ISA Transactions xxx (xxxx) xxx

Contents lists available at ScienceDirect

ISA Transactions

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

Research article

Sintering conditions recognition of rotary kiln based on kernel modification considering class imbalance

Dingxiang Wang [a], Xiaogang Zhang [a,*], Hua Chen [b], Yicong Zhou [c], Fanyong Cheng [d]

[a] College of Electrical and Information Engineering, Hunan University, Changsha 410082, China
[b] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China
[c] Department of Computer and Information Science, University of Macau, Macau 999078, China
[d] College of Electrical Engineering, Anhui Polytechnic University, Wuhu 241000, China

ARTICLE INFO

Article history:
Received 2 November 2019
Received in revised form 6 July 2020
Accepted 7 July 2020
Available online xxxx

Keywords:
Sintering condition recognition
Class imbalance
ODM
Kernel modification

ABSTRACT

Accurate sintering condition recognition (SCR) is an important precondition for optimal control of rotary kilns. However, the occurrence probability of abnormal conditions in the industrial field is much lower than normal, resulting in imbalanced class sintering samples in general. This significantly deteriorates the effectiveness of existing recognition models in abnormal condition detection. In this paper, an integrated framework considering class imbalance is proposed for sintering condition recognition. In the proposed framework, after analysing the characteristics of thermal signals by the Lipschitz method, four discriminant features are extracted to comprehensively describe different sintering conditions. In addition, focusing on the class imbalance of sintering samples, the kernel modification method is introduced to enhance the optimal marginal distribution machine (ODM), and a novel recognition model kernel modified the ODM (KMODM) is proposed for SCR. By constructing a new conformal transformation function to modify the ODM kernel function, KMODM optimizes the spatial distribution of training samples in the kernel space, thereby alleviating the detection accuracy deterioration of the minority class. The experimental results on real thermal signals and standard datasets show that the KMODM model can effectively handle imbalanced data. Based on this, the proposed SCR framework can reduce the misjudgement of abnormal conditions and balance the recognition accuracy of each condition.

1. Introduction

The sintering condition of the rotary kiln directly influences the clinker quality. Taking the alumina rotary kiln as an example, there are three main sintering conditions in kilns: normal, super-chilled and super-heated. Super-chilled and super-heated conditions are abnormal conditions. Under super-chilled conditions, the raw materials calcined insufficiently, resulting in incomplete melting of alumina in the subsequent refining processes. Under super-heated conditions, the clinker is over-calcined and easy to agglomerate, which damage the refractory materials and increase the failure rate of the rotary kiln. Accurate estimation and recognition of the sintering conditions are very important for safe operation and efficient production. It is also the premise for the control system to take appropriate measures to maintain normal production.

Sintering conditions recognition (SCR) is not an easy task at all times. Temperature measured by physical measuring equipment is an important index for SCR. However, it is not enough to estimate the sintering conditions only by temperature. For example, under normal sintering conditions, the sintering temperature (ST) of an alumina rotary kiln may vary from 1000 to 1300 °C according to many complex factors, such as material composition or coal feeding value. In addition, because of the large delay characteristic of the rotary kiln, the change in temperature generally lags behind the change in sintering condition, and ST cannot reflect the current sintering condition in time.

Because the sintering process involves complex physical and chemical reactions, the mechanism model of a rotary kiln is difficult to construct. In fact, in most coal-fired industrial fields, the sintering condition is determined by operators based on experience through the variation in process data. In recent years, research on the automatic recognition of sintering conditions or related parameters using process data has increased. The process data for sintering condition recognition of a rotary kiln include flame images and thermal signals. Flame image-based methods play important roles in SCR owing to their intuition and promptness. Researchers have extracted various vision features [1–3]

* Corresponding author.
E-mail addresses: wangdx@hnu.edu.cn (D. Wang), zhangxg@hnu.edu.cn (X. Zhang), chua@hnu.edu.cn (H. Chen), yicongzhou@um.edu.mo (Y. Zhou), b12090031@hnu.edu.cn (F. Cheng).

https://doi.org/10.1016/j.isatra.2020.07.010
0019-0578/

and developed classifiers [4] based on flame images. However, in pulverized coal fields, smoke and dust disturbances may result in image degradation, which seriously affects the accuracy of the algorithm. Additionally, many researchers have analysed the characteristics of thermal signals and used soft computing techniques for SCR. In [5], a locally linear neuro-fuzzy (LLNF) model was used to establish a power prediction model of the drive motor based on thermal signals and identified the sintering conditions according to the difference between the actual motor power and the predicted value. Similarly, the LLNF-based model was also used to detect the abnormal conditions of rotary kilns in [6]. Other models, such as the hierarchical wavelet TS-type fuzzy inference system (HWFIS) [7] and adaptive threshold monitoring scheme [8], also achieved good performance in recognizing sintering conditions using thermal signals. In [9], the chaotic characteristics of thermal signals were discovered. Based on this, excellent results have been achieved in the prediction of parameters closely related to sintering conditions, such as sintering temperature [10] and coal feeding [11], using thermal signals.

The abovementioned data-driven methods have made great achievements in sintering condition recognition. However, the collected process data are generally class imbalanced because the occurrence probability of abnormal conditions is far less than that of normal conditions, which is generally ignored by existing methods. In the case of class imbalance, the learned separator of the recognition model always skews towards the minority class. As a result, the generalization performance of the recognition model and recognition accuracy of abnormal conditions deteriorates. In the field of the rotary kiln, the misjudgement of abnormal conditions may cause a series of misoperations and cause serious consequences. For instance, if the super-heated condition is misjudged as super-chilled, the control system may carry out a series of operations corresponding to super-chilled conditions such as increasing coal feeding, which will cause a continuous rise of ST, thus reducing the clinker quality and damaging the equipment. This problem of class imbalanced data in rotary kilns usually challenges the designation of sintering condition recognition models.

At present, numerous studies have focused on the improvement of imbalanced classification strategies. The mainstream methods include: data preprocessing, algorithm improvement, and kernel methods. The data preprocessing methods usually refer to resampling the training data using different techniques, for example, generating minority samples in kernel space [12] or using generative adversarial networks (GAN) networks [13], eliminating majority samples from the overlapping region [14]. However, if the spatial distribution of the training data is not revealed, the resampling process may lead to the introduction of noise (oversampling) or elimination of useful information (undersampling). The algorithm improvements mainly refer to cost-sensitive methods. It improves the detection rate of the minority class by assigning larger penalty factors to the minority samples [15,16] or the samples are difficult to recognize [17]. Cost-sensitive methods have achieved excellent performance in the improvements of single-model classification algorithms, such as the large margin distribution machine (LDM) [18] and fuzzy-rough nearest neighbour (FRNN) [19]. Additionally, introducing cost-sensitive strategies into the ensemble framework has also made many achievements, and the most common ensemble frameworks include boosting [20–22], tree augmented naive Bayes (TANB) [23] and error-correcting output codes (ECOC) [24]. The cost-sensitive technologies are usually adjusting the influence weight of the training data for the minimization misclassification costs. In fact, the effect of this method is limited since the influence weight is used only as the upper bound of Lagrange multiplier α according to the Karush–Kuhn–Tucker (KKT) conditions. The adjustment of the influence weight does not necessarily change α. Different from data preprocessing and algorithm improvement methods, the kernel methods attempt to handle imbalanced data by modifying the kernel function of the classifier using conformal transformation [25]. In this way, the distribution of training data can be changed, and the spatial resolution of the class boundary can be enlarged in kernel space to better distinguish different classes. The core of this method lies in the design of a conformal transformation function, which is usually constructed according to the initial margin [26,27] or distribution of boundary samples [28]. However, most of the existing conformal functions do not consider the imbalance rate of training data, and simply relying on increasing the spatial resolution of overlapping regions has limited performance improvement for the classifier in processing imbalanced data.

In this paper, focusing on the class imbalance problem of thermal signals in the SCR task, some effective features of thermal signals are extracted, and a kernel modified optimal margin distribution machine (KMODM) is developed. The features extracted from thermal signals can describe the mechanism of sintering conditions comprehensively, which is conducive to the achievement of accurate recognition of sintering conditions. The optimal margin distribution machine (ODM) is a newly proposed classifier that was originally designed for balanced classification tasks with better generalization performance and strong classification ability [29,30]. By constructing a reasonable conformal function to modify its kernel function, the ability of ODM to deal with imbalanced classification tasks can be effectively improved. This is the first time that the class imbalance issue has been considered in SCR tasks of rotary kilns. It is worth noting that this paper does not simply combine the existing kernel modification methods with ODM but develops an SCR framework according to the characteristics of sintering condition samples. The main contributions of this paper are summarized as follows.

(a) A novel data-driven sintering condition recognition framework for rotary kilns is proposed in this paper. By extracting several discriminate features of thermal signals and modifying the kernel matrix of ODM, the misclassification of abnormal conditions caused by class imbalance is significantly reduced, and the accurate and robust recognition of sintering conditions is achieved.
(b) According to expert experience and the results of thermal signal analysis, four statistical and dynamic features are introduced to distinguish different sintering conditions. These features have high separability and can comprehensively describe the characteristics of different sintering conditions, thereby improving the recognition accuracy of sintering conditions.
(c) For the class imbalance issue of the SCR task, a new conformal transformation function is constructed for kernel modification of ODM. The proposed conformal function can optimize the distribution of the minority samples in kernel space and alleviate the influence of imbalanced data. The experimental results show that the proposed KMODM model can reduce the misjudgement of abnormal conditions and is superior to other baseline models.

The remainder of this paper is organized as follows. The analysis of thermal variables and feature extraction are presented in Section 2. In Section 3, after a brief introduction of the ODM and kernel modification technique, a new conformal function is designed, and an improved classification model KMODM is proposed in this section. In Section 4, the SCR framework is proposed completely, the experimental results are recorded and discussed in detail. Finally, the conclusions and prospects are presented in Section 5.

Please cite this article as: D. Wang, X. Zhang, H. Chen et al., Sintering conditions recognition of rotary kiln based on kernel modification considering class imbalance. ISA Transactions (2020), https://doi.org/10.1016/j.isatra.2020.07.010.

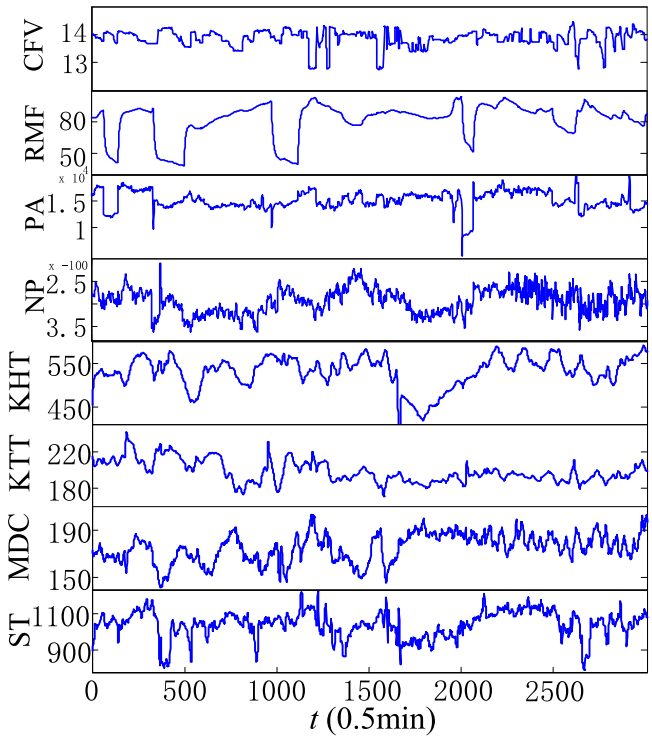

Fig. 1. Thermal signals of the rotary kiln.

Table 1
The thermal variables of the rotary kiln.

Variables	Description	Unit
CFV	The Coal Feed Rate	t/h
RMF	The Material Feed Rate	t/h
PA	The Volume of Primary Air	m^3/h
NP	Negative Pressure of Kiln Tail	Pa
KHT	Kiln Head Temperature	°C
KTT	Kiln Tail Temperature	°C
MDC	Current of Main Motor	A
ST	Temperature of Sintering Zone	°C

2. Thermal signals analysis and feature extraction

2.1. Thermal signals analysis

In rotary kilns, there are eight main thermal signals, including the coal feeding value (CFV), raw meal flow (RMF), primary air (PA), negative pressure (NP), kiln head temperature (KHT), kiln tail temperature (KTT), main driver current (MDC) and ST. A detailed description of these thermal signals is given in Table 1. In Fig. 1, a part of these thermal signals is shown with a length of 100,000 at a 30-s sampling interval.

The noise introduced by sensors is unconducive to the analysis of thermal signals. Therefore, the mean filter is used to smooth the thermal signals in this paper. The definition of the mean filter is as follows:

$$X(t) = \frac{1}{w} \sum_{x(t)\in S} x(t), \tag{1}$$

where S is a sliding window of width w, which is fixed to $w = 10$ in this paper, $x(t)$ is the thermal signal, and $X(t)$ denotes the result after the mean filtering.

Table 2
The analysis results and RPs of thermal signals.

Variables	FRI	LRI	RPs (0.5 min)	
			Start time	End time
CFV	10	44	T-44	T-10
RMF	20	65	T-65	T-20
PA	4	52	T-52	T-4
NP	0	58	T-58	T
KHT	0	55	T-55	T-0
KTT	20	50	T-50	T-20
MDC	10	50	T-50	T-10
ST	4	82	T-82	T-4

2.2. Relevant period analysis of thermal signals

When the rotary kiln is treated as a control system, the ST, which is measured by a colorimetric pyrometer, can be treated as the kiln output, while other thermal signals are input to the system. In our experience, the variation in the kiln input and output has delay characteristics. Generally, it takes a while for the variation in input signals to change the output of the system. For example, after spraying the coal powder into the kiln, it may take a few minutes to affect the ST, and the effect on the sintering condition may take longer. To estimate the variation-relevant time period of the input variable, a model-free approach named the Lipschitz method is adopted. It was initially proposed to determine the order of nonlinear systems [31], and its improvement can be used to estimate the input delay of a complex nonlinear system [32]. The estimation result of the Lipschitz method can reveal the first relevant input (FRI) and last relevant input (LRI); thus, the time period between the first and last output-related inputs can be defined as the relevant period (RP).

To obtain the RP of each input signal, several models with a single thermal signal as the input and the ST as the output are constructed in this paper. The analysis results obtained by the Lipschitz method are shown in Fig. 2 and Table 2. In Fig. 2, the LRI and FRI of each thermal signal can be obtained and used to define RP. Using the thermal signals in the RP to extract sintering condition features, the information redundancy can be reduced effectively. The RP of each thermal signal at current time T is shown in Table 2.

Fig. 3 shows the thermal variables in the RP of a super-chilled sample. In this situation, the values of CFV and PA are relatively small, resulting in a lower ST. Meanwhile, as the NP increases, the flame position moves towards the kiln tail, and thus, the KHT decreases while the KTT increases. It can also be seen from the MDC results that under super-chilled conditions, the material has good fluidity due to insufficient sintering; thus, the power consumption of the motor is low while the MDC is reduced.

In contrast to super-chilled materials, under super-heated conditions, as shown in Fig. 4, the material becomes sticky and sticks more easily to the kiln wall. This generally leads to high power consumption of motors and an increase in the MDC. The kiln operator generally reduces the CFV while increasing the RMF to stabilize the sintering condition.

Compared with two abnormal conditions, the rotary kiln usually works in a relatively stable state under normal conditions, and the change in the thermal signal is relatively small. As shown in Figs. 3–5, the normal and abnormal conditions can be distinguished by signal stability in the RP.

According to the analysis of different conditions, the statistical features and dynamic features of the thermal signals in the RP are important for recognizing sintering conditions in the kiln. Therefore, in this paper, these two types of features are extracted to describe the sintering conditions of a rotary kiln.

Please cite this article as: D. Wang, X. Zhang, H. Chen et al., Sintering conditions recognition of rotary kiln based on kernel modification considering class imbalance. ISA Transactions (2020), https://doi.org/10.1016/j.isatra.2020.07.010.

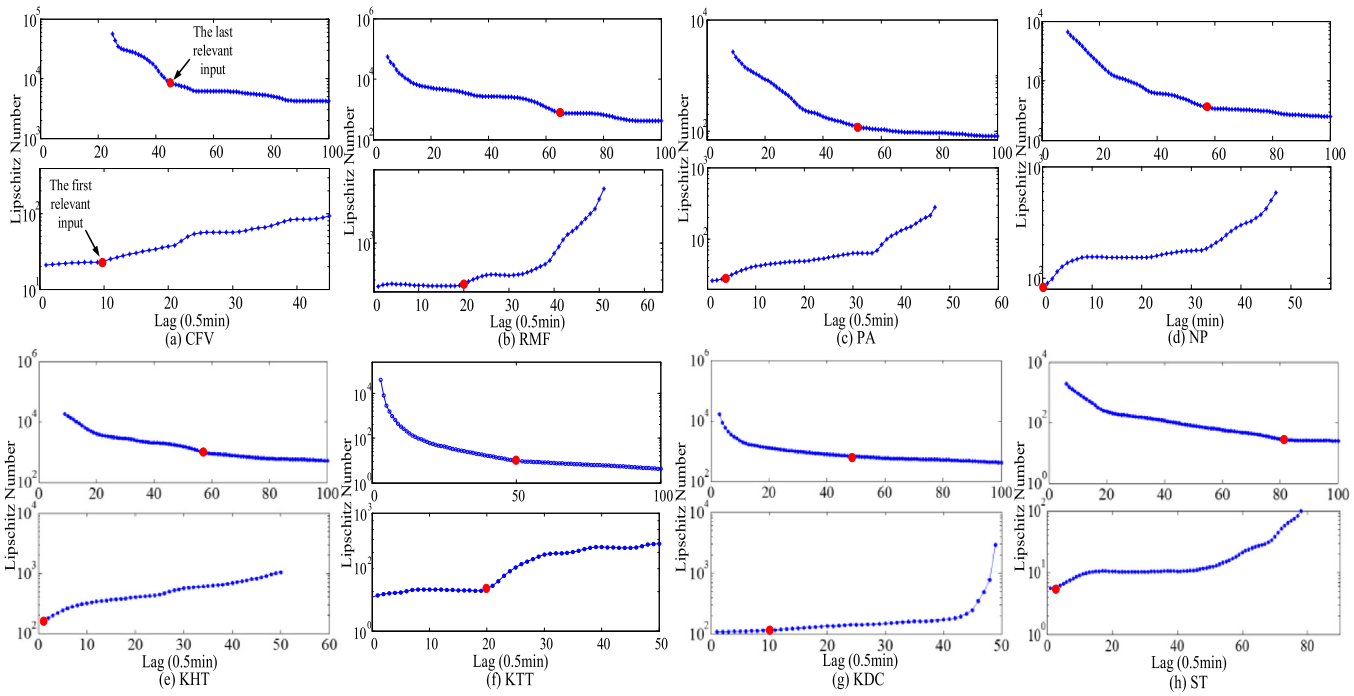

Fig. 2. The analysis results for RP estimation.

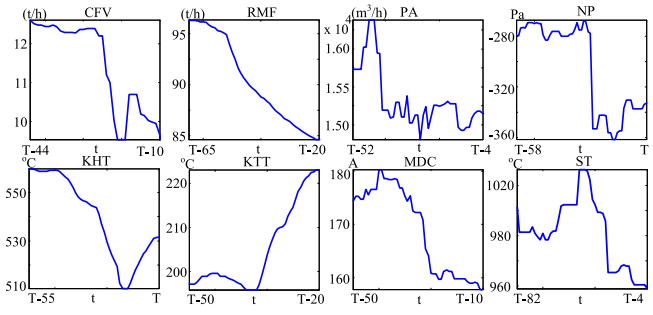

Fig. 3. The thermal signals in the RP of a super-chilled sample.

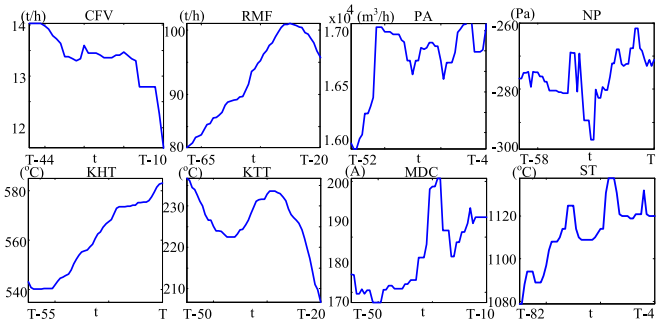

Fig. 4. Thermal signals in the RP of a super-heated sample.

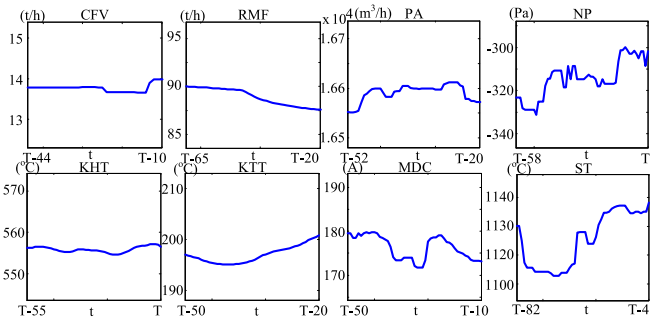

Fig. 5. Thermal signals in the RP of a normal sample.

2.3. Feature extraction

2.3.1. Statistical features

According to previous studies [1], the statistical features, including the mean values and trend features, of thermal signals are extracted as the basis for SCR in this paper.

Mean Value

The mean value of each thermal signal in the RP is closely related to sintering conditions. For example, under super-heated conditions, the means of MDC and KHT are generally higher. The following Eq. (2) can be used to extract the mean value of each thermal signal in the RP:

$$M_D = \frac{1}{L_{DR}} \sum_{i \in DR} X_D(i), \tag{2}$$

where M_D represents the mean feature of the thermal signal, X_D is the thermal signal D after denoising, D $\in$ {CFV, RMF, PA, NP, KHT, KTT, MDC, ST}; DR is the RP of signal D, and L_{DR} is the length of the DR.

Trend Feature

The trend of thermal signals is an unignorable feature for sintering condition recognition. It can be defined as slopes derived from the linear fitting of thermal data. The model of linear fitting is as follows:

$$f(x; a, b) = ax + b, \tag{3}$$

where a and b are the slope and intercept of the fitting line, respectively. The error between the fitted line $f(x)$ and the thermal signal y can be minimized by minimizing the average fitting

Please cite this article as: D. Wang, X. Zhang, H. Chen et al., Sintering conditions recognition of rotary kiln based on kernel modification considering class imbalance. ISA Transactions (2020), https://doi.org/10.1016/j.isatra.2020.07.010.

loss:

$$\arg\min_{a}\min_{b}\frac{1}{L_{DR}}\sum_{i=1}^{L_{DR}}(X_{DR}(i)-(ax(i)+b))^2, \tag{4}$$

where $X_{DR}(i)$ is the ith data of thermal signal D in the RP, and the optimum value of the slope a is the trend feature. The trend feature of thermal signal D is denoted as A_D in this paper.

2.3.2. Dynamic features

Some previous researchers have revealed that the statistical features usually have poor separability. The analysis results in Section 2.2 show that the thermal signal stability is different under different sintering conditions, which is consistent with the situation at the industrial site. Therefore, by introducing dynamic features, the sintering conditions can be described more comprehensively and distinguished more accurately.

Short-time Energy

Short-term energy can be used to distinguish the signal stability in a short time, and it is widely adopted in speech signal processing and recognition [33]. For a series signal $X_D(i)|_{i=1}^{t}$, it can be defined as:

$$E_n(t)=\sum_{i=t-(N-1)}^{t}[x(i)w(t-i)]^2, \tag{5}$$

where $w(t-i)$ is a window function and N is the width of the window. According to operating experience, the stability of thermal signals under different conditions is quite different, the short-term energy of the thermal signal in the RP is an effective feature for SCR. When a rectangular window is defined as:

$$W(n)=\begin{cases}1, & n\in DR\\ 0, & else,\end{cases} \tag{6}$$

the short-time energy of thermal signals in the RP can be computed as:

$$E_D=\sum_{i\in DR}(X_D(i))^2. \tag{7}$$

Sample Entropy Feature

Sample entropy is often used to analyse the complexity of signals or extract related features in fault diagnosis tasks [34,35]. Similarly, the complexity of the thermal signal is different under different sintering conditions, and the sample entropy of the thermal signal in the RP can be calculated to describe it. For a series thermal signal $X_D(i)|_{i=1}^{t}$, a set of m-dimension vectors in sequential order by serial number can be obtained $U_m(i)=\{X_D(i)\}|_{i=1}^{t-m+1}$. The maximum distance between vectors $U_m(i)$ and $U_m(j)$ can be defined as $d[U_m(i),U_m(j)]=\max_{k=1,\ldots,m-1}(|X_D(i+k)-X_D(j+k)|)$. For a given $U_m(i)$, the number of vectors $\{U_m(j)|1\le j\le t-m, j\ne i\}$ meeting the condition that the maximum distance between $U_m(i)$ and $U_m(j)$ is not greater than a given threshold η can be denoted as B_i. For parameter m, the probability that two sequences match for m points can be defined as:

$$B^m(\eta)=\frac{1}{t-m}\sum_{i=1}^{t-m}B_i^m(\eta)=\frac{1}{t-m}\sum_{i=1}^{t-m}\frac{1}{t-m-1}B_i. \tag{8}$$

Then the sample entropy can be calculated by:

$$SE_D=\ln\left[\frac{B^m(\eta)}{B^{m+1}(\eta)}\right]. \tag{9}$$

The above mentioned four features of each thermal signal in the RP are fused to form a sintering condition sample, which can be denoted as $\{M_{CFV},\ldots,M_{ST},\ldots,SE_{CFV},\ldots,SE_{ST}\}$.

3. Kernel modified optimal margin distribution machine for imbalance classification

In this paper, ODM is selected as a recognition model for SCR due to its superior generalization performance and classification ability. However, ODM was originally designed for balanced classification tasks and should be improved for class imbalance tasks.

3.1. Optimal margin distribution machine (ODM)

Generally, a classifier can be denoted as $y=\omega^T\varphi(x)$, which means that the training data can be separated by a linear classifier ω after being mapped to a feature space by the mapping function φ. The margin of a sample (x_i,y_i) can be defined as:

$$f(x_i)=\gamma_i=y_i\omega^T\varphi(x_i),\quad \forall i=1,2,\ldots,n \tag{10}$$

[29] advocates that by optimizing the margin distribution, the better generalization performance can be achieved than optimizing the minimum margin. It formulates the margin mean $\overline{\gamma}$ as:

$$\overline{\gamma}=\sum_{i=1}^{n}y_i\omega^T\varphi(x_i)=\frac{1}{n}(Xy)^T\omega, \tag{11}$$

where $X=[\varphi(x_i),\ldots,\varphi(x_m)]$, $y=[y_1,\ldots,y_m]^T$. After this, by calculating the difference between the margin of each sample and the margin mean, the margin variance can be obtained as follows:

$$\hat{\gamma}=\frac{1}{n}\sum_{i=1}^{n}(y_i\omega^T\varphi(x_i)-\overline{\gamma})^2=\frac{1}{n}\omega^TXX^T\omega-\frac{1}{n^2}\omega^TXyy^TX^T\omega, \tag{12}$$

Considering the lower computation efficiency of Eqs. (11) and (12), the margin mean is set to 1 by scaling $\|\omega\|$ in [29]. Furthermore, by introducing the margin variance in object function, the ODM can be written as

$$\begin{aligned}&\min_{\omega,\xi_i,\varepsilon_i}\frac{1}{2}\omega^T\omega+\frac{1}{n}\sum_{i=1}^{n}(C_1\xi_i^2+C_2\varepsilon_i^2)\\ &s.t.\ \ \gamma_i\ge 1-S-\xi_i,\quad \gamma_i\le 1+S-\varepsilon_i\\ &\qquad \xi_i,\varepsilon_i\ge 0,\ \ i=1,2,\ldots,n,\end{aligned} \tag{13}$$

where γ_i is the margin of sample (x_i,y_i), which is defined in Eq. (10), ξ_i and ε_i are margin deviation between x_i and margin mean, S determines which samples are support vectors, and C_1 and C_2 are the trade-off parameters. The second term of Eq. (13) is an improved representation of the margin variance; if the margin of (x_i,y_i) is greater than $1+S$, then $\xi_i=0$ and $\varepsilon_i>0$; if the margin of (x_i,y_i) is smaller than $1-S$, then $\xi_i>0$, $\varepsilon_i=0$.

The original ODM is designed for balanced data and is not suitable for imbalanced data classification tasks, such as SCR. According to the contribution of former researchers, the mapping ability of classifiers can be improved under class imbalance conditions by adjusting the kernel function or kernel matrix. In this paper, the kernel modification method is introduced to improve the ODM.

3.2. Kernel modification

For a conventional classifier, the kernel function is used to map the linear inseparable training data into a high-dimensional feature space, and it is expected that the training data are linearly separable in the feature space. Therefore, the selection of the kernel function directly affects the performance of the classifier. The kernel functions can also be written as $K(x_i,y_i)=\varphi(x_i)\varphi(y_i)$. The original training data are mapped to a curved Riemannian

Please cite this article as: D. Wang, X. Zhang, H. Chen et al., Sintering conditions recognition of rotary kiln based on kernel modification considering class imbalance. ISA Transactions (2020), https://doi.org/10.1016/j.isatra.2020.07.010.

manifold in kernel space [36]. The metric of the Riemannian manifold can be calculated by:

$$g_{ij}(x) = \left(\frac{\partial^2 K(x,x')}{\partial x_i \partial x_j}\right)_{x=x'}, \tag{14}$$

where $g_{ij}(x)$ is a coefficient that controls the local volume expansion (LVE) of the adjacent area of sample (x, y) in the Riemannian manifold. According to Eq. (14), $g_{ij}(x)$ is determined by the kernel function and can be changed indirectly by adjusting kernel matrix. In [36], the conformal transformation is adopted for kernel modification, which can be written as:

$$\tilde{K}(x,x') = D(x)D(x')K(x,x'), \tag{15}$$

where $D(x)$ is a positive conformal function and $\tilde{K}(x,x')$ is the new kernel derived from Eqs. (14) and (15). The LVE coefficient $\tilde{g}_{ij}(x)$ for $\tilde{K}(x,x')$ is related to $g_{ij}(x)$ by:

$$\begin{aligned}\tilde{g}_{ij}(x) = {} & D(x)^2 g_{ij}(x) + \frac{\partial D(x)}{\partial x_i} K(x,x) \frac{\partial D(x)}{\partial x_j} \\ & + D(x)\{\frac{\partial K(x,x')}{\partial x_i}|_{x'=x} \frac{\partial D(x)}{\partial x_j} + \frac{\partial K(x,x')}{\partial x_j}|_{x'=x} \frac{\partial D(x)}{\partial x_i}\}.\end{aligned} \tag{16}$$

For the selection of $D(x)$, [36] proposed a conformal function considering the distance between support vectors and other samples:

$$D(x) = \sum_{k \in SV} \exp\left(-\frac{|x - x_k|}{\eta \tau_k^2}\right), \tag{17}$$

where SV denotes the set of support vectors and τ_k^2 can be calculated by:

$$\tau_k^2 = \underset{i \in \{\|\varphi(x_i)-\varphi(x_k)\|^2 < M, y_i \neq y_k\}}{AVG} \left(\|\varphi(x_i) - \varphi(x_k)\|^2\right), \tag{18}$$

where M is the average distance between support vectors and $\varphi(x_k)$. Since the specific form of mapping function φ cannot be obtained directly, $\|\varphi(x_i) - \varphi(x_k)\|$ can be calculated by the following kernel trick:

$$\|\varphi(x_i) - \varphi(x_k)\|^2 = K(x_i, x_i) + K(x_k, x_k) - 2K(x_i, x_k). \tag{19}$$

Since the adjacent areas of support vectors are usually the class boundary near the learned separator, the conformal function in Eq. (17) reaches its maximum near the class boundary to amplify the spatial resolution of these areas. However, the conformal function in [36] is sensitive to the samples near class boundary, and its computational complexity is relatively high. Therefore, [28] proposes a conformal function with a simple form:

$$D(x) = e^{-kf(x)^2}, \tag{20}$$

where $f(x)$ represents the margin of sample (x, y) defined in Eq. (10). Therefore, $g_{ij}(x)$ can be magnified in the area support vectors located. It is revealed that the kernel scaling method can asymmetrically enlarge the space around the class boundary and increase the separability of linear inseparable training data [28].

3.3. Kernel modified optimal margin distribution machine (KMODM)

The kernel scaling method generally considers the margin of training data to adjust the kernel of classifier. It ensures that the boundary samples with smaller margins have a larger $g_{ij}(x)$, thus magnifying the spatial resolution of the class boundary. The conformal function is the key factor of the kernel scaling method.

The conformal functions in [26–28,36] were proposed for SVM. Among them, the margin of training samples is considered because the learned separator of SVM is determined by the support vector (SV), which is always near the separator. As seen in Fig. 6, the difference is that the support vectors of ODM are the data with a larger deviation from the margin mean. Therefore, the conformal functions defined in Eqs. (17) and (20) are not suitable for the kernel modification of ODM.

According to Fig. 6(b), the ideal kernel of ODM should be gathering the training data to h_m. In other words, the conformal function $D(x)$ should ensure that the support vectors with small margins have a large expansion coefficient and that the support vectors far away from the initial separator have a small expansion coefficient. At the same time, for imbalanced data, $D(x)$ should reduce the LVE coefficient of minority samples to make most of them become support vectors near the separator after kernel modification. In this way, the learned separator is biased towards the majority class to minimize the loss, thus alleviating the separator skewness caused by imbalanced data. In this paper, a novel conformal function is proposed:

$$D(x) = \begin{cases} e^{-K_n \cdot N/(m - f(x))} & , f(x) \leq (1-S)m \\ e^{K_f \cdot N \cdot (m - f(x))} & , f(x) \geq (1+S)m \\ e^{-K_n \cdot N \cdot m} & , \quad else, \end{cases} \tag{21}$$

where S is the sparse parameter of the KMODM. The margin and margin mean are denoted as $f(x)$ and m, respectively. K_n and K_f are trade-off parameters to control the volume expansion coefficients of different areas in the feature space. N reflects the imbalance ratio (IR) of training samples and can be defined as:

$$N = \begin{cases} 2 \cdot \dfrac{n^-}{n^+} & , x \in X^+ \\ \dfrac{n^+}{n^-} & , x \in X^-, \end{cases} \tag{22}$$

where n^+ and n^- represent the numbers of minority and majority samples, respectively. As seen from Eq. (21), $D(x)$ is a monotonously decreasing function of N. When the margins of the two samples are equal, parameter N defined in Eq. (22) assigns the minority sample a smaller $D(x)$ to reduce the LVE coefficient. Thus, the influence of minority samples on the final separator is enhanced. Furthermore, since the samples in the class boundary and overlapping region are generally close to the ODM separator, controlling K_n and K_f in Eq. (21) allows us to magnify the spatial expansion coefficient of the class boundary and to improve the mapping ability of the ODM.

Generally, the advantages of $D(x)$ constructed in Eq. (21) are as follows:

(i) Optimizing parameters K_n and K_f, $D(x)$ can magnify the LVE coefficient of the area near the initial separator, thus indirectly increasing the spatial resolution of the class boundary. The data separability and generalization performance of the classifier can be improved effectively.

(ii) $D(x)$ can automatically assign a small value to $g_{ij}(x)$ of minority samples considering IR of the training data. In this way, the impact of these minority samples on the final separator is increased. The separator skewness and detection accuracy deterioration of the minority class caused by imbalanced data can be alleviated.

According to Eqs. (15) and (21), the obtained new kernel matrix can be written as:

$$\tilde{\mathbf{K}} = (\tilde{k}_{ij}) = D(x_i)D(x_j)k_{ij}. \tag{23}$$

where $\mathbf{K} = (k_{ij})$ denotes the original kernel matrix. It is easy to prove that the new kernel matrix derived from Eqs. (21) and (23) is positive (semi) definite. This means that if the original kernel matrix of the ODM is valid, the obtained kernel matrix is valid as well.

Please cite this article as: D. Wang, X. Zhang, H. Chen et al., Sintering conditions recognition of rotary kiln based on kernel modification considering class imbalance. ISA Transactions (2020), https://doi.org/10.1016/j.isatra.2020.07.010.

Corollary 1. *When the original kernel matrix of the ODM is valid, the new kernel matrix derived from Eqs. (21) and (23) is a positive (semi) definite and valid kernel matrix.*

Proof. From Eq. (21), it is known that $D(x) \geq 0$. $F(x) = D(x_j)D(x_i)$ is obviously a symmetric function since $D(x_i)D(x_j) = D(x_j)D(x_i)$. For any training samples $x_1, x_2, \ldots, x_n \in X$ and $\alpha_1, \alpha_2, \ldots, \alpha_n \in R$, the following equation can be established:

$$\sum_{i,j=1}^{n} \alpha_i \alpha_j D(x_i)D(x_j) = \sum_{i=1}^{n} \alpha_i D(x_i) \sum_{j=1}^{n} \alpha_j D(x_j) = (\sum_{i=1}^{n} \alpha_i D(x_i))^2 \geq 0. \tag{24}$$

Thus, the conclusion that $F(x)$ is a positive (semi) definite function can be achieved. Denoting $\mathbf{d} = (d_1, d_2, \ldots, d_n)^T$ as an n-dimensional vector with $d_i = D(x_i)$, the matrix $\mathbf{dd}^T$ is a positive (semi) definite matrix. The conformal transformation in Eq. (23) can be rewritten as a Hadamard product of $\mathbf{dd}^T$ and $\mathbf{K}$:

$$\tilde{\mathbf{K}} = \mathbf{dd}^T * \mathbf{K}. \tag{25}$$

Since the original valid kernel matrix $\mathbf{K}$ of the ODM is a positive (semi) definite matrix. According to the Schur product theorem, the conclusion of $\tilde{\mathbf{K}}$ is a positive (semi) definite matrix can also be drawn, and it is thus a valid kernel matrix (Mercer' theorem).

Algorithm 1 shows the pseudocodes of the KMODM. By implementing the standard ODM, the margin and margin mean of training samples can be obtained. The value of conformal function of each sample can be calculated by Eq. (21) and the original kernel matrix $\mathbf{K}$ can be adjusted. The new kernel matrix $\tilde{\mathbf{K}}$ is used to learn the final separator, resulting in the KMODM classifier.

Because the KMODM is designed mainly for binary classification, it cannot be directly used for multi-classification tasks. In this paper, a directed acyclic graph (DAG) is used to combine three KMODM models into the KMODM^{DAG} for sintering condition recognition of rotary kilns.

Algorithm 1 KMODM

1: **Input:** X_{train}, X^+, X^-, $\mathbf{K}$;
2: **Output**: *a modified classifier*
3: **Variables:**
S ; /* *sparse parameter of ODM* */
x; /* *a train instance* */
m; /* *margin mean* */
K_n, K_f ; /* *trade-off parameters of D(x)* */
N; /* *parameter reflecting imbalance rate* */
4: **Begin**
5: $C \leftarrow$ ODM (X_{train}, $\mathbf{K}$); /* *train a standard ODM* */
6: $f(x) \leftarrow$ f (X_{train}, C); /* *extract the margin of training samples* */
7: $m \leftarrow$M ($f(x)$); /* *compute the margin mean* */
9: **for** each $x \in X_{train}$ {
19: $D(x) \leftarrow$D ($f(x)$, S, m, K_n, K_f, N) };/* *compute the conformal function* */
10: **for** each k_{ij} in $\mathbf{K}$ {
11: $\tilde{k}_{ij} = D(x_i) \times D(x_j) \times k_{ij}$ };
12:OC$\leftarrow$ODM (X_{train}, $\tilde{\mathbf{K}}$);/* *train a modified ODM* */
13: *return OC*
14: **End**

4. Experiments and discussion

To assess the effectiveness of the proposed SCR framework, a series of experiments on real thermal signals of a rotary kiln are performed in this section. Furthermore, to verify the applicability of the proposed KMODM model to other imbalance classification tasks, part of the experiment on the UCI standard dataset is also recorded. In this section, the F1 score is adopted to assess the balance degree of detection rates of different classes and can be calculated by:

$$recall = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \quad pre = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \quad F_1 = \frac{2pre \times recall}{pre + recall}$$

where TP denotes the true positive, FN represents false negative and FP refers false positive.

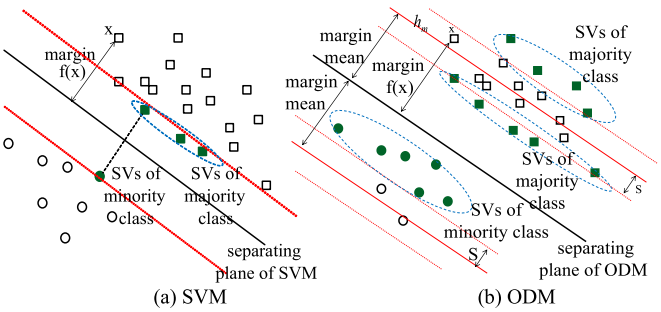

Fig. 6. The difference of support vectors. Green samples are support vectors of each classifier, and h_m denotes the virtual hyperplane of the margin mean. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

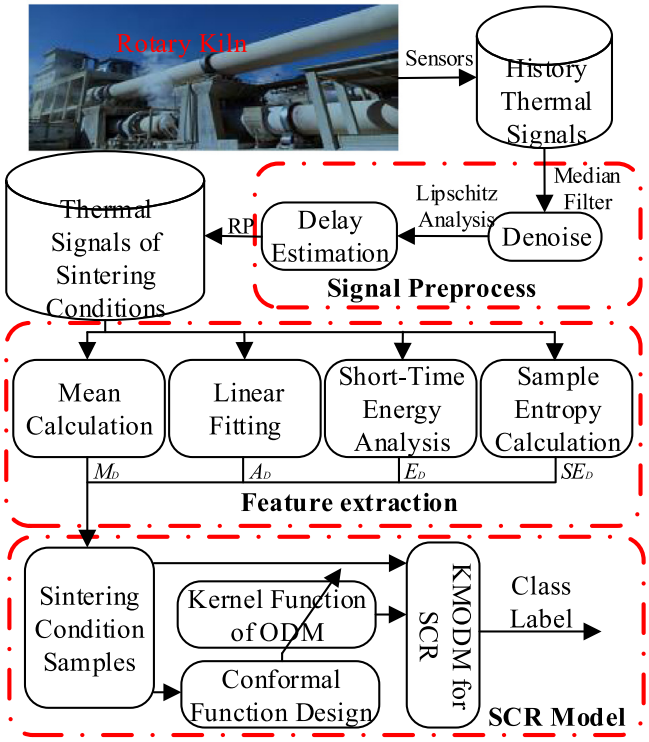

Fig. 7. The proposed sintering condition recognition framework.

4.1. Experiments on real thermal signals

After determining the sintering condition features and recognition model, the whole SCR framework can be constructed, and the main module of the framework is shown in Fig. 7. The real thermal signals are collected from the #3 rotary kiln of the Inner Mongolia DaTang International Recycling Resource Development Co., Ltd. in China.

Please cite this article as: D. Wang, X. Zhang, H. Chen et al., Sintering conditions recognition of rotary kiln based on kernel modification considering class imbalance. ISA Transactions (2020), https://doi.org/10.1016/j.isatra.2020.07.010.

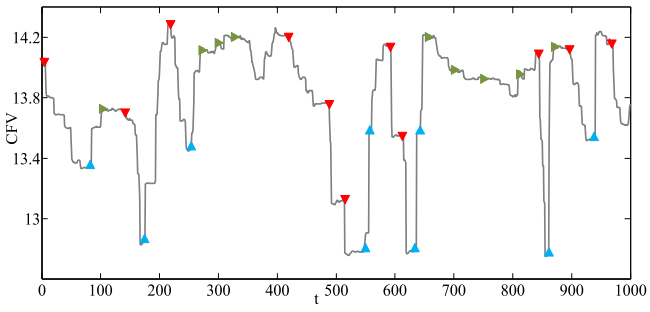

Fig. 8. The inflection points of the CFV. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

CFV is the coal feeding rate, which is manipulated by operators. The operators determine the sintering condition and then adjust the coal feeding to maintain the condition under normal conditions. For example, the operators would increase CFV when the rotary kiln is under super-chill conditions. Therefore, the inflection point of CFV is used to assist in labelling sintering condition samples. Fig. 8 shows the inflection points extracted from a piece of real CFV data. In Fig. 8, the red triangle points are coal reduction points, and the corresponding sintering condition can be preliminarily determined as super-heated. The blue and green triangles represent super-chilled and normal, respectively. After obtaining the inflection points of CFV, their labels are confirmed by two experience kiln operators.

In this way, 4994 samples of different sintering conditions are extracted, including 569 super-heated samples, 1262 super-chilled samples, and 3163 normal samples. Seventy percent of these samples are randomly selected as training data, the rest are used to compose the test data.

4.1.1. Sintering condition recognition results

To assess the effectiveness of the proposed KMODM^{DAG} model, a few classifiers designed for bi-class imbalanced data are selected as baseline models, such as the ODM, the weighted kernel-based synthetic minority oversampling technique (WKSMOTE) [12], and the large cost-sensitive margin distribution machine (LCSDM) [18]. Additionally, eight advanced classifiers for multi-classification are selected and compared with the KMODM^{DAG}. Among them, multi-class SVM (mcSVM) [37] and the multi-class ODM (mcODM) [30] are designed for balanced data, multi-class boosting with costs (BAdaCost) [20] and neural network with focal loss (FocalNN) [17] are the most advanced imbalanced classification algorithms based on cost-sensitive methods, the programs of partially informative boosting (PIBoost) [21], imbalanced fuzzy-rough ordered weighted average nearest neighbour classification (FIECOC) [19], diversified one-against-one (DOVO) [38] and diversified error-correcting output codes (DECOC) [24] can be found at https://github.com/chongshengzhang/multiimbalance. For the ODM, the selection of regularization parameters and sparse parameters are the same as that of the KMODM. λ of the mcODM and C of mcSVM are both selected by 5-fold cross validation from $[2^0, 2^1, \ldots, 2^{10}]$, μ and θ of the mcODM are selected from [0.1, 0.2...,0.9]. For all methods, the RBF kernel $K(x_i, x_j) = \exp(-\delta \left\| x_i - x_j \right\|^2)$ is adopted, and its width is selected by 5-fold cross validation from the set $[2^{-10}, \ldots, 2^5]$.

To prove that the introduction of dynamic features can describe the different sintering conditions more comprehensively, the performance of each recognition model is recorded when using statistical features and fused features. The experiments of each classifier using different features are repeated 30 times, and the means and standard deviations of the F1-score are reported in Table 3. The number in parentheses represents the standard deviation, and the bolded data are the best results.

Comparing the recognition accuracy of each model for sintering conditions using different features, it can be seen that in most cases, the introduction of dynamic features can improve the overall recognition accuracy by more than 2%. This is mainly because the statistical features cannot fully reflect the information about the sintering conditions contained in thermal signals, while the introduction of dynamic features makes the description of sintering conditions more comprehensive, which can effectively improve the separability of the sintering condition samples. In this way, the validity of the extracted dynamic features is proven.

Moreover, from the recognition accuracy of each sintering condition, it can be seen that, due to the low separability between super-chilled and normal samples, the mcSVM obtains the lowest recognition rate of super-chilled condition. By optimizing the margin distribution of training samples, the mcODM, LCSDM ^{DAG} and the KMODM^{DAG} can obtain more reasonable separators and improve the detection rate of super-chilled samples. In particular, the KMODM^{DAG} inherits the strong classification ability of the ODM and improves the ability to deal with class imbalanced data by modifying its kernel function. As a result, its super-chilled detection rate increases by at least 4% over other methods.

It can also be seen in Table 3 that because the influence of imbalanced data is not considered by mcSVM, mcODM and ODM^{DAG}, their recognition rates under normal conditions are much higher than those under abnormal conditions. However, the KMODM ^{DAG} and other classifiers that designed for imbalance data can significantly improve the detection rate of two abnormal conditions and obtain a higher overall accuracy. Among them, FIECOC, DECOC, LCSDM ^{DAG} and KMODM^{DAG} both achieve an overall recognition accuracy of more than 86%, especially the proposed KMODM^{DAG,} which achieves approximately 90% recognition accuracy. A reasonable explanation for this situation is that the parameter N, which reflects the imbalance degree, is embedded in the proposed conformal transformation function. Using this conformal function to modify the kernel of the ODM, the expansion coefficient of the region where the minority class samples are located in the kernel space is compressed, and the classifier is forced to move towards the majority class region, thereby optimizing the margin distribution of the minority class and improving the detection accuracy. Other classifiers that rely on data preprocessing or cost-sensitive methods fail to optimize the spatial distribution of samples in kernel space, and the obtained classification hyperplanes are closer to the minority class, so it is difficult to recognize abnormal conditions in the test dataset with high accuracy.

4.1.2. Parameter optimization

According to Section 3, K_n and K_f are the key parameters of the KMODM that control the LVE coefficient of different areas in the feature space. In this study, the influence of these two parameters on the performance of the KMODM is analysed. The default value of sparse parameter S is set as 0.2, and the kernel function is selected as the RBF. The grid search method is used to find the optimal parameters in [0, 0.01]. It can be seen in Fig. 9 that for all three KMODMs in the SCR model, the optimal value of K_f is generally larger than that of K_n. This causes the area close to the initial separator to have a large spatial resolution and improves the separability of training data. This result is consistent with our analysis in Section 3.

Please cite this article as: D. Wang, X. Zhang, H. Chen et al., Sintering conditions recognition of rotary kiln based on kernel modification considering class imbalance. ISA Transactions (2020), https://doi.org/10.1016/j.isatra.2020.07.010.

Table 3
Average recognition accuracy of different classifiers using different features.

	Statistical feature				Fusion features			
	Super-heated	Super-chilled	Normal	F1 score	Super-heated	Super-chilled	Normal	F1 score
mcSVM	81.64 (2.2)	79.26 (2.7)	91.53 (2.3)	82.07 (2.5)	83.48 (2.1)	81.13 (1.8)	**94.02** (**2.6**)	84.75 (2.0)
mcODM	82.62 (1.9)	80.95 (2.1)	**91.56** (**2.4**)	83.78 (2.2)	85.13 (1.3)	83.58 (1.9)	92.88 (2.0)	85.20 (1.5)
ODM	82.43 (2.3)	80.18 (2.1)	90.94 (2.5)	82.61 (2.2)	84.81 (1.9)	82.50 (2.4)	92.40 (2.2)	84.31 (2.3)
PIBoost	83.45 (2.7)	79.80 (2.9)	89.08 (2.4)	83.50 (2.5)	85.53 (2.4)	82.47 (1.7)	90.51 (1.9)	85.85 (2.5)
FIECOC	85.07 (2.3)	81.38 (2.6)	87.82 (1.7)	84.14 (2.3)	86.14 (2.3)	83.58 (2.3)	89.03 (1.7)	86.10 (2.0)
DOVO	84.52 (1.6)	81.06 (2.5)	87.80 (2.0)	83.88 (2.1)	86.16 (1.7)	83.70 (2.1)	90.45 (2.0)	85.94 (1.9)
DECOC	84.98 (1.8)	81.26 (2.1)	86.48 (1.9)	84.17 (1.8)	86.31 (2.0)	83.14 (2.2)	89.50 (1.8)	86.08 (2.1)
WKSMOTE	84.71 (1.9)	81.29 (2.3)	87.13 (2.2)	83.96 (2.0)	85.95 (1.7)	83.21 (1.5)	90.46 (1.8)	85.71 (1.7)
BAdaCost	84.13 (2.3)	80.47 (2.5)	87.05 (2.0)	83.94 (2.3)	86.25 (1.9)	83.80 (2.2)	90.88 (1.6)	86.12 (2.0)
LCSDM	85.46 (1.4)	80.99 (2.2)	87.68 (1.9)	83.48 (1.9)	87.47 (2.0)	84.42 (1.6)	91.45 (2.5)	87.86 (1.8)
FocalNN	**86.30** (**2.6**)	82.06 (2.7)	87.89 (2.3)	85.21 (2.5)	87.91 (2.7)	84.81 (3.1)	91.07 (2.3)	87.94 (2.7)
KMODM	86.18 (2.1)	**86.07** (**2.3**)	87.97 (1.7)	**86.02** (2.0)	**90.26** (**2.3**)	**88.55** (**1.8**)	92.17 (2.1)	**90.59** (**2.6**)

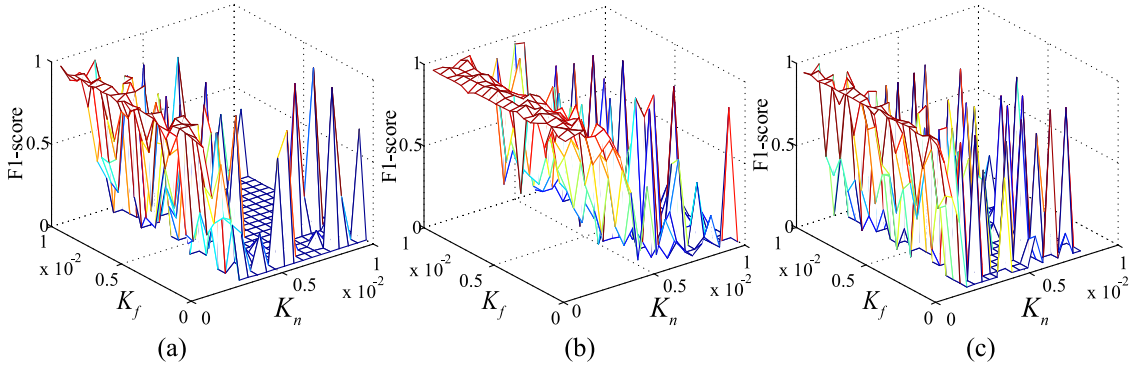

Fig. 9. The optimization of K_n and K_f for the KMODM. (a) Super-chilled vs. super-heated; (b) super-chilled vs. normal; (c) super-heated vs. normal.

4.1.3. Generalization performance analysis

The margin distribution of samples can reflect the generalization performance of the SCR framework. The better the margin distribution is, the better the generalization performance of the model will be. The margin distribution of each sintering condition of the KMODM^{DAG} and the mcODM is shown in Fig. 10. The *x*-axis in the graph represents the margin of training samples obtained by the KMODM^{DAG} and the mcODM, and the *y*-axis represents the statistical frequency of each margin. The margin distribution of abnormal conditions obtained by the KMODM^{DAG} is larger than that of the mcODM. The margin distribution of the normal condition of the KMODM^{DAG} is similar to that of the mcODM. That is, the KMODM^{DAG}-based SCR framework can achieve a better margin distribution and thus will be relatively more accurate for abnormal condition recognition. This conclusion also confirms the analysis of the KMODM in Section 4.1.1.

4.1.4. Operational risk reduction

Using different features, the confusion matrix of the KMODM^{DAG} and the mcODM are shown in Table 4. It can be seen from the table that the recognition accuracy of three sintering conditions obtained by the KMODM^{DAG} using fusion features reaches 90.2%, 88.7% and 91.2%. The super-chilled condition has the lowest recognition accuracy due to the low feature separability between its samples and the normal samples. In this case,

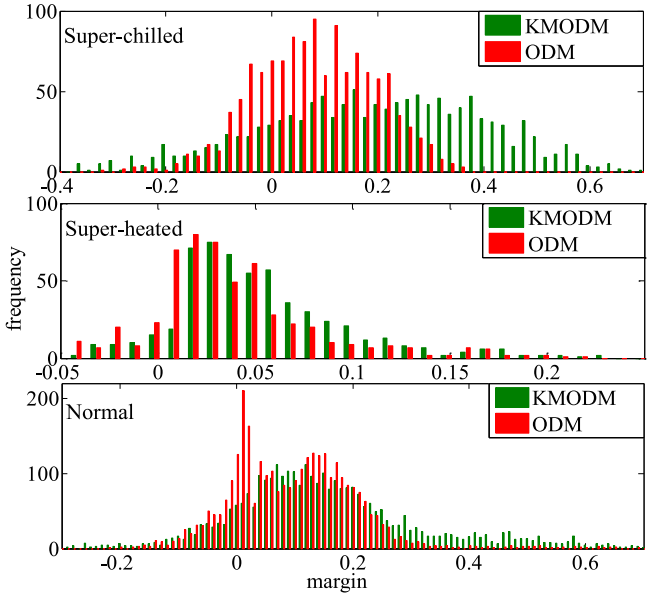

Fig. 10. The margin distribution of each sintering condition obtained by the ODM^{DAG} and KMODM^{DAG}.

the recognition accuracy of the super-chilled condition obtained by the KMODM^{DAG} is still higher than that of the mcODM. Additionally, the super-heated method achieves a relatively high recognition accuracy. This proves that the KMODM^{DAG} is effective in eliminating the influence of imbalanced data and improving the detection rate of the minority class. The accuracies of the mcODM merely using statistical features are 82.4%, 80.2% and 92.2%, which are significantly lower than that of the KMODM^{DAG} considering dynamic features of thermal signals

Furthermore, since the operations corresponding to the two abnormal sintering conditions are quite different, the lower misjudgement rate between two abnormal conditions is essential for operational risk reduction. From the confusion matrix in Table 4, it can be seen that, using fusion features, the KMODM^{DAG} achieves a misjudgement rate of 3.2% between two abnormal sintering conditions, which is approximately one-third of that of the mcODM (10.7%). In fact, it can be seen from Table 4 that even if the same recognition model is adopted, considering the dynamic features of thermal signals can achieve a lower misjudgement rate between two abnormal conditions. That is, by introducing dynamic features of thermal signals and modifying the kernel of the ODM, the proposed the KMODM^{DAG}-based sintering condition recognition framework can significantly improve the recognition rate of abnormal conditions and reduce the operational risk deduced by abnormal sintering condition misclassification.

4.1.5. CPU time cost

In many studies of control algorithms and systems, the CPU time cost is usually used to evaluate the computational efficiency of the algorithm and its application feasibility in industrial fields [39,40]. To further analyse the computation efficiency of the proposed the KMODM^{DAG} model, the average CPU time consumed by the training and testing phases of the above models is shown in Fig. 11. All experiments are performed with MATLAB 2014a on the workstation with 2 $\times$ 3.3 GHz CPUs and 4 GB main memory.

Compared with PIBoost and FIECOC, the KMODM^{DAG} has a significant advantage in training time. Because the DOVO method finds the optimal classifier for each subset, it requires considerable training time. The DECOC method uses ECOC to replace the class decomposition method in DOVO. It trains more classifiers than the DOVO method. Thus, it takes the longest training time. BAdaCost combines the ensemble framework and cost-sensitive method, and its training process also requires considerable CPU time. The core of focalNN is the cost-sensitive idea, and it assigns greater misclassification penalties to the samples that are difficult to classify. The algorithm iterations of focalNN make it not advantageous in terms of training time. Since KMODM also needs iterative calculation to obtain the initial margin of the samples, its calculation efficiency is lower than that of methods such as mcSVM, mcODM and the ODM^{DAG} that learn the classifier only once. However, benefits to the parallel learning of multiple KMODM model in the DAG framework, compared with other classifiers for multi-class imbalanced data, especially PIBoost, DOVO and DECOC, KMODM^{DAG} shows a significantly low CPU time cost. Considering the advantages of the KMODM in abnormal condition recognition, the slight loss of computation efficiency is acceptable. Because the sintering condition of a rotary kiln generally changes slowly, these results show that KMODM^{DAG} is suitable for real-time SCR tasks of a rotary kiln.

4.2. Experiments on standard datasets

To further verify the ability of the KMODM to handle imbalanced data, in this section, the proposed the KMODM is compared with several baseline methods on two-dimensional visualization data and some standard datasets.

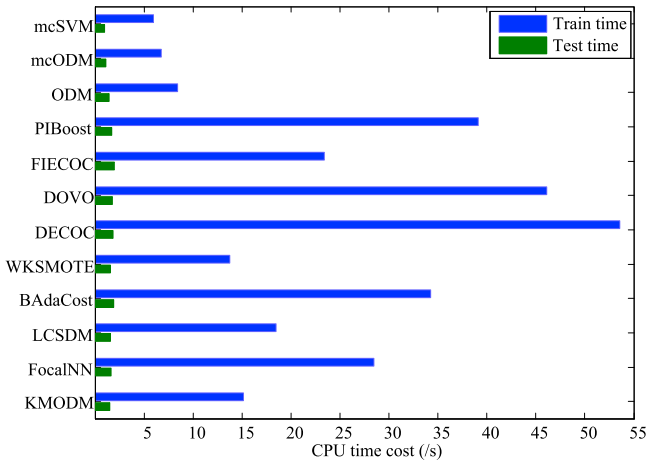

Fig. 11. CPU time cost of each recognition model.

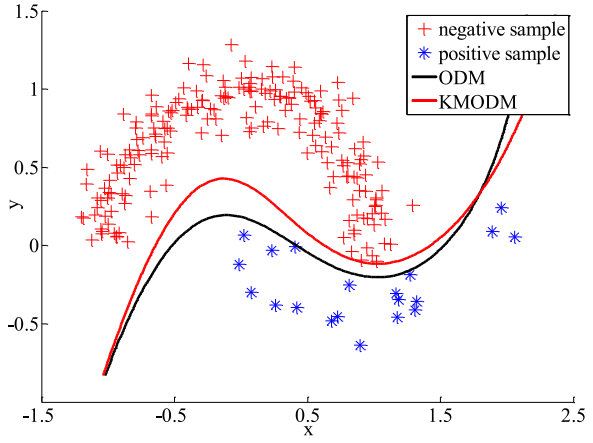

Fig. 12. Imbalanced data and visualized separators of the KMODM and the ODM (IR = 10).

4.2.1. Experiments on visualization data

Fig. 12 shows a comparison of the separators learned by the KMODM and the ODM to illustrate the effectiveness of the proposed model. In this figure, the separator of ODM seriously skews towards the minority class. However, the proposed the KMODM adopts the kernel scaling method and thus can alleviate the skewness of the classifier. As a result, both the margin distribution and the detection rate of the minority class are optimized.

4.2.2. Experiments on UCI datasets

In this part, five multi-class standard UCI datasets are selected to assess the performance of the KMODM. The experimental results are shown in Table 5. It can be seen that the results of the mcSVM, mcODM and ODM^{DAG} are unsatisfactory on most multi-class imbalanced data. Other classifiers designed for imbalanced data can achieve more balanced detection rates and better F1 scores. The proposed KMODM^{DAG} algorithm achieves the best results on most datasets. Analysing the performance of the KMODM on 2-D visualization data and UCI standard datasets, the conclusion that the proposed the KMODM model can effectively deal with imbalanced data classification tasks can be drawn.

5. Conclusion and prospects

Recognition of sintering conditions is a key prerequisite for effective control of rotary kilns. This paper presented a thermal

Please cite this article as: D. Wang, X. Zhang, H. Chen et al., Sintering conditions recognition of rotary kiln based on kernel modification considering class imbalance. ISA Transactions (2020), https://doi.org/10.1016/j.isatra.2020.07.010.

Table 4
Confusion matrix of the KMODM^{DAG} and the mcODM using different features. The data in the confusion matrix are the average of 30 experimental results.

		Statistical Feature			Fusion Features		
		super-heated	super-chilled	normal	super-heated	super-chilled	normal
mcODM	super-heated	140.9 (82.4%)	12.1 (7.1%)	18.0 (10.5%)	147.2 (86.1%)	8.9 (5.2%)	14.9 (8.7%)
	super-chilled	36.0 (9.5%)	304.0 (80.2%)	39.0 (10.3%)	26.9 (7.1%)	316.8 (83.6%)	35.3 (9.3%)
	normal	19.0 (2.0%)	55.0 (5.8%)	875.0 (92.2%)	21.9 (2.3%)	55.0 (5.8%)	872.1 (91.9%)
KMODM	super-heated	146.5 (85.7%)	10.1 (5.9%)	14.4 (8.4%)	154.2 (90.2%)	5.1 (3.0%)	11.7 (6.8%)
	super-chilled	17.5 (4.6%)	329.7 (87.0%)	31.8 (8.4%)	12.5 (3.3%)	336.2 (88.7%)	30.3 (8.0%)
	normal	33.2 (3.5%)	72.1 (7.6%)	843.7 (88.9%)	29.4 (3.1%)	54.1 (5.7%)	865.5 (91.2%)

Table 5
Means and standard deviations of F1 score with RBF kernel on multi-class datasets.

Dataset	Hayes	Newthy	Balance	Car	Thyroid
mcSVM	78.3 ± 2.0	91.5 ± 2.8	65.7 ± 2.5	95.2 ± 2.8	63.1 ± 2.3
mcODM	81.2 ± 1.9	92.3 ± 2.5	67.4 ± 1.9	96.0 ± 2.3	65.4 ± 2.2
ODM	78.7 ± 2.4	90.6 ± 2.9	66.7 ± 1.8	95.0 ± 1.5	63.5 ± 1.6
PIBoost	78.4 ± 2.0	93.3 ± 2.2	68.4 ± 1.9	96.3 ± 2.4	68.6 ± 2.8
FIECOC	78.1 ± 2.3	92.5 ± 3.1	68.8 ± 2.1	97.2 ± 1.7	68.4 ± 2.0
DOVO	80.5 ± 1.7	93.7 ± 2.6	67.4 ± 1.7	97.9 ± 1.3	69.5 ± 1.4
DECOC	81.7 ± 2.1	94.3 ± 3.0	70.6 ± 2.3	98.4 ± 2.7	70.5 ± 2.0
WKSMOTE	80.2 ± 2.6	92.6 ± 3.5	68.2 ± 3.0	97.1 ± 2.2	67.7 ± 2.5
BAdaCost	82.1 ± 2.3	94.6 ± 1.8	**72.0 ± 2.1**	97.7 ± 2.3	69.8 ± 2.6
LCSDM	83.5 ± 1.9	95.1 ± 1.8	69.0 ± 2.4	97.0 ± 1.4	68.9 ± 1.9
FocalNN	82.7 ± 1.5	95.4 ± 2.0	69.5 ± 2.7	98.9 ± 1.9	69.3 ± 2.0
KMODM	**85.9 ± 2.0**	**97.9 ± 2.1**	69.7 ± 1.8	**99.1 ± 1.2**	**72.9 ± 6.3**

signal-based framework considering class imbalance to recognize sintering conditions in kilns. For the separability improvement of sintering samples, this novel framework extracts the dynamic features of thermal signals and combines them with conventional statistical features to describe different sintering conditions. Aiming to address the class imbalance issue of sintering samples, the proposed framework introduces the kernel modification method to improve the selected recognition model. More specifically, a novel imbalance classification model named the KMODM is proposed by constructing a new conformal function to modify the kernel function of the ODM. This is the first time that the class imbalance issue has been considered in sintering condition recognition tasks of rotary kilns. The experimental results on real thermal signals show that the proposed SCR framework can achieve a more balanced detection rate and reduce the misjudgement risk of abnormal conditions. Using our framework, the recognition accuracy of the sintering condition can reach approximately 90%, and the balance and low-risk recognition of the sintering condition have been realized as well.

The proposed SCR method merely relies on the statistical and dynamic features of thermal signals, and it will be worthwhile in the future to introduce other information, such as the features of the flame image and other hidden features extracted by some deep neural networks. Additionally, improving mcODM for multi-class imbalance classification tasks and designing the subsequent control system according to the obtained sintering state recognition results will also be our directions for future research.

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.

Acknowledgements

This work was supported in part by National Key Technologies Research and Development Program of China (2018YFB1305900), by the Hunan Natural Science Foundation (2018JJ2056), by the National Natural Science Foundation of China under Grant 51475157 and 61674054, by the Scientific Research Launch Project of Anhui Polytechnic University (2017YQQ011), by the Macau Science and Technology Development Fund under Grant FDCT/189/2017/A3, and by the Research Committee at University of Macau under Grants MYRG2016-00123-FST and MYRG2018-00136-FST.

References

[1] Chen Hua, et al. Recognition of the temperature condition of a rotary kiln using dynamic features of a series of blurry flame images. IEEE Trans Ind Inf 2015;12(1):148–57.
[2] Matthes J, Waibel P, et al. A new camera-based method for measuring the flame stability of non-oscillating and oscillating combustions. Exp Therm Fluid Sci 2019;105:27–34.
[3] Xiangyu Zhang, et al. Temperature measurement of coal fired flame in the cement kiln by raw image processing. Measurement 2018;129:471–8.
[4] Bai X, Lu G, Hossain M, Szuhánszki J, Daood S, Nimmo W, Yan Y. Multi-mode combustion process monitoring on a pulverised fuel combustion test facility based on flame imaging and random weight network techniques. Fuel 2017;202:656–64.
[5] Makaremi Iman, et al. Identification and abnormal condition detection of a cement rotary kiln. IFAC Proc Vol 2008;41(2):7233–8.
[6] Sadeghian Masoud, Fatehi A. Identification, prediction and detection of the process fault in a cement rotary kiln by locally linear neuro-fuzzy technique. J Process Control 2011;21(2):302–8.
[7] Sharifi A, Shoorehdeli MA, Teshnehlab M. Identification of cement rotary kiln using hierarchical wavelet fuzzy inference system. J Franklin Inst B 2012;349(1):162–83.
[8] Bakdi Azzeddine, Kouadri A, Bensmail A. Fault detection and diagnosis in a cement rotary kiln using PCA with EWMA-based adaptive threshold monitoring scheme. Control Eng Pract 2017;66:64–75.
[9] Zhang Xiaogang, Lv Mingyang, Chen Hua, et al. Chaotic characteristics analysis of the sintering process system with unknown dynamic functions based on phase space reconstruction and chaotic invariables. Nonlinear Dynam 2018;93:395–412.
[10] Zhang Lei, et al. A robust temperature prediction model of shuttle kiln based on ensemble random vector functional link network. Appl Therm Eng 2019;150:99–110.
[11] Zhang Xiaogang, et al. Prediction of coal feeding during sintering in a rotary kiln based on statistical learning in the phase space. ISA Trans 2018;83:248–60.
[12] Mathew J, Pang CK, Luo M, et al. Classification of imbalanced data by oversampling in kernel space of support vector machines. IEEE Trans Neural Netw Learn Syst 2017;99(1):1–12.
[13] Zheng M, Li T, Zhu R, et al. Conditional wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification. Inform Sci 2020;512:1009–23.
[14] Vuttipittayamongkol P, Elyan E. Neighbourhood-based undersampling approach for handling imbalanced and overlapped data. Inform Sci 2020;509:47–70.

[15] Tao Xinmin, et al. Self-adaptive cost weights-based support vector machine cost-sensitive ensemble for imbalanced data classification. Inform Sci 2019;487:31–56.
[16] Wong Man Leung, Seng Kruy, Wong Pak Kan. Cost-sensitive ensemble of stacked denoising autoencoders for class imbalance problems in business domain. Expert Syst Appl 2020;141. 112918.
[17] Lin T, Goyal P, riyal, Girshick R, He K, Dollar P. Focal loss for dense object detection. In *2017 IEEE international conference on computer vision (ICCV)*, Venice, Italy, 2018, p. 2999-3007.
[18] Fanyong Cheng, et al. Large cost-sensitive margin distribution machine for imbalanced data classification. Neurocomputing 2017;224(8):45–57.
[19] Ramentol E, Vluymans S, Verbiest N, et al. IFROWANN: Imbalanced fuzzy-rough ordered weighted average nearest neighbor classification. IEEE Trans Fuzzy Syst 2015;23(5):1622–37.
[20] Fernández-Baldera, Antonio, Buenaposada, José M, Baumela L. BAdaCost: Multi-class boosting with costs. Pattern Recognit 2018;79:467–79.
[21] Fernndez BA, L. Baumela. Multi-class boosting with asymmetric binary weak-learners. Pattern Recognit 2014;47(5):2080–90.
[22] Zelenkov Yuri. Example-dependent cost-sensitive adaptive boosting. Expert Syst Appl 2019;135:71–82.
[23] Gan D, et al. Integrating TANBN with cost sensitive classification algorithm for imbalanced data in medical diagnosis. Comput Ind Eng 2020;140. 106266.
[24] Bi Jingjun, Zhang Chongsheng. An empirical comparison on state-of-the-art multi-class imbalance learning algorithms and a new diversified ensemble learning scheme. Knowl-Based Syst 2018;158(15):81–93.
[25] Zhang XG, Wang DX, et al. Kernel modified optimal margin distribution machine for imbalanced data classification. Pattern Recognit Lett 2019;125:325–32.
[26] Xiong Huilin, et al. Learning the conformal transformation kernel for image recognition. IEEE Trans Neural Netw Learn Syst 2017;28(1):149–63.
[27] Maratea A, Petrosino A, Manzo M. Adjusted f-measure and kernel scaling for imbalanced data learning. Inform Sci 2014;257(2):331–41.
[28] Sun Jiancheng, et al. Scaling the kernel function based on the separating boundary in input space: A data-dependent way for improving the performance of kernel methods. Inform Sci 2012;184(1):140–54.
[29] Zhang T, Zhou ZH. Optimal margin distribution machine. 2016, CoRR, abs/1604.03348.
[30] Zhang Teng, Zhou Zhi-Hua. Multi-class optimal margin distribution machine. In: Proceedings of the 34th international conference on machine learning. PMLR, vol. 70, 2017, p. 4063–71.
[31] He Xiangdong, Asada H. A new method for identifying orders of input-output models for nonlinear dynamic systems. In: American control conference IEEE. 1993, p. 2520–3.
[32] Makaremi Iman, Fatehi A, Araabi BN. Lipschitz numbers: A medium for delay estimation. Ifac Proc 2008;41(2):7468–73.
[33] Leena Mary, et al. Automatic syllabification of speech signal using short time energy and vowel onset points. Int J Speech Technol 2018;21(3):571–9.
[34] Li Xiaoyu, Dai Kangwei, Wang Zhenpo, Han Weiji. Lithium-ion batteries fault diagnostic for electric vehicles using sample entropy analysis method. J Energy Storage 2020;27. 101121.
[35] Wang Zhenya, Yao Ligang, Cai Yongwu. Zhenya wang ligang yao yongwu cai rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine. Measurement 2020;156. 107574.
[36] Wu G, Chang EY. KBA: kernel boundary alignment considering imbalanced data distribution. IEEE Trans Knowl Data Eng 2005;17(6):786–95.
[37] Crammer Koby, Singer Y. On the learnability and design of output codes for multiclass problems. Mach Learn 2002;47(2–3):201–33.
[38] Li Qianmu, Song Yanjun, Zhang Jing, Sheng Victor S. Multiclass imbalanced learning with one-versus-one decomposition and spectral clustering. Expert Syst Appl 2020;147. 113152.
[39] Xie Xiangpeng, Yue Dong, Peng Chen. Relaxed real-time scheduling stabilization of discrete-time takagi–sugeno fuzzy systems via an alterable-weights-based ranking switching mechanism. IEEE Trans Fuzzy Syst 2018;26(6):3808–19.
[40] Zhang Bowen, et al. Real-time action recognition with deeply transferred motion vector CNNs. IEEE Trans Image Process 2018;27(5):2326–39.

Please cite this article as: D. Wang, X. Zhang, H. Chen et al., Sintering conditions recognition of rotary kiln based on kernel modification considering class imbalance. ISA Transactions (2020), https://doi.org/10.1016/j.isatra.2020.07.010.