

基于改进的随机森林算法的烧结状态短期预测

王福斌¹, 王蕊^{1*}, 武晨²¹华北理工大学电气工程学院, 河北 唐山 063210;²唐钢国际工程技术股份有限公司, 河北 唐山 063000

摘要 充分利用烧结机尾断面的火焰图像所蕴含的有效信息,利用随机森林算法对烧结状态进行短期预测,该算法在工程上具有可行性。为改善随机森林中重要性较低的属性对分类结果的影响,提出了一种基于概率决策的随机森林改进算法,实现对烧结机尾断面火焰状态的短期预测。首先,对300张烧结断面火焰图像进行统一预处理,将获得的10个图像几何特征作为输入量;其次,对提取到的10个图像几何特征进行K均值聚类 and 模糊C均值聚类,根据聚类结果的准确率赋予叶子节点处各个类别出现的概率;最后,实验验证了优化的随机森林算法能提高对烧结状态分类的准确性。

关键词 火焰图像; K均值分割; 几何特征; 随机森林

中图分类号 TP391.41

文献标志码 A

DOI: 10.3788/LOP202259.1815018

Short-Term Prediction of Sintering State Based on Improved Random Forest

Wang Fubin¹, Wang Rui^{1*}, Wu Chen²

¹College of Electrical Engineering, North China University of Science and Technology, Tangshan 063210, Hebei, China;

²Tang Steel International Engineering Technology Co., Ltd., Tangshan 063000, Hebei, China

Abstract For the effective use of the valid information contained in the flame image of the tail section of the sintering machine, the random forest algorithm is used to predict the sintering state in a short time. The algorithm is feasible in engineering. To improve the influence of low-importance properties in random forests on classification results, we propose a random forest improvement algorithm using probability decision, making to realize the short-term prediction of the flame state of the sintering machine tail section. First, 300 sintered section flame images were uniformly preprocessed, and geometric features of 10 images were given as input. Second, K-mean and fuzzy C-mean clusterings were performed on the geometric features of the 10 extracted images. Finally, the probability of each category appearing at the leaf node was given according to the accuracy of the clustering results. The experiment proves that the proposed optimized random forest algorithm improves the accuracy for sintering state classification.

Key words flame image; K-mean segmentation; geometric feature; random forest

1 引言

在钢铁生产中,为高炉提供优质的烧结矿是高产、低耗、长寿必不可少的条件。但是由于烧结过程滞后性强,强耦合关系存在于各个变量之间,烧结现场环境恶劣,采用人工看火,条件艰苦,劳动强度大,并且缺乏对烧结终点直接测量的仪器。烧结过程是提前于烧结终点的,因此对烧结终点的预测是必不可少的过程。

同时,影响烧结生产稳定的主要因素为烧结终点的稳定控制,因此准确、稳定的烧结终点不仅可以提高烧结矿的质量,还可以极大地降低对环境的污染。

烧结终点(BTP)的位置一般是通过烧结料层在第几个风箱的位置进行判断的,但是在燃烧的过程中,烧结料内部的热量无法直接测量,因此通过分析与烧结终点相关的可测参数,建立预测模型。模型主要包含风箱废弃温度、风箱废弃成分、负压法、机尾烧结矿

收稿日期: 2021-07-23; 修回日期: 2021-08-03; 录用日期: 2021-09-24

基金项目: 高端钢铁联合研究基金(F2019209323)

通信作者: *18332725629@163.com

断面图像。Patisson 等^[1]通过研究烧结带在燃烧过程中的水分变化,建立烧结带水分迁移的数学模型,进行终点预测;李黎^[2]提出一种基于烧结机速度与烧结终点的模型;吴晓峰^[3]以风箱废弃温度上升点为短期预测依据,利用支持向量机进行短期预测;程佩哲^[4]通过控制配焦比与烧结机速度来进行终点预测,预测较为准确;张宗旺等^[5]建立基于时间序列的终点预报模型;周纪平^[6]设计了基于体积模型的烧结终点位置自适应模糊控制器;汪森辉等^[7]通过分析烧结过程中的参数对烧结终点位置的影响,提出了一种集成算法。

烧结机尾断面以出口的形式展现烧结矿的形成过程,同时还可以反映烧结的过程,因此烧结机尾断面火焰能够最直接、最丰富地反映烧结终点的状态。充分利用烧结机尾图像中所蕴含的有效信息对烧结状态进行短期预测具有可行性及工程实际意义。因此,为减小烧结终点预报的振荡,提高预报的稳定性,采用基于烧结断面火焰几何特征的随机森林方法进行短期烧结终点状态预测。

随机森林^[8]属于半监督学习的一种,运行效率高,使用也较为方便,但是决策树比较容易引起过拟合,使训练得到的学习模型泛化能力较差。为了获得组合的正确性和单棵树的多样性^[9],许多学者做出了许多的研究。刘峻^[10]提出一种基于多模糊核约束的随机森林算法;吴辰文等^[11]通过变量预测和选择提升预测精度;Xia 等^[12]通过提出一种随机森林与代孕算法来解决数据不完整而不归因的分类问题;史金余等^[13]提出一种改进随机森林算法,通过构建模糊决策树,提高分类精度。常规随机森林忽略了属性重要性的高低对分类结果的影响,不同的属性在面对缺失值时的影响也是不尽相同的,在传统的随机森林中每一个属性被选中的概率是一样的,但是实际上在面对缺失值时,有的属性可以较高地判断出缺失值与哪一类更为相似。为了解决这个问题,本文提出了一种基于 K 均值聚类和模糊 C 均值聚类的随机森林算法,利用 K 均值聚类和模糊 C 均值聚类后的相对准确度作为属性在被选择时的概率,降低了准确度低的属性对分类结果的影响,提高了在面对缺失值时的准确率。最后利用在工厂采集到的 300 张烧结火焰图像对算法进行验证。

2 数据来源及数据预处理

2.1 数据来源

采用的图像为唐钢 320 m² 烧结机的 300 张烧结机尾断面火焰图像,采集图片的分辨率为 4032×3024。由于图像分辨率较高,可以提供详细的几何特征,进一步提高随机森林的分类精度。

该研究图像包括正常燃烧、欠烧、过烧三种类别,具体类别数量如表 1 所示,适合进行烧结终点的研究。从 300 张火焰图像中随机抽取 210 张作为训练集,其余 90 张作为测试集。

表 1 类别数量

Table 1 Number of categories

| Category | Normal burning | Underburning | Overheating | Sum |
|----------|----------------|--------------|-------------|-----|
| Quantity | 140 | 80 | 80 | 300 |

2.2 数据预处理

为了减少处理图片所耗费的时间,对图片按像素为 240×120 进行统一裁剪,提取出烧结火焰图像的红火区部分,然后分别采用 K 均值($K=3, 4$)和区域分割的方法对红火区进行分割,区域分割由于会造成红火区的丢失,只适用于红火区为完整连通域的情况。最终,采用 K 均值分割和颜色提取得到最终的红火区,提取红火区的 10 个几何特征数据作为随机森林算法的输入。

3 改进随机森林算法

3.1 随机森林

集成学习就是构建并结合多个机器学习器来完成学习任务的过程^[14]。随机森林是一种以决策树为基学习器的半监督学习算法。算法在对测试样本进行分类划分时,采用多个不同训练样本子集来加大分类模型之间的相异性^[15],利用公平投票机制得到结果。在随机森林中任意两棵树的相关性,以及每棵树自身的分类能力与分类效果是密切相关的^[16]。随机森林方法包括训练与分类两个过程,如图 1 所示。

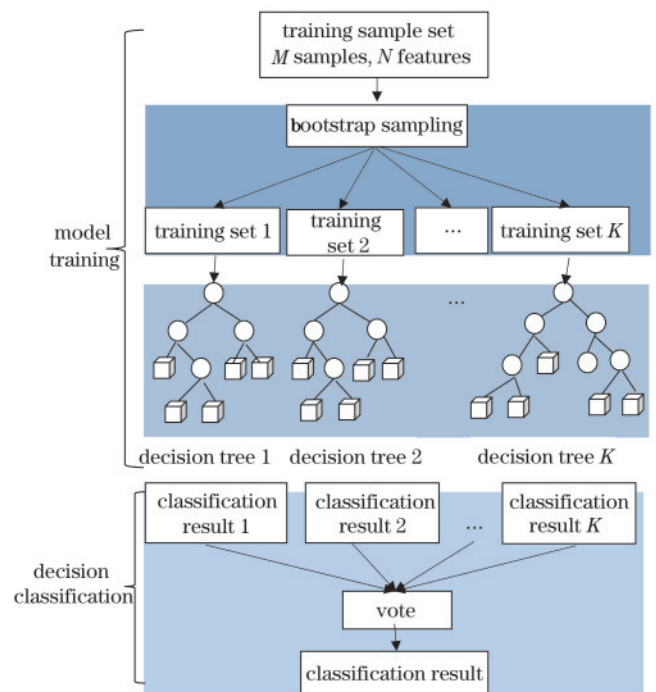


图 1 随机森林算法流程

Fig. 1 Flow chart of random forest algorithm

3.2 改进随机森林

在随机森林中,决策树的属性为条件属性,每个属

性对分类的作用性不同,因此属性的重要性对最终的决策结果有一定的影响^[17]。决策树在进行属性选择的时候概率是一样的,因此重要度小的属性会对分类的精度产生偏差。聚类算法通过分析烧结火焰几何特征,得到同一类烧结图像在某一种几何特征上的相似性,相似性越高的几何特性在面对属性的选择时相较于随机选择属性,越会减小偏差。本文利用 K 均值聚类算法和模糊 C 均值(FCM)算法,得出每个属性聚类的准确度,将其作为选择最优划分属性的准则。在最后投票阶段,每棵树投票的重要度相同,由构建的多棵决策树得到最后的投票结果,通过这种方法,对随机森林算法进行了优化,

分类精度也得到了一定程度的提高。

3.2.1 K 均值聚类

K 均值算法的聚类模型可以描述为一个优化问题^[18],表达式为

$$J = \sum_{m=1}^M \sum_{k=1}^K \lambda_{mk} \| \mathbf{x}_m - \boldsymbol{\mu}_k \|^2, \quad (1)$$

式中: λ_{mk} 是数据点 \mathbf{x}_m 对于聚类中心 k 的归属值; $\| \mathbf{x}_m - \boldsymbol{\mu}_k \|^2$ 为第 m 个样本与第 k 类的聚类中心的欧氏距离。

K 均值聚类结果的准确度是决策树进行分支属性选择的重要因素,通过聚类分析,每个几何特征聚类的准确度如表 2 所示。

表 2 K 均值准确率

Table 2 Accuracy of K-means

| Geometric feature | Area | Major | Minor | Eccentricity | Orientation |
|-------------------|----------------|----------|--------|--------------|----------------|
| Accuracy | 0.49 | 0.42 | 0.64 | 0.47 | 0.59 |
| Geometric feature | Equiv diameter | Solidity | Extent | Perimeter | Thinness ratio |
| Accuracy | 0.51 | 0.35 | 0.51 | 0.47 | 0.37 |

3.2.2 模糊 C 均值聚类

设 $\mathbf{X} = \{ \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M \} \in \mathbf{R}^s$ 表示样本集, s 是样本空间维数。基于 FCM 算法的聚类模型可以描述为一个优化问题,表达式为

$$J_{\text{FCM}}(\mathbf{U}, \mathbf{V}) = \sum_{m=1}^M \sum_{i=1}^I (u_{im})^c \| \mathbf{x}_m - \mathbf{v}_i \|^2, \quad (2)$$

式中: I 为对 \mathbf{X} 进行划分的聚类个数; \mathbf{V} 为聚类中心且 $\mathbf{V} = \{ \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_M \}$; \mathbf{U} 为 $I \times M$ 的模糊划分矩阵; c 是

模糊系数; u_{im} 为第 m 个样本 \mathbf{x}_m 属于第 i 类的隶属度值; $\| \mathbf{x}_m - \mathbf{v}_i \|^2$ 为第 m 个样本与第 i 类的聚类中心的欧氏距离。

在式(2)的基础上,有约束条件: $\sum_{i=1}^I u_{im} = 1, 0 \leq u_{im} \leq 1, 1 \leq i \leq I, 1 \leq m \leq M$ 。通过 FCM 聚类分析,每个几何特征聚类的准确度如表 3 所示。

表 3 FCM 的准确率

Table 3 Accuracy of FCM

| Geometric feature | Area | Major | Minor | Eccentricity | Orientation |
|-------------------|----------------|----------|--------|--------------|----------------|
| Accuracy | 0.51 | 0.43 | 0.64 | 0.5 | 0.64 |
| Geometric feature | Equiv diameter | Solidity | Extent | Perimeter | Thinness ratio |
| Accuracy | 0.51 | 0.48 | 0.39 | 0.47 | 0.46 |

3.2.3 几何特征概率

决策树在进行属性选择时是依据概率进行选择的。首先设置几何特征的准确度为 $\{ v_1, v_2, \dots, v_{10} \}$, 几何特征的概率为 $\{ x_1, x_2, \dots, x_{10} \}$, 则

$$x_i = \frac{v_m}{\sum_{m=1}^{10} v_m}. \quad (3)$$

几何特征的概率如表 4 和表 5 所示。

表 4 K 均值几何特征概率

Table 4 K-means geometric feature probability

| Geometric feature | Area | Major | Minor | Eccentricity | Orientation |
|-------------------|----------------|----------|--------|--------------|----------------|
| Probability | 0.1 | 0.09 | 0.13 | 0.1 | 0.12 |
| Geometric feature | Equiv diameter | Solidity | Extent | Perimeter | Thinness ratio |
| Probability | 0.11 | 0.07 | 0.11 | 0.09 | 0.08 |

3.2.4 所提算法的具体步骤

训练集 $D = \{ (x_1, y_1), \dots, (x_m, y_m), \dots, (x_{300}, y_{300}) \}$

1) 利用 bootstrap 方法在 300 个烧结火焰数据样本中进行重采样, 随机选出 m 个样本, 形成一个自动样

本集, 作为待分裂的样本。

2) 在自动样本集上, 通过以下步骤递归构造决策树, 直到不能再分支时, 停止生长, 生成决策树。从 10 个几何特征中随机选择 p 个属性; 根据 K 均值或模

表 5 FCM 几何特征概率
Table 5 FCM geometric feature probability

| | | | | | |
|-------------------|----------------|----------|--------|--------------|----------------|
| Geometric feature | Area | Major | Minor | Eccentricity | Orientation |
| Probability | 0.1 | 0.08 | 0.13 | 0.1 | 0.13 |
| Geometric feature | Equiv diameter | Solidity | Extent | Perimeter | Thinness ratio |
| Probability | 0.1 | 0.1 | 0.08 | 0.09 | 0.09 |

糊 C 均值算法的准确率计算出每个几何特征的重要度,根据式(3)得到几何特征的概率;根据几何特征的概率在 p 个属性中选择具有最大概率的属性作为最佳分裂点;根据最佳分裂点,将结点拆分成子节点;不进行剪枝,直到不能分裂为止,建立一棵决策树。

3) 重复步骤 2),直到得到 Z 棵决策树。通过 Z 棵决策树进行投票选择,得到最后的分类结果,

$$H(x) = \arg \max_Y \sum_{z=1}^Z I[h_z(x) = Y], \quad (4)$$

式中: $H(x)$ 为随机森林最终分类结果; $h_z(x)$ 为第 z 棵

决策树的类预测; Y 为输出变量; $\arg \max(\cdot)$ 为投票结果的最大值; $I(\cdot)$ 为示性函数^[19]。

4 分类结果

4.1 分割结果

火焰断面图像的红火区中蕴含着有效信息,利用这些信息可以预报烧结终点,通过 K 均值分割方法提高对火焰目标的分割精度及特征提取的准确性。通过多种方法比较,采用 $K=3$ 时的 K 均值分割方法,红火区预处理结果如图 2 所示。

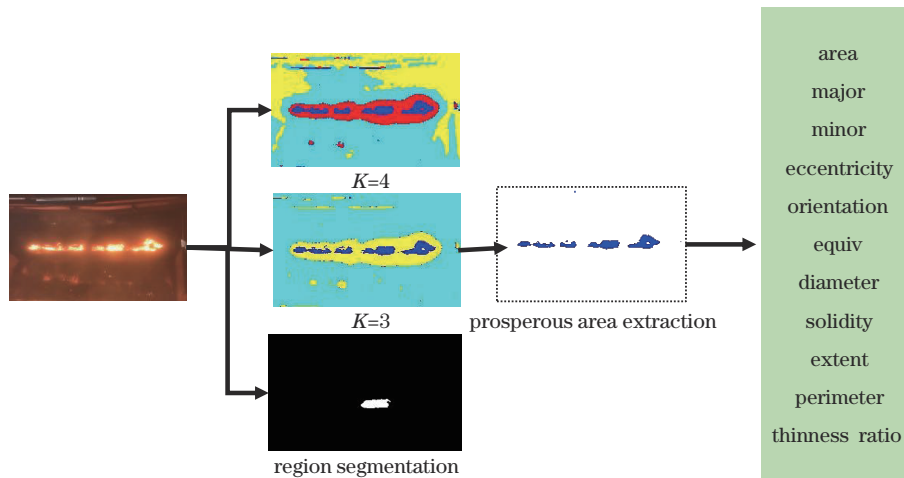


图 2 预处理过程

Fig. 2 Pretreatment process

4.2 特征提取

将烧结火焰的几何特征作为分类器的输入,提取火焰的 10 个几何特征,红火区几何特征数据结果如图 3 所示。

- 1) 面积:为图像前景中为 1 的像素点的个数。
- 2) 长轴:与区域具有相同标准二阶中心矩的椭圆的长轴长度(像素意义下)。
- 3) 短轴:与区域具有相同标准二阶中心矩的椭圆的短轴长度(像素意义下)。
- 4) 离心率:与区域具有相同标准二阶中心矩的椭圆的离心率。
- 5) 方向:与区域具有相同标准二阶中心矩的椭圆的长轴与 x 轴的交角(度)。
- 6) 等效直径:与区域具有相同面积的圆的直径。
- 7) 坚固性:同时在区域和其最小凸多边形中的像素比例。

- 8) 幅度:同时在区域和其最小边界矩形中的像素比例。
- 9) 周长:区域边界像素总和。
- 10) 薄度比:为短轴与长轴的比值。

4.3 分类

4.3.1 参数选择

随机森林中决策树的数量影响最后分类的准确度,由于决策树的数量不同,计算机的运行周期不同,因此需要找一个合适的决策树的数量。为了寻找合适的决策树的数量,在固定其他参数不变的情况下,对烧结火焰图像进行分类实验,分类精度与决策树的数量关系如图 4 所示。图 4 表明,决策树数量小于 500 之前时,准确率不变,但随着决策树数量增加,虽然正确率稍有提升,但是运行效率降低。因此本文的决策树棵数选为 500。

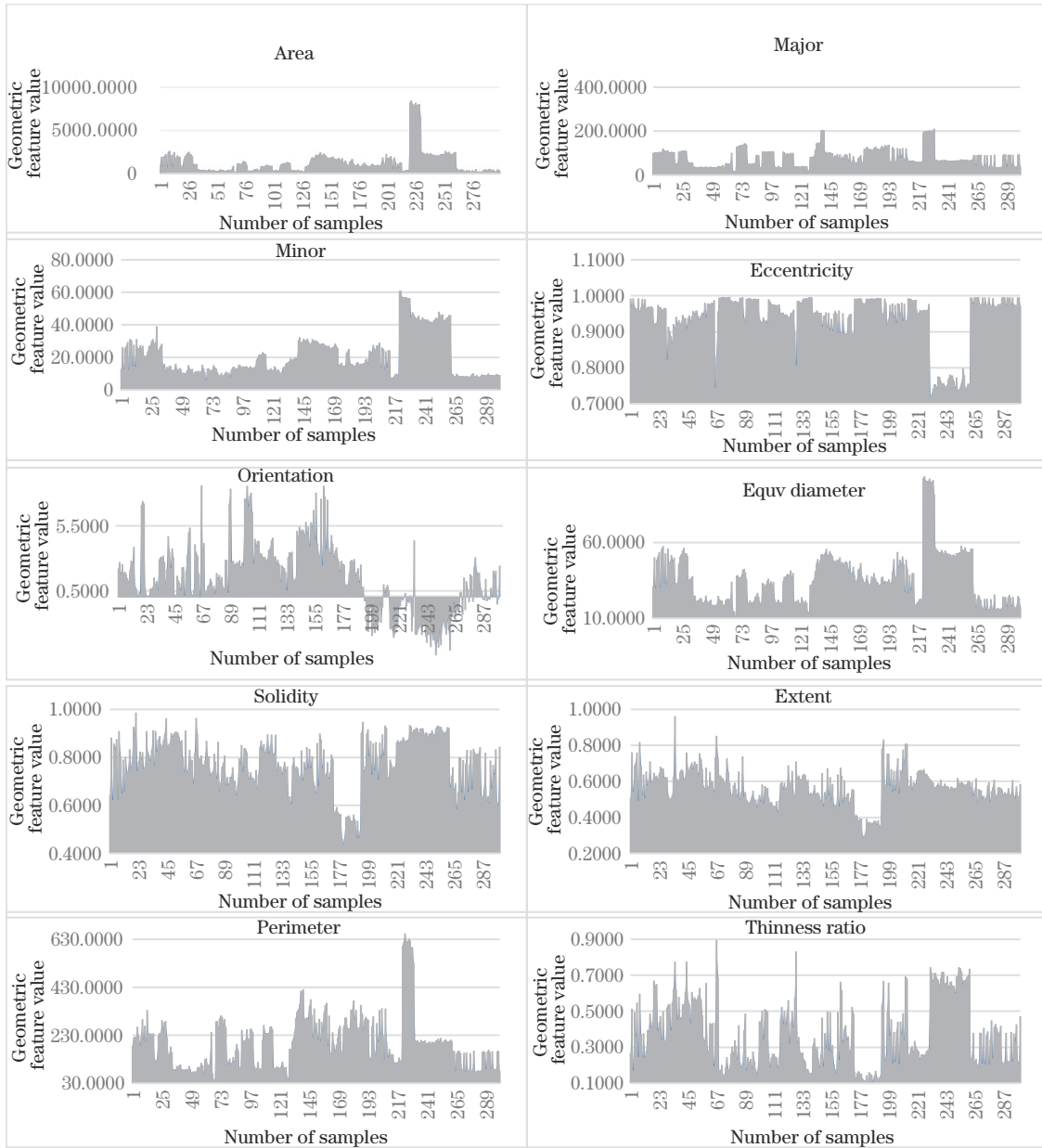


图 3 几何特征

Fig. 3 Geometric features

4. 3. 2 实验结果分析

为了方便比较和分析,设置决策树的棵数为 500,抽取属性个数为 7。为了验证改进随机森林算法的有效性,分别计算了测试集的准确率(accuracy)、精确率(precision)、召回率(recall)、F1 分数,公式分别为

$$P_{\text{precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (5)$$

$$P_{\text{recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (6)$$

$$F_1 = \frac{2 \times P_{\text{precision}} \times P_{\text{recall}}}{P_{\text{precision}} + P_{\text{recall}}}, \quad (7)$$

式中:TP 为真正例,FP 为假正例,TN 为真反例,FN 为假反例。改进随机森林算法对三个类别的精确率和

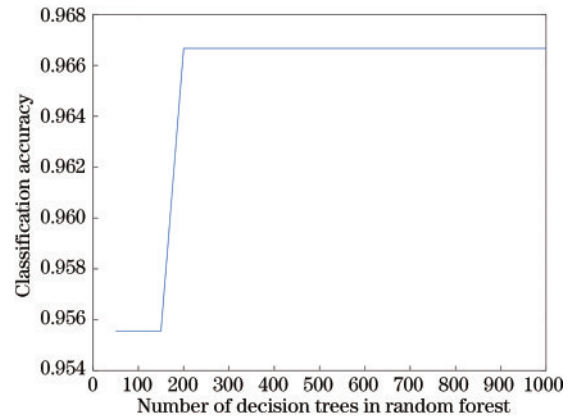


图 4 分类正确率与决策树数量的关系

Fig. 4 Relationship between classification accuracy and number of decision trees

召回率如表 6 和表 7 所示。总体测试集结果对比如表 8 所示。结果表明,相较于常规随机森林与支持向量机(SVM)(RBF 核函数),改进随机森林算法(K 均值)与改进随机森林算法(FCM)在数据集上取得了最优分类正确率,充分展示了改进随机森林算法对烧结火焰图像的分类能力。

表 6 改进随机森林算法对三类别图像的精确率

Table 6 Accuracy of improved random forest algorithm for three categories of images unit: %

| Method | Normal burning | Underburning | Overheating |
|--------------------------|----------------|--------------|-------------|
| Conventional method | 85.71 | 91.67 | 91.67 |
| SVM(RBF kernel function) | 95.24 | 91.67 | 91.67 |
| Improvement (K-means) | 97.62 | 95.83 | 100 |
| Improvement (FCM) | 97.56 | 100 | 95.83 |

表 7 改进随机森林算法对三类别图像的召回率

Table 7 Recall of improved random forest algorithm for three categories of images unit: %

| Method | Normal burning | Underburning | Overheating |
|--------------------------|----------------|--------------|-------------|
| Conventional method | 94.74 | 81.48 | 88 |
| SVM(RBF kernel function) | 97.56 | 88 | 91.67 |
| Improvement (K-means) | 97.62 | 95.83 | 100 |
| Improvement (FCM) | 97.56 | 92.31 | 100 |

表 8 总体结果对比

Table 8 Comparison of overall result

| Method | Accuracy / % | Precision / % | Recall / % | F1 value / % |
|--------------------------|--------------|---------------|------------|--------------|
| Conventional method | 88.89 | 89.68 | 88.07 | 88.87 |
| SVM(RBF kernel function) | 93.3333 | 92.86 | 92.41 | 92.63 |
| Improvement (K-means) | 97.78 | 97.82 | 97.82 | 97.82 |
| Improvement (FCM) | 96.67 | 97.80 | 96.62 | 97.21 |

在 300 张火焰图像中随机抽取 210 张作为训练集,剩下 90 张作为测试集。SVM(RBF 核函数)算法的分类结果较好,真实类别与预测类别精度较好,但是相较于 SVM(RBF 核函数),所提改进随机森林算法的精度得到一定程度的提升。常规随机森林与改进随机森林性能分析如图 5~7 所示,data 1 为正常燃烧,data 2

为过烧,data 3 为欠烧,category 1 指分类正确的类别,category 2 指分类错误的类别。从图中看出:常规随机森林错误分类较多,且最后投票阶段较为分散;改进随机森林错误分类稍少,投票集中,性能较好。分类结果表明,所提基于 K 均值和 FCM 的随机森林算法在一定程度上提高了随机森林分类正确率。

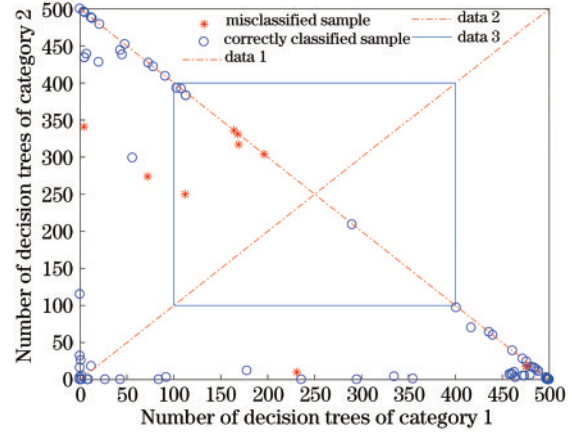


图 5 常规随机森林性能分析

Fig. 5 Performance analysis of conventional random forest

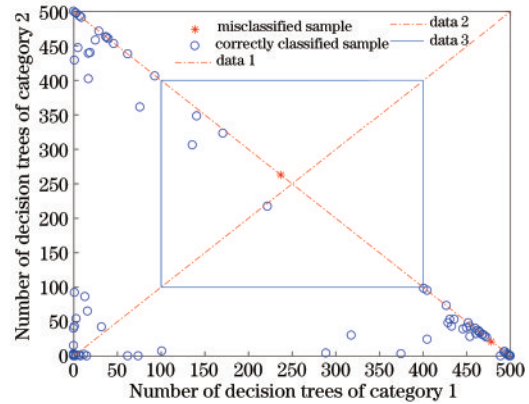


图 6 改进随机森林性能分析(K 均值)

Fig. 6 Performance analysis of improved random forest (K-means)

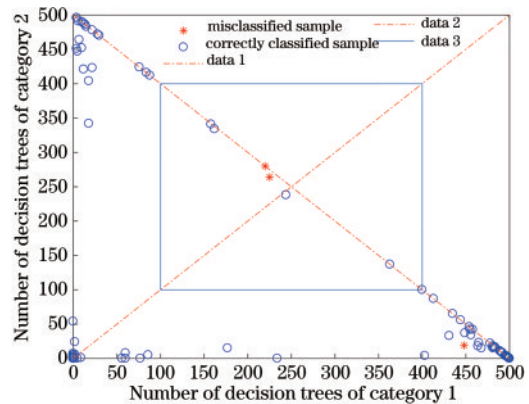


图 7 改进随机森林性能分析(FCM)

Fig. 7 Performance analysis of improved random forest(FCM)

5 结 论

烧结终点的预测是与烧结矿的产量、质量、成本和能源消耗密切相关的重要过程。由于烧结终点的滞后性及烧结终点无法直接测量,通过分析烧结图像中的几何特性,并通过随机森林算法对烧结终点进行预测。将烧结火焰图像的几何特征作为改进随机森林的输入,对烧结火焰图像进行分类预测,以达到预测烧结终点的目的。改进随机森林算法通过 K 均值聚类和 FCM 算法得到不同属性的重要度,并在分支过程中赋予相应的属性,构建新的决策树,降低了重要度较低的属性对整个模型的影响,提高了传统随机森林算法分类的准确率。

参 考 文 献

- [1] Patisson F, Bellot J P, Ablitzer D. Study of moisture transfer during the strand sintering process[J]. Metallurgical Transactions B, 1990, 21(1): 37-47.
- [2] 李黎. 基于数值模拟的烧结终点在线控制[D]. 长沙: 中南大学, 2011.
Li L. On line controlling of burning through point based on numerical simulation[D]. Changsha: Central South University, 2011.
- [3] 吴晓峰. 基于支持向量机的烧结终点预报与控制研究[D]. 上海: 上海大学, 2009.
Wu X F. Research on prediction and control of burning through point based on support vector machines in the sintering process[D]. Shanghai: Shanghai University, 2009.
- [4] 程佩哲. 基于监督学习的多样化推荐算法研究[D]. 济南: 山东大学, 2017.
Cheng P Z. Research on diversified recommendation algorithm based on supervised learning[D]. Jinan: Shandong University, 2017.
- [5] 张宗旺, 伯飞虎, 周凡, 等. 基于时间序列的烧结终点自适应预报模型[J]. 世界有色金属, 2018(18): 264-266.
Zhang Z W, Bo F H, Zhou F, et al. Self adaptive modeling of burn through point in sintering process based on time series[J]. World Nonferrous Metals, 2018(18): 264-266.
- [6] 周纪平. 基于体积模型的烧结终点预测及模糊控制研究[J]. 安徽工业大学学报(自然科学版), 2019, 36(1): 62-67.
Zhou J P. Research on forecasting and fuzzy control of sintering endpoint based on volume model[J]. Journal of Anhui University of Technology (Natural Science), 2019, 36(1): 62-67.
- [7] 汪森辉, 李海峰, 张永杰, 等. 基于改进的 AdaBoost.RS 算法的烧结终点预报分析[J]. 中国冶金, 2019, 29(10): 13-19.
Wang S H, Li H F, Zhang Y J, et al. Prediction and analysis of burning through point base on modified AdaBoost.RS algorithm[J]. China Metallurgy, 2019, 29(10): 13-19.
- [8] Rahman R, Otridge J, Pal R. IntegratedMRF: random forest-based framework for integrating prediction from different data types[J]. Bioinformatics, 2017, 33(9): 1407-1410.
- [9] 牛志华, 屈景怡, 吴仁彪. 基于 Spark 的分层子空间权重树随机森林算法[J]. 信号处理, 2017, 33(10): 1301-1307.
Niu Z H, Qu J Y, Wu R B. Random forest algorithm using stratified subspaces and weighted trees based on Spark[J]. Journal of Signal Processing, 2017, 33(10): 1301-1307.
- [10] 刘峻. 基于单图像超分辨率的约束随机森林算法[J]. 计算机工程与设计, 2017, 38(4): 970-975.
Liu J. Constrained random forest algorithm for single image super-resolution[J]. Computer Engineering and Design, 2017, 38(4): 970-975.
- [11] 吴辰文, 梁靖涵, 王伟, 等. 一种顺序响应的随机森林: 变量预测和选择[J]. 小型微型计算机系统, 2017, 38(8): 1762-1766.
Wu C W, Liang J H, Wang W, et al. Random forest algorithm for sequential response: prediction and selection of variables[J]. Journal of Chinese Computer Systems, 2017, 38(8): 1762-1766.
- [12] Xia J, Zhang S Y, Cai G L, et al. Adjusted weight voting algorithm for random forests in handling missing values[J]. Pattern Recognition, 2017, 69(9): 52-60.
- [13] 史金余, 杨泽宇, 谢兄. 基于模糊决策的随机森林算法[J]. 计算机工程与设计, 2020, 41(8): 2207-2212.
Shi J Y, Yang Z Y, Xie X. Algorithm of random forest based on fuzzy decision[J]. Computer Engineering and Design, 2020, 41(8): 2207-2212.
- [14] 王其, 曾万聘, 夏志平, 等. 基于随机森林算法的食源性致病菌拉曼光谱识别[J]. 中国激光, 2021, 48(3): 0311002.
Wang Q, Zeng W D, Xia Z P, et al. Recognition of foodborne pathogenic bacteria by Raman spectroscopy based on random forest algorithm[J]. Chinese Journal of Lasers, 2021, 48(3): 0311002.
- [15] 王猛, 张新长, 王家耀, 等. 结合随机森林面向对象的森林资源分类[J]. 测绘学报, 2020, 49(2): 235-244.
Wang M, Zhang X C, Wang J Y, et al. Forest resource classification based on random forest and object oriented method[J]. Acta Geodaetica et Cartographica Sinica, 2020, 49(2): 235-244.
- [16] Adnan M N, Islam M Z. Optimizing the number of trees in a decision forest to discover a subforest with high ensemble accuracy using a genetic algorithm[J]. Knowledge-Based Systems, 2016, 110: 86-97.
- [17] Drakakis G, Moledina S, Chomenidis C, et al. Decision trees for continuous data and conditional mutual information as a criterion for splitting instances[J]. Combinatorial Chemistry & High Throughput Screening, 2016, 19(5): 423-428.
- [18] Zhao W L, Deng C H, Ngo C W. K-means: a revisit[J]. Neurocomputing, 2018, 291: 195-206.
- [19] Xu Y G. Pattern recognition of traffic flows in elevator group control systems based on SVM[J]. Journal of South China University of Technology, 2005, 33(6): 32-35.