与浓度预测模型

为满足稀土联动萃取过程对多元素组分含量检测需求,和探索多元素组分含量之间及组分含量与浓度之间的共性,本文提出一种基于多任务学习的稀土多元素组分含量与浓度预测模型,对所采集稀土混合萃取溶液图像中的稀土多元素组分含量与浓度在共享层网络中同时训练,并设置多个特定任务层网络同时输出组分含量与浓度多个任务的预测结果.多任务学习模型结构如图3所示,该模型中共享层网络由除去末尾全连接层的ResNet18网络构成;每个特定任务层网络头部都与共享层网络相连,实现参数共享,并各设置一个全连接层用于输出特定任务预测值.

该模型建立在基于多图像级分类任务的硬参数共享结构之上,在经过对单稀土元素与浓度同时训练的比较实验中选取了对数据集预测结果拟合效果较好的ResNet18作为共享主干网络.其中ResNet18<sup>[28]</sup>网络中的残差模块由卷积层、非线性激活函数层、批处理归一化层构成,将批处理归一化(batch normalization)层取代dropout层,对网络隐含层输出进行标准化处理,在保留原有数据的同时使数值更加稳定,并防止了过拟合,网络结构参数如表1所示,其中Conv层实

现卷积操作后接批处理归一化操作和ReLU非线性激活函数,第6,10和14层额外进行下采样操作后用于特征短接,特征短接操作使模型可以学习到更深层网络的特征。

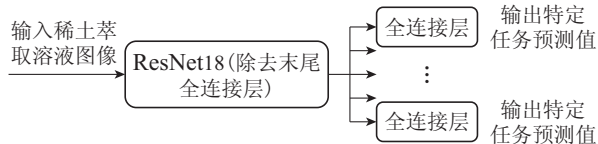


图3 基于多任务学习的稀土多元素组分含量与浓度预测模型结构

Fig. 3 Structure of the multi-task learning-based model for predicting the content and concentration of rare earth multi-element components

表1 ResNet18(除去末尾全连接层)网络结构参数

Table 1 ResNet18 (excluding the end fully connected layer) network structure parameters

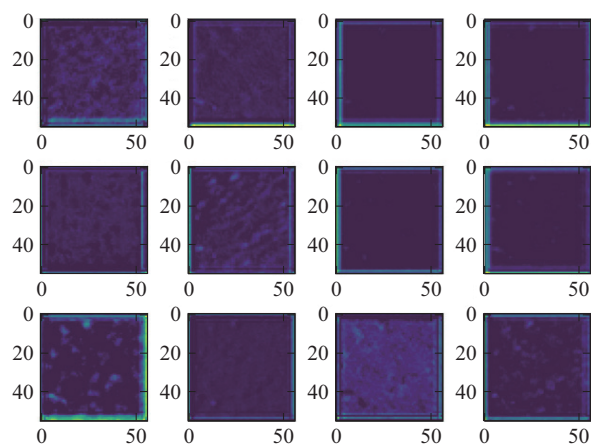
层数	层名称	通道数	滤波器大小	步长
1	Conv1	64	7 × 7	2 × 2
2	Conv2_1.1	64	3 × 3	1 × 1
3	Conv2_1.2	64	3 × 3	1 × 1
4	Conv2_2.1	64	3 × 3	1 × 1
5	Conv2_2.2	64	3 × 3	1 × 1
6	Conv3_1.1	128	3 × 3	2 × 2
7	Conv3_1.2	128	3 × 3	1 × 1
8	Conv3_2.1	128	3 × 3	1 × 1
9	Conv3_2.2	128	3 × 3	1 × 1
10	Conv4_1.1	256	3 × 3	2 × 1
11	Conv4_1.2	256	3 × 3	1 × 1
12	Conv4_2.1	256	3 × 3	1 × 1
13	Conv4_2.2	256	3 × 3	1 × 1
14	Conv5_1.1	512	3 × 3	2 × 2
15	Conv5_1.2	512	3 × 3	1 × 1
16	Conv5_2.1	512	3 × 3	1 × 1
17	Conv5_2.2	512	3 × 3	1 × 1
18	Avgpool	512	-	-

特别地,为获得单一准确的稀土元素组分含量与浓度预测值,本文将特定任务层网络中正向传播过程中的全连接层输出与Softmax函数<sup>[29]</sup>相连,利用Softmax函数对特定任务各类别概率的输出与各类别真实值通过线性回归损失函数建立回归模型,再将损失函数值进行反向传播,优化网络参数.线性回归模型中常用的损失函数有L1Loss和L2Loss,分别计算平均绝对误差(mean absolute error, MAE)(式1)和均方误差(mean square error, MSE)(式2),即

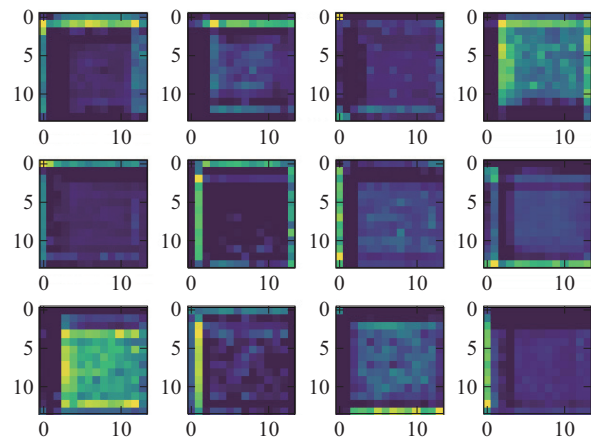
$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}, \quad (1)$$

$$\text{MSE} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}, \quad (2)$$

其中:  $y_i$ 为模型预测值,  $\hat{y}_i$ 为所预测数据点对应任务的真实标签值,  $n$ 为所预测数据点个数.相较于L1Loss, L2Loss损失函数将误差平方化,会加剧忽略微小的误差.例如,某混合萃取溶液图像样本真实元素组分含量为0.001,其预测值为0.011.虽然误差是真实值的10倍,但数值较小.若使用L2Loss,会进一步增加修改该误差的难度.因此,本文采用L1Loss作为多任务学习模型各任务的损失函数<sup>[30]</sup>.



(a) 网络第5层卷积仿真结果



(b) 网络第13卷积仿真结果

图4 ResNet18网络第5和第13卷积层在Pr组分含量为42.86%的稀土混合萃取溶液图像中前12个通道的仿真结果

Fig. 4 Simulation results of the first 12 channels in the image of rare earth mixed extraction solution with Pr component content of 42.86% in the 5th and 13th convolutional layer of ResNet18 network

### 3 稀土多元素组分含量与浓度预测多目标优化算法

在稀土多元素组分含量与浓度预测多任务学习模型的训练过程中,多个特定任务之间可能会发生冲突,导致共享层网络参数偏向某一个特定任务的情况.本

文在受到Sener启发后, 尝试将该情况通过多目标优化算法来解决, 将多个特定任务共同训练过程视为多目标优化过程, 通过找到多个特定任务之间的 Pareto 解来平衡各任务之间的竞争, 使共享层网络参数不会偏向某个特定任务, 从而提高对稀土多元素组分含量与浓度预测的精度。

### 3.1 多目标优化定义

给定稀土多元素组分含量与浓度预测的多任务学习问题中输入空间 $\mathcal{X}$ 为所采集稀土混合萃取溶液图像,  $\{\mathcal{Y}^t\}_{t \in [T]}$ 为一组稀土混合萃取溶液图像中对应的组分含量和浓度真实值空间, 则整个稀土混合萃取溶液图像数据集中的数据点为 $\{x_i, y_i^1, \dots, y_i^T\}_{i \in [N]}$ , 其中:  $T$ 为所需预测的稀土元素组分含量与浓度任务个数,  $N$ 为稀土混合萃取溶液图像数据点的个数,  $y_i^t$ 为第 $i$ 个数据点对应第 $t$ 个组分含量或浓度预测任务的真实值标签. 再进一步, 在多任务学习模型中考虑每个特定任务的参数化假设为 $f^t(x; \theta^{\text{sh}}, \theta^t): \mathcal{X} \rightarrow \mathcal{Y}^t$ , 其中:  $\theta^{\text{sh}}$ 为共享层网络参数,  $\theta^t$ 为特定任务层网络参数. 则可将稀土元素组分含量或浓度预测各任务的经验性最小化公式表示为式(3):

$$\min_{\theta^1, \dots, \theta^T} \sum_{t=1}^T c^t \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t), \quad (3)$$

其中:  $c^t$ 为特定任务被衡量所计算出的权重;  $\hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)$ 为第 $t$ 个预测任务的经验损失函数, 具体可被定义为

$$\hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t) \triangleq \frac{1}{N} \sum_i \mathcal{L}(f^t(x_i; \theta^{\text{sh}}, \theta^t), y_i^t), \quad (4)$$

假设式3中有两组参数解 $\theta$ 和 $\bar{\theta}$ , 在对于有两个特定任务 $t_1$ 和 $t_2$ 的条件下可能会产生如下问题:  $\hat{\mathcal{L}}^{t_1}(\theta^{\text{sh}}, \theta^{t_1}) < \hat{\mathcal{L}}^{t_1}(\bar{\theta}^{\text{sh}}, \bar{\theta}^{t_1})$ , 且 $\hat{\mathcal{L}}^{t_2}(\theta^{\text{sh}}, \theta^{t_2}) > \hat{\mathcal{L}}^{t_2}(\bar{\theta}^{\text{sh}}, \bar{\theta}^{t_2})$ , 即参数解 $\theta$ 更偏向于任务 $t_1$ , 而 $\bar{\theta}$ 更偏向于任务 $t_2$ . 为解决该问题, 本文将多个任务共同学习视为一个多目标优化问题, 寻找平衡多个任务之间的Pareto解, 使网络参数不偏向任何一个特定任务. 则可将上述问题转化为求下式参数解:

$$\begin{aligned} & \min_{\theta^{\text{sh}}, \theta^1, \dots, \theta^T} \mathbf{L}(\theta^{\text{sh}}, \theta^1, \dots, \theta^T) = \\ & \min_{\theta^{\text{sh}}, \theta^1, \dots, \theta^T} (\hat{\mathcal{L}}^1(\theta^{\text{sh}}, \theta^1), \dots, \hat{\mathcal{L}}^T(\theta^{\text{sh}}, \theta^T))^T, \end{aligned} \quad (5)$$

式(5)由包含稀土元素组分含量与浓度预测多个任务对应损失函数的向量 $\mathbf{L}$ 组成, 目的在于寻求平衡多个任务之间的Pareto解. 下面给出在多目标优化中多个任务之间Pareto解的定义:

**定义 1** 对所有任务 $t$ 有 $\hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t) \leq \hat{\mathcal{L}}^t(\bar{\theta}^{\text{sh}}, \bar{\theta}^t)$ ,  $\forall t \in (1, 2, \dots, T)$ , 且 $\mathbf{L}(\theta^{\text{sh}}, \theta^1, \dots, \theta^T) \neq \mathbf{L}(\bar{\theta}^{\text{sh}}, \bar{\theta}^1, \dots, \bar{\theta}^T)$ , 则称参数解 $\theta$ 支配 $\bar{\theta}$ .

**定义 2** 如果不存在参数解 $\theta$ 支配 $\theta^*$ , 则称参数解 $\theta^*$ 为Pareto最优解.

### 3.2 基于优化上界的多重梯度下降算法

本文提出一种基于优化上界的多重梯度下降算法, 用于寻求稀土元素组分含量与浓度预测多任务学习中的 Pareto 最优解, 通过多任务学习模型训练中的梯度下降方式完成多目标优化. 下面将使用 Karush-Kuhn-Tucker(KKT)条件实现多目标优化, 该条件也是多目标优化的必要条件. 现将基于多任务学习的稀土多元素组分含量与浓度预测模型中共享层网络参数与特定任务层网络参数所需的KKT条件表示如下:

$$1) \text{ 存在 } \alpha^1, \dots, \alpha^T \geq 0, \text{ 并有 } \sum_{t=1}^T \alpha^t = 1, \text{ 且}$$

$$\sum_{t=1}^T \alpha^t \nabla_{\theta^{\text{sh}}} \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t) = 0;$$

2) 对于任意稀土元素组分含量与浓度预测任务 $t$ ,  $\nabla_{\theta^t} \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t) = 0$ . 满足上述条件的任何解都被称为 Pareto 平稳点, 其中本文模型在完成上述优化条件中所使用损失函数为L1Loss, 在第2.2节中被提及. 再考虑如下优化问题:

$$\begin{aligned} & \min_{\alpha^1, \dots, \alpha^T} \{ \|\sum_{t=1}^T \alpha^t \nabla_{\theta^{\text{sh}}} \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)\|_2^2 \mid \sum_{t=1}^T \alpha^t = 1, \\ & \alpha^t \geq 0, \forall t \}. \end{aligned} \quad (6)$$

Désidéri等表明: 该优化问题的解为0时, 其解满足KKT条件; 否则, 该解给定了一个下降方向, 优化稀土元素组分含量与浓度所有预测任务. 然而, 该问题描述的算法需要计算 $\nabla_{\theta^{\text{sh}}} \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)$ , 则需为每个特定任务关联的共享层网络参数进行反向传播, 因此, 各特定任务的梯度计算需要在 $T$ 次反向传播得到, 再进行正向传播, 大大增加了内存和时间的消耗. 对此, 本文提出基于优化上界的多重梯度下降算法, 不同于多重梯度下降算法, 该算法优化了目标上界, 仅需一遍反向传播就可在正向传播中得到所有特定任务的梯度, 节约了内存和时间上的消耗. 优化上界需要将共享的表示函数与特定任务的决策函数相结合, 可以将假设类约束定义为下式:

$$\begin{aligned} f^t(x; \theta^{\text{sh}}, \theta^t) &= (f^t(\cdot; \theta^t) \circ g(\cdot; \theta^{\text{sh}}))(x) = \\ & f^t(g(x; \theta^{\text{sh}}); \theta^t), \end{aligned} \quad (7)$$

其中:  $g$ 为所有任务共享的表示函数;  $f^t$ 为将该表示作为输入的特定任务函数, 若将该表示函数表示为 $Z = (Z_1, \dots, Z_N)$ , 其中 $Z_i = g(x_i; \theta^{\text{sh}})$ , 则可将上界表示为下式, 该式为链式法则的直接结果:

$$\begin{aligned} & \|\sum_{t=1}^T \alpha^t \nabla_{\theta^{\text{sh}}} \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)\|_2^2 \leq \\ & \|\frac{\partial Z}{\partial \theta^{\text{sh}}}\|_2^2 \|\sum_{t=1}^T \alpha^t \nabla_Z \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)\|_2^2, \end{aligned} \quad (8)$$

其中  $\|\frac{\partial Z}{\partial \theta^{\text{sh}}}\|_2^2$  是  $Z$  相当于  $\theta^{\text{sh}}$  的雅可比矩阵范数. 该上界的两个理想性质为

1)  $\nabla_Z \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)$  可以在一次反向传播中计算所有特定任务的梯度;

2)  $\|\frac{\partial Z}{\partial \theta^{\text{sh}}}\|_2^2$  不是一个关于  $\alpha^1, \dots, \alpha^T$  的方程, 因此, 当它作为优化目标时可被移除. 为得到近似优化解, 考虑到上述两个理想性质, 可将式(6)中的  $\sum_{t=1}^T \alpha^t \times \|\nabla_{\theta^{\text{sh}}} \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)\|_2^2$  由式(8)中的上界替代, 并删除  $\|\frac{\partial Z}{\partial \theta^{\text{sh}}}\|_2^2$  项, 则将多任务优化转化为求解下式:

$$\begin{aligned} \min_{\alpha^1, \dots, \alpha^T} \{ & \|\sum_{t=1}^T \alpha^t \nabla_Z \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)\|_2^2 \mid \sum_{t=1}^T \alpha^t = 1, \\ & \alpha^T \geq 0, \forall t\}, \end{aligned} \quad (9)$$

求解该式的算法则称为基于优化上界的多重梯度下降算法, Sener提出当  $\frac{\partial Z}{\partial \theta^{\text{sh}}}$  为满秩且  $\alpha^1, \dots, \alpha^T$  为式(9)解的情况下, 下面两种条件之一能被满足:

1)  $\sum_{t=1}^T \alpha^t \nabla_{\theta^{\text{sh}}} \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t) = 0$ , 且当前多任务学习模型参数为Pareto平稳;

2)  $\sum_{t=1}^T \alpha^t \nabla_{\theta^{\text{sh}}} \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)$  是所有目标任务的下降方向, 则该算法找到了一个Pareto平稳点, 且计算开销可忽略不计.

式(9)属于具有线性约束的凸二次问题, 求解该优化问题等价于在输入点集的凸包中找到最小范数点. 本文先考虑基于两个任务求解的情况, 优化问题可被定义为

$$\min_{\alpha \in [0,1]} \|\alpha \nabla_Z \hat{\mathcal{L}}^1(\theta^{\text{sh}}, \theta^1) + (1-\alpha) \nabla_Z \hat{\mathcal{L}}^2(\theta^{\text{sh}}, \theta^2)\|_2^2, \quad (10)$$

这是一个具有解析解的关于  $\alpha$  的一维二次函数, 下面将  $\nabla_Z \hat{\mathcal{L}}^1(\theta^{\text{sh}}, \theta^1)$  简化为  $\theta$ , 将  $\nabla_Z \hat{\mathcal{L}}^2(\theta^{\text{sh}}, \theta^2)$  简化为  $\bar{\theta}$ , 则对式(10)求导, 解如下式(11):

$$\alpha = \begin{cases} 0, & \theta^T \bar{\theta} \geq \bar{\theta}^T \bar{\theta}, \\ 1, & \theta^T \bar{\theta} \geq \theta^T \theta, \\ (\bar{\theta} - \theta)^T \bar{\theta} / \|\theta - \bar{\theta}\|_2^2, & \theta^T \bar{\theta} < \bar{\theta}^T \bar{\theta} \text{ 且 } \theta^T \bar{\theta} < \theta^T \theta. \end{cases} \quad (11)$$

虽然式(11)只适用于两个任务的求解, 但本文将使用Jaggi<sup>[31]</sup>提出的 Frank-Wolfe 算法, 将式(10)作为线性搜索的子进程, 从而解决两个及以上任务的约束优化问题. Frank-Wolfe 算法流程如图5. 所求得  $\alpha^1, \dots, \alpha^T$  即为基于优化上界的多重梯度下降算法的解, 可通过此解优化所有特定任务的下降方向, 求得稀土多元素组分含量与浓度预测多目标的Pareto最优.

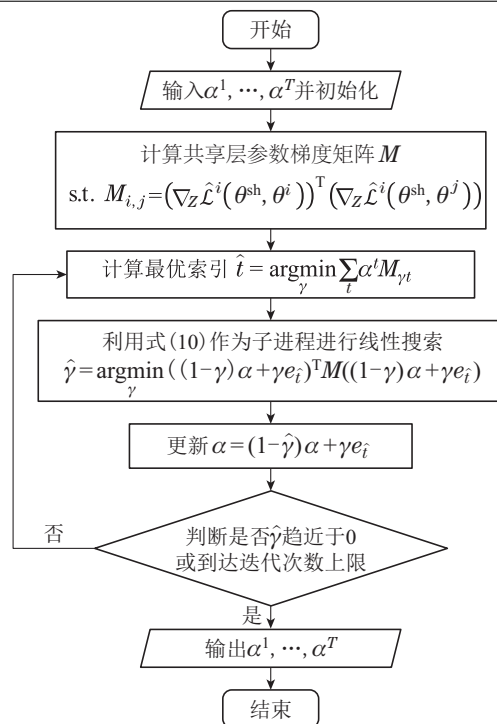


图5 Frank-Wolfe算法流程

Fig. 5 Frank-Wolfe algorithm flow

### 3.3 稀土多元素组分含量与浓度预测多目标优化算法流程

在对于稀土多元素组分含量与浓度预测多个目标进行优化过程中, 本文采用基于优化上界的多重梯度下降算法求解各任务在模型训练中的梯度, 该算法优化了目标上界, 因此, 只需经过一遍反向传播, 便可在正向传播过程中得到所有特定任务的梯度. 此后, 再将各特定任务梯度代入Frank-Wolfe算法中求多目标优化解, 得到关于稀土多元素组分含量与浓度预测多个任务的Pareto平稳点, 实现多目标优化. 详细步骤描述如下:

**输入:** 将标准化预处理后的稀土混合萃取溶液图像输入基于多任务学习的稀土多元素组分含量与浓度预测模型, 加载图像样本对应真实标签值. 初始化共享层网络参数  $\theta^{\text{sh}}$ , 特定任务层网络参数  $\theta^t$ ; 设置损失函数、最大迭代次数、批大小、优化器和学习率  $\eta$ , 在共享层网络中求解多目标优化问题.

**输出:** 在各特定任务层输出稀土元素组分含量或浓度预测值.

**步骤1** 根据式(9)初始化多目标优化上界  $\nabla_Z \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)$ ;

**步骤2** 在优化上界下求得每个特定任务的梯度  $\nabla_{\theta^t} \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)$ , ( $t \in T$ );

**步骤3** 将所有特定任务的梯度代入 Frank-Wolfe 算法中, 求多目标优化解, 得到关于稀土多元素组分含量与浓度预测多个任务的Pareto平稳点  $\alpha^1$ ,

$\dots, \alpha^T$ ;

**步骤4** 利用 $\alpha^1, \dots, \alpha^T$ 更新共享层网络参数  
 $\theta^{\text{sh}} = \theta^{\text{sh}} - \eta \sum_{t=1}^T \alpha^t \nabla_{\theta^{\text{sh}}} \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)$ ;

**步骤5** 更新特定任务层网络参数 $\theta^t = \theta^t - \eta \nabla_{\theta^t} \hat{\mathcal{L}}^t(\theta^{\text{sh}}, \theta^t)$ , ( $t \in T$ ), 优化所有特定任务的下降方向;

**步骤6** 判断是否达到最大迭代次数, 是则进入步骤7, 否则返回步骤1;

**步骤7** 根据优化后的共享层网络参数 $\theta^{\text{sh}}$ 和特定层网络参数 $\theta^t$  ( $t \in T$ ), 对需检测的稀土混合萃取溶液图像中组分含量与浓度进行预测。

#### 4 对比实验及误差分析

为体现本文所提多目标优化算法对稀土多元素组分含量与浓度预测的有效性, 本节将在不同预测任务上从以下两个角度, 对单任务学习与不同多任务优化方法进行比较: 1) 验证集的损失函数值变化、训练过程单次迭代平均时长; 2) 10个测试集样本预测值与真实值的误差评价指标值。

##### 4.1 制备稀土混合萃取溶液图像数据集及实验环境

本文对基于多目标优化算法的稀土多元素组分含量与浓度预测以Pr(苹果绿色)、Nd(紫红色)两种具备离子颜色特征的稀土萃取溶液为例, 进行仿真实验, 所使用的Pr/Nd混合萃取溶液经由如下途径获得:

1) 原溶液组分含量测量: 于赣南某稀土公司分别购买纯度均为99.9%的1 L 1.8275 mol/L PrCl<sub>3</sub>和1 L 2.063 mol/L NdCl<sub>3</sub>萃取纯溶液, 其稀土浓度和配分均由国家钨与稀土产品质量监督检验中心提供。

2) 溶液稀释: 将两种元素原溶液的浓度分别稀释为0.01~0.50 mol/L各11种不同浓度且透光性良好的稀土萃取纯溶液。

3) 溶液混合: 将两种稀土元素不同浓度的溶液各50 ml分别相互混合, 共得到121组不同组分含量与浓度的Pr/Nd混合萃取溶液。

由上述步骤得到Pr/Nd各元素组分含量由1.96%~98.04%变化、浓度由0.005~0.25 mol/L变化的121组稀土混合萃取溶液, 并将各组分含量与浓度不同的稀土混合萃取溶液倒入收集皿中密封保存。由于实验室制备的稀土混合萃取溶液具备对光线呈现一定反射性和折射性的实际特点, 透光性较好, 满足基于机器视觉实验稀土元素组分含量软测量的光学成像条件。同时, 在充满稀土混合溶液的容器中形成的“离子颜色特征带”为采用快速、准确且可连续检测的图像识别技术提供可行途径。

实验图像采集时将稀土混合萃取溶液倒入至长宽

高为150 × 5 × 170 mm的石英容器中, 直至溶液充满容器。并随后将容器置于60 cm摄影棚中。其中, 摄影棚中搭建两处LED光源, 光源输出电压为24 V, 功率为48 W, 灯珠流明最大为15000 LM; 背景为纯白色; 图像采集设备为NIKON D700相机; 最终采集图像为4256 × 2832分辨率的JPG格式图片。

由于拍摄得到的Pr/Nd混合萃取溶液图像(图6)中包含石英容器边缘外非溶液部分及边缘内颜色不均匀部分, 因此, 将拍摄图像中充满颜色特征的部分进行裁剪, 每组溶液图像依照从上至下、从左至右的顺序裁剪10张图片, 共获取1210张颜色均匀的稀土混合萃取溶液图像。将稀土混合萃取溶液图像按照Pr, Nd各元素的组分含量与浓度不同归为121个类别, 并将每个类别分别标记上Pr, Nd组分含量与浓度的真实标签值, 将其中70%划分为训练集, 20%划分为验证集, 10%划分为预测集, 制备成完整数据集, 用于基于多任务学习的稀土多元素组分含量与浓度预测模型构建。

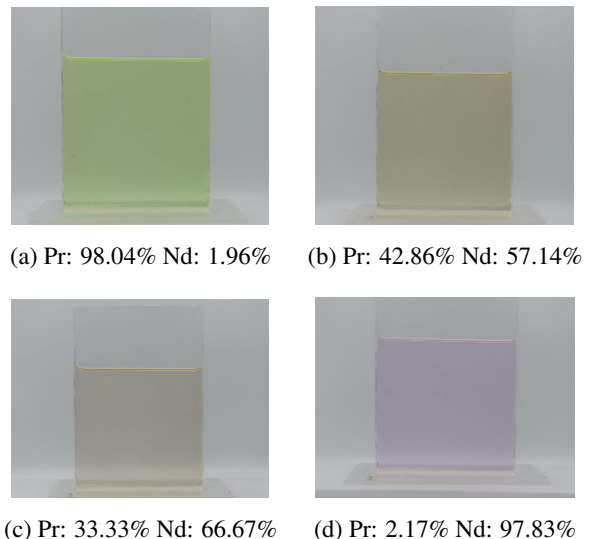


图6 部分不同组分含量Pr/Nd混合溶液图像

Fig. 6 Images of some mixed solutions of Pr/Nd with different component contents

所有模型训练及预测实验都在以下实验环境中实现: 硬件环境为Windows10操作系统, CPU Intel Core i7-12700F (12核), GPU RTX3080 (内存为10 G); 软件环境为在PyCharm上使用的PyTorch深度学习框架。

##### 4.2 对比实验

为验证本文所提基于多目标优化算法的稀土多元素组分含量与浓度预测方法有效性, 下面将本文所用多目标优化算法(基于优化上界的多重梯度下降)与单任务学习、多重梯度下降及其他多任务优化主流方法, 即梯度归一化、同方差不确定性, 从验证集的损失函数值变化、训练过程单次迭代平均时长的角度在以下多个预测任务上进行比较:

- 1) 双预测任务: 预测Pr, Nd组分含量(图7);
- 2) 三预测任务: 预测Pr, Nd组分含量和Pr浓度(图8);
- 3) 四预测任务: 预测Pr, Nd组分含量和Pr, Nd浓度(图9).

Nd浓度预测的验证集损失值较高, 可能是由于Pr, Nd组分含量与浓度两种预测任务的量级不一所导致.

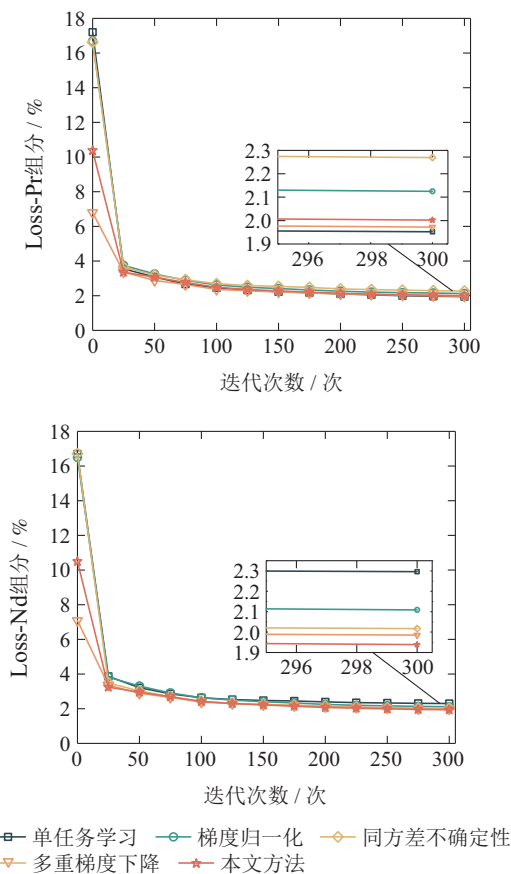


图 7 各方法在Pr, Nd 组分含量双预测任务中验证集损失函数值随迭代次数变化

Fig. 7 Variation of validation set loss function values with the number of iterations for each method in dual-prediction task of Pr and Nd component content

为保证对比实现的公平性, 本文所有多任务优化方法都在基于多任务学习的稀土多元组分含量与浓度预测模型中实现, 且将上述模型与单任务学习模型设置统一的模型主要参数如表2所示, 这些主要参数是提升模型整体性能的一般性参数. 因此, 这对于对比实验中所有优化方法都是公平的.

由图7-9可以看出本文方法在双、三、四预测任务中Nd组分含量预测任务的验证集损失函数值最低; 在三预测任务中Pr组分含量预测任务及四预测任务中Pr, Nd浓度预测任务中仅次于损失函数值最低的单任务学习; 在其他多预测任务中Pr组分、Pr, Nd浓度预测任务的损失函数值均与最低值接近, 且优于大部分优化方法. 而多重梯度下降在所有预测任务中也呈现出较优的性能, 损失函数值普遍低于其他主流多目标优化方法. 特别地, 梯度归一化在各多预测任务中对 Pr,

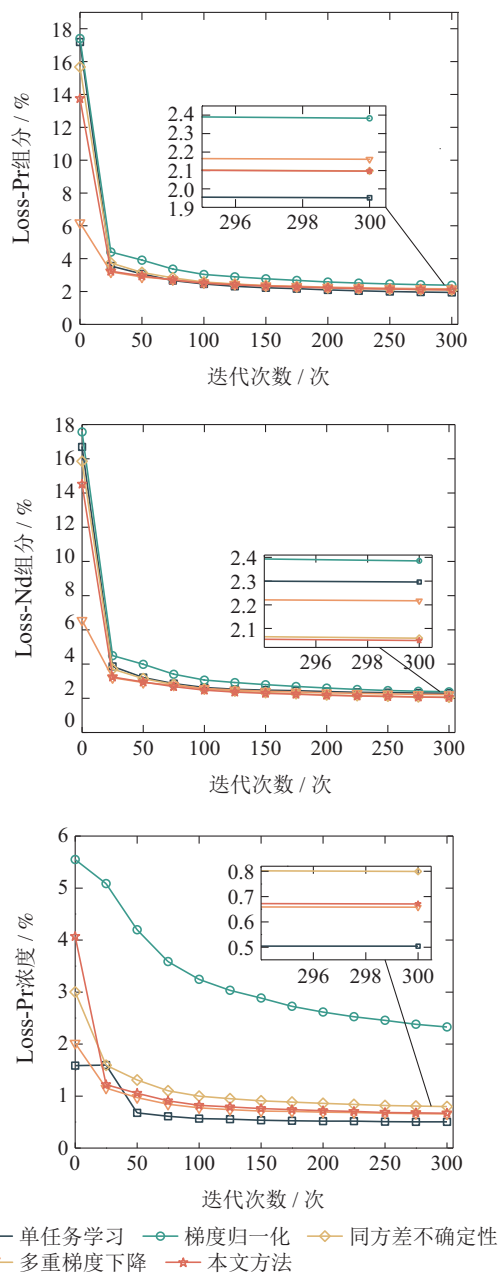
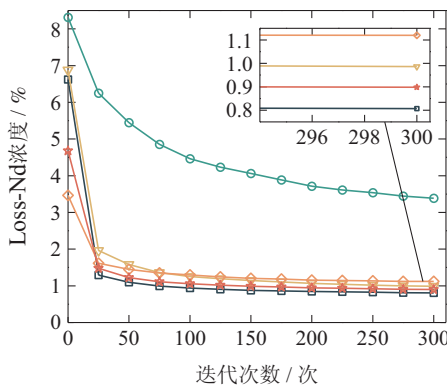
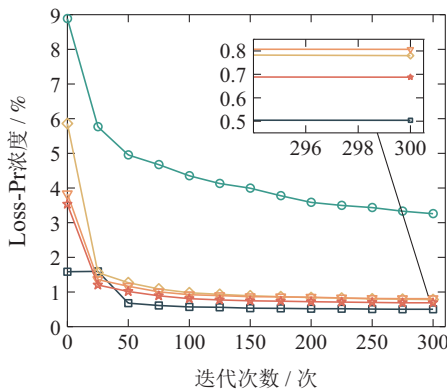
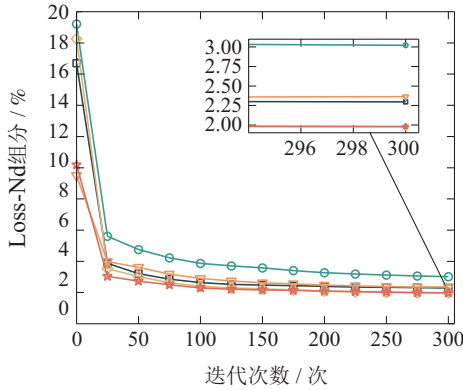
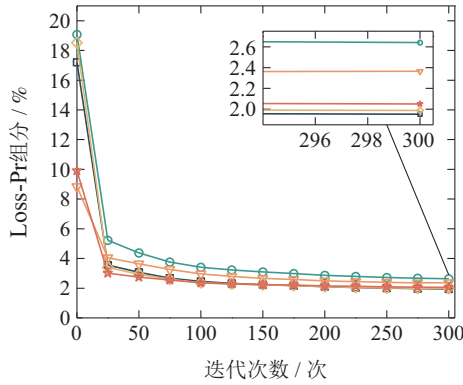


图 8 各方法在Pr, Nd 组分含量和Pr浓度三预测任务中验证集损失函数值随迭代次数变化

Fig. 8 Variation of validation set loss function values with the number of iterations for each method in the three-prediction task of Pr, Nd component content and Pr concentration

下面将各优化方法在多个任务训练中的单次迭代平均时长进行比较, 其中单任务学习训练中的单次迭代平均时长如表3, 不同多任务优化方法在多任务学习训练中的单次迭代平均时长如表4. 表3中各任务的单次迭代平均时长与表4中各方法同时对所有任务训练不同, 是由单任务学习模型对各任务进行独立训练所得.





—■— 单任务学习    —○— 梯度归一化    —◇— 同方差不确定性  
—▽— 多重梯度下降    —★— 本文方法

图9 各方法在Pr, Nd 组分含量和Pr, Nd浓度四预测任务中验证集损失函数值随迭代次数变化

Fig. 9 Variation of validation set loss function values with the number of iterations for each method in the four-prediction task of Pr and Nd component content and Pr and Nd concentration

表2 主要参数设置

Table 2 Main parameter settings

主要参数	参数设置
迭代次数	300
批大小	32
优化器	Adam
学习率	初始为 $1e-4$ , 每迭代30次减半

表3 单任务学习训练中的单次迭代平均时长

Table 3 Average time duration of single iteration in single-task learning training

优化方法	平均单次迭代时间/s			
	任务-Pr组分	任务-Nd组分	任务-Pr浓度	任务-Nd浓度
单任务学习	4.8720	4.8815	4.9512	4.9957

表4 不同多任务优化方法在多个预测任务训练中的单次迭代平均时长

Table 4 Average time of single iteration for different multi-task optimization methods in multiple prediction task training

多目标优化方法	平均单次迭代时间/s		
	双任务	三任务	四任务
梯度归一化	7.5722	7.6443	7.6755
同方差不确定性	4.9501	4.9493	4.9590
多重梯度下降	11.9842	15.7859	19.5779
本文方法	5.7961	5.7732	5.8732

由表3-4可以得知在多任务优化方法中同方差不确定性训练耗时最短, 与单任务学习训练时长接近; 多重梯度下降耗时最长, 且训练时间会随任务数量的增加而增长; 本文方法的训练时长仅高于同方差不确定性与单任务学习, 优于其他多目标优化方法。

综上可得本文方法在各多预测任务中损失函数值普遍较低, 在个别预测任务中损失值仅高于当前任务的最低值, 基本优于其他多任务优化方法; 本文方法在训练过程中迭代时长与耗时最低的单任务学习和同方差不确定性接近, 远远低于多重梯度下降方法, 且训练时长不会随着任务数量的增加而增长; 由此证明, 本文所提基于优化上界的多重梯度下降算法在训练时间上相较于多重梯度下降算法的较大优势, 在降低训练耗时的同时降低了各预测任务的损失函数值。

### 4.3 误差分析

为避免基于多任务学习的稀土多元素组分含量与浓度预测模型训练过拟合情况, 本小节将对测试集样

本进行仿真预测, 将预测值与真实值的误差进行比较与分析, 判断模型是否在稀土联动萃取过程中对稀土混合萃取图像中组分含量预测误差的最大容错范围内. 下面将本文所用多目标优化算法(基于优化上界的多重梯度下降)与单任务学习、多重梯度下降及其他多任务优化主流方法, 即: 梯度归一化、同方差不确定性, 从10个测试集样本预测值与真实值的误差评价指标角度在以下多个预测任务上进行比较:

- 1) 双预测任务: 预测Pr, Nd组分含量(图10、表5);
- 2) 三预测任务: 预测Pr, Nd组分含量和Pr浓度(图11、表6);
- 3) 四预测任务: 预测Pr, Nd组分含量和Pr, Nd浓度(图12、表7).

由图10-12可以从纵坐标明显看出各优化方法在对Pr, Nd组分含量预测任务中相对误差相差不大, 但在对Pr, Nd浓度预测任务中相对误差较大, 由其是梯度归一化的误差波动幅度较大, 该现象与梯度归一化在验证集损失函数值较高表现一致.

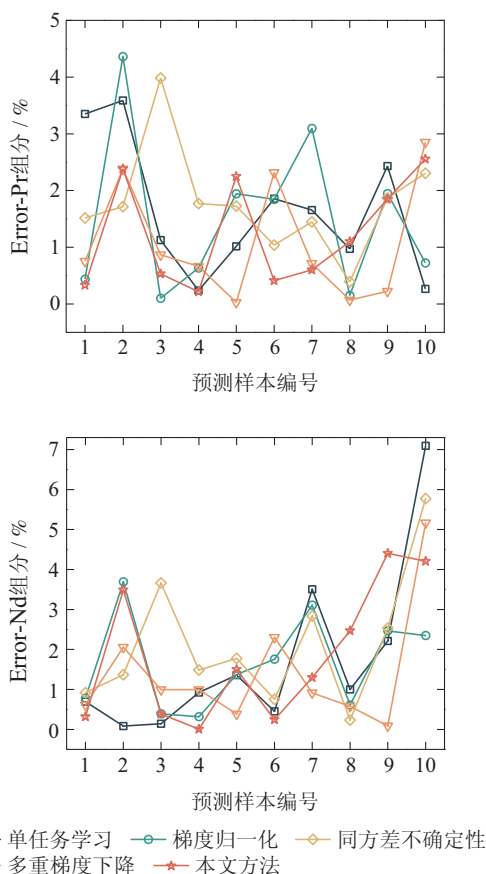


图 10 各方法在Pr, Nd组分含量双预测任务下对10个预测集样本预测值与真实值相对误差绝对值

Fig. 10 Absolute values of relative errors between predicted and true values for each method for 10 prediction set samples under dual-prediction task of Pr and Nd component content

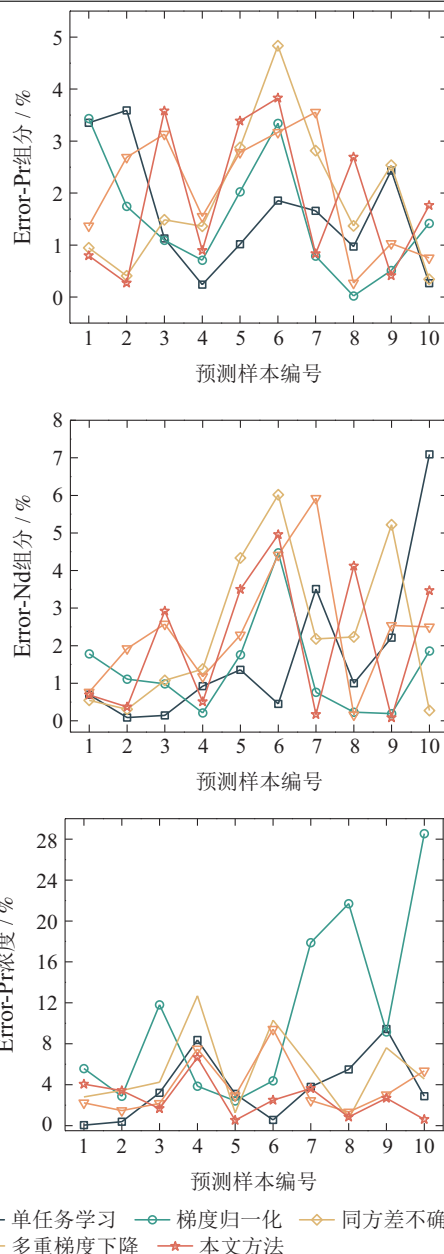
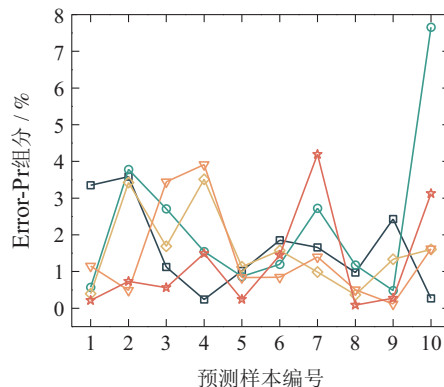


图 11 各方法在Pr, Nd组分含量和Pr浓度三预测任务下对10个预测集样本预测值与真实值相对误差绝对值

Fig. 11 Absolute values of relative errors between predicted and true values for each method for 10 prediction set samples under the three-prediction task of Pr, Nd component content and Pr concentration



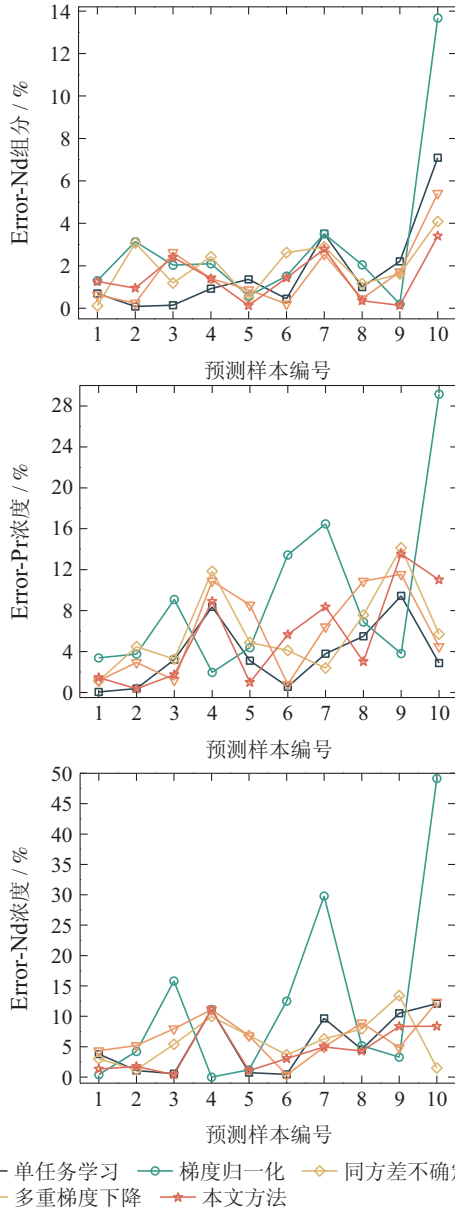


图 12 各方法在Pr, Nd组分含量和Pr, Nd浓度四预测任务下对10个预测集样本预测值与真实值相对误差绝对值

Fig. 12 Absolute values of relative errors between predicted and true values of samples from 10 prediction sets under four-prediction task of Pr and Nd component content and Pr and Nd concentration by each method

由表5-7可以看出如下6点:

1) 在双预测任务中, 将本文方法与单任务学习比较: 本文方法在Pr组分预测任务的平均相对误差、均方根误差、最大相对误差绝对值与Nd组分最大相对误差绝对值都低于单任务学习, 在Nd组分预测任务的平均相对误差、均方根误差与单任务学习接近, 整体优于单任务学习; 与其他多任务优化方法比较: 本文方法在Pr组分含量预测任务的平均相对误差、均方根误差仅次于多重梯度下降, Pr组分预测任务的最大相对误差值为所有优化方法中最低, Nd组分预测任务的最大相对误差值仅低于梯度归一化. 因此, 本文方

法在双预测任务中整体性能仅次于多重梯度下降.

2) 在三预测任务中, 本文方法在Pr浓度预测任务的平均相对误差、均方根误差、最大相对误差绝对值均为所有优化方法中最低. 将本文方法与单任务学习比较: 除上述最低外其他误差评价指标均高于单任务学习, 整体次于单任务学习; 将本文方法与其他多任务优化方法比较: 本文方法在除上述最低外, 其余误差评价指标整体仅次于梯度归一化, 优于其他多任务优化方法, 但梯度归一化在Pr浓度预测中误差较高, 不适用于在量级不一的任务中进行训练及预测. 因此, 本文方法在三预测任务中整体性能仅次于单任务学习.

3) 在四预测任务中, 本文方法在Nd组分、Nd浓度预测任务的平均相对误差、均方根误差、最大相对误差绝对值与Pr组分预测任务中的平均相对误差均低于其他优化方法, 整体性能较优; 将本文方法与单任务学习比较: 除上述最低外, 其余误差评价指标值均高于单任务学习, 但整体性能优于单任务学习; 与其他多任务优化方法比较, 除上述最低外, 在Pr浓度预测任务的平均相对误差、均方根误差均低于其他优化方法. 因此, 在四预测任务中, 本文方法整体性能优于其他优化方法.

4) 将单任务学习与多任务优化方法比较: 单任务学习仅在四预测任务中Pr浓度预测任务的各项性能评价指标最低, 其余指标在所有预测任务中均高于多任务优化方法的最低值, 说明多任务学习相较于单任务学习, 能找到组分与浓度各任务中的相关性, 共同提升模型整体性能.

5) 将多个预测任务之间的结果进行比较, 首先, 将双、三、四预测任务中的Pr, Nd组分预测任务误差评价指标值进行比较, 可以发现三预测任务中的各Pr, Nd组分预测任务误差均高于双、四预测任务, 说明只有将各元素组分同时训练及预测或将各元素组分与浓度同时训练及预测对提高组分含量预测精度有效; 其次, 将三、四预测任务中的Pr浓度预测任务进行比较, 可以发现三预测任务中Pr浓度预测任务的误差评价指标值普遍低于四预测任务, 因此, 将元素组分与浓度两种任务同时训练及预测只能提升组分预测的精度, 并不能提升浓度预测精度; 最后, 将双、四预测任务中的组分预测任务进行比较, 可以发现双预测任务的各误差评价指标比四预测任务稍低, 但基本接近, 说明将组分任务同时训练及预测与将组分与浓度两种任务同时训练及预测对组分预测精度的提升不明显.

6) 将本文方法在各多预测任务的误差评价指标值中进行比较, 可以得出与第5点一致的结论, 而本文方法在双、三、四中的Pr, Nd组分最大相对误差绝对

值依次为, 双任务: 2.5549%, 4.4033%; 三任务: 3.8298%, 4.9564%; 四任务: 4.1913%, 3.4028%, 均满足 ±5%<sup>[9]</sup> 以内的要求。

表 5 各优化方法在双预测任务下对 10 个测试集样本预测值与真实值的误差评价指标值

Table 5 Error evaluation index values of each optimization method for the predicted and true values of the 10 test set samples under dual-prediction task

优化方法	平均相对误差/%		均方根误差		最大相对误差绝对值/%	
	Pr组分	Nd组分	Pr组分	Nd组分	Pr组分	Nd组分
单任务学习	1.6500	1.7466	0.0095	0.0098	3.5883	7.0937
梯度归一化	1.5250	1.6826	0.0101	0.0097	4.3641	<b>3.6881</b>
同方差不确定性	1.7769	2.1345	0.0102	0.0112	3.9852	5.7753
多重梯度下降	<b>1.0878</b>	<b>1.4087</b>	<b>0.0083</b>	<b>0.0080</b>	2.8520	5.1684
本文方法	1.2232	1.8313	0.0084	0.0105	<b>2.5549</b>	4.4033

表 6 各优化方法在三预测任务下对 10 个测试集样本预测值与真实值的误差评价指标值

Table 6 Error evaluation index values of each optimization method for the predicted and true values of the 10 test set samples under three-prediction task

优化方法	平均相对误差/%			均方根误差			最大相对误差绝对值/%		
	Pr组分	Pr组分	Pr浓度	Pr组分	Pr组分	Pr浓度	Pr组分	Pr组分	Pr浓度
单任务学习	1.6500	1.7466	3.7233	0.0095	0.0098	0.0090	3.5883	7.0937	9.4410
梯度归一化	<b>1.5089</b>	<b>1.3362</b>	10.8166	<b>0.0092</b>	<b>0.0083</b>	0.0186	<b>3.4356</b>	<b>4.4730</b>	28.5365
同方差不确定性	1.8979	2.3569	5.3189	0.0128	0.0135	0.0120	4.8321	6.0180	12.6839
多重梯度下降	2.0301	2.4259	3.7813	0.0119	0.0129	0.0084	3.5557	5.9204	9.4275
本文方法	1.8454	2.0778	<b>2.6511</b>	0.0122	0.0121	<b>0.0054</b>	3.8298	4.9564	<b>6.6489</b>

表 7 各优化方法在四预测任务下对 10 个测试集样本预测值与真实值的误差评价指标值

Table 7 Error evaluation index values of each optimization method for the predicted and true values of the 10 test set samples under four-prediction task

优化方法	平均相对误差/%		均方根误差		最大相对误差绝对值/%		平均相对误差/%		均方根误差		最大相对误差绝对值/%	
	Pr组分	Nd组分	Pr组分	Nd组分	Pr组分	Nd组分	Pr浓度	Nd浓度	Pr浓度	Nd浓度	Pr浓度	Nd浓度
单任务学习	1.6500	1.7466	0.0095	0.0098	3.5883	7.0937	<b>3.7233</b>	5.4495	<b>0.0090</b>	0.0099	<b>9.4410</b>	12.0852
梯度归一化	2.2715	3.0090	0.0185	0.0175	7.6585	13.6795	9.2327	12.1586	0.0154	0.0153	29.1582	49.1503
同方差不确定性	1.6017	1.9650	0.0093	0.0103	<b>3.5122</b>	4.0896	5.9410	5.9103	0.0134	0.0112	14.1193	13.4016
多重梯度下降	1.4311	1.6114	<b>0.0092</b>	0.0089	3.9119	5.4141	5.8832	6.6867	0.0132	0.0117	11.5413	12.3623
本文方法	<b>1.2370</b>	<b>1.4235</b>	0.0107	<b>0.0079</b>	4.1913	<b>3.4028</b>	5.4961	<b>4.4585</b>	0.0119	<b>0.0091</b>	13.5243	<b>11.0931</b>

### 5 结论

针对稀土联动萃取流程对多稀土元素组分含量预测的需求, 以及探索多元素组分含量间及组分含量与浓度间的共性, 本文首先提出基于多任务学习的稀土多元素组分含量与浓度预测模型用于多个任务的同时训练及预测; 接着, 为防止多个任务在模型中竞争与冲突的问题, 提出一种稀土多元素组分含量与浓度预测多目标优化算法—基于优化上界的多重梯度下降算法, 通过此算法搜索模型中各任务间的Pareto最优, 提高预测精度; 最后, 从验证集的损失函数值变化趋

势、训练过程单次迭代平均时长与各优化方法在 10 个测试集样本预测值与真实值的误差评价指标, 这两个角度在多个预测任务中呈现出的性能进行多组对比实验, 得到以下结论:

- 1) 单任务学习在验证集损失函数值上较低, 但是在仿真验证中并未得到良好的效果, 出现了过拟合现象, 而多任务学习通过多个任务的联合训练, 在仿真实验中取得了优异的效果, 证明多任务学习能够找到多个任务之间的共性, 提升模型整体泛化力和鲁棒性;
- 2) 本文方法在训练过程中单次迭代平均时长与耗

时最短的单任务学习和同方差不确定性接近, 远远低于多重梯度下降方法, 且训练时长不会随着任务数量的增加而增长, 在多任务优化方法中具有训练耗时短的优势, 弥补了双预测任务中整体误差高于多重梯度下降的不足; 而本文方法在三预测任务中整体性能仅次于单任务学习; 在四预测任务中整体性能优于其他优化方法. 因此, 本文方法相较其他优化方法, 在训练时长和预测精度上整体占优, 更适用于稀土多元素组分含量与浓度预测;

3) 将稀土元素组分含量与浓度两种任务联合训练只能提升组分含量的预测精度, 并不能明显提升浓度的预测精度; 而四预测任务(Pr, Nd组分含量, Pr, Nd浓度预测)仅比双预测任务(Pr, Nd组分含量预测)的组分含量预测精度稍高, 不能明显提升组分含量的预测精度. 由此说明组分含量与浓度两种任务共性不足, 联合训练效果欠佳; 而多种元素的组分含量之间具有良好的共性, 其联合训练能够共同提升各元素组分含量的预测精度;

4) 本文方法在各预测任务中对组分含量预测的最大相对误差均在稀土萃取生产中对组分含量检测最大相对误差在 $\pm 5\%$ 要求以内, 预测时间成本在3 s以内, 能满足稀土联动萃取过程中对多元素组分含量在线软测量的精确性与实时性, 为稀土元素组分含量软测量方法提供了新思路.

## 参考文献:

- [1] FENG Zongyu, WANG Meng, ZHAO Longsheng, et al. Current status and outlook of the development of rare earth element extraction and separation purification technology. *Chinese Journal of Rare Earths*, 2021, 39(3): 469 – 478.  
(冯宗玉, 王猛, 赵龙胜, 等. 稀土元素萃取分离提纯技术发展现状与展望. 中国稀土学报, 2021, 39(3): 469 – 478.)
- [2] ZENG Xiaoqin, CHEN Yiwen, WANG Jingya, et al. New advances in the research of high performance rare earth magnesium alloys. *Chinese Journal of Nonferrous Metals*, 2021, 31(11): 2963 – 2975.  
(曾小勤, 陈义文, 王静雅, 等. 高性能稀土镁合金研究新进展. 中国有色金属学报, 2021, 31(11): 2963 – 2975.)
- [3] LIU T, CHEN J. Extraction and separation of heavy rare earth elements: A review. *Separation and Purification Technology*, 2021, 276: 119263.
- [4] ASADOLLAHZADEH M, TORKAMAN R, TORAB-MOSTAEDI M. Extraction and separation of rare earth elements by adsorption approaches: Current status and future trends. *Separation & Purification Reviews*, 2021, 50(4): 417 – 444.
- [5] DEBLONDE G J P, MATTOCKS J A, PARK D M, et al. Selective and efficient biomacromolecular extraction of rare-earth elements using lanmodulin. *Inorganic Chemistry*, 2020, 59(17): 11855 – 11867.
- [6] LIAO Chunsheng, CHENG Fuxiang, WU Sheng, et al. Development trends and related advances in string-level extraction theory and rare earth separation technology. *Chinese Science: Chemistry*, 2020, 50(11): 1730 – 1736.  
(廖春生, 程福祥, 吴声, 等. 串级萃取理论和稀土分离技术的发展趋势及相关进展. 中国科学: 化学, 2020, 50(11): 1730 – 1736.)
- [7] CHAI Tianyou, YANG Hui. Research status and development trend of automatic control of rare earth extraction and separation process. *Chinese Journal of Rare Earths*, 2004, 22(4): 427 – 433.  
(柴天佑, 杨辉. 稀土萃取分离过程自动控制研究现状及发展趋势. 中国稀土学报, 2004, 22(4): 427 – 433.)
- [8] YANG H, XU Y, WANG X. Component content soft-sensor based on RBF neural network in rare earth countercurrent extraction process. *The 6th World Congress on Intelligent Control and Automation*. Dalian, China: IEEE, 2006, 1: 4909 – 4912.
- [9] LU Rongxiu, HE Quanheng, YANG Hui, et al. Multi-component content prediction of rare earth mixed solutions based on GA-ELM. *Computer Engineering*, 2021, 47(1): 284 – 290, 297.  
(陆荣秀, 何权恒, 杨辉, 等. 基于GA-ELM的稀土混合溶液多组分含量预测. 计算机工程, 2021, 47(1): 284 – 290, 297.)
- [10] LU Rongxiu, RAO Yunchun, YANG Hui, et al. Prediction of praseodymium/neodymium component content based on improved instantaneous learning algorithm. *Control Theory & Applications*, 2020, 37(8): 1846 – 1854.  
(陆荣秀, 饶运春, 杨辉, 等. 基于改进即时学习算法的镨/钕元素组分含量预测. 控制理论与应用, 2020, 37(8): 1846 – 1854.)
- [11] YIN Gang, CHEN Gen, HE Wen, et al. Anode effect prediction for 300 kA aluminum electrolyzer based on deep learning. *Chinese Journal of Nonferrous Metals*, 2021, 31(1): 161 – 170.  
(尹刚, 陈根, 何文, 等. 基于深度学习的300 kA铝电解槽阳极效应预测. 中国有色金属学报, 2021, 31(1): 161 – 170.)
- [12] WANG Liguan, CHEN Sijia, JIA Mingtao, et al. Deep learning-based image recognition method for wolframite ore beneficiation. *Chinese Journal of Nonferrous Metals*, 2020, 30(5): 1192 – 1201.  
(王李管, 陈斯佳, 贾明滔, 等. 基于深度学习的黑钨矿图像识别选矿方法. 中国有色金属学报, 2020, 30(5): 1192 – 1201.)
- [13] NIE Shanyu, HE Guichun, SHI Yan, et al. Research progress on data-driven predictive modeling of flotation metrics. *The Chinese Journal of Nonferrous Metals*, 2023, 33(7): 2330 – 2338.  
(聂善煜, 何桂春, 石岩, 等. 基于数据驱动的浮选指标预测建模研究进展. 中国有色金属学报, 2023, 33(7): 2330 – 2338.)
- [14] O'SHEA K, NASH R. An introduction to convolutional neural networks. *ArXiv Preprint*, 2015, arXiv: 1511.08458.
- [15] COLLOBERT R, WESTON J. A unified architecture for natural language processing: Deep neural networks with multitask learning. *Proceedings of the 25th International Conference on Machine Learning*. Helsinki, Finland: ACM, 2008: 160 – 167.
- [16] VANDENHENDE S, GEORGIOULIS S, VAN GANSBEKE W, et al. Multi-task learning for dense prediction tasks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(7): 3614 – 3633.
- [17] KOKKINOS I. Ubernet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE, 2017: 6129 – 6138.
- [18] TEICHMANN M, WEBER M, ZOELLNER M, et al. Multinet: Real-time joint semantic reasoning for autonomous driving. *Intelligent Vehicles Symposium (IV)*. Changshu, China: IEEE, 2018: 1013 – 1020.
- [19] RUDER S, BINGEL J, AUGENSTEIN I, et al. Latent multi-task architecture learning. *Proceedings of the AAAI Conference on Artificial Intelligence*. Honolulu, HI: AAAI, 2019, 33(1): 4822 – 4829.
- [20] LIU S, JOHNS E, DAVISON A J. End-to-end multi-task learning with attention. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA: IEEE, 2019: 1871 – 1880.
- [21] CHEN Z, BADRINARAYANAN V, LEE C Y, et al. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. *International Conference on Machine Learning*. Stockholm, Sweden: PMLR, 2018: 794 – 803.

- [22] KENDALL A, GAL Y, CIPOLLA R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT: IEEE, 2018: 7482 – 7491.
- [23] DÉSIDÉRI J A. Multiple-gradient descent algorithm (MGDA) for multiobjective optimization. *Comptes Rendus Mathématique*, 2012, 350(5/6): 313 – 318.
- [24] SENER O, KOLTUN V. Multi-task learning as multi-objective optimization. *ArXiv Preprint*, 2019, arXiv:1810.04650v2.
- [25] ZHANG Bo, XU Liming, HUANG Zhiwei, et al. Pareto optimal solution algorithm for multi-objective GANs with gradient strategy. *Computer Engineering and Applications*, 2021, 57(9): 89 – 95.  
(张波, 徐黎明, 黄志伟, 等. 梯度策略的多目标GANs帕累托最优解算法. 计算机工程与应用, 2021, 57(9): 89 – 95.)
- [26] ZHOU Xiaojun, GAO Yuan, LI Chaojie, et al. An end-to-end license plate recognition method based on multi-objective optimization with multi-task learning. *Control Theory & Applications*, 2021, 38(5): 676 – 688.  
(周晓君, 高媛, 李超杰, 等. 基于多目标优化多任务学习的端到端车牌识别方法. 控制理论与应用, 2021, 38(5): 676 – 688.)
- [27] ZHAO Jiaming, ZHOU Xiaojun, GAO Yuan, et al. Multi-objective learning-based identification of working conditions in froth flotation process. *Journal of Central South University (Natural Science Edition)*, 2022, 53(6): 2071 – 2079.  
(赵佳铭, 周晓君, 高媛, 等. 基于多目标学习的泡沫浮选过程工况识别. 中南大学学报(自然科学版), 2022, 53(6): 2071 – 2079.)
- [28] YU X, WANG S H. Abnormality diagnosis in mammograms by transfer learning based on ResNet18. *Fundamenta Informaticae*, 2019, 168(2/4): 219 – 230.
- [29] WANG M, LU S, ZHU D, et al. A high-speed and low-complexity architecture for softmax function in deep learning. *The Asia Pacific Conference on Circuits and Systems (APCCAS)*. Chengdu, China: IEEE, 2018: 223 – 226.
- [30] ZHANG Shuiping, ZHANG Qihan, WANG Bi, et al. Deep machine vision-based content prediction of rare earth element components. *Nonferrous Metal Science and Engineering*, 2023, 14(4): 587 – 596.  
(张水平, 张奇涵, 王碧, 等. 基于深度机器视觉的稀土元素组分含量预测. 有色金属科学与工程, 2023, 14(4): 587 – 596.)
- [31] JAGGI M. Revisiting frank-wolfe: Projection-free sparse convex optimization. *International Conference on Machine Learning*. Atlanta, GA, USA: The MIT Press, 2013: 427 – 435.

### 作者简介:

**张水平** 教授, 硕士生导师, 目前研究方向为智能矿山、智能计算,

E-mail: 870208231@qq.com;

**张奇涵** 硕士研究生, 目前研究方向为深度学习、多任务学习,

E-mail: zhangqihan0722@163.com;

**王碧** 博士, 目前研究方向为强化学习, E-mail: wangbi@jxust.edu.cn.