

一种基于鲁棒随机向量函数链接网络的磨矿粒度集成建模方法

李德鹏¹⁾, 代伟^{1 2)}✉, 赵大勇³⁾, 黄罡¹⁾, 马小平¹⁾

1) 中国矿业大学信息与控制工程学院, 徐州 221116 2) 东北大学流程工业综合自动化国家重点实验室, 沈阳 110819

3) 中国科学院沈阳自动化研究所, 沈阳 110016

✉ 通信作者, E-mail: weidai@cumt.edu.cn

摘要 作为磨矿过程的主要生产质量指标, 磨矿粒度是实现磨矿过程闭环优化控制的关键。将磨矿粒度控制在一定范围内能够提高选别作业的精矿品位和有用矿物的回收率, 并减少有用矿物的金属流失。由于经济和技术上的限制, 磨矿粒度的实时测量难以实现。因此, 磨矿粒度的在线估计显得尤为重要。然而, 目前我国所处理的铁矿石大多数为性质不稳定的赤铁矿, 其矿浆颗粒存在磁团聚现象, 所采集的数据存在大量异常值, 使得利用数据建立的磨矿粒度模型存在较大误差。同时, 传统前馈神经网络在磨矿粒度数据建模过程中存在收敛速度慢、易于陷入局部最小值等缺点, 且单一模型泛化性能较差, 现有的集成学习在异常值干扰下性能严重下降。因此, 本文在改进的随机向量函数链接网络(random vector functional link networks, RV-FLN)的基础上, 将Bagging算法与自适应加权数据融合技术相结合, 提出一种基于鲁棒随机向量函数链接网络的集成建模方法, 用于磨矿粒度集成建模。所提方法首先通过基准回归问题进行了实验研究, 然后采用磨矿工业实际数据进行验证, 表明其有效性。

关键词 磨矿粒度; 随机向量函数链接网络; 集成学习; 鲁棒性; 数据融合

分类号 TP18

Grinding process particle size modeling method using robust RVFLN-based ensemble learning

LI De-peng¹⁾, DAI Wei^{1 2)}✉, ZHAO Da-yong³⁾, HUANG Gang¹⁾, MA Xiao-ping¹⁾

1) School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

2) State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110819, China

3) Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

✉ Corresponding author, E-mail: weidai@cumt.edu.cn

ABSTRACT As a key production quality index of grinding process, particle size is of great importance to closed-loop optimization and control. This is because controlling particle within a proper range can improve the concentrate grade, enhance the recovery rate of useful minerals, and reduce the loss of metal in the sorting operation; thus, the particle size determines the overall performance of the grinding process. In fact, it is not easy to optimize or control the practical industrial process because the optimal operation largely depends on a good measurement of particle size of grinding process; however, it is difficult to realize the real-time measurement of particle size because of limitations of economy or technique. Employing soft sensor techniques is necessary to solve the problem of particle size estimation, which is particularly important for the actual grinding processes. Considering that soft sensors are applicable in many fields,

收稿日期: 2018-07-07

基金项目: 国家自然科学基金青年资助项目(61603393, 61741318); 江苏省自然科学基金青年基金资助项目(BK20160275); 中国博士后科学基金资助项目(2015M581885, 2018T110571); 流程工业综合自动化国家重点实验室开放课题资助项目(PAL-N201706); 江苏省研究生科研与实践创新计划资助项目(SJCX17_0524)

the data-driven soft sensor will be a useful tool for achieving particle size estimation. However, most of the iron ores processed in China are characterized by hematite with unstable properties, and the slurry particles exhibit magnetic agglomeration, giving rise to a large number of outliers in the collected data. In this case, there are gross errors in the particle size estimation model constructed based on the data and thus unreliable measurements. Meanwhile, the traditional feedforward neural networks have the disadvantages of slow convergence speed and easily fall into local minimum during the prediction process. A single model tends to lack superiority in sound generalization, and the performance of existing ensemble learning methods will be worse under outlier interference. Therefore, in this study, based on the improved random vector functional link networks (RVFLN), the Bagging algorithm is incorporated into an adaptive weighted data fusion technique to develop an ensemble learning method for particle size estimation of grinding processes. Experimental studies were first conducted through benchmark regression issues and then validated by the samples collected from an actual grinding process, indicating the effectiveness of the proposed method.

KEY WORDS particle size; random vector functional link networks; ensemble learning; robustness; data fusion

磨矿过程作为矿石选矿过程中最为关键的工序起着承上启下的重要作用。该过程主要是将矿石原料粉碎到合适的粒度尺寸,从而能够将有用矿物与脉石单体解离,或者将不同有用矿物相互解离,为后续选别作业提供原料。其中,磨矿粒度是磨矿过程中表征生产质量的关键运行指标。因此,在磨矿过程中,实现磨矿粒度的检测具有重要意义。粒度的闭环优化控制是矿物加工行业多年来的研究热点^[1-2]。不幸的是,在线粒度分析仪对粒度的实时测量往往受到磨矿过程中矿石磁团聚特性的影响,导致测量的不稳定和不准确。因此,在缺乏可靠测量的情况下实施闭环优化控制通常很困难,甚至是不可行的。为了解决这个问题,粒度估计的相关研究被广泛开展。传统方法是基于一些理想假设的第一原理模型^[3-4],往往导致较大的估计误差。因此,采用数据驱动的建模技术来解决粒度估计问题已被广泛接受^[5-7]。

单隐含层前馈网络(single-hidden layer feedforward networks, SLFN)由于具有逼近能力,被广泛应用于解决回归、分类等问题^[8-9]。传统单层前馈神经网络的训练基于反向传播(back propagation, BP)算法^[10]。BP算法由于自身的迭代过程导致收敛速度慢,并易于陷入局部最小值等缺点,其速度和精度难以满足在线应用的要求。为了解决这个问题,文献[11]提出了一种随机向量函数链接网络(random vector functional link networks, RVFLN)。Igel'nik和Pao^[8]证明了随机向量函数链接网络的随机参数(输入权值和偏置)产生于一个均匀分布的范围内时,该网络可以逼近连续函数。文献[12]表明,与传统的单层前馈神经网络相比,随机向量函数链接网络学习速度快、更加易于实现等。因此,可以使用随机向量函数链接网络来建立磨矿粒度估计模型。

研究表明,与单一模型相比,在绝大多数情况下通过对多个模型集成可以显著提高原来模型的精度

与泛化性能^[13]。即单个随机向量函数链接网络模型之间往往会存在模型参数和样本选取上的差异,而集成随机向量函数链接网络可以有效提高模型的预测精度与泛化能力,增强算法的稳定性。集成学习是一种新的机器学习范式,通过构建并结合多个基模型来完成学习任务。1988年,Kearns与Valiant^[14]首先提出了将多个弱学习器提升为强学习器的设想。1990年,Schapire^[15]证明了多个弱学习器可以构成一个强学习器。常用的集成学习策略有:Boosting^[16]、Bagging^[17]等。然而,实际工业数据中往往包含异常值。尽管集成学习通过组合多个基模型在一定程度上可以提高模型的鲁棒性,但是当异常值水平较高时,仍然会导致不准确的估计,使得传统的集成学习在异常值干扰下性能下降。因此,如何提高集成建模的鲁棒性是目前亟待解决的问题。本文将鲁棒随机向量函数链接网络作为基模型,根据Bagging的采样方式,以自适应加权数据融合技术作为基模型的组合方法,从而提出一种基于鲁棒随机向量函数链接网络的集成学习方法用于磨矿粒度估计。

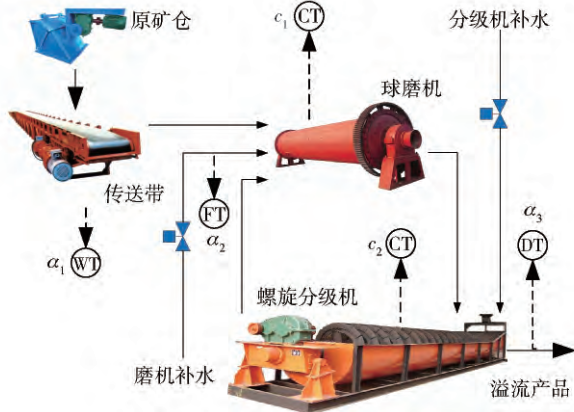
本文对集成建模的研究首先从其基模型的角度出发,针对如何提高所集成基模型的鲁棒性进行深入研究,并力图获得训练速度快、差异性大、精确度高、泛化能力好的基模型,从而保证最终的集成模型可以拥有良好的性能。然后,通过基准回归问题和磨矿生产过程的实际工业数据,验证了所提集成建模方法的有效性。

1 磨矿过程与特性分析

1.1 磨矿过程工艺描述

典型一段磨矿过程如图1所示,主要由原矿仓、传送带、球磨机、螺旋分级机以及若干仪表和电动执行机构组成。首先,原矿仓中的矿石经皮带输送至球磨机进行研磨,同时在球磨机入口加入与给矿量

成比例的水,以稳定球磨机内部矿浆浓度,保证球磨机保持最佳工作状态。然后,矿浆经球磨机出口处排出送入螺旋分级机,在叶片搅拌作用下分成两类矿浆,即溢流矿浆和底流矿浆。分级机补水用于调节分级浓度以使溢流矿浆保持适当比例的粒度。含有较大粒度的底流矿浆返回至球磨机中再研磨,而含有较小粒度的溢流矿浆由螺旋分级机下端溢流口流出,形成磨矿产品以供后续的选别作业。



α_1 —球磨机给矿量, $t \cdot h^{-1}$; α_2 —球磨机入口给水量, $m^3 \cdot h^{-1}$; α_3 —螺旋分级机溢流质量分数; c_1 —球磨机电流, A; c_2 —螺旋分级机电流, A; T—检测仪; W—质量; F—流量; C—电流; D—密度

图1 磨矿过程工艺流程

Fig. 1 Flow diagram of grinding process

通常,磨矿粒度过大,容易使已单体解离的有用矿物过粉碎,难以回收,且增加能耗;而矿石粒度过小,矿物不能达到单体解离,难以分选。因此,选别作业要求过粗和微细粒度均要少,而中间易选粒级要多。在实际磨矿过程中,通常采用磨矿粒度 r 作为评价磨矿产品质量的重要工艺指标。所谓磨矿粒度,是指矿石产物中直径小于 0.074 mm 的颗粒在该产物的占比。

1.2 磨矿过程特性分析

为提高选别作业的精矿品位和有用矿物的回收率,减少有用矿物的金属流失,要求磨矿粒度 r 控制在一定范围内,然而,由于我国所处理的铁矿石大多数为赤铁矿,其矿石性质不稳定,矿浆颗粒存在磁团聚现象,难以采用在线粒度检测仪表实现磨矿粒度指标的准确测量,只能靠实验室人工化验的方法获得,无法满足磨矿粒度闭环优化控制的要求,为此需要采集过程数据,利用影响因素间接估计磨矿粒度 r ,建立其软测量模型。影响磨矿粒度 r 的因素包括磨机给矿量 α_1 ,磨机入口给水流量 α_2 ,分级机溢流质量分数 α_3 等操作参数,以及矿石粒度 B_1 和可磨性 B_2 等矿石性质,各影响因素具体分析如下。

(1) 磨机给矿量 α_1 对磨矿粒度 r 的影响。

在磨矿粒度 r 合格的前提下,通常希望磨机给矿量 α_1 越高越好,但单独增加磨机给矿量 α_1 将降低矿粒与钢球碰撞并破裂的几率。由粉磨动力学方程,即

$$R_t = R_0 \exp(-kt^\mu) \quad (1)$$

可知,由于 k 是与矿粒破裂几率相关的参数, μ 是与矿石性质相关的参数, t 是研磨时间,因此对于粗粒级 R_0 一定的矿石, k 变小,则从磨机出口排出的粒度将随粗粒级 R_t 的增加而变粗。此时,分级机返砂量也随之增多,导致进入磨机的矿浆总量增加,加快矿浆在磨机内的流动速度,从而减少矿物的研磨时间 t 。而 t 越小, R_t 越大,磨矿粒度 r 将变大。

(2) 磨机入口给水流量 α_2 对磨矿粒度 r 的影响。

磨矿过程是由矿物的破裂和物料运输两个相互制约的子过程组合而成的,磨矿浓度通过影响磨矿黏度来同时影响这两个过程的参数,其主要通过磨机入口给水流量 α_2 来调节。图2给出了磨矿速率和磨矿浓度的关系,即分别在磨矿低浓度A区,中浓度B区和高浓度C区,由于钢球周围黏着的矿粒量不同,磨矿速率也不同,且成非线性变化。此外,当磨矿浓度改变时,矿浆流速变化,磨矿时间 t 随之变化,由粉磨动力学方程(1)可知,磨矿粒度 r 将发生改变。磨矿浓度与相对黏度呈非线性关系,而且随着磨矿粒度的变化而变化。因此磨机入口给水流量 α_2 与磨矿粒度 r 之间具有复杂非线性特性。

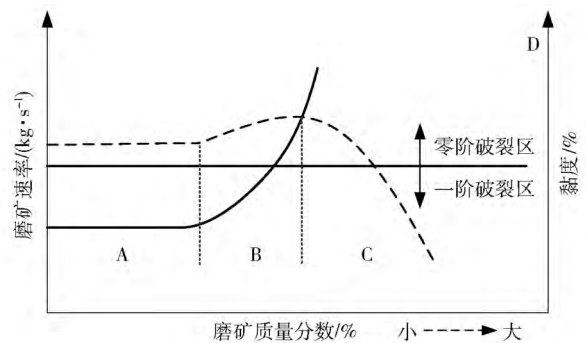


图2 批次磨矿试验下的磨矿速率与磨矿浓度关系

Fig. 2 Relationship between grinding rate and mill density based on batch grinding experiment

(3) 分级机溢流质量分数 α_3 对磨矿粒度 r 的影响。

由重力选矿原理可知,分级机是按不同粒级的矿粒在流体中沉降速率的差异而进行分级的,而沉降速率直接影响着分级机的分级效率,由于分级机矿浆黏度是干涉沉降速率的主要因素,因此决定矿

浆黏度的溢流质量分数 α_3 是影响磨矿粒度 r 的关键过程变量,图 3 给出了分级机溢流质量分数与磨矿粒度 r 之间的非线性关系.

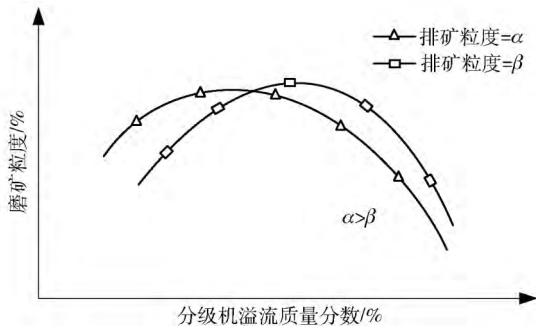


图 3 分级机溢流质量分数与磨矿粒度的关系

Fig. 3 Relationship between classifier overflow concentration and particle size

(4) 矿石粒度 B_1 和可磨性 B_2 对磨矿粒度 r 的影响.

矿石粒度 B_1 表示原矿中所有矿粒大小的总体分布. 在相同的操作参数下,磨矿粒度 r 与矿石粒度 B_1 成正相关的变化趋势,即矿石粒度 B_1 粗则磨矿粒度 r 粗,矿石粒度 B_1 细则磨矿粒度 r 细.

矿石的可磨性 B_2 由矿石的硬度和韧性共同决定. 矿石硬度大则难磨,反之则易磨. 同时矿石韧性大,也不易磨碎. 通常采用邦德(Bond)功指数来衡量矿石的可磨性,即

$$W_1 = W / \left(\frac{10}{\sqrt{Q}} - \frac{10}{\sqrt{F}} \right) \quad (2)$$

其中 Q 和 F 分别为给矿和磨矿产品中按 80% 矿粒通过筛孔的尺寸, μm ; W 为磨碎功耗, $\text{kW} \cdot \text{h} \cdot \text{t}^{-1}$; W_1 即为邦德功指数. 从上式可以看出,当给矿粒度 B_1 不变,即 Q 不变时,对于相同的功耗 W ,矿石的邦德功指数 W_1 不同,则 F 不同,即磨矿粒度 r 不同.

由于矿石粒度 B_1 和可磨性 B_2 难以在线标定与检测,因此无法作为磨矿粒度估计模型的辅助变量. 为此需要寻找能够反映矿石粒度 B_1 和可磨性 B_2 变化的数据. 由特性分析可知,当给矿粒度 B_1 和可磨性 B_2 发生变化时,磨机排矿粒度也随之发生改变. 由于分级机的溢流粒度和底流粒度均与磨机排矿粒度密切相关,通常磨机排矿粒度变粗或变细,分级机的溢流粒度和底流粒度也将发生相同的变化. 而底流粒度的粗细决定了分级机返砂矿浆中的固体含量,其变化将直接导致分级机返砂量发生改变. 文献[18]指出电流随着返砂量的增加而增加,两者呈近似线性关系. 因此分级机电流 c_2 可在一定程度上反映出矿石性质的变化. 同理,磨机电流 c_1 也可在一定程度上获知矿石性质的变化. 因此利用磨机电

流 c_1 、分级机电流 c_2 、磨机给矿量 α_1 、磨机入口给水流量 α_2 、分级机溢流质量分数 α_3 等数据可以预测磨矿粒度,即实现以下的映射关系:

$$\{\hat{r}\} \leftarrow f\{c_1, c_2, \alpha_1, \alpha_2, \alpha_3\} \quad (3)$$

其中 f 为未知的非线性函数,难以用精确地数学模型来表示,但可以采用数据驱动的方法对其进行逼近. 考虑到磨矿过程中受矿石磁团聚特性影响,导致测量的不稳定和不准确,影响建模精度,本文提出了一种鲁棒随机向量函数链接网络集成建模方法,用于降低甚至消除异常值影响.

2 基于鲁棒随机向量函数链接网络的集成建模

2.1 鲁棒集成建模策略

本文所提出的鲁棒集成学习策略如图 4 所示. 首先基于 Bagging 的采样方式,获得 s 组子样本数据集;然后将这些不同的样本数据集用于训练并构建 s 组鲁棒基模型;以自适应加权数据融合策略作为基模型的组合方法,对不同基模型的输出分配权重. 其中,鲁棒模型基于传统随机向量函数链接网络,并针对异常值数据,将加权最小二乘(weighted least squares, WLS)技术与非参数核密度估计(non-parametric kernel density estimation, NKDE)^[19]方法相结合;而集成建模策略基于 Bagging 的采样方式,将其与数据融合技术相结合,一方面保证了基模型之间互相独立能够并行建模,另一方面自适应加权数据融合方法有效评估了基模型对于集成建模的贡献度.

2.2 基于鲁棒随机向量函数链接网络的基模型

作为一种随机学习算法,随机向量函数链接网络的特点是两步训练过程:随机分配隐含层参数(输入权值和偏置);通过最小二乘法求解线性方程来评估输出权值. 从而,随机向量函数链接网络在学习速度、建模精度和算法实现上都有优势^[8,12]. 但是,实际工业过程中的样本数据往往存在异常值,这种情况下传统最小二乘难以实现输出权值的稳健估计.

为此,本文采用加权最小二乘方法来计算成本函数,并引入非参数核密度估计来计算惩罚权值,从而得到改进的鲁棒随机向量函数链接网络,具体模型如下.

鲁棒随机向量函数链接网络的结构如图 5 所示,其输出包括两部分:输入到输出的直接连接(direct links)和非线性映射. 考虑给定的多输入单输出训练样本数据集 $S = \{(x_i, y_i) | i = 1, \dots, N\} \subset$

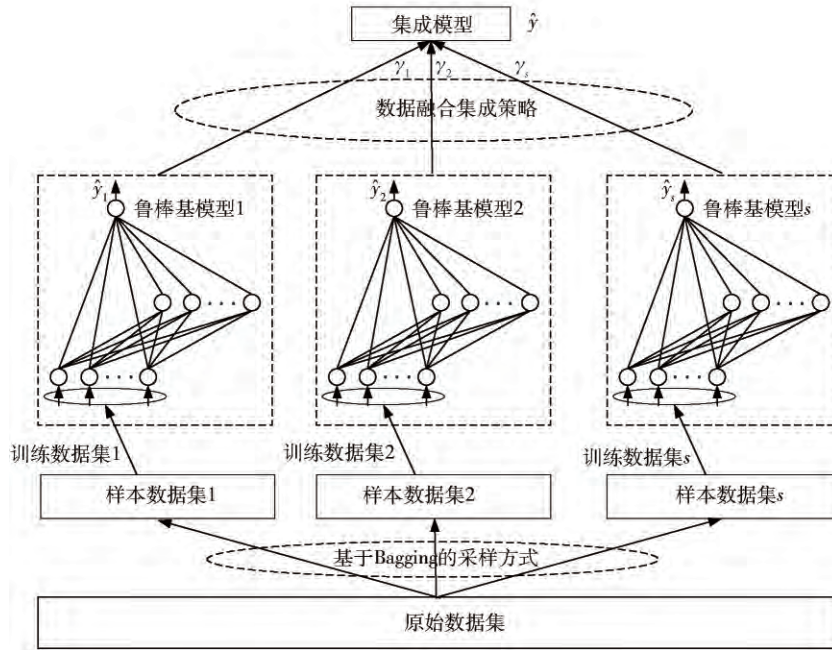


图4 鲁棒集成学习策略的结构图

Fig.4 Structural diagram of ensemble learning strategy

$\mathbf{R}^d \times \mathbf{R}$ 那么带有 L 个隐含层节点的鲁棒随机向量函数链接网络模型的输出可以表示为:

$$y(\mathbf{x}) = \sum_{k=1}^d w_{1k} \mathbf{x}_k + \sum_{j=1}^L w_{2j} h_j(\mathbf{v}_j, b_j, \mathbf{x}) \quad (4)$$

其中 w_{1k} ($k=1, 2, \dots, d$) 表示从输入到输出的直接连接权值; w_{2j} ($j=1, 2, \dots, L$) 表示隐含层与输出节点之间的权值; 激活函数 h_j 表示隐含层特征映射, 通常采用径向基函数或者 sigmoid 曲线函数; \mathbf{v}_j 和 b_j 分别表示从输入层到隐含层的输入权值和偏置, 即隐含层随机参数.

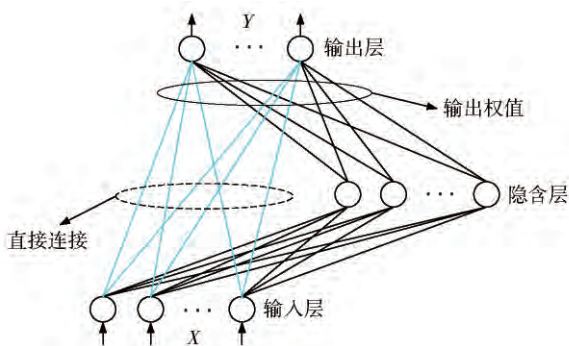


图5 鲁棒随机向量函数链接网络的结构图

Fig.5 Structural diagram of robust RVFLN

以鲁棒随机向量函数链接网络基模型输出误差构造如下成本函数:

$$J = \sum_{i=1}^N p_i \left\| \sum_{k=1}^d w_{1k} x_{ik} + \sum_{j=1}^L w_{2j} h_j(\mathbf{v}_j, b_j, \mathbf{x}_i) - y_i \right\|^2 = (\mathbf{A}\boldsymbol{\beta} - \mathbf{Y})^T \mathbf{P} (\mathbf{A}\boldsymbol{\beta} - \mathbf{Y}) \quad (5)$$

其中 $\mathbf{P} = \text{diag}\{p_1, \dots, p_i, \dots, p_N\}$ 是惩罚权值矩阵, p_i

为第 i 个训练样本对于建模的贡献度, y_i 表示第 i 个训练样本输出值, 拓展输入矩阵 $\mathbf{A} = [\mathbf{X} \ \mathbf{H}]_{N \times (d+L)}$ 包含输入矩阵 \mathbf{X} 和隐含层输出矩阵 \mathbf{H} :

$$\mathbf{H} = \begin{bmatrix} h(\mathbf{v}_1, b_1, \mathbf{x}_1) & \dots & h(\mathbf{v}_L, b_L, \mathbf{x}_1) \\ \vdots & & \vdots \\ h(\mathbf{v}_1, b_1, \mathbf{x}_N) & \dots & h(\mathbf{v}_L, b_L, \mathbf{x}_N) \end{bmatrix}_{N \times L} \quad (6)$$

输出权值 $\boldsymbol{\beta} = [\mathbf{W}_1 \ \mathbf{W}_2]^T$, 模型输出 $\mathbf{Y} = [Y_1, \dots, Y_N]^T$ 其中 $\mathbf{W}_1 = \{w_{1k}\} (k=1, 2, \dots, d)$, $\mathbf{W}_2 = \{w_{2j}\} (j=1, 2, \dots, L)$. 于是有: $\mathbf{Y} = \mathbf{A}\boldsymbol{\beta} = [\mathbf{X} \ \mathbf{H}] [\mathbf{W}_1 \ \mathbf{W}_2]^T = \mathbf{X}\mathbf{W}_1 + \mathbf{H}\mathbf{W}_2$.

由此可建立如下的权值优化模型:

$$\begin{aligned} \min: & \frac{1}{2} \sum_{i=1}^N p_i \|\boldsymbol{\varepsilon}_i\|^2 \\ \text{s. t.} & \sum_{k=1}^d w_{1k} x_{ik} + \sum_{j=1}^L w_{2j} h_j(\mathbf{v}_j, b_j, \mathbf{x}_i) = y_i - \boldsymbol{\varepsilon}_i, \forall i \end{aligned} \quad (7)$$

其中 $\boldsymbol{\varepsilon}_i$ 为训练过程中第 i 个样本输出与模型输出之间的残差. 将上述优化模型转化为以下对偶优化问题:

$$\begin{aligned} J_{\frac{1}{2}}(\boldsymbol{\beta}, \boldsymbol{\varepsilon}, \boldsymbol{\alpha}) = & \frac{1}{2} \sum_{i=1}^N \|\boldsymbol{\varepsilon}_i\|^2 - \\ & \sum_{i=1}^N \alpha_i \left(\sum_{k=1}^d w_{1k} x_{ik} + \sum_{j=1}^L w_{2j} h_j(\mathbf{v}_j, b_j, \mathbf{x}_i) - y_i + \boldsymbol{\varepsilon}_i \right) \end{aligned} \quad (8)$$

其中 α_i 表示第 i 个样本的拉格朗日乘子. 根据 Karush-Kuhn-Tucker (KKT) 理论, 对上述凸二次优化

问题进行求解:

$$\frac{\partial J_{\frac{\lambda}{2}}}{\partial \beta} = 0 \rightarrow A^T \alpha = 0 \quad (9)$$

$$\frac{\partial J_{\frac{\lambda}{2}}}{\partial \varepsilon} = 0 \rightarrow \alpha = P\varepsilon \quad (10)$$

$$\frac{\partial J_{\frac{\lambda}{2}}}{\partial \beta} = 0 \rightarrow A\beta - Y - \varepsilon = 0 \quad (11)$$

由公式 (9) 和 (10) 可得:

$$A^T P\varepsilon = 0 \quad (12)$$

将等式 (11) 两边同时左乘 $A^T P$ 则:

$$A^T P A \beta - A^T P Y - A^T P \varepsilon = 0 \quad (13)$$

于是, 得出优化后的权值如下:

$$\beta^* = (A^T P A)^{\dagger} A^T P Y \quad (14)$$

其中, $(A^T P A)^{\dagger}$ 为矩阵 $A^T P A$ 的 Moore-Penrose 广义逆^[20].

以上为鲁棒随机向量函数链接网络基模型框架, 这里的惩罚权值 P 可以通过非参数核密度估计来获得, 它代表了不同样本对于建模的贡献度. 由于异常值的残差分布总是偏离整体样本的残差分布, 即异常值的残差密度较低, 因此可以根据残差的概率密度函数来评估样本的可靠性. 从而, 异常值所对应样本的贡献度很小, 甚至接近零. 具体过程描述如下:

先初始化 P 为单位对角阵, 根据公式 (14) 计算输出权值, 得到鲁棒随机向量函数链接网络训练过程残差的概率分布, 即:

$$\varepsilon_i = \sum_{k=1}^d w_{1k} x_{ik} + \sum_{j=1}^L w_{2j} h_j(v_j, b_j, x_i) - y_i \quad (15)$$

然后, 通过非参数核密度估计方法得到残差的概率密度函数:

$$f(x) = \frac{1}{hN} \sum_{l=1}^N \phi\left(\frac{x - \varepsilon_l}{h}\right) \quad (16)$$

其中, N 为训练样本总个数, $h = 1.06\hat{\sigma}N^{-1/5}$ 是窗口宽度, $\hat{\sigma}$ 是残差的标准差, ϕ 是高斯核函数:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \quad (17)$$

因此, 可求出惩罚权值 p_i :

$$p_i = f(\varepsilon_i) = \frac{1}{hN} \sum_{l=1}^N \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\varepsilon_i - \varepsilon_l}{h}\right)^2} \quad (18)$$

注 1: 鲁棒随机向量函数链接网络引入加权最小二乘求解成本函数, 并通过非参数核密度估计来获得惩罚权值, 从而提高模型的鲁棒性. 可靠性高的样本具有较高的惩罚权值, 反之亦然, 从而减弱异常值对建模的影响. 然而, 单个鲁棒随机向量函数

链接网络基模型的泛化能力较差. 为解决这个问题, 本文提出了一种基于鲁棒随机向量函数链接网络的磨矿粒度集成建模方法.

2.3 基于 Bagging 与自适应加权数据融合算法的鲁棒集成建模方法

Boosting 和 Bagging 是两种常用的集成策略. 其中, Boosting 是一种序列化方法: 其首先将所有原始样本用于第一个基模型的构建, 并根据该模型精度将未达到容忍误差时所对应的样本筛选出来, 处理后用于下一个基模型的构建; 而 Bagging 是一种并行化方法: 通过均匀采样, 即随机地从原始样本中有放回地多次抽取, 每次抽取相同个数的样本, 因此各个子样本之间是相互独立的, 从而使得各个基模型之间相互独立, 可以通过并行训练快速建模. 此外, 针对实际工业中具有异常值的建模样本, 若采用 Boosting, 上一个基模型会把包含异常值的样本留给下一个基模型, 则基模型的性能会越来越差, 因此 Boosting 不适用于异常值存在的条件. 鉴于此, 本文以 Bagging 的集成策略构建基模型.

基模型之间的组合可以采用自适应加权数据融合算法来实现. 自适应加权数据融合的基本思想: 在最优条件下, 即总方差最小时, 根据基模型的输出, 以自适应的方式寻找它们对应的最优加权因子, 使得融合后的结果达到最优. 自适应加权数据融合模型见图 4, 其中 $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_s$ 表示 s 个独立基模型的输出, 不同基模型对应的加权因子分别为 $\gamma_1, \gamma_2, \dots, \gamma_s$, \hat{y} 则是数据融合的结果. 具体过程描述如下.

设 s 个鲁棒随机向量函数链接网络基模型的方差分别为 $\sigma_1^2, \sigma_2^2, \dots, \sigma_s^2$, 真实值为 y , 则融合后的模型输出 \hat{y} 满足下面的关系:

$$\begin{cases} \hat{y} = \sum_{a=1}^s \gamma_a \hat{y}_a \\ \sum_{a=1}^s \gamma_a = 1 \end{cases} \quad (19)$$

总方差 σ^2 为:

$$\begin{aligned} \sigma^2 &= E[(y - \hat{y})^2] = \\ &E\left[\left(\sum_{a=1}^s \gamma_a y - \sum_{a=1}^s \gamma_a \hat{y}_a\right)^2\right] = \\ &E\left[\left(\sum_{a=1}^s \gamma_a (y - \hat{y}_a)\right)^2\right] = \\ &E\left[\sum_{a=1}^s \gamma_a^2 (y - \hat{y}_a)^2 + 2 \sum_{\substack{a=1 \\ a \neq b}}^s \gamma_a \gamma_b (y - \hat{y}_a)(y - \hat{y}_b)\right] \end{aligned} \quad (20)$$

其中 E 表示数学期望. 由于 $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_s$ 彼此相互独立, 且为 \hat{y} 的无偏估计, 则有:

$$E(y - \hat{y}_a)(y - \hat{y}_b) = 0 \quad (a=1, \dots, s; b=1, \dots, s; a \neq b) \quad (21)$$

因此 σ^2 可以写成:

$$\sigma^2 = E \left[\sum_{a=1}^s \gamma_a^2 (y - \hat{y}_a)^2 \right] = \sum_{a=1}^s \gamma_a^2 \sigma_a^2 \quad (22)$$

其中 σ_a^2 ($a=1, 2, \dots, s$) 是第 a 个鲁棒随机向量函数链接网络基模型测试输出的方差。

从 (22) 式可知, 总方差 σ^2 是加权因子的多元二次函数, 存在最小值 σ_{\min}^2 , 求解如下:

$$\begin{cases} \sigma_{\min}^2 = \min \left(\sum_{a=1}^s \gamma_a^2 \sigma_a^2 \right) \\ \sum_{a=1}^s \gamma_a = 1 \end{cases} \quad (23)$$

根据多元函数求极值理论, 可求得 σ_{\min}^2 时第 a 个鲁棒随机向量函数链接网络基模型所对应的最优加权因子:

$$\gamma_a^* = \frac{1}{\sigma_a^2 \sum_{a=1}^s \frac{1}{\sigma_a^2}}, \quad (a=1, \dots, s) \quad (24)$$

相应地, 总方差的最小值为

$$\sigma_{\min}^2 = \frac{1}{\sum_{a=1}^s \frac{1}{\sigma_a^2}} \quad (25)$$

2.4 算法步骤

给定一个训练样本集 $\{(x_i, y_i) \mid i=1, \dots, N\} \subset \mathbf{R}^d \times \mathbf{R}$.

步骤 1: 设置隐含层节点数 L , 在区间 $[-\lambda, +\lambda]^d$ 和 $[-\lambda, +\lambda]$ 内分别随机生成输入权值 v_j 和偏置 b_j , 其中 $j=1, 2, \dots, L$;

步骤 2: 计算隐含层输出矩阵 H , 得到拓展输入矩阵 A ;

步骤 3: 根据公式 (15) 计算残差, 代入公式 (18) 求 p_i 并构造对角矩阵 P ;

步骤 4: 根据公式 (14) 来计算输出权值 β^* ;

步骤 5: 重复步骤 1~4, 构建一组 s 个鲁棒随机向量函数链接网络, 并求出每个基模型输出的方差;

步骤 6: 根据公式 (24) 求出每一个鲁棒随机向量函数链接网络基模型所对应的最优加权因子;

步骤 7: 根据公式 (19) 求出集成模型数据融合后的输出。

注 2: 集成时需要基模型个数进行设定, 文献 [21] 提出了一种基于进化算法尺寸控制选择性集成, 给出了一个重要的结论: 最佳基模型的个数范围为 2~8。

2.5 性能评估

为了验证本文所提集成建模方法的有效性, 在

基模型个数 s 为 8 的条件下, 将其与直接平均法的集成随机向量函数链接网络和数据融合的集成随机向量函数链接网络进行了对比实验。实验结果基于一个函数近似和三个基准数据回归问题, 仿真在 MATLAB 2016a 环境下运行, 所用 PC 的 CPU 为 i5, 3.4 GHz, 内存为 8 GB RAM。

2.5.1 函数近似

采用以下非线性复合函数:

$$f(x) = 0.2e^{-(10x-4)^2} + 0.5e^{-(80x-40)^2} + 0.3e^{-(80x-20)^2}, \quad x \in [0, 1] \quad (26)$$

首先随机产生 5000 个不同的样本数据作为原始数据集, 然后根据 Bagging 的采样方法, 每轮有放回地从中随机抽取 500 个训练数据, 得到 10 组训练数据; 再随机生成 200 个不同的样本, 作为集成模型的测试数据。为了验证随机向量函数链接网络的鲁棒性, 在训练样本中加入了异常值。本实验中对于每个训练数据集加入的异常值水平 ξ 分别为 0%、10%、20% 和 30%。

在数据预处理阶段, 输入和输出数据均被归一化至区间 $[-1, 1]$, 所有的实验结果均基于 50 次独立试验。比较实验中, 集成建模性能包括模型精度和训练时间。其中, 模型精度使用 50 次实验均方根误差 (root mean squares error, RMSE) 的均值 mean 表示; 训练时间用 t (s) 表示。相关参数的设置通过交叉验证法在具体的实验中给出。

表 1 为 50 次独立实验的结果, 不同异常值水平下三种集成建模方法中最佳的模型性能用粗体表示。由表 1 可知, 在没有异常值干扰下, 即异常值水平 $\xi=0$ 时, 三种方法的模型精度基本相同。随着 ξ 的不断增大, 虽然三种方法的建模精度均有所下降, 但本文所提方法受其影响最小。与其他两种方法相比, 在稍微增加训练时间的情况下, 数据融合鲁棒随机向量函数链接网络集成方法具有更高的模型精度。此外, 图 6 为额外的一次独立实验, 它给出了三种集成建模方法在异常值水平 $\xi=20\%$ 时的测试结果。从图 6 中可以看出, 在该异常值水平下, 直接平均和数据融合随机向量函数链接网络的预测值已经出现了明显的波动, 表明模型误差较大; 而本文所提方法的预测值能够很好的拟合真实值, 表明了其有效性。考虑到文章的可读性, 其他异常值水平下的结论与 $\xi=20\%$ 时相同, 且与表 1 中的实验结果一致, 在此省略。

需要指出的是, 三种集成建模方法基模型的构建方式均为并行, 本文所提方法在训练速度上稍微

表 1 函数近似集成建模性能比较

Table 1 Performance comparison of ensemble learning for function approximation

数据集	集成建模方法	模型参数, 模型参数,		mean t/s			
		L	λ	异常值水平 0	异常值水平 10%	异常值水平 20%	异常值水平 30%
非线性 复合函数	直接平均随机向量函数 链接网络集成	500	20	8.2×10^{-4} , 0.143	0.008, 0.138	0.011, 0.156	0.013, 0.147
	数据融合随机向量函数 链接网络集成	500	20	8.1×10^{-4} , 0.143	0.007, 0.140	0.010, 0.145	0.012, 0.144
	数据融合鲁棒随机向量 函数链接网络集成	500	50	8.9×10^{-4} , 0.408	0.003 , 0.409	0.005 , 0.411	0.008 , 0.419

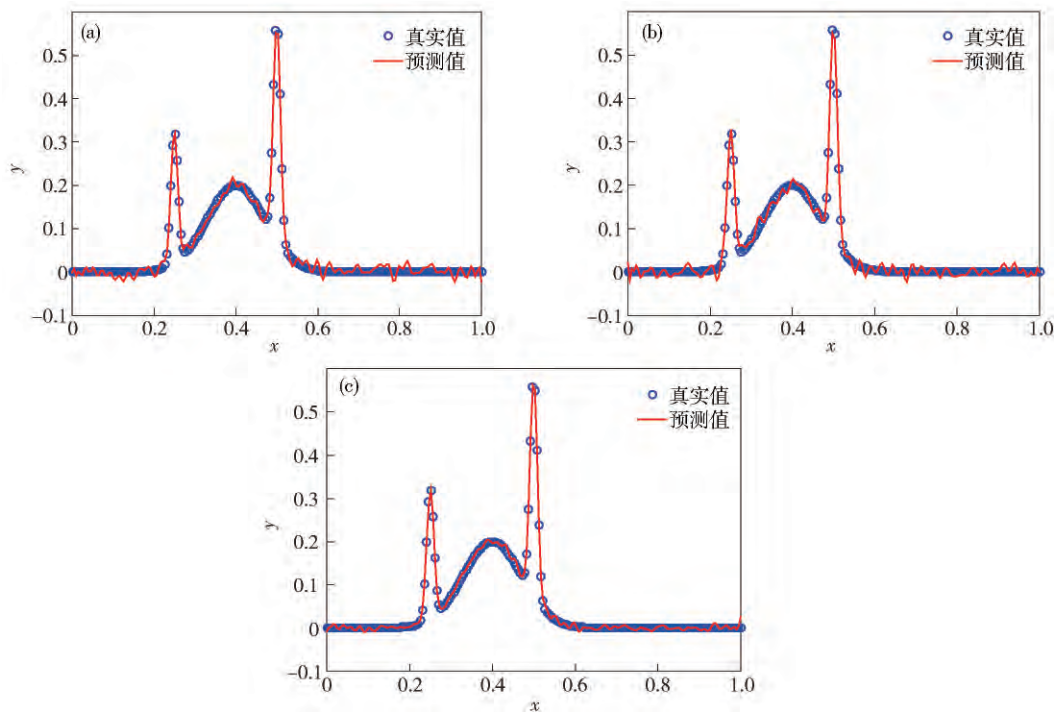


图 6 $\xi = 20\%$ 时函数近似比较试验。(a) 直接平均随机向量函数链接网络集成; (b) 数据融合随机向量函数链接网络集成; (c) 数据融合鲁棒随机向量函数链接网络集成

Fig. 6 Comparison experiments of function approximation at $\xi = 20\%$: (a) RVFLN-based direct average ensemble learning; (b) RVFLN-based data fusion ensemble learning; (c) robust RVFLN-based data fusion ensemble learning

逊色于其他两种方法的原因为: 在构建基模型时, 鲁棒随机向量函数链接网络比传统随机向量函数链接网络多了非参数核密度估计评估样本数据对于建模贡献度的过程; 在基模型组合方式上, 直接加权法虽然简单, 在一定程度上提高算法的稳定性, 但分配给每个基模型相同的权重是不合理的, 而自适应加权数据融合的方法则是根据每个基模型精度的差异, 并保证集成后模型总方差最小的情况下, 给基模型分配不同的权值。实验结果表明, 本文所提方法能够以较快的速度集成建模, 有效提高模型的鲁棒性。

2.5.2 基准数据

这部分给出了来自加州大学欧文(University of

California at Irvine, UCI) 机器学习数据集的三个实际基准回归案例, 包括 combined cycle power plant, concrete 和 wine。基于 50 次独立实验, 图 7 给出了三种方法在不同异常值水平下的模型性能。这里的异常值水平是相对于原始数据而言, 并在其基础上所额外添加的离群点比例。实验中, 针对三个实际数据集, 集成模型参数(L, λ) 的设置通过交叉验证法得出, 分别为(1000, 50)、(40, 1)和(50, 0.5)。如图 7 所示, 随着异常值水平的增加, 测试均方根误差均呈现出整体递增的趋势。不同之处在于, 本文所提方法受异常值影响较小, 能够在一定程度上有效降低异常值的影响, 且具有良好的泛化性能, 对于实际的复杂工业过程建模有着重要意义。

本节基于一个函数近似和三组基准数据回归问题,将所提方法与直接平均法的集成随机向量函数链接网络和数据融合的集成随机向量函数链接网络

进行了对比实验. 实验结果验证了基于鲁棒随机向量函数链接网络集成建模方法的有效性,并将其用于实际工业过程进行磨矿粒度估计.

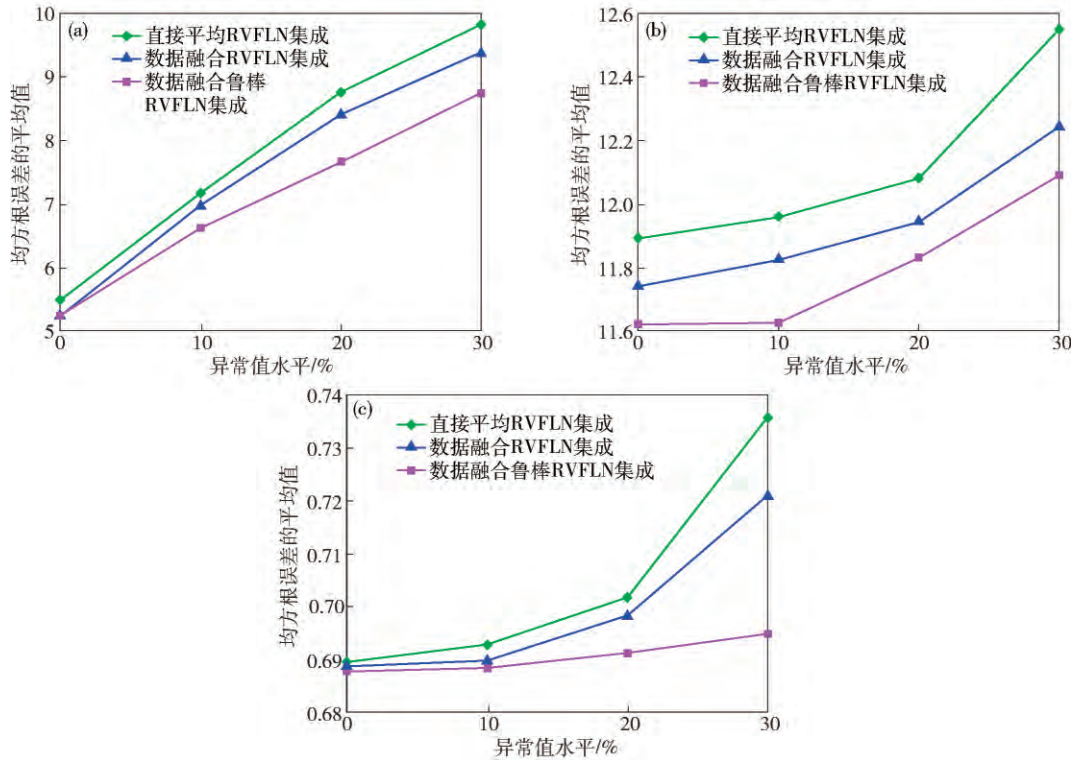


图7 基准回归比较试验. (a) combined cycle power plant; (b) concrete; (c) wine

Fig.7 Comparison experiments of benchmark regression: (a) combined cycle power plant; (b) concrete; (c) wine

3 磨矿粒度估计

通过使用实际工业赤铁矿磨矿过程的数据,将本文所提方法应用于磨矿粒度估计. 采样时,从原始数据中随机选取8组,每组500个作为训练数据;300个作为测试数据. 具体的磨矿过程和特性分析见第1章. 同2.4节性能评估,这部分在不同异常

值水平下进行了50次独立实验,结果见表2. 考虑到文章的可读性,图8只给出了 $\xi = 30\%$ 时三种集成方法的模型性能. 需要指出的是,这里的异常值水平是相对于原始数据而言,为了进一步体现所提方法的鲁棒性,在原始数据集上额外添加的不同比例的离群点.

表2 磨矿粒度集成建模性能比较

Table 2 Performance comparison of ensemble learning for particle size of grinding process

数据集	集成建模方法	模型参数, 模型参数,		mean t/s			
		L	λ	异常值水平 0	异常值水平 10%	异常值水平 20%	异常值水平 30%
实际工业 磨矿过程	直接平均随机向量函数 链接网络集成	50	1	0.095, 0.063	0.183, 0.063	0.212, 0.062	0.231, 0.063
	数据融合随机向量函数 链接网络集成	50	1	0.086, 0.063	0.168, 0.062	0.201, 0.623	0.224, 0.063
	数据融合鲁棒随机向量 函数链接网络集成	50	1	0.009 , 0.118	0.035 , 0.113	0.072 , 0.120	0.106 , 0.119

由表2可知,虽然数据融合鲁棒随机向量函数链接网络集成建模在训练时间上有轻微增加,但在不同异常值水平下,其均方根误差均最低;而且随着

异常值水平的增加,本文所提方法的建模性能波动最小,体现了良好的鲁棒性. 其中 $\xi = 30\%$ 时实验结果见图8,三种集成方法用于磨矿粒度的估计值

被分别给出. 虽然三种方法在一定程度上都能实现粒度的估计, 但从粒度的实际值与估计值的拟合程度来看, 数据融合鲁棒随机向量函数链接网络集成效果最好, 数据融合随机向量函数链接网络集成次之, 直接平均随机向量函数链接网络集成较差. 通

过实际工业过程的磨矿粒度估计结果可知, 本文所提方法取得了良好的效果. 再次表明, 数据融合鲁棒随机向量函数链接网络集成建模方法具有良好的泛化性能和预测精度, 并因其较好的鲁棒性适用于异常值干扰的环境下.

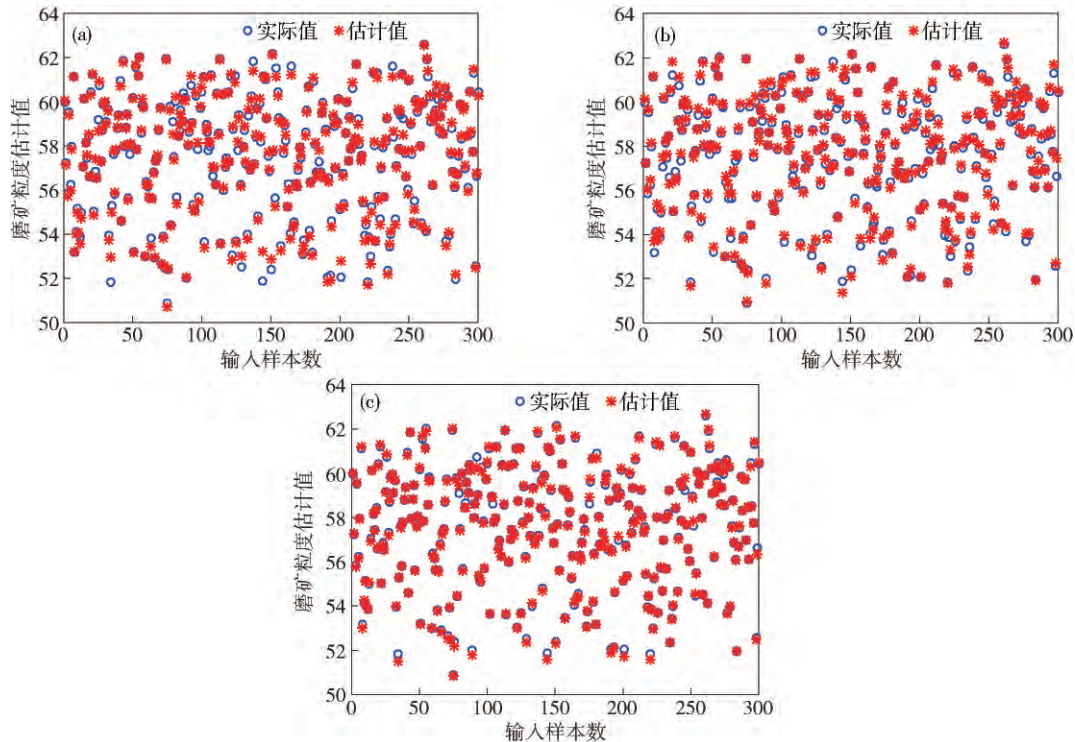


图 8 $\xi = 30\%$ 时磨矿粒度估计. (a) 直接平均随机向量函数链接网络集成; (b) 数据融合随机向量函数链接网络集成; (c) 数据融合鲁棒随机向量函数链接网络集成

Fig. 8 Particle size estimation of grinding process at $\xi = 30\%$: (a) RVFLN-based direct average ensemble learning; (b) RVFLN-based data fusion ensemble learning; (c) robust RVFLN-based data fusion ensemble learning

4 结论

(1) 以鲁棒随机向量函数链接网络作为基模型, 它通过随机分配隐含层参数, 并将加权最小二乘技术与非参数核密度估计方法相结合, 能够确保模型在异常值干扰的条件下的逼近能力, 并且在学习速度和建模精度上均比传统前馈神经网络有优势.

(2) 与单一模型相比, 在绝大多数情况下通过对多个模型集成可以显著提高原来模型的精度与泛化性, 且传统的集成学习方法在异常值干扰下性能下降. 因此, 本文提出了一种基于鲁棒随机向量函数链接网络的自适应数据融合集成建模方法. 实验结果表明, 与其他两种集成建模方法相比, 所提方法稳定性好、精度高, 并且具有良好的鲁棒性.

(3) 实际工业过程中的样本数据往往存在异常值. 在本文实验中, 将所提方法采用磨矿生产过程的实际工业数据进行集成建模. 结果表明, 基于鲁棒随机向量函数链接网络的自适应数据融合集成建

模方法在异常值水平为 30% 时仍能取得较好的结果. 从而, 本文所提方法能够应用到实际磨矿过程中实现粒度估计.

参 考 文 献

- [1] Chen X S, Li Q, Fei S M. Supervisory expert control for ball mill grinding circuits. *Expert Syst Appl*, 2008, 34(3): 1877
- [2] Zhou P, Dai W, Chai T Y. Multivariable disturbance observer based advanced feedback control design and its application to a grinding circuit. *IEEE Trans Control Syst Technol*, 2014, 22(4): 1474
- [3] Wang X L, Gui W H, Yang C H, et al. Wet grindability of an industrial ore and its breakage parameters estimation using population balances. *Int J Miner Process*, 2011, 98(1-2): 113
- [4] Meyer E J, Craig I K. The development of dynamic models for a dense medium separation circuit in coal beneficiation, *Miner Eng*, 2010, 23(10): 791
- [5] Sun Z, Wang H G, Zhang Z K. Soft sensing of overflow particle size distributions in hydrocyclones using a combined method. *Tsinghua Sci Technol*, 2008, 13(1): 47

- [6] Wang X H , Gui W H , Wang Y L , et al. Prediction modeling for particle size of grinding circuit of mixture kernels SVM. *Comput Eng Appl* ,2010 ,46(12) : 207
(王新华,桂卫华,王雅琳,等.混合核函数支持向量机的磨矿粒度预测模型.计算机工程与应用,2010,46(12):207)
- [7] Qiao J H , Chai T Y. Soft measurement model and its application in raw meal calcination process. *J Process Control* ,2012 ,22(1) : 344
- [8] Igel'nik B , Pao Y H. Stochastic choice of basis functions in adaptive function approximation and the functional-link net. *IEEE Trans Neural Netw* ,1995 ,6(6) : 1320
- [9] Huang G B , Chen Y Q , Babri H A. Classification ability of single hidden layer feedforward neural networks. *IEEE Trans Neural Netw* ,2000 ,11(3) : 799
- [10] Zhang S Y , Bao Y P , Zhang C J , et al. Prediction model of aluminum consumption with BP neural networks in IF steel production. *Chin J Eng* ,2017 ,39(4) : 511
(张思源,包燕平,张超杰,等. BP神经网络IF钢铝耗的预测模型.工程科学学报,2017,39(4):511)
- [11] Pao Y H , Phillips S M , Sobajic D J. Neural-net computing and the intelligent control of systems. *Int J Control* ,1992 ,56(2) : 263
- [12] Scardapane S , Wang D H. Randomness in neural networks: An overview. *WIREs Data Min Knowl Disc* ,2017 ,7(2) : e1200
- [13] Ditterich T G. Machine learning research: four current directions. *Artif Intell Mag* ,1997 ,4: 97
- [14] Kearns M J , Valiant L G. *Learning Boolean Formulae or Finite Automata is as Hard as Factoring*. Cambridge: Harvard University , Center for Research in Computing Technology , Aiken Computation Laboratory ,1988
- [15] Schapire R E. The strength of weak learnability. *Mach Learn* ,1990 ,5(2) : 197
- [16] Schwenk H , Bengio Y. Boosting neural networks. *Neural Comput* ,2000 ,12(8) : 1869
- [17] Martínez-Muñoz G , Suárez A. Out-of-bag estimation of the optimal sample size in bagging. *Pattern Recognit* ,2010 ,43(1) : 143
- [18] Dai W , Chai T Y , Yang S X. Data-driven optimization control for safety operation of hematite grinding process. *IEEE Trans Ind Electron* ,2015 ,62(5) : 2930
- [19] Dai W , Liu Q , Chai T Y. Particle size estimate of grinding processes using random vector functional link networks with improved robustness. *Neurocomputing* ,2015 ,169: 361
- [20] Rao C R. *Generalized Inverse of Matrices and its Applications*. New York: Wiley ,1971
- [21] Wang D H , Alhamdoosh M. Evolutionary extreme learning machine ensembles with size control. *Neurocomputing* ,2013 ,102: 98