Available online at www.sciencedirect.com
ScienceDirect
IFAC PapersOnLine 51-21 (2018) 123–128

Feature selection in froth flotation for production condition recognition ⋆

Qi'an Wang, Miao Huang, Xiaojun Zhou

School of Information Science and Engineering, Central South University, Hunan 410083, China (e-mail: michael.x.zhou@csu.edu.cn).

Abstract: At present, a lot of features have been extracted to characterize the froth flotation, but there exist redundant and irrelevant features which may degrade the performance of a classifier and influence the production condition recognition. In this study, a feature selection strategy based on the minimal-redundancy-maximal-relevance criterion (mRMR) is used to find the most useful but less redundant features. Additionally, the least squares support vector machine (LSSVM) optimized by state transition algorithm is proposed to serve as the classifier in feature selection. It is found that hue and energy of high frequency play significant roles in classification of flotation froth images. Experimental results show the effectiveness of the proposed method.

Keywords: Froth flotation process, Feature selection, mRMR, LSSVM.

1. INTRODUCTION

Froth flotation is an important industrial process which is aiming to extract valuable minerals from raw ore. More than 90% of nonferrous metals in China are processed by froth flotation. Because of the complex physical and chemical process of flotation, the control of flotation depends mainly on the workers' observation of the froth (Han et al. (2016)). Considering the workers' arbitrariness, it is necessary to extract the characteristic features from froth image objectively based on machine vision. Various machine learning method have been studied to realize the condition classification (Xu et al. (2016)). At the same time, more and more features for the flotation have been extracted for recognizing the condition better (Gui et al. (2013)). However, these features may comprise redundant, irrelevant and relevant features, which would increase the difficulty and cost of training, eventually influence the accuracy of classification.

Feature selection is one of the most common way to reduce the number of features. Irrelevant, redundant and noisy features can be removed by feature selection. However, faced with the feature set with n features, finding the optimal feature subset by the exhaustive method would exist 2^n possible combinations, and if n is large, it is computationally expensive to search all possible feature subset. Instead of the exhaustion of all possibilities, there are two popular feature selection methods, wrapper algorithm and filter algorithm (Estvez et al. (2009)). The wapper algorithm utilizes the classifier to evaluate each subset's quality of features, it can ensure high classification accuracy for a particular classifier but need expensive computation cost. On the contrary, the filter algorithm is independent of classifier, it need fewer cost but its performance is determined by its evaluate criteria. The minimal-redundancy-maximal-relevance criteria (mRMR) is an practical criterion which takes the class relevance and dependence between features into consideration simultaneously, and thus has been widely applied in feature selection (Peng et al. (2005)). In this paper, the mRMR criterion is used to constitute a filter-wrapper hybrid feature selection strategy to find the optimal feature subset for characterizing the flotation froth image.

As a variant of support vector machine (SVM), the least squares support vector machine (LSSVM) has exhibited excellent generalization performance and lower computational complexity in many data classification application (Mitra V (2007)). However, its performance highly depends on the adequate setting of parameters, such as penalty coefficient and kernel parameters. To address the selection of the hyper-parameters, many evolutionary algorithms such as ant colony algorithm (ACO), particle swarm optimization (PSO) has been introduced to solve the hyper-parameters optimization problem in SVMs (Zhang et al. (2010); Bao et al. (2013)). In recent few years, a novel intelligent optimization method named state transition algorithm (STA), has exhibited excellent performance for solving various optimization problems compared with GA and PSO (Zhou et al. (2012); Han et al. (2017)). In this paper, LSSVM with the STA-based hyperparameters optimization (STA-LSSVM) will be proposed to serve as the basic classifier for the product condition recognition arising in froth flotation.

The remainder of the paper is organized as follows. Section 2 lists the froth image's features in froth flotation process. Section 3 introduces a two-stage feature selection algorithm based on mRMR criteria. In Section 4, a STA-

⋆ This work was supported by the National Natural Science Foundation of China (Grant No. 61503416), Hunan Provincial Natural Science Foundation of China (Grant No. 2018JJ3683), the 111 Project (Grant No. B17048) and the Innovation-Driven Plan in Central South University.

2405-8963
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2018.09.403

based hyperparameters optimization method for LSSVM is proposed. Section 5 demonstrates the effectiveness of the proposed approach, and finally, a conclusion is given in Section 6.

2. FEATURES OF FLOTATION FROTH IMAGES

Recently, various features have been studied for recognizing the production condition in froth flotation (Han et al. (2016); He et al. (2013); Cao et al. (2013)). The features are collected in this paper as follows.

Mean of red: A RGB image has 3 color channels, red, green and blue. A pixel's every channel is donated by a number from 0 to 255 and the numerical value presents the color level. Mean of red is calculated by Eq.(1)

$$R_{mean} = \frac{\sum_{i=1}^{N_p} R_i}{N_p} \tag{1}$$

where R_i is the red channel's value of pixel i. N_p is the number of pixels in the whole image.

The mean of green: It can be calculated by replacing R_i to G_i in Eq.(1). G_i is the Green channel's value of pixel i.

The mean of blue: It can be calculated by replacing R_i to B_i in Eq.(1). B_i is the blue channel's value of pixel i.

The mean of gray: Gray level shows the luminance of an image. Instead of 3 channels in RGB image, gray image has only one channel for each pixel, ranging from 0 to 255. In fact, the mean of gray is equal to the value of luminance Y in YUV color space. Y can be calculated as

$$Y_i = 0.299R_i + 0.587G_i + 0.114B_i \tag{2}$$

where Y_i is the gray value of pixel i. Then the mean value of gray can be gotten by replacing R_i to Y_i in Eq.(1).

Hue: Due to the color difference of froth and pulp in flotation process, hue is an important color feature to reflect the froth image in HSI (Hue, Saturation, Intensity) color space. A RGB image can be converted into a HSI one by Eq.(3).

$$\begin{cases} H_i = \begin{cases} \theta & , G_i \geq B_i \\ 2\pi - \theta, & G_i < B_i \end{cases} \\ S_{a,i} = 1 - \frac{3}{(R_i+G_i+B_i)}[\min(R_i, G_i, B_i)] \\ I_i = \frac{1}{3}(R_i + G_i + B_i) \end{cases} \tag{3}$$

where $\theta = \arccos \frac{(R_i-G_i)+(R_i+B_i)}{2[(R_i-G_i)^2+(R_i-B_i)(G_i-B_i)]^{1/2}}$ and H_i,S_i, I_i are the HSI value of pixel i.

Relative red component: Another color feature based on RGB color space can be calculated as follows.

$$R_p = \frac{R_{mean}}{Y_{mean}} \tag{4}$$

Y_{mean} is the mean of gray.

Froth speed: Froth speed can reflect the mobility and wind pressure of the flotation process. Macro block tracking technique is used in this study to get the feature of froth speed. Velocity vector can be calculated with its unit as pixel per second as follows.

$$v = \frac{d}{t} \tag{5}$$

d is the displacement of a macro block, and t is the time horizon between two successive frames.

Mean of bubble size: Another series of important features are derived from the bubble size in froth image. In this study, a watershed segmentation algorithm is used to segment the flotation froth image to get the edge of each bubble, and Fig. 1 shows the segmentation result of an image. The number of bubbles N_b can be counted through the segmentation image and the size of bubbles can be calculated by counting the number of pixels. The mean bubble size can be described by Eq.(6).

$$S_{mean} = \frac{N_p}{N_b} \tag{6}$$

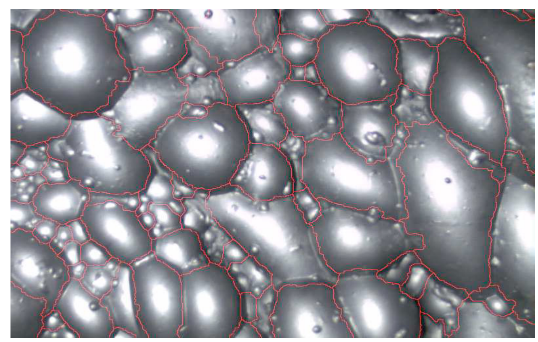

Fig. 1. The froth image segmented by watershed algorithm

Bubble size distribution: This is a statistical feature of bubble size in an froth image. Fig.2 shows the probability density function (PDF) of bubble size. Note that the bubble size is normalized.

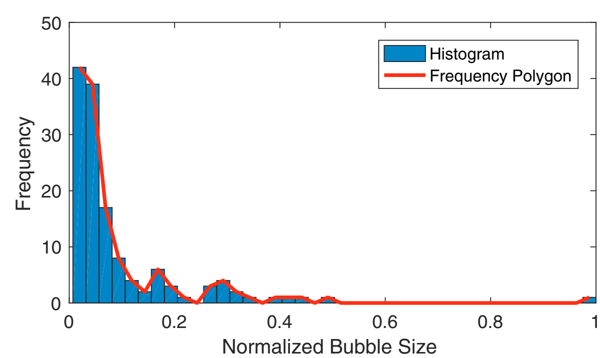

Fig. 2. The PDF of the bubble size in an froth image

Length-width ratio: Based on the segmentation image, the morphological character of the flotation bubble, length-width ratio, can be obtained by

$$R_{lw} = \frac{L}{W} \tag{7}$$

where L and W are the length of major axis and minor axis of each connected region, respectively.

Standard deviation: This is a statistical feature of bubble size. It can be calculated by

$$\sigma = \sqrt{\frac{1}{N_b}\sum_{i=1}^{N_b}(S_{b,i} - S_{mean})^2} \tag{8}$$

where $S_{b,i}$ is the size of bubble i.

Kurtosis: This is a statistical feature of bubble size. It can be calculated by

$$\gamma_2 = \frac{\frac{1}{N_b}\sum_{i=1}^{N_b}(S_{b,i} - S_{mean})^4}{\sigma^4} - 3 \tag{9}$$

Skewness: This is a statistical feature of bubble size. It can be calculated by

$$S_c = \frac{\sum_{i=1}^{N_b}(S_{b,i} - S_{mean})^3}{N_b} \tag{10}$$

Froth load: It is the description of attached mineral amount of froth. The mineral amount can influence the reflectance of bubbles, i.e., the higher reflectance is, the less attached mineral amount the bubble has. Meanwhile, the bubble which has few attached mineral amount always has a total reflectance point. Therefore, the value of froth load can be calculated by

$$L_f = \frac{S_h}{N_p} \tag{11}$$

S_h is the sum area of bubbles without a total reflectance point.

Coarseness: Coarseness is a texture feature of an image. In flotation process, the surface of froth changes with industrial conditions. To calculate the coarseness, we should change the gray image into neighboring gray-level dependence matrix (NGLDM) (Peng et al. (2016)). Then, coarseness can be calculated by

$$C_o = \frac{\sum_{i=1}^{N_{rn}}\sum_{j=1}^{N_{cn}}(j^2 M_{i,j})}{\sum_{i=1}^{N_{rn}}\sum_{j=1}^{N_{cn}} M_{i,j}} \tag{12}$$

where N_{rn} and N_{cn} are the row and column number of NGLDM, $M_{i,j}$ is the value of NGLDM at point (i, j).

Energy: Energy value is the second moment of NGLDM, which reflects the uniformity of gray scale of an image. The more uniform the gray scale of the image is, the lower energy value it has.

$$E = \frac{\sum_{i=1}^{N_{rn}}\sum_{j=1}^{N_{cn}} M_{i,j}^2}{\sum_{i=1}^{N_{rn}}\sum_{j=1}^{N_{cn}} M_{i,j}} \tag{13}$$

Non-uniformity: It is another texture feature describing the uniform level of the gray scale in an image. Its calculation formula is

$$N_u = \frac{\sum_{i=1}^{N_{cn}}(\sum_{j=1}^{N_{rn}} M_{i,j})^2}{\sum_{i=1}^{N_{rn}}\sum_{j=1}^{N_{cn}} M_{i,j}} \tag{14}$$

Coarseness of low frequency image: The low frequency image can be gotten by wavelet transformation. The coarseness of low frequency image can be calculated by its NGLDM through Eq.(12).

Energy of high frequency image: The high frequency image can be gotten by wavelet transformation as well. Then, the energy of high frequency image can be calculated by its NGLDM through Eq.(13).

3. FEATURE SELECTION BASED ON MRMR

3.1 Max-Relevance and Min-Redundancy (mRMR)

The mutual-information-based mRMR criterion has been applied to feature selection successfully and exhibits excellent performance in terms of classification accuracy and computation complexity (Peng et al. (2005)). In this paper, we use mRMR criterion to find a set of features for characterizing the flotation froth image completely.

Given the input data D which has N samples with M features $F = \{f_i, i = 1, ..., M\}$, and the corresponding class label c, the mRMR method aims at finding a maximally relevant and minimally redundant subset S with m features for discriminating classes. Firstly, the following Max-Relevance condition can be used to find the feature set S, which jointly has the largest dependency on the target class c.

$$\max D(S, c), D = \frac{1}{|S|}\sum_{f_i \in S} I(f_i; c) \tag{15}$$

where $|S|$ is the cardinality of the set S and $I(f_i; c)$ is the mutual information values between two random variables. While the feature subset selected by Max-Relevance may comprise redundant features which has high dependency with each other. Therefore, the Minimal-Redundancy condition can be employed to find the individual features.

$$\min R(S, c), R = \frac{1}{|S|^2}\sum_{f_i, f_j \in S} I(f_i; f_j) \tag{16}$$

Eventually, by combining the above two conditions, the maximally relevant and minimally redundant feature subset can be obtained by maximizing the following formula:

$$\max \Phi(D, R), \Phi = D - R \tag{17}$$

Once the first feature has been selected according to the Max-Relevance directly, the remaining optimal feature defined by $\Phi(\cdot)$ can be found by the incremental search method one by one. Given the set with $m - 1$ features, S_{m-1}, the mth feature can be determined by the following condition:

$$\max_{f_j \in F - S_{m-1}} \left[I(f_j; c) - \frac{1}{m-1}\sum_{f_i \in S_{m-1}} I(f_j; f_i) \right] \tag{18}$$

Note that, since the difficulty in calculating the mutual information when at least one variable is continuous, we use the Parzen window method to estimate the mutual information as suggested in (Kwak and Choi (2002)).

3.2 Feature selection algorithm

In this paper, a hybrid feature selection algorithm based on mRMR criterion is used. Instead of using all features with wrapper selectors directly, we intend to apply mRMR criterion to find a candidate feature set firstly, whose size is much smaller than the original one. Then, a wrapper method, the backward selection, is employed to search a compact feature subset from the candidate feature set by minimizing the cross-validate classification error directly.

The complete feature selection algorithm is given.

step 1: Initialization: Set $k = 1$ and S_k as the subset of selected feature with k features.
step 2: First selection: Set $S_1 \leftarrow \{f_1\}$, where f_1 can maximize the mutation information between individual feature f_i and class c. Then calculate the cross-validate classification error e_1 with subset S_1.
step 3: Next selection: Set $k = k + 1$, then choose the next feature f_k according to the Eq.(18) and set $S_k \leftarrow \{f_1, ..., f_k\}$.
step 4: Candidate subset: Calculate the error e_k with the sequential feature sets S_k. If the respective error e_k is consistently small, then output the candidate subset S_k and continue; otherwise, go to step 3.
step 5: Backward selection: Remove the features whose removal do not bring any classification error addition.
step 6: Termination: For each possible configuration, the respective classification error e_{k-1} is larger than e_k, that is to say, there is no gain in either classification accuracy or feature dimension reduction, therefore terminate the whole algorithm and the subset S_k is what we ask for.

Note that, in this paper, we use the LSSVM classifier to calculate the cross-validate classification error of the subsets for the wrapper selector in step 4 and step 5.

4. MULTIPLE CLASSIFIER: STA-LSSVM

4.1 LSSVM

In recent years, the LSSVM multiple classifier has been extensively used in classification applications owning to its excellent generalization performance and low computational complexity (Suykens and Vandewalle (1999); Wu and Peng (2015)).

Given the training set $(\boldsymbol{x}_i, y_i), i = 1, ..., N$, where $\boldsymbol{x}_i \in R^m$ is the m-dimensional input vector and $y_i \in \{-1, 1\}$ is the corresponding class label. In the LSSVM framework, the input data $\boldsymbol{x}_i$ will be first mapped into a new higher dimensional feature space through a nonlinear feature mapping function $\phi(\boldsymbol{x}_i)$, because of its nonlinear separability in the original input space. In the nonlinear feature space, a linear classification model can be constructed as follows:

$$y = \text{sign}[\boldsymbol{\omega}^T \phi(\boldsymbol{x}) + b] \tag{19}$$

where $\boldsymbol{\omega} \in R_N$ is the normal vector of the separating hyperplane of two different classes in the feature space, b is the corresponding bias term. Then, according to the structural risk minimization principle, LSSVM is to find the optimal separating hyperplane which maximizes the separating margins between two classes while also minimizes the fitting error. Therefore, the resulting optimization problem of LSSVM can be formulated as

$$\min L_1(\boldsymbol{\omega}, b, \boldsymbol{\zeta}) = \frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\omega} + C\frac{1}{2}\sum_{i=1}^{N}\zeta_i^2 \tag{20}$$

$$\text{s.t. } y_i(\boldsymbol{\omega} \cdot \phi(\boldsymbol{x}_i) + b) = 1 - \zeta_i, i = 1, ..., N$$

where $C > 0$ is the regularization parameter to balance complexity and approximation precision of the model, and ζ_i is the training error between the true class label and the predict one.

Based on the Karush-Kuhn-Tucker (KKT) theorem (Peressini et al. (1988)), the primal optimization problem L_1 can be solved by transforming it into a dual form L_2 with Lagrange multipliers $\boldsymbol{\alpha}$:

$$\begin{aligned} L_2(\boldsymbol{\omega}, b, \boldsymbol{\zeta}; \boldsymbol{\alpha}) = & \frac{1}{2}\boldsymbol{\omega}^T\boldsymbol{\omega} + \frac{1}{2}C\sum_{i=1}^{N}\zeta_i^2 \\ & - \sum_{i=1}^{N}\alpha_i(t_i(\boldsymbol{\omega} \cdot \phi(\boldsymbol{x}_i) + b) - 1 + \zeta_i) \end{aligned} \tag{21}$$

where the Lagrange multipliers α_i can be either positive or negative. And the optimal solution can be obtained by solving the following linear equations:

$$\begin{bmatrix} 0 & \boldsymbol{Y}^T \\ \boldsymbol{Y} & C^{-1}\boldsymbol{I} + \boldsymbol{Z}\boldsymbol{Z}^T \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \overrightarrow{\mathbf{1}} \end{bmatrix}, \tag{22}$$

where $\boldsymbol{Z} = [\phi(x_1)y_1, ..., \phi(x_N)y_N]^T$, $\boldsymbol{Y} = [y_1, y_2, ..., y_N]^T$, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, ..., \alpha_N]^T$, and $\overrightarrow{\mathbf{1}} = [1, 1, ..., 1]^T$. Finally, the decision function of LSSVM classifier can be reformulated as

$$y = \text{sign}(\sum_{i=1}^{N}\alpha_i y_i K(\boldsymbol{x}, \boldsymbol{x}_i) + b) \tag{23}$$

where $K(\boldsymbol{x}, \boldsymbol{x}_i) = \phi(\boldsymbol{x}) \cdot \phi(\boldsymbol{x}_i)$ is the kernel function, satisfying the Mercer's condition, so that it can avoid the computational difficulty in high dimension feature space. The radial basis function (RBF) is commonly used in LSSVM, denoted as

$$K(\boldsymbol{u}, \boldsymbol{v}) = \exp(-\gamma\|\boldsymbol{u} - \boldsymbol{v}\|^2), \gamma > 0 \tag{24}$$

where γ is the kernel parameter. Since the RBF kernel has various advantages, such as super nonlinearly mapping capability, less hyperparameters and fewer numerical difficulties (Hsu et al. (2003)), it is introduced to construct the LSSVM classification model in this paper.

Note that, in LSSVM, the regularization parameter C serves to balance the model complexity and training error, and the kernel parameter γ directly affects the distribution characteristics of the training data in feature space. Therefore, these two hyper-parameters are critical for the model's performance and it is necessary to find the optimal hyper-parameters so as to obtain a good classifier with LSSVM.

4.2 STA based hyper-parameters optimization of LSSVM

Since the hyper-parameters optimization problem in SVMs is characterized as nonlinear and multi-modal (Bao et al. (2013)), we introduce a global optimization algorithm, state transition algorithm (STA) to find the global solution of this complex problem.

STA is a global stochastic optimization algorithm proposed by Zhou (Zhou et al. (2012)), and it has already shown excellent performance for solving multi-modal and nonlinear optimization problems (Zhou et al. (2018)). A unified framework for generating candidate solutions in STA can be described as follows:

$$\begin{cases} \boldsymbol{x}_{k+1} = A_k \boldsymbol{x}_k + B_k \boldsymbol{u}_k \\ y_{k+1} = f(\boldsymbol{x}_{k+1}) \end{cases} \tag{25}$$

where $\boldsymbol{x}(t) \in \Re^n$ stands for a state, which denotes a solution to the problem; $\boldsymbol{u}_k$ is a function of $\boldsymbol{x}_k$ and historical states; A_k and B_k are state transition matrices, which are usually some state transformation operators; $f(\cdot)$ is the fitness function, and y_{k+1} is the function value of individual $\boldsymbol{x}_{k+1}$.

In order to ensure the accuracy of the solution, a two-stage optimization method based on STA is proposed. At first, we apply the STA responsible for the global search. Then, if the fitness function, the k-fold cross-validation misclassification rate, is not equal to 0, the simplex method (Nelder and Mead (1965)) will be employed for further fine-tuning. The flowchart of the STA-based hyper-parameter selection algorithm for the LSSVM is shown in Fig. 3.

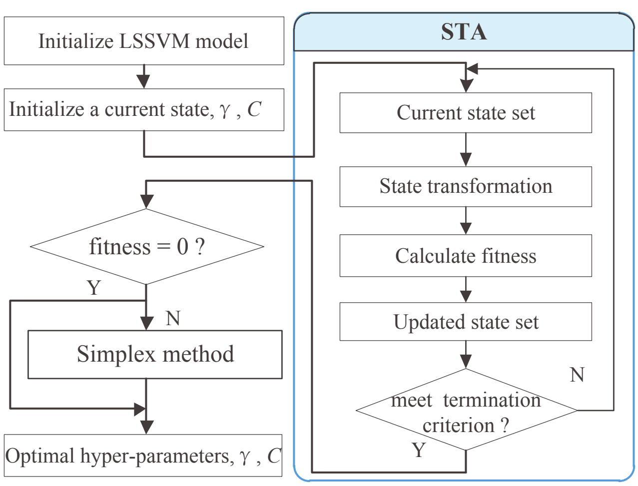

Fig. 3. STA-based parameters optimization of LSSVM

5. EXPERIMENTAL VALIDATION

Since the features characterizing the froth flotation may comprise redundancy and irrelevance, we adopt a feature selection strategy based on mRMR criteria to find the most useful but less redundant features, and use the proposed STA-LSSVM to recognize the production condition to verify the effectiveness of the selected feature subset.

The experiment data is shown in Table 1. A total of 200 samples are collected from a gold-antimony flotation process, classified into 3 levels: good (65 samples), medium (66 samples) and poor (69 samples). According to Section 2, we extract 19 features to characterize the froth image. Notice that, the bubble size distribution curve is described by 20 piecewise constants in this study. Therefore, the original feature set consists of 38 features. The whole dataset is split into 2 disjoint sets: train (50%) and test (50%). Features and hyper-parameters are both selected and evaluated using 5-fold Cross-Validation (CV) in the training process.

Table 1. Experiment Data

Class	Train Sample	Test Sample	Feature variable
3	100	100	38

Fig. 4(a) shows the 5-fold CV misclassification rate curves of the train data with the sequential feature subsets selected by mRMR and the x-axis represents the size of the subsets. And for comparison, the Fig. 4(b) shows the corresponding predict accuracy of the test data with the sequential feature subsets. It can be seen that only according to the first four features selected by the mRMR criteria, the LSSVM classifier can ensure 100% accuracy in both training and testing process. In addition, the fluctuation in the Fig. 4(b) reveals that the redundant features may cause over-fitting in training process and degrade the generalization performance of the classifier.

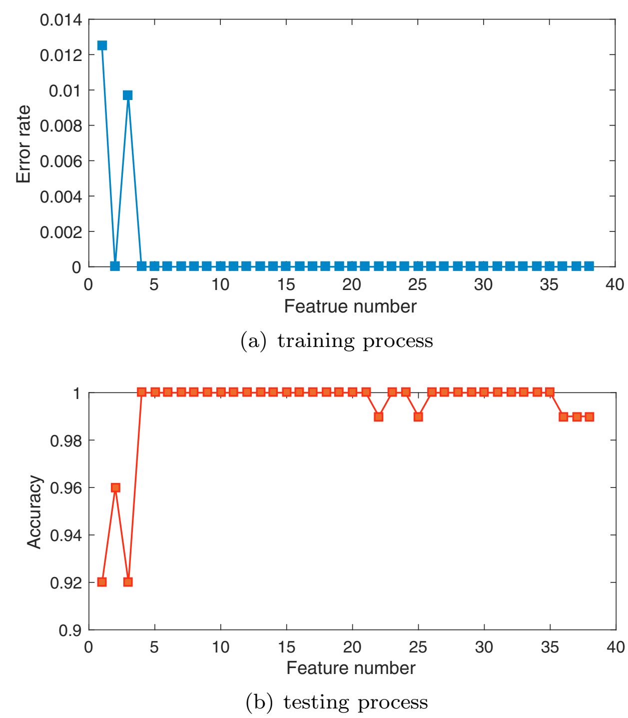

Fig. 4. The training(a) and testing(b) accuracy of STA-LSSVM with the sequential mRMR feature subsets

Note that, in practice, once the respective error of the sequential feature subsets is consistently small, the mRMR filter terminates and outputs the candidate subset directly. Here we chose the first six features for the next wrapper selector to find a compact feature set. Finally, there are only two important features remain in the selected feature set, Hue and Energy of high frequency image. The sample distribution with the selected features is shown in Fig. 5. Obviously, dataset has been divided into three distinct parts. The gaps between each two parts are large enough, which can facilitate the following classification.

To verify the effectiveness of the selected features, we conduct the before-and-after experiment to compare the accuracy and efficiency of the STA-LSSVM learned from the original feature set and the selected subset. For fairness and reproducibility, all of these calculations are carried on MATLAB (Version R2016b) software platform using

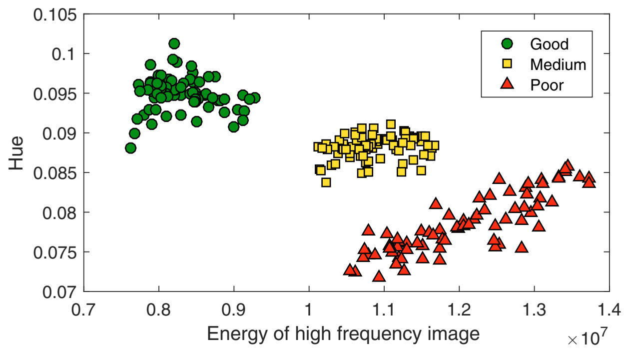

Fig. 5. The sample distribution with two selected features

3.4GHz Intel i7 PC with 8G RAM, and the accuracy and processing time is presented in terms of mean values over the 20 runs in Table 2. It is obvious that the accuracy of the selected subset is stable at 100%, and because of the dimensionality reduction, the processing time is less than the original one, which is critical to on-line recognition. Therefore, it can be concluded that the aforementioned feature selection can not only improve the classification accuracy, but accelerate the recognition process well.

Table 2. "Before and after" comparison

	Number of features	Accuracy(%)	Processing time(s)
Before	38	99.1	0.8836
After	2	100.0	0.7023

6. CONCLUSION

Feature selection is significant for production condition recognition. Appropriate feature selection can improve accuracy and reduce computational cost of the classification. In this paper, a feature selection strategy based on mRMR criterion has been used to find the most useful but less redundant features, Hue and Energy of high frequency image. Meanwhile, a novel classifier STA-LSSVM was proposed to demonstrate the effectiveness of the selected features. The experiments shows that the aforementioned feature selection can not only improve the classification accuracy, but accelerate the recognition process well for the product condition recognition arising in froth flotation. In the future, we will classify the production condition to more classes, and different classifiers will be taken into consideration.

REFERENCES

Bao, Y., Hu, Z., and Xiong, T. (2013). A PSO and pattern search based memetic algorithm for svms parameters optimization. *Neurocomputing*, 117, 98–106.

Cao, B., Xie, Y., Gui, W., Wei, L., and Yang, C. (2013). Integrated prediction model of bauxite concentrate grade based on distributed machine vision. *Minerals engineering*, 53, 31–38.

Estvez, P.A., Tesmer, M., Perez, C.A., and Zurada, J.M. (2009). Normalized mutual information feature selection. *IEEE Transactions on Neural Networks*, 20(2), 189–201.

Gui, W., Liu, J., Yang, C., Chen, N., and Liao, X. (2013). Color co-occurrence matrix based froth image texture extraction for mineral flotation. *Minerals Engineering*, 46, 60–67.

Han, J., Yang, C., Zhou, X., and Gui, W. (2016). Entropy-based estimation of bubble size distributions in froth flotation using B-spline functions. *IFAC-PapersOnLine*, 49(20), 69–101.

Han, J., Yang, C., Zhou, X., and Gui, W. (2017). Dynamic multi-objective optimization arising in iron precipitation of zinc hydrometallurgy. *Hydrometallurgy*, 173, 134–148.

He, M., Yang, C., Wang, X., Gui, W., and Wei, L. (2013). Nonparametric density estimation of froth colour texture distribution for monitoring sulphur flotation process. *Minerals Engineering*, 53, 203–212.

Hsu, C.W., Chang, C.C., Lin, C.J., et al. (2003). A practical guide to support vector classification.

Kwak, N. and Choi, C. (2002). Input feature selection by mutual information based on Parzen window. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12), 1667–1671.

Mitra V, Wang C J, B.S. (2007). Text classification: A least square support vector machine approach. *Applied Soft Computing*, 7(3), 908–914.

Nelder, J.A. and Mead, R. (1965). A simplex method for function minimization. *Commput J*, 7(4), 308–313.

Peng, H., Long, F., and Ding, C.H.Q. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), 1226–1238.

Peng, X., Peng, T., Zhao, L., Song, Y., and Gui, W. (2016). Working condition recognition based on an improved NGLDM and interval data-based classifier for the antimony roughing process. *Minerals Engineering*, 86, 1–9.

Peressini, A.L., Sullivan, F.E., and Uhl, J.J. (1988). *The mathematics of nonlinear programming*. Springer-Verlag New York.

Suykens, J.A. and Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters*, 9(3), 293–300.

Wu, Q. and Peng, C. (2015). Wind power grid connected capacity prediction using LSSVM optimized by the bat algorithm. *Energies*, 8(12), 14346–14360.

Xu, D., Chen, Y., Chen, X., Xie, Y., Yang, C., and Gui, W. (2016). Multi-model soft measurement method of the froth layer thickness based on visual features. *Chemometrics and Intelligent Laboratory Systems*, 154, 112–121.

Zhang, X.L., Chen, X.F., and He, Z.J. (2010). An ACO-based algorithm for parameter optimization of support vector machines. *Expert Systems with Applications*, 37(9), 6618–6628.

Zhou, X., Shi, P., Lim, C.C., Yang, C., and Gui, W. (2018). A dynamic state transition algorithm with application to sensor network localization. *Neurocomputing*, 273, 237–250.

Zhou, X., Yang, C., and Gui, W. (2012). State transition algorithm. *Journal of Industrial & Management Optimization*, 8(4), 1039–1056.