Minerals Engineering 153 (2020) 106201

Contents lists available at ScienceDirect

Minerals Engineering

journal homepage: www.elsevier.com/locate/mineng

A hybrid feature selection method for production condition recognition in froth flotation with noisy labels

Xiaojun Zhou[a,b,*], Qi'an Wang[a], Rundong Zhang[a], Chunhua Yang[a]

[a] School of Automation, Central South University, Changsha 410083, China
[b] The Peng Cheng Laboratory, Shenzhen, Changsha 518000, China

ARTICLE INFO

Keywords:
Feature selection
Froth flotation
Binary STA
Machine vision
Production condition recognition

ABSTRACT

Production condition recognition plays a major role in the froth flotation process control, and its accuracy has great influence on the performance of froth flotation. In order to improve the performance, lots of features are extracted from froth images; however, not all of these features are useful. There are some redundant and irrelevant features among them, which may increase the computational cost and the difficulty of training classifier, or even reduce the recognition accuracy. In order to resolve this issue, a feature selection method named mRMR-BSTA is proposed in this paper to find the best feature subset with most useful and less redundant features. This proposed method is comprised of two phases: the filter phase based on minimal-redundancy-maximal-relevance (mRMR) criterion and the wrapper phase based on binary state transition algorithm (BSTA), and it is applied to the production condition recognition in gold-antimony froth flotation. Least squares support vector machine (LSSVM) is applied to recognition and used as a black box to evaluate the quality of selected features in wrapper phase. Especially, for removing the noisy data with wrong labels caused by the subjectivity of workers, a pre-processing approach based on fuzzy C-means (FCM) is proposed in this paper. Finally, the most efficient feature subset in froth flotation is selected, including hue, mean of blue, relative red component, coarseness, and high frequency energy. The production condition is classified into eight classes with high accuracy successfully, and the effectiveness of the proposed method for production condition recognition in froth flotation is validated.

1. Introduction

Mineral resources are the material basis of human society for survival and development, which can supply plenty of energy, industrial materials and agricultural capital goods. However, material resources cannot be utilized directly, so mineral processing is necessary. Beneficiation is an essential part of mineral processing (Ejtemaei et al., 2014), and froth flotation is widely used in beneficiation to extract valuable minerals from raw ore (Wang et al., 2018). However, due to the complex physic-chemical separation process of flotation, there lacks generally useful mathematical models. Even worse, in the actual industry process, the key indicators reflecting the flotation performance like concentrate grade are hard to be detected (Gui et al., 2013). It is generally known that the flotation production condition is directly reflected on the froth surface appearance, and the characteristics of the froth surface is changing with the variation of the flotation production condition states. Recently, it is known as an important indicator reflecting the flotation performance. In the industry process, experienced workers adjust the practical operation variables of flotation process like the chemical reagent relying on their observation of the features of the froth surface.

However, considering the subjectivity of workers and the lack of objective criteria for judging the froth states, which always leads to low levels of labor productivity and high levels of labor intensity, it is difficult to achieve the optimal production condition. Therefore, it usually results in the low utilization of the ore, huge waste of reagents, fluctuation of production conditions and unqualified productions. Recently, machine vision-based method is the development trend of automation for flotation process (Xu et al., 2016). It can utilize the features extracted from froth images to achieve the recognition of froth states objectively, and then monitor the production, so as to improve the labor productivity. Fig. 1 shows froth images in different levels, and it is clear that differences exist among the features of froth surface under different production conditions.

Recently, in order to improve the performance of production condition recognition in froth flotation, more and more features have been extracted from froth images for production condition recognition in various studies (Xu et al., 2016; Zhao et al., 2015; Liu et al., 2013; Peng

* Corresponding author.
E-mail addresses: michael.x.zhou@csu.edu.cn (X. Zhou), wangqi_an@csu.edu.cn (Q. Wang), ychh@csu.edu.cn (C. Yang).

https://doi.org/10.1016/j.mineng.2020.106201
Received 20 April 2019; Received in revised form 20 November 2019; Accepted 11 January 2020
Available online 01 May 2020
0892-6875/

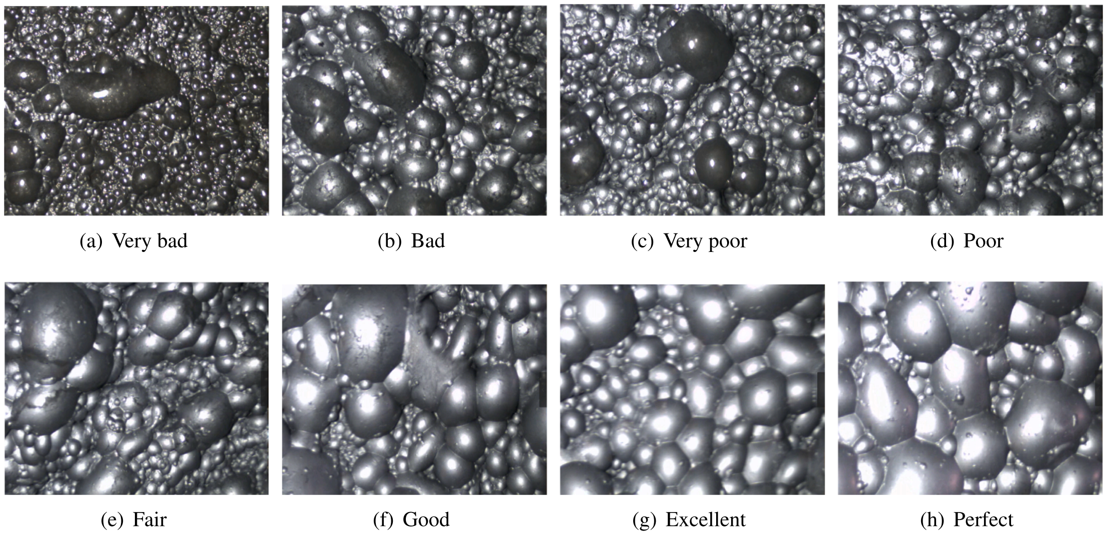

Fig. 1. Froth images under different production conditions.

et al., 2016; Zhao et al., 2017; Wang et al., 2018). So far, there is no attempt to apply all these extracted features to recognize production condition. However, so many features of froth flotation are extracted, such as the mean of red, the mean of green, the mean of blue, the mean of gray and so on. They contain some redundant and irrelevant features, which could increase the difficulty of model training and reduce the performance of production condition recognition in froth flotation (Huang et al., 2018).

Feature selection (FS) could deal with these problems well. FS aims to find a feature subset with most relevant and least redundant features from the original feature set, so as to avoid degradation of learning performance (Jović et al., 2015; Li et al., 2018). FS could bring two apparent advantages for the production condition recognition in froth flotation: (1) it can select the most useful features which can improve the accuracy of the recognition; (2) it can remove the relevant, irrelevant and redundant features and help learning algorithm to build a simpler and more comprehensible model with less training time, and also reduce the time of recognition. Machine vision-based method recognizes flotation production condition according to the features of froth. It is clear that, if the feature set used is effective, it would be easier to classify different production conditions. FS helps to obtain a good feature set for flotation to improve the performance of recognition. However, it is still a challenge to obtain the best feature subset. When facing a feature set with n features, there are 2^n possible combinations. and if n is large, finding all possible feature subset is computationally expensive. Thus, searching a global optimal feature subset is non-deterministic polynomial hard (NP-hard) (Inbarani et al., 2014).

At present, many approaches are proposed to handle FS problems. According to their selection strategy, they are broadly categorized into two kinds (Li et al., 2018): the wrapper based method and the filter based method. Wrapper based method utilizes the performance of classifier (classification accuracy) to evaluate the quality of each subset, hence it can ensure high classification accuracy for the particular classifier. However, it needs expensive computational cost. Alternatively, the filter based method requires an objective criterion, and it evaluates features by this criterion. Hence, it need much less cost than wrapper based method. However, the filter based FS methods select feature subsets without using the information of classifier, and it usually results in the uncertainty of classification accuracy using the selected feature subset. In order to improve the performance of FS and combine the advantages of filter method and wrapper method, a hybrid FS method is proposed in this paper. It combines a filter method based on minimal-redundancy-maximal-relevance criterion (mRMR) and a wrapper method based on binary state transition algorithm (BSTA), because of the comprehensive consideration of mRMR (Li et al., 2012), and the strong global search ability of BSTA (Zhou et al., 2018). BSTA is a specialized form of state transition algorithm (STA) (Zhou et al., 2012) which has shown the excellent performance in practical problems (Zhou et al., 2016; Huang et al., 2018; Zhou et al., 2018; Han et al., 2018; Zhou et al., 2018). Meanwhile, in order to improve the performance of BSTA in FS problems, a new substitute operator of BSTA is proposed in this paper.

It is worth noting that the dataset which comes from factory is labeled by workers. In practice, the workers always estimate the condition with a high subjectivity. Inevitably, there are some wrong labels in dataset which may influence the model training and reduce the accuracy of the final classification. So far, researchers habitually ignored the possibility of the existing of wrong labels in gold-antimony flotation process dataset. In this paper, a preprocessing for noisy dataset based on fuzzy C-means (FCM) clustering method is proposed to remove the outliers in the dataset to improve the classification accuracy of recognition, due to the strong performances in data mining field of FCM (Havens et al., 2012). Currently, preprocessing method could be used after FS or before FS (Yin and Gai, 2015). However, considering that the noisy features may cause a high difficulty of removing the outliers, the preprocessing method is adopted after FS (before classification) in this paper.

In summary, the general process of gold-antimony froth flotation is shown in Fig. 2, and the processes in the rectangles highlight the main works in this paper, including FS, preprocessing and classification. The contributions of this paper are summarized as follows:

(1) A hybrid FS method based on mRMR and BSTA is proposed to find an optimal feature subset. Meanwhile, in order to improve the global search ability and robustness performance of BSTA, a new substitute operator of BSTA is proposed in this paper, which is based on the feature weights standing for the importance of features.
(2) In order to remove the noisy labels in dataset, a data preprocessing method based on FCM clustering is proposed in this paper.
(3) The proposed FS method, mRMR-BSTA, is applied into the production condition recognition of gold-antimony froth flotation to reduce the redundancy and noise of the extracted 38 features. 5 efficient features are selected and the production condition recognition of gold-antimony froth flotation is categorized into 8 classes with high classification accuracy successfully.

The rest of paper is organized as follows. Section 2 gives a brief

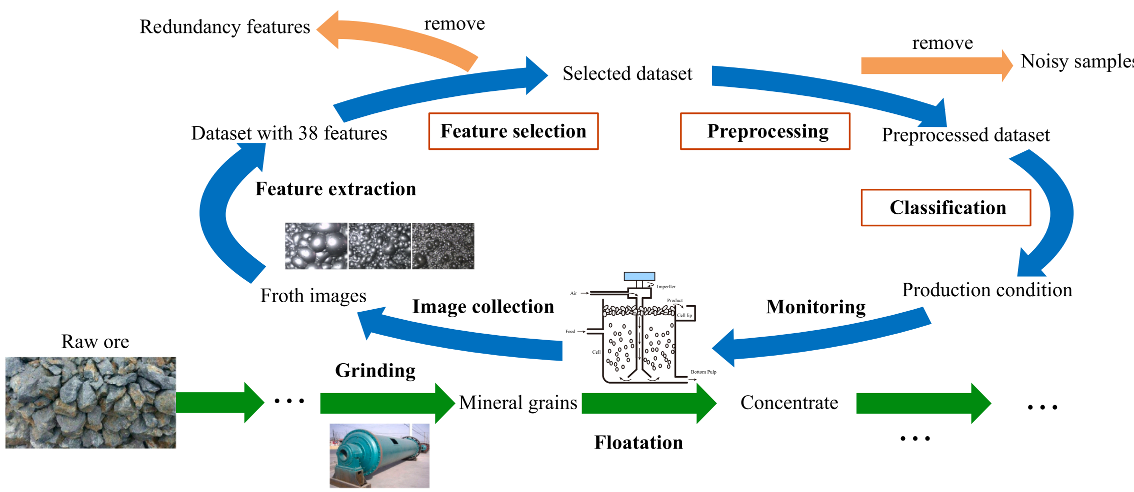

Fig. 2. The framework of the production condition recognition in froth flotation.

description of the gold-antimony flotation process. Section 3 shows the proposed FS method, mRMR-BSTA. Section 4 presents the proposed preprocessing method based on FCM. Section 5 contains the experimental results and analysis. Finally, Section 6 concludes the paper.

2. Description of the gold-antimony flotation process

In order to separate gold and antimony from the raw ore, a preferential-bulk flotation technique is employed during the gold-antimony flotation process. It contains two parts (Li and Gui, 2018): (1) separating gold in a weak-alkaline environment; (2) recovering antimony under a weak-acid condition. There is an impeller fitted with each flotation bank, which aims to agitate air into the pulp to generate air bubbles. When the generated air bubbles rise to the top, the hydrophobic particles will attach to them forming a froth layer.

The industrial gold-antimony flotation process includes a gold-roughing bank, two gold-concentrating banks I and II, an antimony-roughing bank, two antimony-concentrating banks I and II, two antimony-scavenging banks I and II. The flow sheet is shown in the Fig. 3.

Gold concentrate and antimony concentrate are produced by the froth layer of gold-concentrating bank II and antimony-concentrating bank II respectively, which are scraped by mechanical froth scrapers. In this flotation process, the final tailing is the underflow of antimony-scavenging bank II. The chemical reagents are fed into the gold-roughing bank and antimony-roughing bank to help extracting the gold and antimony.

According to the workers with rich experience, the reagent addition control plays a major role in the whole flotation process and influences the final product directly. Naturally, the roughing banks where the most reagents are added, are considered as the main control target. The main ingredient of reagents added is a kind of mixture which is composed of aerofloat, xanthate, copper sulphate and pine oil at a certain proportion. The effect for the flotation process of the reagents addition can be reflected on the features of the froth in the roughing banks. Better recognition of production condition is the key to adjust the reagent addition. Machine vision-based method classifies different production conditions through the features extracted from froth images. Hence, it is necessary to collect and analyze the froth images of the roughing banks, and FS could find a good feature set to improve the performance of classification.

A platform for collecting and processing high quality froth images has built. It is composed of two lights, two cameras, a computer, and

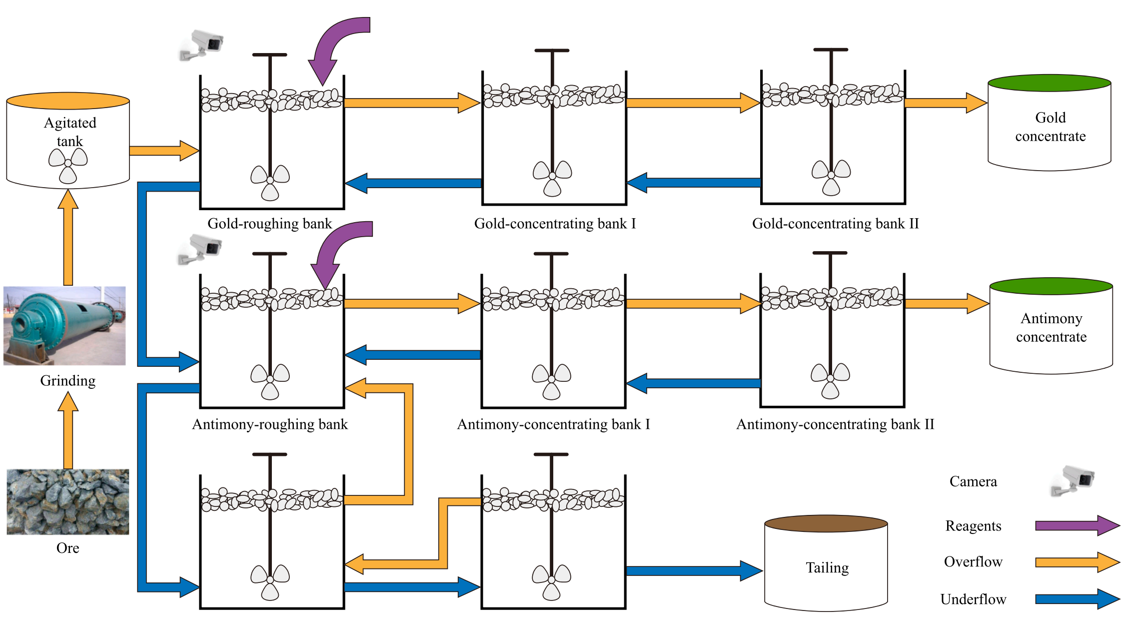

Fig. 3. Flow sheet of the gold-antimony flotation process.

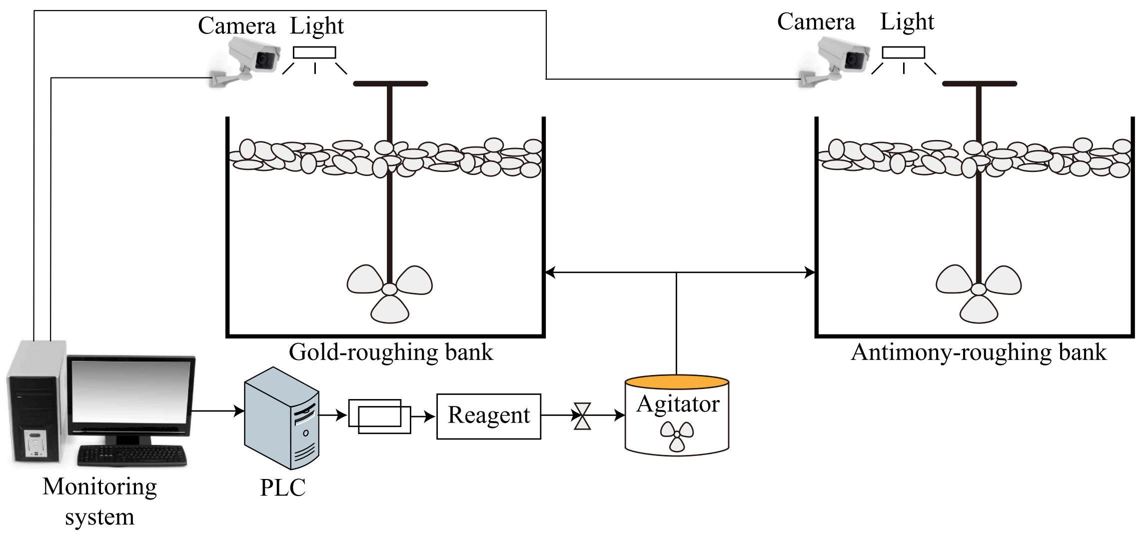

Fig. 4. The image collecting platform.

other appurtenances. Fig. 4 shows the main hardware structure of this platform. It is worth noting that for getting high quality froth images, the high-definition CCD (Charge Coupled Device) color cameras are installed above the roughing banks of 110 cm, and the lights are set next to the cameras, which aims to generate bright points on the top of the froth to help the subsequent image processing. The froth images collected by these cameras are transmitted to image acquisition card using optical fiber and IEEE 1394 interface. Then these images are processed by machine vision-based method to realize production condition recognition instead of workers. It reduces the labor intensity and improves efficiency. In order to achieve recognition, 38 features are extracted from each image collected through the platform. That is to say, an image sample with 600×800 pixels can be characterized by a vector with only 38 values. However, there are still some redundant and noisy features among them, and the accuracy of classification and the cost of calculations are still affected by the redundant and noise features in the dataset. Therefore, FS is necessary.

3. Proposed feature selection method: mRMR-BSTA

FS methods are broadly categorized as two kinds from the perspective of selection strategy: (1) wrapper method; (2) filter method. Wrapper method selects features depending on the performances of classification, so that the selected feature subset always has a good performance. However, it always consumes a large amount of time. In contrast, filter methods cost little, but the classification accuracy cannot be guaranteed using the feature subset selected by them. Therefore, the proposed method combines the advantages of these two kinds of method. It is described in detail in this section.

3.1. Framework of proposed method

The FS method proposed in this paper mainly contains two phases: (1) the filter phase based on mRMR criterion; (2) the wrapper phase based on BSTA. The purpose of the combination of these two approaches is taking advantages of the filter method and wrapper method. The advantage of former is less computation and the advantage of latter is high accuracy. How the proposed hybrid method works is shown in the flowchart Fig. 5. The major works of filter phase is expressed in three aspects: (1) to reduce the search space (the space of all feasible solutions in FS); (2) to calculate the ranking of candidate features based on mRMR criterion; (3) to provide information of features weights for the next wrapper method.

In short, FS problems can be described as selecting N_s features from a M features for a certain dataset, which makes the classification performance best with the numbers of N_s as small as possible. In this problem, a solution can be represented by a binary encoding vector composed of one and zero, where one means the feature is selected and zero means the feature is not selected (Huang et al., 2018). If a dataset has four features, vector $\boldsymbol{x_1}$ [1,1,1,1] means all the features are selected,

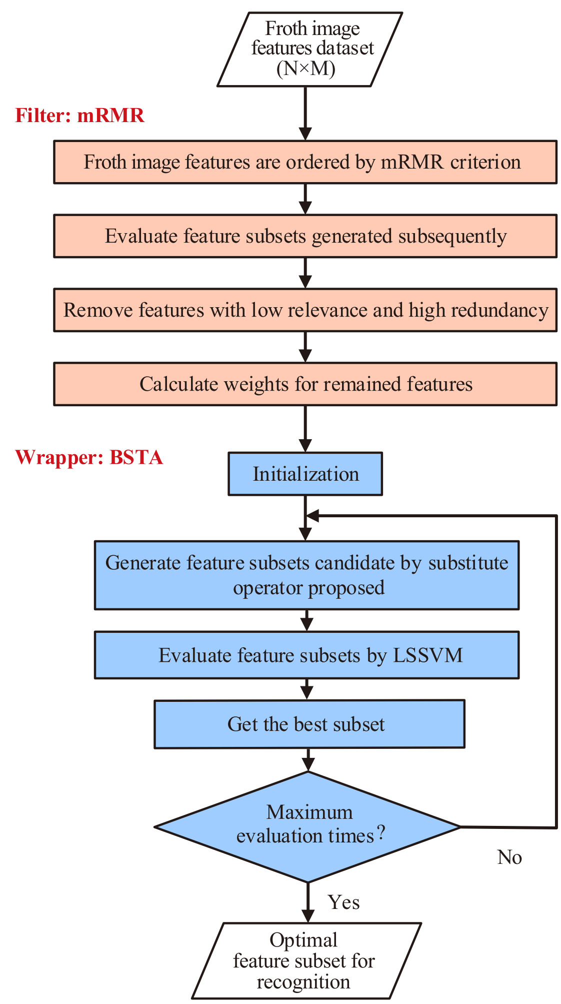

Fig. 5. The flowchart of the proposed feature selection method mRMR-BSTA.

vector x_2 [1,0,1,0] means the first and third features are selected, and vector x_3 [0,0,0,0] means all the features are not selected.

Therefore, the FS problems can be transformed into a problem to search a best vector with highest classification accuracy and least number of "1". In other words, the problem has been transformed into a binary optimization problem. So, BSTA is employed in the wrapper phase as the feature subset search strategy to find the best feature subset including all relevant features and no redundant and irrelevant features for classification. Fig. 5 shows the flowchart of the proposed FS method mRMR-BSTA.

3.2. Primary feature selection based on mRMR

As mentioned above, the wrapper methods need large calculation and time consumption, especially when the features number becomes large. Hence, in this phase, we take full advantage of filter method to reduce the search space so that the next phase using wrapper method can select the best feature subset swiftly. The core of filter method is evaluation criteria. The normal filter methods include Relief and ReliefF (Robnik-Šikonja and Kononenko, 2003), Correlation-based feature selection (CFS) (Hall and Smith, 1999), Fast correlation-based filter (FCBF)(Yu and Liu, 2003), etc. The filter method utilized in this phase is a mutual-information-based method called mRMR which considers both the relevance between features and classes and the redundancy among features. mRMR has been widely applied in the FS problems and has shown excellent performance (Li et al., 2017).

This criterion evaluates each feature via the information of two aspects mentioned above. A feature subset S^* can be selected from a dataset having N samples with M features $F = \{f_i, i = 1, \ldots, M\}$ by mRMR method, according to the relevance and redundance among features and classes. Firstly, max-relevance condition shown in Eq. (1) is used to find a feature set S including features with the largest dependency of the target classes c.

$$\max D\left(S, c\right), D = \frac{1}{|S|}\sum_{f_i \in S} I\left(f_i; c\right) \tag{1}$$

where $|S|$ is the number of the selected features, and $I(f_i;c)$ is the mutual information values between f_i and c. $I(\cdot)$ is the function used to calculate the mutual information. However, feature subset selected only by Max-Relevance always contains features having high dependency with classes but also high redundancy with other features. Hence, the minimal-Redundancy condition is used to remove these features.

$$\min R(S), R = \frac{1}{|S|^2}\sum_{f_i, f_j \in S} I\left(f_i; f_j\right) \tag{2}$$

Eventually, a comprehensive criterion combining the above conditions can be defined as follow:

$$\max \Phi(D, R), \Phi = D - R \tag{3}$$

By maximizing the Eq. (3), the best feature subset can be obtained. However, in fact, it is difficult and costs too much to select feature subset using the method mentioned above. Peng et al. (2005), proposed an incremental search method to simplify the calculation. It first selects a feature having highest dependency with target classes according to the Max-Relevance directly. Once the first feature has been selected, the feature remained can be added into the optimal feature subset one by one using the incremental search method. Given the feature subset S_{m-1} which includes $m-1$ features, the mth feature can be obtained by the following formula:

$$\max_{f_j \in F - S_{m-1}} \Phi(D, R)' = \left[I\left(f_j; c\right) - \frac{\sum_{f_i \in S_{m-1}} I\left(f_j; f_i\right)}{m-1} \right] \tag{4}$$

It is difficult to calculate the mutual information with at least one continuous variable. Therefore, a method called Parzen window (Kwak and Choi, 2002) is employed in this study to estimate the mutual information.

When calculations are completed, an order of features can be obtained on the basis of sequence that when features are selected. Drawing on this order, first m features are selected subsequently, and $m = 1, 2, 3, \ldots, M$. Therefore, M feature subsets are generated and they will be evaluated by the learning machine. The feature subset with highest accuracy of classification will be selected. If it is the mth feature subset, it means the first m features are selected. Noteworthily, if there are multiple feature subsets possess the same highest accuracy, we will select the feature subset containing more features.

After this step, a rough feature subset S' can be selected which is much smaller than the original feature set, and $S' \subset S$. The reduction of features in this phase vastly reduces the cost and difficulty for finding the best feature subset in the next wrapper method. For each feature, there is an evaluation given by mRMR criterion, and it is the weight for this feature, which can reflect the importance of this feature.

The weight of feature f_i is denoted as w_i.

$$w_i = \Phi_i'(D, R) \tag{5}$$

In order to make the next wrapper phase convenient, the weights need be normalized after the mRMR processes. The normalization expression is given as follows:

$$w_i' = \frac{w_i - \min(w)}{\max(w) - \min(w)} \tag{6}$$

where w_i' is the normalized value, and $\min(w)$ and $\max(w)$ is the minimum value and maximum values of feature weights w.

3.3. Compact feature selection based on BSTA

In this subsection, the wrapper phase based on BSTA is described in details. As mentioned above, since the FS problems can be modeled as optimization problems, BSTA is employed in this phase. BSTA is a specialized form of discrete state transition algorithm (DSTA) for solving boolean integer optimization problems, and it is a new intelligent optimization algorithm in the framework of state transition algorithm (STA) (Zhou et al., 2019; Huang et al., 2019). Before explaining the FS method in this subsection, there is a brief introduction of DSTA. DSTA is a global stochastic optimization algorithm proposed by Zhou et al. (2016), which aims to solve integer optimization problems. For the traveling salesman problem (Yang et al., 2013), staff assignment problem (Dong et al., 2016) and community detection problem (Zhou et al., 2019), DSTA has shown excellent performance. The unified framework of generation of candidate solutions in DSTA is shown as follows:

$$\begin{cases} \mathbf{x}_{k+1} = A_k(\mathbf{x}_k) \oplus B_k(\mathbf{u}_k) \\ \mathbf{y}_{k+1} = f(\mathbf{x}_{k+1}) \end{cases} \tag{7}$$

where $\mathbf{x}_k \in \mathbb{Z}^n$ stands for a state corresponding to a solution of a specific optimization problem, and in this study it is the binary encoding vector; $\mathbf{u}_k$ is a function of the historical states and current state; $A_k(\cdot)$ and $B_k(\cdot)$ denote state transformation operators, which are usually some transition matrixes; $\oplus$ is an operation, which is admissible to operate on two states; $f(\cdot)$ denotes the fitness function, and $\mathbf{y}_{k+1}$ is the fitness of $\mathbf{x}_{k+1}$. In this study, the evaluation of wrapper phase for each feature subset according to the classification accuracy is equivalent to the fitness function.

The wrapper phase of mRMR-BSTA is mainly comprised of three procedures: (1) initialization; (2) generation of feature subset candidates; (3) evaluation and selection.

The initialization is important to this method, since a good initial solution can lead to a significant reduction of the time cost and calculations. The higher the feature weight is, the more important the feature is. However, since the previous method based on mRMR cannot effectively remove the features whose relevance and redundancy are both high. We do not expect that all the features holding high weights are selected. Meanwhile, it is not expected that the initial solution leads to a local optima easily. Hence, an initialization approach based on the information of feature ranking is employed here. It is shown as follows:

$$x_i = \begin{cases} 1, & \textit{if}\ \ \text{rand}() < p \\ 0, & \textit{otherwise} \end{cases}, \textit{when}\ S(x_i) \leqslant n * P_d \tag{8}$$

$$x_i = \begin{cases} 1, & \textit{if}\ \ \text{rand}() < q \\ 0, & \textit{otherwise} \end{cases}, \textit{when}\ S(x_i) > n * P_d \tag{9}$$

Three parameters, P_d, p, and q, are used in the initialization to control the performance of the initial solution, and they are specified by user. Here, $P_d \in [0, 1]$ signifies a user-specified percentage and $M \times P_d$ denotes the number of the features with high weights. n is the number of the selected features via the previous part. The probability of choosing a high weight feature is denoted as p, and $p \in [0, 1]$, but it is always close to 1 to obtain a better performance. Correspondingly, q means the probability of choosing a feature with a lower weight, and $q \in [0, 1]$.

The generation of candidate features subset plays a vital role in the wrapper method based on BSTA, and their diversity is essential to find the optimal feature subset. Therefore, according to the characteristic of FS problems, a substitute operator based on the feature weights is proposed, which can generate more feature subsets in unknown areas of the search space. It works as follows:

$$x_i = \begin{cases} 1 - x_i, & \textit{if}\ \ \text{rand}() \leqslant r_i \\ x_i, & \textit{otherwise} \end{cases} \tag{10}$$

$$r_i = r_t \times r_{w,i} \tag{11}$$

where r_i is the mutation rate of feature i, which is the major factor influencing the possibility for each feature changing their state in this operator. It comprises two parameters, r_t and $r_{w,i}$, and they are described as follows:

$$r_t = \alpha \times \frac{T - t + 1}{T} \tag{12}$$

$$r_{w,i} = e^{-\frac{(w'^{i} - 0.5)^2}{2\sigma_r^2}} \times \left(L_u - L_b \right) + L_b \tag{13}$$

where r_t is mainly effected by the iteration t, and T is the maximum number of iteration, which is a preset parameter. r_w is a deformation of a normal distribution, where the parameters L_u and L_b control the maximum value and minimum value of r_w respectively. The relation of r_w and w' is shown in Fig. 6.

The evaluation criterion of a feature subset in wrapper method is the performance of the classification using a specific classifier. In this paper, the multi-classifier called least squares support vector machine (LSSVM) is employed, since it has an excellent performance for gold-antimony production condition recognition in our previous study (Wang et al., 2018). However, in fact, FS problems have two main conflicting objectives, the maximum classification accuracy and the minimum selected features number. Hence, a relative dominance-based selection strategy is utilized in this paper to find the current optimal feature subset. For gold-antimony production condition recognition, compared with less number of features, higher accuracy of the classification is more important. In order to make the description easy to follow, for a certain feature subset candidate $\boldsymbol{x_i}$, the two objectives are

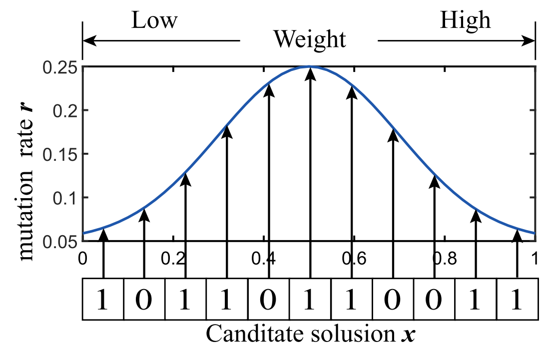

Fig. 6. The illustration of the proposed substitute operator ($\alpha = 0.5$, $T = 50$, $t = 1$, $L_u = 0.5$, $L_b = 0.1$, $\sigma_r = 0.2$).

denoted as $Acc(\boldsymbol{x_i})$ and $Num(\boldsymbol{x_i})$ respectively, where $Acc(\boldsymbol{x_i})$ means the accuracy of the classification using $\boldsymbol{x_i}$, and $Num(\boldsymbol{x_i})$ means the features number of $Num(\boldsymbol{x_i})$. The current best feature subset is donated as $\boldsymbol{x_{best}}$. There are two conditions about changing $\boldsymbol{x_{best}}$ as follows:

(1) $Acc(\boldsymbol{x_i}) > \boldsymbol{Acc}(\boldsymbol{x_{best}})$;
(2) $Acc(\boldsymbol{x_i}) = \boldsymbol{Acc}(\boldsymbol{x_{best}})$ and $Num(\boldsymbol{x_i}) < \boldsymbol{Num}(\boldsymbol{x_{best}})$.

When $\boldsymbol{x_i}$ is compared with $\boldsymbol{x_{best}}$, the candidate $\boldsymbol{x_i}$ overwrites $\boldsymbol{x_{best}}$ if $\boldsymbol{x_i}$ meets any condition above. When the termination condition is met, the $\boldsymbol{x_{best}}$ is the optimal solution.

3.4. Evaluation and Classification

There are eight classes for gold-antimony froth flotation. The classifier used to achieve the production condition recognition is LSSVM. As mentioned above, the evaluation of wrapper method is the classification accuracy. In order to ensure the recognition performance, the LSSVM is also utilized to evaluate the solutions in the wrapper phase. LSSVM is an alternate formulation of support vector machines (SVM) (Suykens and Vandewalle, 1999). It replaces the e-insensitive loss function of SVM with the classical squared loss function (Chauchard et al., 2004), which increases the accuracy and reduces the calculation a lot (Yan and Chowdhury, 2013). Eq. (14) is the model of LSSVM.

$$\min_{w_s, b, e} L_{P_{LSSVM}} = \frac{1}{2} \left|\left| w_s \right|\right|^2 + \frac{1}{2} \gamma \sum_{i=1}^{m} e_i^2 \tag{14}$$

$$\text{s. t. } y_i(\langle w_s^T, \phi(x_i, \sigma) \rangle + b) = 1 - e_i,\ i = 1, 2, 3, \ldots, m$$

As shown in Eq. (14), the parameter γ and an implicit parameter σ in kernel function influence the performance of LSSVM. The values of these two parameters are real numbers ranging from 0 to ∞. In order to improve the performance of LSSVM, it is necessary to seek a good parameter combination consisting of γ and σ. Therefore, LSSVM is improved in our previous study (Wang et al., 2018), and it is used in this study.

4. Preprocessing using fuzzy C-means clustering method

In order to train the model of LSSVM, training samples need be labeled. Therefore, a dataset is collected from plant and labeled by workers with rich experience. However, when faced with such a big dataset, there is still a high probability of the existence of some wrong labels, because of the subjectivity of these workers. The data with wrong label is always manifested as an outlier in the data subset which has same labels with it. So a preprocessing method based on FCM clustering is proposed in this paper aiming to remove the data with wrong labels. There is a description of fuzzy C-means (FCM) clustering method in the first part of this section, and the method of removing outliers is shown in the next part.

4.1. Fuzzy C-means clustering method

An effective tool in solving data mining problems called fuzzy c-means (FCM) clustering method has draw widely attention in decade (Bai et al., 2019; Salmeron et al., 2017). The aim of clustering is trying to distinguish a set of n objects into N_v groups ($N_v \leqslant n$) (Pei et al., 2017). Different from the conventional clustering method, FCM method escapes from distinguishing objects in the same set by the simple Euclidean distance. Combining with the theory of the fuzzy set, FCM method adopts the membership function as the similarity strategy (Gu et al., 2010). Considering a dataset with n samples and can be divided into N_v clusters and each cluster is a set whose clustering center can be denoted as $v_j (j \leqslant N_v)$, the unity form of the FCM method can be described as follows:

$$\min J_m\left(u, v\right) = \sum_{j=1}^{N_v} \sum_{i=1}^{n} u_{ij}^{m} d_{ij} \tag{15}$$

$$\text{s. t.} \begin{cases} \sum_{j=1}^{c} u_{ij} = 1 \ \ , \ i = 1, 2, \cdots, n \\ 0 \leqslant u_{ij} \leqslant 1 \ \ , \ i = 1, 2, \cdots, n; j = 1, 2, \cdots, N_v \end{cases}$$

where the membership function u_{ij} which represents the degree of ith node belonging to jth cluster v_j. $d_{ij} = \|x_i - v_j\|$ is the distance between the ith node and jth cluster center. m is called the fuzziness index which can be set as 2 if no additional conditions are need. These indexes can be updated by the following formula:

$$u_{ij} = \frac{1}{\sum_{k=1} \left(d\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} \quad and \quad v_j = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_i}{\sum_{i=1}^{n} u_{ij}^{m}} \tag{16}$$

With the iterative process of minimizing the target $J_m(u, v)$, the membership function u_{ij} and the clustering center v_j can be updated repeatedly. This method will provide some excellent cluster centers for the dataset, and they will be used in the next step to get rid of the outliers in the dataset.

4.2. Removing of data with wrong labels

As mentioned above, the FCM is utilized to provide some good clusters center for the proposed preprocess method. It is worth noting that the proposed method does not deal with all data in the dataset at once. In fact, it handles data with same label each time, and outliers are removed in batches.

Outliers are removed mainly based on the distance information of these data points and the cluster centers. The main steps are as follows:

Step 1: calculate the distances among the cluster centers V, and find the mean of it ($\overline{d_V}$);

Step 2: calculate the numbers of data points in hyperspheres whose radiuses are $\overline{d_V}$ and the centers are these cluster centers respectively;

Step 3: find the hyperspheres, by which the number of data points surrounded is less than a set point T_n, and then find the corresponding cluster centers;

Step 4: remove the data points which are attached to the cluster centers found in step 3.

The number of the cluster centers N_v should be set in advance in FCM method.

5. Experimental result and discussion

In this section, we will demonstrate the performance of the proposed FS method, mRMR-BSTA, and the proposed preprocessing method based on FCM. The dataset used in this section is collected from real production environment, and it contains 38 features extracted from froth images gained from gold-antimony froth flotation process. The experimental machine is a personal computer which has Core i7-7700 3.60 GHz CPU and 16 GB RAM.

5.1. Dataset description

Each data in the industrial dataset used in this section comprises of 38 features. These features are extracted from froth images gained from gold-antimony froth flotation process, and the images are collected by the image collecting platform mentioned above in a gold-antimony froth flotation plant. The camera used to collect froth images is positioned about 110 cm above the flotation cell. The resolution of the camera is 600 × 800 pixels, i.e. each image images collected by it has 240 thousand pixels. It can be difficulty used directly to achieve production condition recognition, therefore, the dataset used in this section is the extracted 38 features.

The dataset includes 672 samples, and they are extracted from 672 images which are randomly selected from eight kinds of different classes. In the experiments, about 2/3 of the data are utilized as the training set, and the remained 1/3 are used as the test set. It is worth noting that the test set will remain unchanged, thought the some samples in training set will be removed by the preprocessing method. In this paper, 480 samples training set has been reduced to 435 samples after the preprocessing process.

5.2. Experimental results

To validate effectiveness of the proposed mRMR-BSTA method and preprocessing method based on FCM, several comparative experiments are conducted in this study, and the results are shown in Table 1. Since the proposed FS method is contained by two relative independent phases, the experimental data obtained in these two phases are both worthy of observation. Fig. 7 shows the change of classification accuracy with the added features number increasing in the filter phase based on mRMR, and Fig. 8 shows the iterative curves of the selected features number and classification accuracy of the dataset in the wrapper phase based on BSTA.

As shown in Fig. 7, the classification accuracy increased with the increment of the features number at first. However, when the feature number reaches a certain value, the accuracy of classification stops rising or even decreasing. It further proves that there are some redundant and noisy features in the feature set, and they pull down the classification accuracy. According to the data obtained in experiments, in the filter phase, the highest classification accuracy always appears when the feature number added to 10 to 13.

The parameters of the proposed FS and preprocessing method is shown in the Table 2, where SE is used to represent the number of generated candidate solutions in each iteration, and $N_{c,i}$ is the total number of data points in the dataset with the class c_i.

In Table 1, the feature number of dataset (d) is shown in column 3. The average values in Table 1 are calculated by running the method 30 times for each dataset. Columns 4 and 5 display the best classification accuracy (Acc^*) and the average classification accuracy ($\overline{Acc}$) respectively. $\overline{time}$ is the average time of selected feature subset to train classifier. As is shown in Table 1, after FS, whether the dataset is preprocessed or not, their corresponding best classification accuracy both reach 95.83%. However, by using the preprocessed dataset, the average

Table 1
Classification accuracy in different situation.

Dataset	Training set samples	Feature selection	d	Acc^* (%)	$\overline{Acc}$ (%)	$\overline{time}$ (s)
Original	480	No	38	91.67	89.24	5.1614
		Yes	5	95.83	93.73	3.8199
Preprocessed	432	Yes	5	95.83	94.64	2.4246

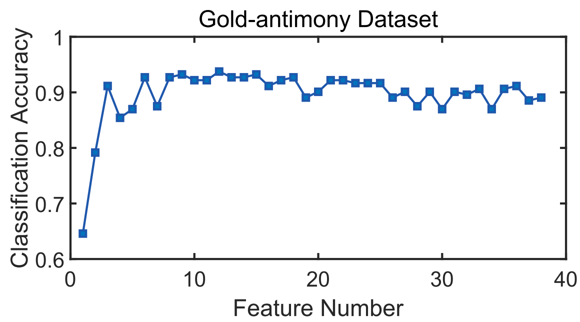

Fig. 7. The *Acc* with the sequential mRMR feature subsets.

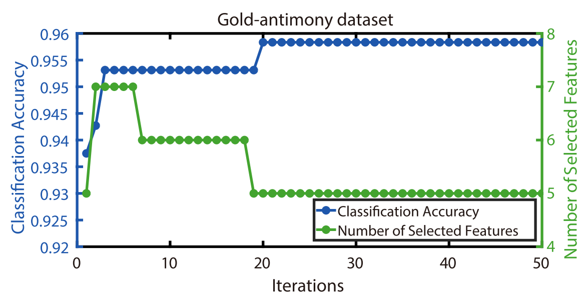

Fig. 8. The *Acc* and *d* with respect to the number of iterations.

Table 2
Parameter settings

Method	Parameters
Preprocessing	$\delta_T = 0.017$, $N_v = 11$, $T_n = \frac{1.8}{N_v} N_{c,i}$
mRMR-BSTA	$SE = 20$ $P_d = 0.6$ $p = 0.8$, $q = 0.3$ $T = 100$, $L_u = 0.5$, $L_b = 0.1$ $\sigma_r = 0.2$, $\alpha = 0.5$

classification accuracy achieved is higher and the training time used is less. The average selected features number is 5 with running the mRMR-BSTA 30 times. The least number of features is 4, however the corresponding accuracy is 95.31% which is lower than 95.83%. The best solution with accuracy 95.83% includes 5 features, hue, mean of blue, relative red component, coarseness and high frequency energy, and the corresponding sequence numbers of these features are 3,7,10,14 and 18.

5.3. Discussion

Obviously, FS brings a significant improvement for the accuracy of classification. However, it is worth noting that the classification accuracy trained by all 38 features is even lower than the classification accuracy classifying using only some morphological features (89.3%) (Wang et al., 2018). It also shows that there are some redundance and noise in the feature set. The time cost for training the classifier using the selected feature subset is also less than using all 38 features.

In order to improve the performance of LSSVM, a preprocessing method is employed before the final classification process. As shown in the experimental results, though the preprocessing method reduces the samples number of training set, the performance of trained classifier with preprocessed training set always achieves a higher accuracy. Whether the dataset is preprocessed or not, the highest classification accuracies that LSSVM has achieved are same (95.83%) in this study. But the average values are very different. It can be concluded that the noisy data points may cause overfitting of LSSVM, since the error rates of trained classifier with preprocessed training set and original training set are similar. It should be noted that classifier LSSVM has a certain capability of noise immunity. However, for an industrial dataset with large of noise, it does not always work very well, and the preprocessing method helps it to train a more accurate model. At the same time, FS and preprocessing both reduce the time cost of training the LSSVM.

Fig. 9 shows the feature values of the best solution in different classes of froth images. In order to observe the differences conveniently, the value of each feature has been normalized respectively, and features are represented by their own sequence number. As is shown in Fig. 9, there are huge differences among the features of different classes froth images. This is the reason why the selected features can distinguish different production conditions effectively, and it can help researchers to get a deeper understanding about the flotation process.

6. Conclusion

To improve the performance of the production condition recognition in froth flotation, a hybrid FS method and a data preprocessing method are proposed in this paper. The proposed hybrid FS method mainly contains two phases. In the first phase, a rough feature subset is selected by mRMR, removing the features with high redundancy and noise which affect classification accuracy negatively. At the same time, the weights of features are also calculated in this phase by mRMR criterion. In the second phase, BSTA is utilized as the optimization technique to search for a high quality feature subset which possesses highest classification accuracy and less features. The candidate solutions in each iteration of BSTA are generated by a new substitute operator proposed in this paper based on the feature weights calculated in the first phase. It should be noted that there is an initialization to generate the initial solution of BSTA, which can select more features with high relevance and less features having high redundancy, and it is also based on the feature weights. When selecting the best solution in each iteration, this method follows the principle that the classification accuracy is prior compared with the number of selected features. The preprocessing method based on FCM removed the outliers of real dataset. The preprocessed dataset can help the learner training a model with high classification accuracy more easily. In this study, several experiments have been conducted to analyze the effectiveness of FS method mRMR-BSTA and the proposed data preprocessing method. More importantly, five efficient features, hue, mean of blue, relative red component, coarseness and high frequency energy, are selected in this paper. The average classification accuracy of production condition in gold-antimony froth flotation is improved to 94.64%.

On the other hand, there is a crucial process variable due to the real-time froth surface level. Hence, in the future, we will consider applying BSTA based online feature selection method in production condition recognition in froth flotation. At the same time, the influence of class imbalance problems in production condition will be considered in the future work. Moreover, using regression instead of classification is also worth doing in the future.

CRediT authorship contribution statement

Xiaojun Zhou: Writing - review & editing, Conceptualization, Investigation. **Qi'an Wang:** Writing - review & editing, Software. **Rundong Zhang:** Writing - review & editing, Validation. **Chunhua Yang:** Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

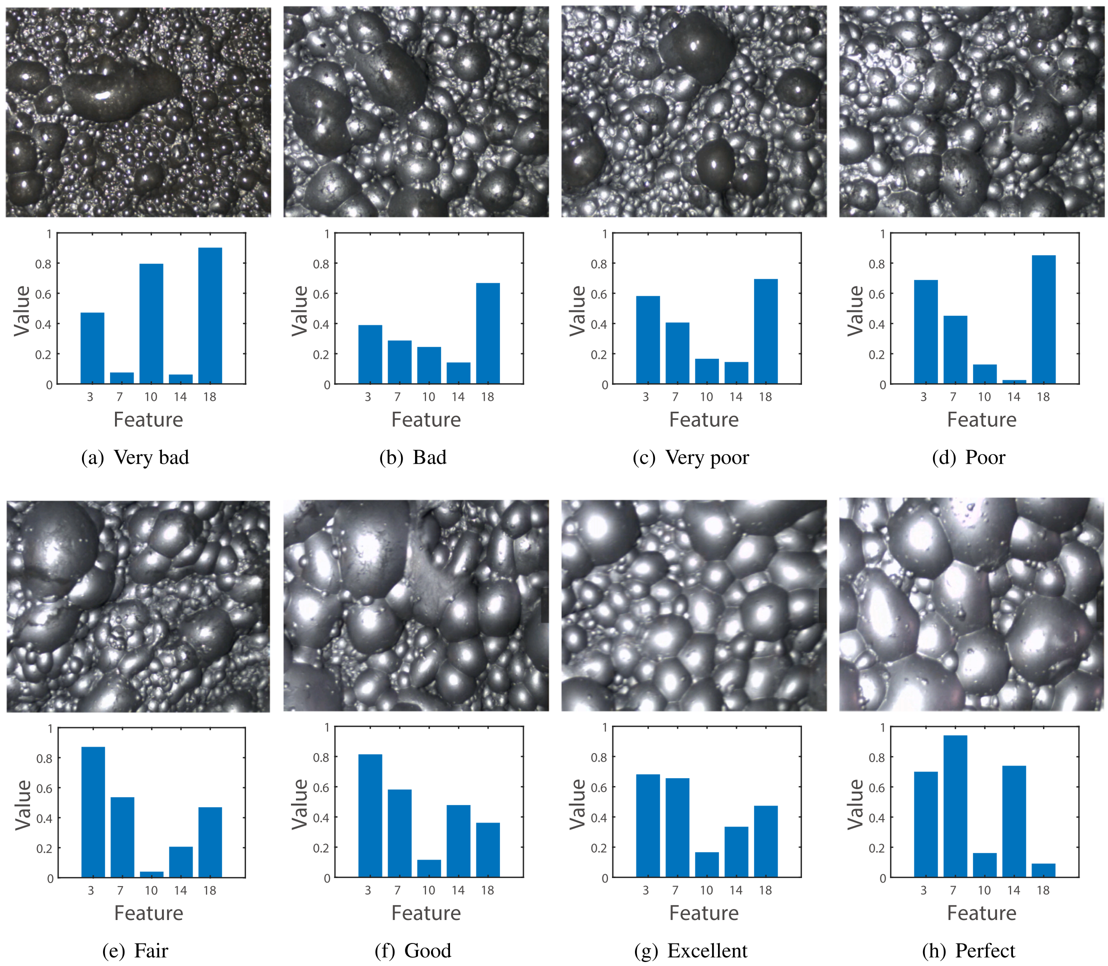

Fig. 9. Froth images and their corresponding features value in best solution for eight typical conditions.

Acknowledgements

The authors would like to acknowledge the National Natural Science Foundation of China (Grant Nos. 61873285, 61533021, 61621062), the Innovation-Driven Plan in Central South University (Grant No. 2018CX12), the 111 Project (Grant No. B17048), the Hunan Provincial Natural Science Foundation of China (Grant No. 2018JJ3683), and the Fundamental Research Funds for the Central Universities of Central South University (Grant No. 2019zzts956) for their funding support.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.mineng.2020.106201.

References

Bai, X., Sun, C., Sun, C., 2019. Cell segmentation based on fopso combined with shape information improved intuitionistic fcm. IEEE J. Biomed. Health Informatics 23 (1), 449–459.

Chauchard, F., Cogdill, R., Roussel, S., Roger, J., Bellon-Maurel, V., 2004. Application of ls-svm to non-linear phenomena in nir spectroscopy: development of a robust and portable sensor for acidity prediction in grapes. Chemometrics Intell. Lab. Syst. 71 (2), 141–150.

Dong, T., Yang, C., Zhou, X., Gui, W., 2016. A novel discrete state transition algorithm for staff assignment problem. Control Theory Appl. 33 (10), 1378–1388.

Ejtemaei, M., Gharabaghi, M., Irannajad, M., 2014. A review of zinc oxide mineral beneficiation using flotation method, advances in Colloid and Interface. Science 206, 68–78.

Gu, S.M., Zhao, J.W., He, L., 2010. An improved fcm clustering method for interval data. In: International Conference on Rough Sets and Knowledge Technology. Springer, pp. 545–550.

Gui, W., Liu, J., Yang, C., Chen, N., Liao, X., 2013. Color co-occurrence matrix based froth image texture extraction for mineral flotation. Miner. Eng. 46–47, 60–67.

Hall, M.A., Smith, L.A., 1999. Feature selection for machine learning: comparing a correlation-based filter approach to the wrapper. In: FLAIRS Conference. vol. 1999, pp. 235–239.

Han, J., Yang, C., Zhou, X., Gui, W., 2018. A two-stage state transition algorithm for constrained engineering optimization problems. Int. J. Control Autom. Syst. 16 (2), 522–534.

Havens, T.C., Bezdek, J.C., Leckie, C., Hall, L.O., Palaniswami, M., 2012. Fuzzy c-means algorithms for very large data. IEEE Trans. Fuzzy Syst. 20 (6), 1130–1146.

Huang, Z., Yang, C., Zhou, X., Huang, T., 2018. A hybrid feature selection method based on binary state transition algorithm and relieff. IEEE J. Biomed. Health Informatics 23 (5), 1888–1898.

Huang, Z., Yang, C., Zhou, X., Gui, W., 2018. A novel cognitively inspired state transition algorithm for solving the linear bi-level programming problem. Cognitive Comput. (1), 1–11.

Huang, Z., Yang, C., Zhou, X., Yang, S., 2019. Energy consumption forecasting for the nonferrous metallurgy industry using hybrid support vector regression with an adaptive state transition algorithm. Cognitive Comput. doi:10.1007/s12559-019-09644-0.

Inbarani, H.H., Azar, A.T., Jothi, G., 2014. Supervised hybrid feature selection based on pso and rough sets for medical diagnosis. Comput. Methods Programs Biomed. 113 (1), 175–185.

Jović, A., Brkić, K., Bogunović, N., 2015. A review of feature selection methods with applications. In: 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, pp. 1200–1205.

Kwak, N., Choi, C.H., 2002. Input feature selection by mutual information based on parzen window. IEEE Trans. Pattern Anal. Mach. Intell. (12), 1667–1671.

Li, Z., Gui, W., 2018. The method of reagent control based on time series distribution of bubble size in a gold-antimony flotation process. Asian J. Control 20 (6), 2223–2236.

Li, B.Q., Hu, L.L., Chen, L., Feng, K.Y., Cai, Y.D., Chou, K.C., 2012. Prediction of protein domain with mrmr feature selection and analysis. PLoS One 7 (6), e39308.

Li, Y., Yang, Y., Li, G., Xu, M., Huang, W., 2017. A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mrmr feature selection. Mech. Syst. Signal Process. 91, 295–312.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J., Liu, H., 2018. Feature selection: a data perspective. ACM Comput. Surv. (CSUR) 50 (6), 94.
Liu, J., Gui, W., Tang, Z., Yang, C., Zhu, J., Li, J., 2013. Recognition of the operational statuses of reagent addition using dynamic bubble size distribution in copper flotation process. Miner. Eng. 45, 128–141.
Pei, H.X., Zheng, Z.R., Wang, C., Li, C.N., Shao, Y.H., 2017. D-fcm: Density based fuzzy c-means clustering algorithm with application in medical image segmentation. Procedia Comput. Sci. 122, 407–414.
Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. (8), 1226–1238.
Peng, X., Peng, T., Zhao, L., Song, Y., Gui, W., 2016. Working condition recognition based on an improved ngldm and interval data-based classifier for the antimony roughing process. Miner. Eng. 86, 1–9.
Robnik-Šikonja, M., Kononenko, I., 2003. Theoretical and empirical analysis of relieff and rrelieff. Mach. Learn. 53 (1–2), 23–69.
Salmeron, J.L., Rahimi, S.A., Navali, A.M., Sadeghpour, A., 2017. Medical diagnosis of rheumatoid arthritis using data driven pso–fcm with scarce datasets. Neurocomputing 232, 104–112.
Suykens, J.A., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural Process. Lett. 9 (3), 293–300.
Wang, X., Song, C., Yang, C., Xie, Y., 2018. Process working condition recognition based on the fusion of morphological and pixel set features of froth for froth flotation. Miner. Eng. 128, 17–26.
Wang, Q., Huang, M., Zhou, X., 2018. Feature selection in froth flotation for production condition recognition. IFAC-PapersOnLine 51 (21), 123–128.
Xu, D., Chen, Y., Chen, X., Xie, Y., Yang, C., Gui, W., 2016. Multi-model soft measurement method of the froth layer thickness based on visual features. Chemometrics Intell. Lab. Syst. 154, 112–121.
Yan, X., Chowdhury, N.A., 2013. Mid-term electricity market clearing price forecasting: a hybrid lssvm and armax approach. Int. J. Electrical Power Energy Syst. 53, 20–26.
Yang, C., Tang, X., Zhou, X., Gui, W., 2013. A discrete state transition algorithm for traveling salesman problem. Control Theory Appl. 30 (8), 1040–1046.
Yin, H., Gai, K., 2015. An empirical study on preprocessing high-dimensional class-imbalanced data for classification. In: IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems. IEEE, pp. 1314–1319.
Yu, L., Liu, H., 2003. Feature selection for high-dimensional data: a fast correlation-based filter solution. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03), pp. 856–863.
Zhao, L., Peng, T., Zhao, L., Xia, P., Zhao, Y., Song, Y., 2015. Fault condition recognition based on multi-scale texture features and embedding prior knowledge k-means for antimony flotation process. IFAC-PapersOnLine 48 (21), 864–870.
Zhao, L., Peng, T., Xie, Y., Yang, C., Gui, W., 2017. Recognition of flooding and sinking conditions in flotation process using soft measurement of froth surface level and qta. Chemometrics Intell. Lab. Syst. 169, 45–52.
Zhou, X., Yang, C., Gui, W., 2012. State transition algorithm. J. Industr. Manage. Optim. 8 (4), 1039–1056.
Zhou, X., Gao, D.Y., Simpson, A.R., 2016. Optimal design of water distribution networks by a discrete state transition algorithm. Eng. Optim. 48 (4), 603–628.
Zhou, X., Gao, D.Y., Yang, C., Gui, W., 2016. Discrete state transition algorithm for unconstrained integer optimization problems. Neurocomputing 173, 864–874.
Zhou, X., Yang, C., Gui, W., 2018. A statistical study on parameter selection of operators in continuous state transition algorithm. IEEE Trans. Cybernet. (99), 1–9.
Zhou, X., Shi, P., Lim, C., Yang, C., Gui, W., 2018. A dynamic state transition algorithm with application to sensor network localization. Neurocomputing 273, 237–250.
Zhou, X., Zhou, J., Yang, C., Gui, W., 2018. Set-point tracking and multi-objective optimization-based pid control for the goethite process. IEEE Access 6, 36683–36698.
Zhou, X., Huang, M., Huang, T., Yang,, Gui, W., 2019. Dynamic optimization for copper removal process with continuous production constraints. IEEE Trans. Industr. Informatics. doi:10.1109/TII.2019.2943500.
Zhou, X., Yang, K., Xie, Y., Yang, C., Huang, T., 2019. A novel modularity-based discrete state transition algorithm for community detection in networks. Neurocomputing 334 (8), 89–99.