

doi: 10.3969/j.issn.2095-1744.2023.01.011

基于 CFS-KNN 的双能 X 射线废金属特征优选与分类

陈煜昊¹, 叶文华¹, 徐祥¹, 符杰²

(1. 南京航空航天大学机电学院, 南京 210016;

2. 江苏华宏科技股份有限公司, 江苏无锡 214400)

摘要: 在基于双能 X 射线透射技术的废金属分选中, 识别效果很大程度上受所选物料特征影响。已有废金属物料识别算法所用特征不够全面且各特征数据间冗余性较大, 识别准确度不高。针对该问题, 充分考虑特征间冗余性和特征与类别间关联性提出相关性特征选择方法(CFS), 在众多物料特征中确定由特征 I_H 、 I_L 、 x 、 T_{ML} 和 T_{MH} 组成的最优特征子集。通过采用不同物料特征组合方法进行对比实验, 并选择最优特征子集下 K-最近邻(KNN)作为最优分类模型。实验结果表明: 基于相关性特征选择 K-最近邻(CFS-KNN)的废有色金属物料识别分类方法与已有分类方法相比用相对较少的特征获得更高的识别准确度(96.13%)。

关键词: 特征选择; 相关性; K-最近邻; 废金属; 分类

中图分类号: TP274

文献标志码: A

文章编号: 2095-1744(2023)01-0086-07

Feature Optimization and Classification of Dual Energy X-ray Scrap Metal Based on CFS-KNN

CHEN Yuhao¹, YE Wenhua¹, XU Xiang¹, FU Jie²

(1. School of Mechatronic Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;

2. Jiangsu Huahong Technology Co., Ltd., Wuxi 214400, China)

Abstract: In the separation of scrap metal based on dual-energy X-ray transmission technology, the recognition effect is largely affected by the features of the selected materials. The features used in the existing scrap metal material recognition algorithms are not comprehensive enough, and the redundancy among feature data is large, so the recognition accuracy is not high. To solve this problem, this paper gives full consideration to the redundancy between features and the correlation between features and categories, and puts forward a correlation-based feature selection method(CFS) to determine the optimal feature subset consisting of features I_H 、 I_L 、 x 、 T_{ML} and T_{MH} among many material features. Comparative experiments were carried out by using different material feature combination methods, and K nearest neighbor(KNN) under the optimal feature subset was selected as the optimal classification model. The experimental results show that: compared with the existing classification methods, the identification and classification method of waste nonferrous metals materials based on the correlation-based feature selection and K nearest neighbor(CFS-KNN) has higher identification accuracy(96.13%) with relatively few features.

Key words: feature selection; correlation; K Nearest neighbor; scrap metal; classification

收稿日期: 2022-07-01

基金项目: 江苏省重点研发计划项目(BE2020779)

Fund: Supported by the Primary Research & Development Plan of Jiangsu Province(BE2020779)

作者简介: 陈煜昊(1999—), 男, 硕士研究生, 主要从事有色金属分选方面的研究。

通信作者: 叶文华(1965—), 男, 博士, 教授, 主要从事智能制造、数控装备技术、现代集成制造系统等研究。

引用格式: 陈煜昊, 叶文华, 徐祥, 等. 基于 CFS-KNN 的双能 X 射线废金属特征优选与分类[J]. 有色金属工程, 2023, 13(1): 86-92.

CHEN Yuhao, YE Wenhua, XU Xiang, et al. Feature Optimization and Classification of Dual Energy X-ray Scrap Metal Based on CFS-KNN[J]. Nonferrous Metals Engineering, 2023, 13(1): 86-92.

随着国内有色金属资源供需矛盾日益突出和碳达峰、碳中和目标不断推进,我国迫切需要一种新型自动化分选技术对报废量急速增长的有色金属进行种类分选,以此提高金属资源回收利用率。双能 X 射线透射分选技术^[1-2]由于具有不受金属废料形状和表面状况影响、检测厚度大等优点,是目前先进的废金属识别分选技术之一。

双能 X 射线透射分选技术主要基于单能谱射线的 R 值计算原理,即通过高、低能 X 射线透射物料后的能量强度和原始射线强度等参数,计算得一个与有效原子序数相关的 R 值^[3],并以此确定物料种类。该项技术最先运用于安检和矿石分选行业,如王学武^[4-5]通过 R 值曲线与特征量 x 分辨危险品种类,葛丁等^[6]利用 R - D 特征检测矿石金属含量,杨晨光^[7]基于 R 值与矿物成像特征对煤和矸石进行分选。

废金属分选系统中的双能 X 射线源多为连续能谱,其低能谱段能量更易被物料吸收,透射后射线平均能量偏高,存在射线硬化效应以及多频谱效应。此外,点光源 X 射线存在的扇形效应使得射线强度分布不均。这些因素都会影响物料的特征表达及 R 值计算。

为此,许多学者通过在 R 值基本算法的基础上引入其他物料特征来提高物料识别准确性。罗宗彬^[8]依据双能 R 值与低能 X 射线透射强度 I_L 特征,拟合金属铜与铝的 R - I_L 曲线,识别率达 90% 以上。王祺奥^[9]利用复化求积方法对 R - I_L 曲线拟合算法进行了优化。黄河等^[10]基于低能信号特征和双能 R 值提出区域分块识别法,提高了算法在废金属分选领域的适用范围。MAZOOCHI 等^[11]考虑厚度特征,利用复合辛普森数值方法改进 R 值算法,减小材料厚度对铝和塑料识别效果的影响。李伟毅等^[12]引入位置特征 x 并结合 BP 神经网络算法优化 R - I_L 曲线,减少了扇形效应的影响,识别率达 92.19%。

上述研究虽然使用了多种特征组合,不同程度地提高了物料识别率,但尚没有找到最优的特征组合,其识别率也有待提高。由于射线多频谱效应、硬化效应、扇形效应对不同特征的影响不同,特征间存在不同程度的耦合与相关,关系复杂。因此,本文采用基于相关性特征选择方法(Correlation-based Feature Selection,简称 CFS)寻找最优特征组合,同时采用在不同类别存在交叉或重叠情况下仍具有较高分类准确率的 K-最近邻法(K Nearest Neighbor,简称 KNN)进行废有色金属的识别分类。

1 CFS-KNN 特征选择与分类方法

1.1 基于 CFS 的最优特征选择

1.1.1 CFS 特征选择原理

基于相关性的特征选择算法(CFS)是一种用于机器学习的有效方法^[13]。CFS 算法基于相关性的启发式评估函数对特征进行评估,通过计算特征与类别、特征与特征之间的相关性建立最优特征子集,筛选对分类影响很小或者产生负面影响的特征。图 1 给出了 CFS 算法特征选择的原理。

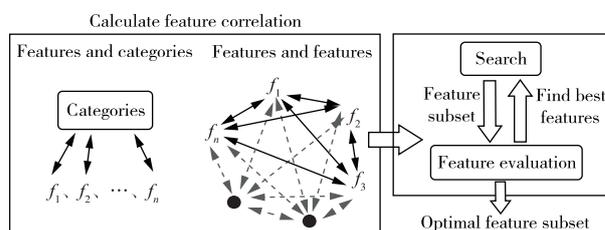


图 1 基于 CFS 算法的特征选择原理

Fig. 1 Principle of feature selection based on CFS algorithm

其中,特征评估使用如式(1)的启发式方程对每一个特征进行评估:

$$V = \frac{n\bar{r}_{cf}}{\sqrt{n+n(n-1)\bar{r}_{ff}}} \quad (1)$$

式中, \bar{r}_{cf} 为特征与类别之间的平均相关系数, \bar{r}_{ff} 为特征与特征之间的平均冗余性系数, V 是包含 n 个特征的特征子集的启发式最优值, V 值越大,表明特征与类别的相关程度越高,特征与特征之间的不相关程度越高。本文研究对象为废金属物料特征,该评价指标能有效地给出各特征对于金属分类的贡献度,并对特征进行筛选。通过采用对称不确定性度量方法(SU)来衡量特征之间的相关性,基于 SU 计算相关系数 r 公式如下:

$$r = \frac{2IG(X,Y)}{H(X)+H(Y)} \quad (2)$$

式中 X, Y 是两个特征变量, $IG(X, Y)$ 是特征变量 X, Y 之间的互信息, $H(X), H(Y)$ 分别是 X, Y 的信息熵。 $H(X), IG(X, Y)$ 计算公式如下:

$$H(X) = - \sum_{x \in X} p(x_i) \log p(x_i) \quad (3)$$

$$IG(X, Y) = \sum_{i=1}^m \sum_{j=1}^n p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \quad (4)$$

其中信息熵 $H(X)$ 取决于每个金属特征变量 X 的概率分布,可通过计算废金属特征数据 $\{x_1, x_2, \dots, x_i\}$ 中各特征值出现的概率 $p(x_i)$ 求得。同时,

互信息有效衡量了两个金属特征变量间共有的信息含量。当给定废金属特征 X, Y , 通过计算两金属特征变量的联合概率分布 $p(x_i, y_i)$ 和边缘概率密度 $p(x_i)$ 可求得 $IG(X, Y)$ 。若两金属特征变量相互独立, 他们互信息最小为 0。反之, 两特征相互依赖程度越高, 互信息值也越大。

1.1.2 CFS 特征选择流程

CFS 算法可通过前向选择、后向消除和最佳优先三种特征搜索方式实现^[14]。前两种方式特征集分别始于空集和全集, 通过不断增加或减少特征数量直至得到最优特征集。而最佳优先特征搜索方式是对前两者的改进。为获得较好的废金属特征选择效果, 本文采用基于前向选择的最佳优先搜索方式实现 CFS 算法。记 $F = \{f_1, f_2, \dots, f_m\}$ 为废金属特征全集, $C = \{c_1, c_2, \dots, c_k\}$ 为样本类别(铜、铝), $\{a_i, c_i \mid i = 1, 2, \dots, N\}$ 为样本容量为 N 的数据集, S 为已选特征子集, T 为备选特征子集, $R = F - S$ 为剩余特征集, 其中 a_i 是具有 m 维特征的样本。特征选择时, S 开始于空集, 从 F 中选择 V 值最大的一个特征放入 S 中, 根据公式(1)和(2)可得:

$$S = \Phi U \operatorname{argmax}_{f_i \in F, n=1, c_i \in C} (V(c_i, f_i, n)) \quad (5)$$

然后, 从 R 中选出 V 值最大的特征放入 S 中获得 T , 比较 V_S 和 V_T 的大小, 若 V_T 大于 V_S , 则将 S 更新为 T 。其中, V_S 和 V_T 分别表示特征集合为 S 和 T 时由公式(1)计算得到的启发式最优值。CFS 算法整体流程如图 2 所示。

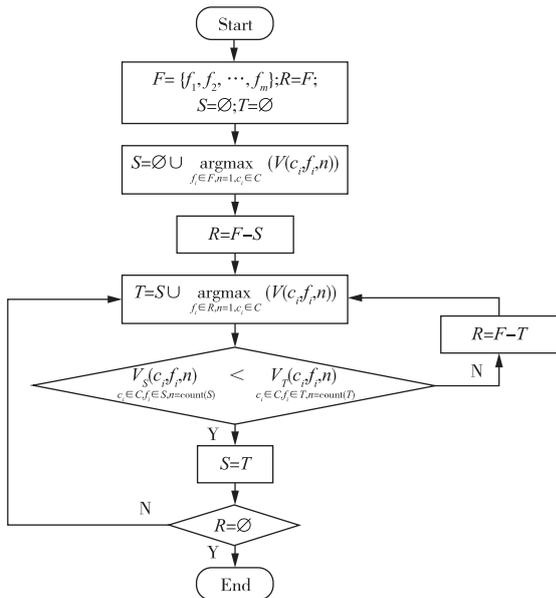


图 2 CFS 算法流程图

Fig. 2 Flowsheet of CFS algorithm

1.2 基于 KNN 的分类算法

KNN 算法是基于原始已知标签样本集的机器学习算法^[15], 根据距未知标签新样本最近的 k 个已知标签样本, 确定新样本所属类别。影响 KNN 分类算法性能的因素包括 k 值、距离计算方法和判断准则。其中, k 值对分类性能有直接影响, 然而其取值目前并无统一依据, 本文选择从 $k=1$ 开始增加, 同时采取交叉验证法获得最优的 k 值。样本在特征空间中的距离通常采用欧氏距离。判断准则一般使用投票表决策策略, 即在所得最邻近的 k 个样本中, 属于哪个标签类别的数量最多, 被测样本即属于哪一类。

通过 CFS 算法获得样本的最优特征子集 $S_R = \{f_1, f_2, \dots, f_s\}$, 作为样本的 s 维特征, 并结合 KNN 分类算法, 获得基于 CFS-KNN 的分类模型。记 l 为未知类别标签的待分类样本, l_c 为该样本实际属于类别的标签值, $\{l_i, c_i \mid i = 1, 2, \dots, N_{\text{label}}\}$ 为样本容量为 N_{label} 的已知类别标签的数据集, 且 l 和 l_i 均是具有 S_R 所包含的 s 维特征的样本。CFS-KNN 分类方法的实现过程如图 3 所示。

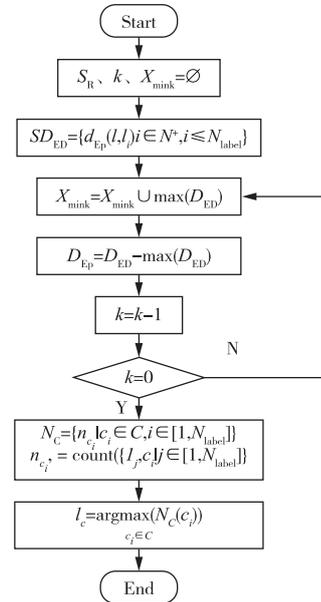


图 3 CFS-KNN 分类流程

Fig. 3 CFS-KNN classification process

2 基于废金属物料的特征选择与分类模型建立

2.1 样本数据和分析流程

2.1.1 废金属物料样本特征全集

根据已有的识别算法和文献研究, 通过对初始

特征量进行组合变换,得到一系列具有实际物理意义的特征量,表 1 给出了每个样本中各个特征的含义。

2.1.2 废金属物料数据集建立

使用机器学习中的分类模型对废有色金属废料进行识别时,为提高模型识别准确率,往往需要大量样本数据对模型进行训练。为提高单张图像的样本

采集效率,同时充分考虑物料位置特征 x 及点光源 X 射线扇形效应对结果的影响,本文通过人工摆放物料的方式,将不同形状、大小的铝、铜两类金属物料摆放于探测器不同位置处进行图像采集,如图 4 所示。最终获得了 19 525 个有效样本数据,其中铜金属的样本有 9 341 个,铝金属样本有 10 184 个,两种样本数量比接近 1 : 1,样本总体表现均衡。

表 1 废金属物料样本特征说明

Table 1 Description of features of scrap metal samples

Exp. No.	Features	Description
1	I_L	Transmission value of low-energy X-ray
2	I_H	Transmission value of high energy X-ray
3	$\Delta I_L = I_{L0} - I_L$	Attenuation of low-energy X-ray intensity
4	$\Delta I_H = I_{H0} - I_H$	Attenuation of high-energy X-ray intensity
5	$\alpha_H = \mu_{mH}\rho t = \ln(I_{H0}/I_H)$	α features of high energy X-ray
6	$\alpha_L = \mu_{mL}\rho t = \ln(I_{L0}/I_L)$	α features of low energy X-ray
7	$\alpha_D = \alpha_H - \alpha_L$	α difference between high and low energy X-ray
8	$R = \alpha_L/\alpha_H$	R value
9	$R_L = \alpha_D/\alpha_L$	R value of low energy X-ray
10	$R_H = \alpha_D/\alpha_H$	R value of high energy X-ray
11	$S_H = (I_{H0} - I_H)/I_{H0}$	Transparency of high energy X-ray
12	$S_L = (I_{L0} - I_L)/I_{L0}$	Transparency of low energy X-ray
13	$T_{MH} = (I_H - I_{H0})$	Transmittance of high energy X-ray
14	$T_{ML} = (I_L - I_{L0})$	Transmittance of low energy X-ray
15	x	Material location



图 4 铝、铜废金属物料样本图

Fig. 4 Sample diagram of aluminum and copper scrap metal materials

2.2 废金属物料识别分类

2.2.1 废金属物料的特征选择

基于 CFS 算法对上一小节建立的废金属物料数据集进行特征选择。根据公式(1)计算特征与特征、特征与类别之间的对称不确定性相关系数。表 2 记录了基于 CFS 算法的特征选择计算过程。从表 2 中可知,单个特征 V_s 值最大的为 I_H 特征,值为 0.100 8, CFS 算法最终得到 V_s 的最大值为 0.110 7,其对应的特征子集由 I_H 、 I_L 、 x 、 T_{ML} 和 T_{MH} 共 5 个特征组成,其中, I_H 、 I_L 、 x 分别为低能、高能

X 射线透射值、物料位置, T_{ML} 和 T_{MH} 为低能、高能 X 射线的透射比。

2.2.2 CFS-KNN 识别分类效果

通过 CFS 算法确定了最优特征子集,还需要确定 KNN 算法的参数 k 、距离计算方法和判断准则。

对于参数 k ,本文采取交叉验证法最终确定 k 的值为 3;样本在特征空间中的距离采用欧氏距离计算方法;在所建立的废铜、铝金属物料数据集中,铜和铝的废料样本比例接近 1 : 1,因此采用投票表决策策略作为模型的判断准则。

表 2 基于相关性的特征选择过程

Table 2 Feature selection process based on correlation

Number	Feature subset	\bar{r}_{cf}	\bar{r}_{ff}	V_s	Number	Feature subset	\bar{r}_{cf}	\bar{r}_{ff}	V_s
1	I_H	0.100 8	1.000 0	0.100 8	1	α_D	0.036 9	1.000 0	0.036 9
1	I_L	0.096 4	1.000 0	0.096 4	2	I_H, I_L	0.098 6	0.831 7	0.103 0
1	x	0.085 9	1.000 0	0.085 9	3	I_H, I_L, x	0.094 3	0.610 8	0.109 6
1	ΔI_H	0.080 6	1.000 0	0.080 6	4	$I_H, I_L, x, \Delta I_H$	0.090 9	0.687 1	0.103 9
1	T_{ML}	0.077 8	1.000 0	0.077 8	4	$I_H, I_L, x, \Delta I_{Hx}$	0.090 6	0.671 8	0.104 4
1	S_L	0.077 8	1.000 0	0.077 8	4	I_H, I_L, x, T_{ML}	0.090 2	0.558 4	0.110 3
1	T_{MH}	0.077 5	1.000 0	0.077 5	5	I_H, I_L, x, T_{ML}, S_L	0.087 7	0.586 8	0.107 2
1	S_H	0.077 5	1.000 0	0.077 5	5	$I_H, I_L, x, T_{ML}, T_{MH}$	0.087 7	0.534 5	0.110 7
1	ΔI_L	0.074 6	1.000 0	0.074 6	6	$I_H, I_L, x, T_{ML}, T_{MH}, S_H$	0.086 0	0.560 6	0.108 0
1	α_H	0.073 8	1.000 0	0.073 8	6	$I_H, I_L, x, T_{ML}, T_{MH}, \alpha_H$	0.085 4	0.581 2	0.105 8
1	α_L	0.071 8	1.000 0	0.071 8	6	$I_H, I_L, x, T_{ML}, T_{MH}, R$	0.082 3	0.517 0	0.106 4
1	R	0.055 3	1.000 0	0.055 3	6	$I_H, I_L, x, T_{ML}, T_{MH}, R_H$	0.081 9	0.517 0	0.106 0
1	R_H	0.053 3	1.000 0	0.053 3	6	$I_H, I_L, x, T_{ML}, T_{MH}, R_L$	0.081 8	0.504 4	0.106 8
1	R_L	0.052 6	1.000 0	0.052 6	6	$I_H, I_L, x, T_{ML}, T_{MH}, \alpha_D$	0.079 2	0.527 7	0.101 7

为了降低样本选取造成的随机偏差,通过对数据集样本采用 10 次十折交叉验证法计算分类准确度的均值。具体过程为:在一次十折交叉过程中,将数据集随机划分为 10 个等份,任选 9 份作为训练集,剩余 1 份作为验证集,如此可获得 10 组不同的训练-验证集;最后,重复 10 次十折交叉验证过程,对获得的 100 个分类准确度取平均值作为模型最终的验证结果。表 3 给出了基于 CFS-KNN 模型分类结果,其中,分类时间是对单个样本进行分类所耗费的平均时间。

表 3 基于 CFS-KNN 模型分类结果

Table 3 Classification results based on CFS-KNN model

Model	Classification accuracy/%	Time of classification/s
CFS-KNN	96.13	0.004 9

3 实验验证与对比讨论

为了验证基于 CFS-KNN 的分类方法在废有色金属物料识别分类中的优越性,本节分别将已有的废有色金属物料识别方法中使用到的特征集合、表 1 给出的物料特征全集和基于 CFS 算法选择的最优特征子集作为特征组合,使用常见机器学习分类模型进行训练,评估各个方法对废有色金属物料的分类能力,选出最优的特征子集与分类模型,并将其与现有废金属物料分选方法进行比较。

实验包括三个部分:1)比较不同的特征子集在相同模型(KNN)下的分类效果,验证基于 CFS 算法选择的特征子集是否能够提升废有色金属物料的分类性能、最优特征子集是否具最高的识别准确度;

2)基于最优特征子集废有色金属物料识别的常见机器学习分类算法比较,验证 KNN 分类模型用于最优特征子集较其他分类算法具有最优分类性能;3)比较 CFS-KNN 分类算法与已有的废有色金属识别分类算法,验证基于 CFS-KNN 的废有色金属物料识别分类模型具有更高的分类准确率。

3.1 不同特征子集在 KNN 模型下的分类效果比较

本节根据已有的常见特征,结合特征全集和基于 CFS 选择出的特征子集,给出表 4 所示的 7 组特征子集作为分类模型的输入。

同样,对数据集样本采用 10 次十折交叉验证法计算分类准确度的均值,用以降低因样本选择带来的随机误差,获得 7 种特征子集在 KNN 模型下的分类准确度,如表 5 所示。分析表 5 可得以下结论:

1)比较组合 2 与 3,可以发现位置特征 x 减少了 X 射线扇形效应对金属识别的影响,显著地提高了模型分类精度;

2)在前六组特征子集中,组合 6 即全部的 15 个特征,具有最高的分类准确度(95.22%)。然而,组合 3 和组合 4 仅使用少量特征,分类准确度同样高达 95.15%和 94.18%。因此,特征全集中可能包含一定数量的冗余性特征;

3)组合 7 为基于 CFS 算法选择出最优特征子集,在七组特征子集中具有最高分类准确度(96.13%)。相比于组合 6,CFS 算法去除了特征全集中对分类性能贡献较小的特征,提高了模型分类准确度。

表 4 废金属物料特征的几种组合

Table 4 Several combinations of features of waste metal materials

Exp. No.	Number of features	Features
1	2	R, I_L
2	2	I_H, I_L
3	3	I_H, I_L, x
4	4	I_H, I_L, x, T_{ML}
5	6	$I_H, I_L, x, T_{ML}, T_{MH}, \alpha_D$
6	17	All features
7	5	$I_H, I_L, x, T_{ML}, T_{MH}$

表 5 7 种特征子集在 KNN 模型下的分类准确度

Table 5 Classification accuracy of seven feature subsets under KNN model

/%

Classification model	1	2	3	4	5	6	7
KNN	85.73	87.77	95.15	94.18	93.89	95.22	96.13

3.2 基于最优特征子集在不同分类模型下的分类效果比较

3.1 节结果显示在 KNN 分类模型下第 7 组最优特征子集拥有最高的分类准确度。使用随机森林

(RF)算法、朴素贝叶斯(NB)算法、支持向量机算法(SVM)和 BP 神经网络(BPNN)算法这另外四种机器学习算法计算最优特征子集下模型分类准确度,结果如表 6 所示。

表 6 最优特征子集下 5 种模型分类准确度

Table 6 Classification accuracy of five models under optimal feature subset

/%

Feature subset	RF	KNN	NB	SVM	BPNN
Optimal feature subset	92.63	96.13	86.61	86.29	92.91

在五种典型的分类模型中, KNN 模型分类准确度最高达 96.13%, 这表明 KNN 模型较其他分类模型在对废金属物料特征数据样本的分类上具有较优的性能。

3.3 废有色金属物料识别算法比较

使用同样的数据集对现有基于双能 X 射线的废有色金属识别分类算法进行试验, 记录各个现有算法的分类准确度, 并与基于 CFS-KNN 的分类算法进行对比, 如表 7 所示。

表 7 五种物料识别分类算法比较

Table 7 Comparison of five classification algorithms for material identification

Recognition algorithm	Features	Classification accuracy/%
$R - I_L$ curve fitting algorithm ^[8]	R, I_L	81.32
$R_\psi - I_L$ curve fitting algorithm ^[9]	R_ψ, I_L	83.27
Regional block recognition algorithm ^[10]	R, I_L	81.03
BPNN classification algorithm ^[12]	α_H, α_D, x	91.70
CFS-KNN classification algorithm	$I_H, I_L, x, T_{ML}, T_{MH}$	96.13

分析表 7 可以得出以下结论:

1) 在 $R - I_L$ 曲线拟合算法^[8]基础上, 基于复化求积的 $R_\psi - I_L$ 曲线拟合算法^[9]通过数值积分的方式, 小幅提高了算法的分类准确度, 然而其算法的复杂度较高;

2) 区域分块识别算法^[10]的分类准确度稍低于 $R - I_L$ 曲线拟合算法^[8], 这是由于用于划分各分块区域的临界值存在一定的误差, 导致分类精度下降;

3) 在五种废金属物料识别分类算法中, 基于 CFS-KNN 模型分类算法具有最高的分类准确度。

综合表 5、6 和表 7 分析可知, 通过对比不同特征子集在 KNN 分类算法下的分类性能、最优特征子集下不同分类模型的识别分类效果以及基于 CFS-KNN 模型分类算法与其他现有废有色金属物料的分类准确度差异, 发现基于 CFS-KNN 模型分类算法可通过较少的特征获得更高的分类性能。

4 结论

在传统的废金属物料识别分类方法的基础上, 针对现有方法无法有效选择物料最优种类特

征造成分类准确度不够理想这一现状,通过使用最佳优先搜索策略的 CFS 算法对物料特征进行优选,选择出特征间冗余性较低的特征子集,并结合 KNN 分类算法,提出了一种基于 CFS-KNN 的识别分类方法。在对废铜、铝金属的识别分类过程中,CFS 确定了 I_H 、 I_L 、 x 、 T_{ML} 和 T_{MH} 五个特征作为最优特征子集,结合 KNN 分类算法,以较短的运算时间获得了 96.13% 的识别准确度;并通过对比使用最优特征子集训练的 NB、BPNN、SVM 和 RF 四种模型的分类性能、对比不同的废有色金属物料识别算法,验证了基于 CFS-KNN 分类方法的有效性和优越性。

参考文献:

- [1] MACOVSKI A, ALVAREZ R E, CHAN L H, et al. Energy dependent reconstruction in X-ray computerized tomography[J]. Computers in Biology and Medicine, 1976, 6(4): 0-336.
- [2] ALVAREZ R E, MACOVSKI A. Energy-selective reconstructions in X-ray computerized tomography[J]. Canadian Metallurgical Quarterly, 1976, 21(5): 733-744.
- [3] 原培新,赵秉军,孙丽娜,等. X 射线安检设备中双能量方式下 R 值的提取[J]. 机械与电子, 2005(9): 25-27.
YUAN Peixin, ZHAO Bingjun, SUN Lina, et al. Obtain R on dual-energy in X-ray security inspection equipment [J]. Machinery & Electronics, 2005 (9): 25-27.
- [4] WANG X W, LI J M, KANG K J, et al. Material discrimination by high-energy X-Ray dual-energy imaging[J]. High Energy Physics & Nuclear Physics and Nuclear Physics, 2007, 31(11): 1076-1081.
- [5] 郭小平,王学武,谢亚丽,等. 基于贝叶斯决策的双能 X 射线材料分辨算法研究[J]. 中国体视学与图像分析, 2003, 8(3): 162-165.
WU Xiaoping, WANG Xuewu, XIE Yali, et al. A practical material discrimination method with dual X-ray energy based on bayesian decision [J]. Chinese Journal of Stereology and Image Analysis, 2003, 8(3): 162-165.
- [6] 葛丁,梁殿印. 基于 X 射线透射的矿石品位检测方法研究[J]. 有色金属(选矿部分), 2019(4): 87-93.
GE Ding, LIANG Dianyin. Research on detection method of ore grade based on X-ray transmission[J]. Nonferrous Metals (Mineral Processing Section), 2019(4): 87-93.
- [7] 杨晨光. 煤矸中矿物组分在 X 射线下的识别规律[D]. 淮南:安徽理工大学, 2020.
- YANG Chenguang. Recognition law of mineral components in coal gangue under X-ray[D]. Huainan: Anhui University, 2020.
- [8] 罗宗彬. 报废汽车破碎铜铝 X 射线识别及其分选装备研究[D]. 昆明:昆明理工大学, 2020.
LUO Zongbin. Research on X-ray recognition and sorting equipment of scrapped automobile broken copper and aluminum [D]. Kunming: Kunming University of Science and Technology, 2020.
- [9] 王祺奥. 基于双能 X 射线透射的废金属识别与分离控制技术研究[D]. 南京:南京航空航天大学, 2017.
WANG Qiao. Research on scrap metal identification and separation control technology based on dual energy X-ray transmission[D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2017.
- [10] 黄河,叶文华,熊田忠,等. 基于双能 X 射线透射的区域分块废有色金属识别算法[J]. 机械制造与自动化, 2019, 48(4): 26-29.
HUANG He, YE Wenhua, XIONG Tianzhong, et al. Zone division scrap non-ferrous metal recognition algorithm based on dualenergy X-ray transmission[J]. Machine Building & Automation, 2019, 48(4): 26-29.
- [11] MAZOOCHI A, RAHMANI F, DAVANI F A, et al. A novel numerical method to eliminate thickness effect in dual energy X-ray imaging used in baggage inspection[J]. Nuclear Instruments & Methods in Physics Research, 2014, 763(1): 538-542.
- [12] 李伟毅,叶文华,熊田忠. 基于 BP 神经网络的双能 X 射线透射的金属识别算法[J]. 有色金属工程, 2020, 10(8): 124-130.
LI Weiyi, YE Wenhua, XIONG Tianzhong. Metal identification algorithm based on BP neural network for dual energy X-ray transmission [J]. Nonferrous Metals Engineering, 2020, 10(8): 124-130.
- [13] HALL M A. Correlation-based feature selection for machine learning [J]. Morgan Kaufmann Publishers, 2000(29): 359-366.
- [14] 刘华文. 基于信息熵的特征选择算法研究[D]. 长春:吉林大学, 2010.
LIU Huawen. A study on feature selection algorithms using information entropy [D]. Changchun: Jilin University, 2010.
- [15] GUO G, WANG H, BELL D, et al. KNN model-based approach in classification [C]//OTM Confederated International Conferences "On the Move to Meaningful Internet Systems", Springer, Berlin, Heidelberg, 2003.