Chinese Journal of Chemical Engineering 23 (2015) 2020–2028

Contents lists available at ScienceDirect

Chinese Journal of Chemical Engineering

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

Selective ensemble modeling based on nonlinear frequency spectral feature extraction for predicting load parameter in ball mills☆

Jian Tang [1,2], Tianyou Chai [2,]*, Zhuo Liu [2], Wen Yu [3]

[1] Unit 92941, PLA, Huludao 125001, China
[2] State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 110004, China
[3] Departamento de Control Automatico, CINVESTAV-IPN, Av.IPN 2508, México D.F. 07360, México

ARTICLE INFO

Article history:
Received 25 May 2015
Received in revised form 17 July 2015
Accepted 3 August 2015
Available online 11 November 2015

Keywords:
Nonlinear latent feature extraction
Kernel partial least squares
Selective ensemble modeling
Least squares support vector machines
Material to ball volume ratio

ABSTRACT

Strong mechanical vibration and acoustical signals of grinding process contain useful information related to load parameters in ball mills. It is a challenge to extract latent features and construct soft sensor model with high dimensional frequency spectra of these signals. This paper aims to develop a selective ensemble modeling approach based on nonlinear latent frequency spectral feature extraction for accurate measurement of material to ball volume ratio. Latent features are first extracted from different vibrations and acoustic spectral segments by kernel partial least squares. Algorithms of bootstrap and least squares support vector machines are employed to produce candidate sub-models using these latent features as inputs. Ensemble sub-models are selected based on genetic algorithm optimization toolbox. Partial least squares regression is used to combine these sub-models to eliminate collinearity among their prediction outputs. Results indicate that the proposed modeling approach has better prediction performance than previous ones.

1. Introduction

The objective of industrial process optimization and feedback control is to improve the production quality and efficiency indices continuously, while keeping consumption indices as low as possible [1]. For ball mills, accurate measurement of some key process parameters, such as load parameters of wet ball (material to ball volume ratio (MBVR), pulp density (PD) and charge volume ratio (CVR)), is one of the key factors [2]. Lots of approaches have been proposed, but MBVR model always gives the worst prediction accuracy. Mechanical vibration and acoustic frequency spectra of a ball mill contain useful information [3], which can be used to determine mill load parameters indirectly [4,5]. Ball mill vibration signal based data-driven soft measuring methods are a new focus due to the high sensitivity [6,7], but they need further study [8,9].

Using frequency spectrum to construct an effective model, we have to face the Hughes phenomenon and dimensionality problem [10]. Thus dimension reduction is the first task [11,12]. Principal component analysis (PCA) and kernel PCA have been used to extract frequency spectral features in our previous study [13], but the extracted PCs and kernel PCs cannot take into account the correlation between frequency spectrum and MBVR [14]. Partial least squares (PLS) algorithm overcomes this problem by extracting latent feature [15]. Recently, a general framework of feature extraction based on PLS deflation is proposed [16]. Kernel PLS (KPLS) extends nonlinear item to the original input matrix and constructs nonlinear model with a small number of latent scores [17]. Thus KPLS algorithm is suitable to extract nonlinear frequency spectral latent feature.

Statistical inference techniques and machine learning techniques have been widely employed in data-driven soft sensor [18,19]. The most commonly used machine learning based methods are artificial neural network (ANN) and support vector machines (SVM). Although ANN is widely used for developing soft sensor model, it suffers from problems of local optima, uncontrolled convergence speed and overfitting. Traditional ANN based methods are considered to have shallow architectures (less than three layers of computation units), lacking a powerful representation efficacy on approximating highly varying functions [20]. Recently, deep learning technique has been used to train deep neural networks successfully in speech recognition and difficulty-to-measure process for industrial parameter measurement [21]. Deep neural networks include two phases, unsupervised pretraining phase and supervised back-propagation phase [22]. All available massive process data can be utilized in most of industrial processes. However, in practical mineral grinding process, useful shell vibration and acoustic modeling data can be obtained on some special phases only, such as ball mill re-starting and stopping. Therefore, the number of training samples is small. Structural risk minimization modeling criterion based SVM has been proved effective to tackle

☆ Supported partially by the Post Doctoral National Science Foundation of China (2013M532118,2015T81082), the National Natural Science Foundation of China (61573364, 61273177, 61503066), the State Key Laboratory of Synthetical Automation for Process Industries, the National High Technology Research and Development Program of China (2015AA043802), and the Scientific Research Fund of Liaoning Provincial Education Department (L2013272).
* Corresponding author.
E-mail address: tychai@mail.neu.edu.cn (T. Chai).

http://dx.doi.org/10.1016/j.cjche.2015.10.006
1004-9541/

small samples. However, quadratic program problem has to be solved in term of long learning time. The least square-support vector machines (LS-SVM) avoid this problem by solving a set of linear equations [23]. It integrates with latent feature extraction technique based on KPLS and is suitable to model MBVR. The relationships between multi-source frequency spectral features and mill load parameters are very complex, so single-model based soft measuring approach always leads to weak generalization ability [13].

Selective ensemble (SEN) modeling can improve generalization, validity and reliability of data-driven model [24]. Ensemble model based on KPLS and entropy [25], SEN model based on KPLS, branch and bound (BB) and adaptive weighted fusion (AWF) and its on-line version have been developed to measure mill load parameters [26,27]. However, prediction accuracy of MBVR model is unsatisfactory. The above SEN models fuse multi-source features selectively by using "manipulating input features" based ensembles construction approach. Contribution of different training samples (history samples) with multi-condition is not effectively fused. In industrial practice, the domain experts always use their cumulative knowledge to infer MBVR status. It may be realized by using the useful information in multi-source features and multi-condition history samples jointly.

Genetic algorithm based SEN (GASEN) [28] uses "sub-sample training samples" method to construct ensembles, which can fuse information in different history samples effectively. However, its candidate sub-models are built based on back propagation neural network (BPNN). And its ensemble sub-models are combined with simple average method. Normally, combination strategy of effective ensemble sub-models can improve prediction performance. The above SEN modeling methods cannot eliminate collinearity among prediction outputs of these ensemble sub-models.

SEN model needs more computation than single model because of two reasons. The first one is the size of ensemble sub-model. For kernel based learning model, numbers of the training samples and input features are two key factors. This paper faces with a high dimensional small sample modeling problem, so KPLS based feature extraction is employed to reduce the dimension. However, the selection of the number of kernel latent variables (KLVs) for different spectral segments with acceptable prediction performance and computation load is an open issue. Another reason is ensemble size of SEN model. Large ensemble size means more computation. The optimal range of ensemble size is 2– 8 [29], so a compromise should be made to SEN model for industrial applications.

This paper aims to develop an improved soft sensor technique to measure MBVR. Shell vibrations and acoustical spectral segments with different physical interpretations are used as inputs. Multi-source nonlinear latent features are extracted from these spectral segments by using KPLS. They are fed into bootstrap and LS-SVM to construct candidate sub-models with different training samples. Genetic algorithm (GA) and partial least squares regression (PLSR) are used to select and combine ensemble sub-models to obtain the final SEN model. Useful information in multi-source spectral segments and multi-condition training samples are fused selectively.

Relations among different modeling methods of our study are shown in Fig. 1.

2. SEN Modeling Strategy Based on Nonlinear Latent Frequency Spectral Feature Extraction

A new nonlinear latent spectral feature based SEN modeling strategy is proposed, which consists of data processing module based on fast Fourier transform and frequency spectrum automatic partition, nonlinear latent feature extraction module based on KPLS, candidate sub-models construction and selection module based on LS-SVM and GA, and combination module of ensemble sub-models based on PLSR. Its structure is shown in Fig. 2, where superscripts t and f represent time domain and frequency domain; subscripts V and A represent vibration and acoustic; $x_{\rm V}^{\rm t}$ and $x_{\rm A}^{\rm t}$ are the original time-domain signals; $\boldsymbol{x}_{\rm V}^{\rm f}$ and $\boldsymbol{x}_{\rm A}^{\rm f}$ are shell vibration and acoustic frequency spectra; $\boldsymbol{x}_{{\rm V}d}^{\rm f}$ and $\boldsymbol{x}_{{\rm A}d}^{\rm f}$ are the $d_{\rm v}$th and $d_{\rm A}$th spectral segments; $D_{\rm V}(D_{\rm A})$ is the number of vibration (acoustic) spectral segments; $\boldsymbol{z}$ is the extracted frequency spectral latent features; $J_{\rm sel}$ is the number of ensemble sub-models, *e.g.*, ensemble size; $\hat{y}_{j_{\rm sel}}$ is the prediction output of the $J_{\rm sel}$th ensemble sub-model; $\hat{y}$ and y are the prediction output of SEN model and true value.

The modeling process is described as follows. Firstly, transform the shell vibrations and acoustic signals into frequency spectra using fast Fourier transform, and then partition these frequency spectra to several spectral segments. Secondly, extract nonlinear latent features from these spectral segments using KPLS. Thirdly, construct and select ensemble sub-models with LS-SVM and GA. Finally, PLSR is used to combine prediction outputs of these ensemble sub-models.

3. Realization of the Proposed SEN Modeling Approach

Data processing method in this paper is same as that in [13]. The final frequency spectra can be denoted as

$$\begin{aligned}\boldsymbol{X}^{\rm f} &= \left\{\boldsymbol{X}_{\rm V}^{\rm f}, \boldsymbol{X}_{\rm A}^{\rm f}\right\} \\ &= \left\{\boldsymbol{X}_{\rm V1}^{\rm f}, \cdots, \boldsymbol{X}_{{\rm V}d_{\rm V}}^{\rm f}, \cdots, \boldsymbol{X}_{{\rm V}D_{\rm V}}^{\rm f}, \boldsymbol{X}_{\rm A1}^{\rm f}, \cdots, \boldsymbol{X}_{{\rm A}d_{\rm A}}^{\rm f}, \cdots, \boldsymbol{X}_{{\rm A}D_{\rm A}}^{\rm f}\right\} \\ &= \left\{\boldsymbol{X}_1^{\rm f}, \cdots, \boldsymbol{X}_d^{\rm f}, \cdots, \boldsymbol{X}_D^{\rm f}\right\}\end{aligned} \tag{1}$$

where $\boldsymbol{X}^{\rm f} \in \Re^{k\times p}$ is the training samples with number k and dimension p, and $\boldsymbol{X}_d^{\rm f} \in \Re^{k\times p_d}$ is the dth spectral segment. Relations among dimensions of different spectral matrices can be represented as

$$p = p_{\rm V} + p_{\rm A} = \sum_{d_{\rm V}=1}^{D_{\rm V}} p_{d_{\rm V}} + \sum_{d_{\rm A}=1}^{D_{\rm A}} p_{d_{\rm A}} = \sum_{d=1}^{D} p_d \tag{2}$$

where $p_{\rm V}$ and $p_{\rm A}$ are the dimensions of shell vibration and acoustic spectra, and $D = D_{\rm V} + D_{\rm A}$ is the number of spectral segments.

3.1. Nonlinear latent feature extraction based on KPLS

Take the dth spectral segment $\boldsymbol{X}_d^{\rm f} = \{(\boldsymbol{x}_d^{\rm f})_l\}_{l=1}^k$ as example. $\{(\boldsymbol{x}_d^{\rm f})_l\}_{l=1}^k$ is non-linearly transformed to the feature space first. Kernel trick

$$\boldsymbol{K}_d = \left(\left(\boldsymbol{x}_d^{\rm f}\right)_l\right)^{\rm T} \Phi\left(\left(\boldsymbol{x}_d^{\rm f}\right)_m\right), \quad l, m = 1, 2, \cdots k \tag{3}$$

is used to realize this mapping. The kernel matrix $\boldsymbol{K}_d \in \Re^{k\times k}$ is centralized as

$$\tilde{\boldsymbol{K}}_d = \left(\boldsymbol{I} - \frac{1}{k}\boldsymbol{1}_k\boldsymbol{1}_k^{\rm T}\right)\boldsymbol{K}_d\left(\boldsymbol{I} - \frac{1}{k}\boldsymbol{1}_k\boldsymbol{1}_k^{\rm T}\right) \tag{4}$$

where $\boldsymbol{I}$ is a unite matrix of k dimension and $\boldsymbol{1}_k$ is a vector with value l and length k. Linear PLS algorithm is performed in this feature space. The extracted latent features from $\{(\boldsymbol{x}_d^{\rm f})_l\}_{l=1}^k$ can be represented as

$$\boldsymbol{Z}_d = \left[\boldsymbol{t}_{d_1}, \boldsymbol{t}_{d_2}, \cdots, \boldsymbol{t}_{d_{h_{\rm klv}}}\right] \tag{5}$$

where $\boldsymbol{Z}_d \in \Re^{k\times h_{\rm klv}}$ and $h_{\rm klv}$ is the number of KLVs.

We extract latent features from every spectral segment with the same number of KLV. All the extracted features can be represented as

$$\boldsymbol{Z} = [\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_d, \cdots, \boldsymbol{Z}_D] \tag{6}$$

where $\boldsymbol{Z} \in \Re^{k\times(D\cdot h_{\rm klv})}$ and $D \cdot h_{\rm klv}$ in superscript is the number of extracted latent features.