Component Content Soft-Sensor of SVM Based on Ions Color Characteristics

Lu Rongxiu$^{1,2,3}$, Yang Hui*$^{2,3}$, Zhang Kunpeng$^{2,3}$

$^1$School of Electrical Mechanical & Electrical Engineering, Nanchang University, Nanchang, Jiangxi, 330031, China
$^2$School of Electrical and Electronic Engineering, East China Jiaotong University, Nanchang, Jiangxi, 330013, China, Ph/Fax: (+86 791) 87046216/87046184
$^3$Key Laboratory of Advanced Control & Optimization of Jiangxi Province, Nanchang, Jiangxi, 330013, China

*corresponding author, e-mail: ecjtu_rxlu@163.com, yhshuo@263.net*, ecjtu.zhangkunpeng@163.com

Abstract

In consideration of different characteristic colors of Ions in the P507-HCL Pr/Nd extraction separation system, ions color image feature H, S, I that closely related to the element component contents are extracted by using image processing method. Principal Component Analysis algorithm is employed to determine statistics mean of H, S, I which has the stronger correlation with element component content and the auxiliary variables are obtained. With the algorithm of support vector machine, a component contents soft-sensor model in Pr/Nd extraction process is established. Finally, simulations and tests verify the rationality and feasibility of the proposed method. The research results provide theoretical foundation for the online measurement of the component content in Pr/Nd countercurrent extraction separation process.

Keywords: Pr/Nd extraction, image processing, PCA, SVM

1. Introduction

Rare earth is a non-renewable resource and Pr/Nd are widely used elements among all rare-earth elements and play an important role in the rare-earth field. The process flow of the P507-HCL Pr/Nd extraction separation system has been widely used to extract high purity Pr or Nd oxide, of which the element component content is one of the key process parameters. The online measurement of the component content must be solved firstly to realize the automatic control in the extraction process because that the Pr/Nd extraction process has the features of long process flow, complex mechanism and many influence factors. But most of the online measurement devices in use currently have defects such as high cost, low stability and reliability and measurement lag [1-3].

There are some visual information related to working condition in the Pr/Nd extraction separation process, and Pr and Nd ions feature color image information can be used to forecast the element component content [4]. The soft-sensor method based on ions feature color have been used successfully in the industry of the full-range optical PH sensor with the intuitive, fast, and cheap advantages [5]. However, in the Pr/Nd extraction field, the application is less. In [6], Image processing method is adopted to extract rare-earth ions color features and the relationship between component content and ions color features is described quantificationally, a model between component content and H components statistics mean is built by the least square method. However, the model is only suitable for the situation that single component content plays a leading role. When purity of the element component content is low, its prediction error is larger. Support Vector Machine (SVM) possesses the ability of handling some practical problems such as a small sample, multivariable, nonlinear and local minimum etc. It is widely used in industrial process modeling [7-8]. And it can be used to provide a new way for solving online detection of the element component contents in the Pr/Nd extraction process [9].

In this paper, in view of the ion characteristic color in the Pr/Nd extraction process, color characteristics of ions color image are extracted and the auxiliary variables are got with Principal Component Analysis (PCA) [10-11], and a soft-sensor model of the element component content in the Pr/Nd extraction process is set up with SVM algorithm.
2. Description of Pr/Nd Extraction Process

Figure 1 shows the flow chart of Pr/Nd extraction. The whole production flow is composed of an extraction stage and a washing stage. The mixed liquor of Pr/Nd, extraction solvent P507 and washing liquid HCL are added into extractor from the last stage, the first stage of the extraction stage and the last stage of the scrubbing stage, respectively. After exchange and purification between each extraction stage and washing one, extract liquid of Pr whose purity is $y_1$ is obtained from the first extraction stage, similarly, extract liquid of Nd whose purity is $y_2$ is gotten from the last scrubbing stage.

Due to the multistage characters, regulating control effects of extraction liquid, washing liquid or feed usually affect purity of two outlet production long after several hours' transfer. For the complex mechanism of the extraction process, element component content is hard to measure online and there is lack of an effective online analysis method and automatic control model [1,12-13]. However, the ions of Pr and Nd can be taken on characteristic color for their unique electron structure in the Pr/Nd extraction separation process. Namely, the ion of Nd is enriched in the washing stage and emerges gradually its feature color, that is purple color, as well as the ion of Pr shows its one of apple green in the extraction stage. So the element component contents of each extractor in the extraction process can be judged according to the change of the ions feature color.

3. Features Extraction of Pr and Nd Ions Image

Taking the Pr/Nd extraction separation production line of one company as the research object, producing field of extraction process has been simulated. By collocating the mixed solution of Pr/Nd and 38 pictures that can reflect the color change of each extractor in the washing stage of the Pr/Nd extraction process are photographed using the industrial camera. Then we can extract the characteristic value in these characteristic color images using extracting color characters method. Because the outlet production of the washing stage is Nd, the paper is focus on the characteristic color of Nd whose content range is 49.7% ~ 96.1%.

3.1 Color Features Extraction of Nd Ions color Image

Compared with other visual features, color features have some advantages of good stability and strong robustness. For the mixed solution of Pr and Nd, color features are the most direct one among all visual features [6]. To describe the ions features color in the mixed solution of Pr and Nd, HSI color space which accords better with our visual properties will be adopted. Where, H represents hue, S represents saturation and I represents intensity. Those three
characteristic values are mutually independent [14]. Therefore, the calculation can be greatly simplified in HSI color space. These characteristic values of HSI can be transformed from the most basic RGB values, because the two are different representations of the same physical quantities.

The formula of H is the following one:

$$H = \begin{cases} \theta , & B \leq G \\ 360 - \theta , & B > G \end{cases}$$

where,

$$\theta = \arccos \left( \frac{1}{2} \left[ \frac{(R - G) + (R - B)}{[(R - G)^2 + (R - G)(G - B)]^{1/2}} \right] \right)$$

The calculation formula of S and I are given, respectively:

$$S = 1 - \frac{3}{R + G + B} \left[ \min(R, G, B) \right]$$

$$I = \frac{1}{3} (R + G + B)$$

According to the above formula, the ions feature color change of Nd can be transferred to information of HSI space. Because the component value of H, S or I which is extracted from those ion color pictures is a very large dataset, the statistical characteristic of ions color information in the Pr/Nd extraction process may be described by means of the statistic mean, namely

$$\mu_i = \frac{1}{N} \sum_{j=1}^{N} P_{ij}$$

Where, $$\mu_i (i=1,2,3)$$ is a statistic mean of a certain component (H, S or I), N is its total statistic number, $$P_{ij}$$ is its statistic value.

### 3.2 Selection of the Instrumental Variables based on PCA

In order to determine the influence level of the characteristic values H, S or I of ions color information to the element component contents, namely, to select input variables and the corresponding number of variables for the element component contents soft-sensor model, the Principal Component Analysis (PCA) is employed to analyze the statistic mean of three characteristic values.

The PCA is a multivariate statistical method which studies the correlation among many numerical variables and finds out the main influence factors. On the basis of the above 38 images, ions color information feature values can be extracted and the corresponding statistic mean can be expressed as $$X = \{\mu_1, \mu_2, \mu_3\}$$, $$i = 1, 2, \ldots, 38$$ . Because error inevitably exists in the process of data collection, it is required to preprocess the dataset.

The pretreatment of dataset includes two parts. Firstly, it is the gross error processing to statistic mean of feature components. The Pauta Criterion of statistics assessment method is used to reject the abnormal values, and 37 groups of data are remained. Secondly, there is different between the order of magnitudes of the different characteristics components, so, it’s need to make a standardized treatment by Eq.(5) [18].

$$\mu_i = \frac{\bar{\mu}_i - \mu_i}{\sqrt{\text{var}(\mu_i)}} \quad (k = 1,2,\ldots,37, i = 1,2,3)$$

Where, $$\bar{\mu}_i$$ and $$\sqrt{\text{var}(\mu_i)}$$ are the average and standard deviation of $$i^{th}$$ characteristics component, respectively. And after standardized treatment, the average of each characteristics component’s statistic mean is 0 and the corresponding standard deviation is 1. The above processed data can be expressed as the matrix of variable $$X = \{\mu_{m1}, \mu_{m2}, \mu_{m3}\}$$, $$k = 1,2,\ldots,37, X \in (37 \times 3)$$ . By using the PCA algorithm, the X is analyzed as following:

1. Calculating the X’s correlation coefficient matrix
Where,  

\[ R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \]  

(6)

\[ r_{ij} = \frac{\sum_{k=1}^{3} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{3} (x_{ik} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^{3} (x_{jk} - \bar{x}_j)^2}}, \quad \bar{x}_i, \bar{x}_j \text{ are expectation values of the } i^{th} \text{ and } j^{th} \text{ argument, respectively.} \]

(2) Calculating the characteristic value and feature vector of coefficient matrix R, and sorting the characteristic value from big to small.

(3) Computing the contribution and the cumulative contribution of the principal component according to the Eq. (7) and (8):

The contribution is accounted for:

\[ \lambda_i = \frac{\lambda_i}{\sum_{i=1}^{3} \lambda_i}, \quad i = 1, 2, 3 \]  

(7)

The cumulative contribution is accounted for:

\[ \frac{\sum_{i=1}^{3} \lambda_i}{\sum_{i=1}^{3} \lambda_i}, \quad i = 1, 2, 3 \]  

(8)

According to the above steps, the principal component analysis of X is implemented and contribution of main components is shown in Table 1.

<table>
<thead>
<tr>
<th>Component</th>
<th>Characteristic value</th>
<th>Contribution /%</th>
<th>Cumulative contribution /%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_H )</td>
<td>0.1161</td>
<td>92.70</td>
<td>92.70</td>
</tr>
<tr>
<td>( \mu_S )</td>
<td>0.0084</td>
<td>6.67</td>
<td>99.37</td>
</tr>
<tr>
<td>( \mu_I )</td>
<td>0.0008</td>
<td>0.63</td>
<td>100</td>
</tr>
</tbody>
</table>

Generally, the first few variables are considered as the main component ones as long as their accumulative contribution rate reaches 85~95% [15]. From table 1, it is known that the first two variables of the accumulative contribution rate reached 99.37%, which shows that the two eigenvalues H and S have larger influence to the element component content. For this reason, the statistic mean \( \mu_H \) and \( \mu_S \) can be selected as the input variables of the soft-sensor model. According to the mechanism analysis in the Pr/Nd extraction separation process, the corresponding relation between element component content \( y \) and input variables of \( \mu_H \) and \( \mu_S \) can be expressed as following:

\[ y(\mu) = f(\mu_H, \mu_S) \]  

(9)

Where, \( f(\cdot) \) denotes nonlinear functional relationship between input and output variables.

4. SVM Model Based on the Characteristic Component

The support vector machine is a new machine learning method proposed by Vapnik et al in the early 90s of last century [16-17, 19], which is based on a limited sample of information to find the best compromise between model complexity and ability to learn in order to obtain the best generalization ability. Its basic idea is that it maps practical problems to a high dimensional feature space through the nonlinear transformation and then fitting function in this high-dimensional space. In the process, the complexity of the algorithm is independent of the sample dimension; the global optimal solution of limited sample information is available.
According to the nonlinear function Eq.(9) which describes the relationship between the element component content and main characteristic components, a soft-sensor model based on SVM is established whose system construction is shown in Figure 2. In which, $\mu_H$, $\mu_S$ are input vectors, and $SV_1$, $SV_2$...$SV_N$ are support vectors, $N$ is the support vector number; $\alpha - \alpha^*$ is the network weight.

Then the modeling process for the soft-sensor model structure in Figure 2 will be analyzed. Assumed that there is $m$ observed values, the input and output dataset can be expressed as $\{\mu, v\} = \{\mu_i, \mu_j, y\} = \{\mu_{hi}, \mu_{si}, y\}, i = 1, 2, \cdots, m$. Under the control of the accuracy error that is $\varepsilon$, SVM can be fitted by the follow equation and unknown function also can be estimated.

$$y(\mu) = w \cdot \phi(\mu) + b$$

(10)

Where, nonlinear function $\phi(\cdot): R^2 \rightarrow R$ stands for mapping input space into high dimensional feature space.

According to the principle of structural risk minimization (SRM), SVM can be defined as a followed convex quadratic optimization problem.

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} (\xi_i + \xi^*_i)$$

(11)

$$s.t \quad \begin{cases} 
  y_i - w \cdot \mu_i - b \leq \varepsilon + \xi_i, \\
  w \cdot \mu_i + b - y_i \leq \varepsilon + \xi^*_i, \\
  \xi_i, \xi^*_i \geq 0 
\end{cases}$$

(12)

Where, $\xi$, $\xi^*$ are slack variables which are adopted to process these data when function $y$ can’t estimate under the accuracy error that is $\varepsilon$. C is penalty factor and means that the penalty to the samples beyond the error is stronger with greater C.

Lagrange multiplier is introduced to solve the quadratic optimization problem. According to KKT condition, dual optimization problem is obtained as follows:

$$\max_{\alpha, \alpha'} W(\alpha, \alpha') = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} (\alpha_i - \alpha^*_i)(\alpha_j - \alpha^*_j)K(\mu_i, \mu_j) - \varepsilon \sum_{i=1}^{m} (\alpha_i + \alpha^*_i) + \frac{\sum_{i=1}^{m} y_i (\alpha_i - \alpha^*_i)}{2}$$

(13)

$$s.t. \quad \begin{cases} 
  \sum_{i=1}^{m} (\alpha_i - \alpha^*_i) = 0 \\
  0 \leq \alpha_i \leq C \quad i = 1, 2, \cdots, m \\
  0 \leq \alpha^*_i \leq C 
\end{cases}$$

(14)
Solving equation (13) with constraint of Eq. (14) to get Lagrange multiplier $\alpha, \alpha^\ast$. That is:

\[
\begin{align*}
    w &= \sum_{i=1}^{m} (\alpha - \alpha^\ast) \cdot \mu_i \\
    b &= -\frac{1}{2} \sum_{i=1}^{m} [(\alpha - \alpha^\ast) (K(\mu_i, \mu_j) + K(\mu_i, \mu))] 
\end{align*}
\]

(15)

The fitting function of the element component content based on ions color feature components can be gotten by substituted Eq.(15) into Eq.(8). Where, $K(\mu_i, \mu_j) = \phi(\mu_i) \cdot \phi(\mu_j)$ is kernel function satisfying the Mercer condition.

5. The Experimental Results and Analysis

The residual 37 group statistic mean data of ion color feature components and their corresponding element component content form an input and output data matrix which can be expressed as $(\mu, y) = (\mu_{ik}, y_{ik}), k = 1, 2, \ldots, 37$ after the normalized and standardized treatment. In order to verify the generalization ability of this model aiming to the data of ion color features, 25 groups sample data are randomly selected for model fitting and the rest 12 groups for prediction.

The selection of kernel influences the model performance of the fitting and forecasting or generalization in a large degree. According to the SVM modeling concept, when the specific forecast model is established, it is usually directly determined the kernel function that satisfies the Mercer condition without knowing the specific form of the nonlinear transformation; then with a selected kernel function, the best model parameters is able to be optimized. In this paper, the commonly used Gaussian Radial Basis Function (RBF) function is used as a kernel function:

\[
k(\mu, \mu) = \exp\left(-\frac{||\mu - \mu||^2}{2\sigma^2}\right)
\]

(16)

In the entire SVM modeling process, the parameters that should be adjusted are insensitive coefficient $\epsilon$, penalty coefficient $C$ and kernel function parameter $\sigma$. To seek the best collocation of these parameters is the problem to select best model.

Grid searching method is a simpler one among all methods used to determine parameters of SVM model, which is suitable to the condition that the sample data are less. Because the sample data for modeling is 25 in this paper which is relatively less, the grid searching method is appropriate and adopted to determine the optimal model parameters. Its optimizing principle is as follows. When $C$ and $\sigma$ are respectively taken N and M different values, there have $N \times M$ different types of $[C, \sigma]$ combination. The SVM model is trained to each combination. When the root mean square error (RMSE) is minimum, the corresponding combination for $[C, \sigma]$ is the best combination for parameters.

<table>
<thead>
<tr>
<th>Table 2. Prediction performance of component content based on SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modeling performance(%)</td>
</tr>
<tr>
<td>PMRE</td>
</tr>
<tr>
<td>1.8283</td>
</tr>
</tbody>
</table>

Figure 3. Modeling result of component content based on SVM
Usually, a larger step has been taken firstly in a wide range by the grid search means roughly, and the preliminary optimal parameter combination \([C', \sigma']\) will be gotten. Then a detailed search is selected and adopted at this point nearby with the same way, the optimal parameter combination \([C, \sigma]\) will be obtained. According to this idea, these selected optimal parameter values are \(\varepsilon = 0.008, C = 300, \sigma = 0.6\). The component content soft-sensor model based on SVM in the Pr/Nd extraction process is established with these parameter values. To obtain the generalization performance index of the newly founded model, a test to this model will be done by the residual 12 group sample data. The results of testing show in the Figure.3.

To measure generalization capability of a soft-sensor model, the most representative performance indexes which are the positive maximum relative error (PMRE), the negative maximum relative error (NMRE) and the RMSE are usually chosen. The prediction performance indexes of the proposed model listed in Table 2. It is known from the Figure 3 and Table 2 that the absolute values of the relative maximum error among the prediction indexes of the model is all less than 2%, and the RMSE is less than 1%, which shows the element component content soft-sensor model based on SVM in the Pr/Nd extraction process have high prediction precision, whose input variables are the feature components of ion color image information processed by PCA.

6. Conclusion

In this paper, a component content soft-sensor model based on ion color Characteristics is proposed. With the Principal Component Analysis method, the statistic mean of H and S components that have greater influence to the element component content in the Pr/Nd extraction process are chosen as the input variables of the forecasting model. The component content soft-sensor model is established by means of SVM algorithm. And simulation test shows that the proposed model has stronger generalization ability and is a good way to solve the online forecasting of component content in the Pr/Nd extraction process.

Acknowledgment

The authors acknowledged the National Natural Science Foundation of China (51174091), the Jiangxi Natural Science Foundation of China(2010GQS0032), the Jiangxi Project of Education Department of China(GJJ10137) and East China Jiaotong University Research Fund of China(09DQ07) for having provided the financial support for this work.

References


