Soft sensor for ball mill load based on multi-view domain adaptation learning

Guo, Xuqi; Yan, Fei; Pang, Yusong; Yan, Gaowei

Citation (APA)
Guo, X., Yan, F., Pang, Y., & Yan, G. (2019). Soft sensor for ball mill load based on multi-view domain adaptation learning. In *Proceedings of the 31st Chinese Control and Decision Conference (CCDC 2019)* (pp. 6082-6087). IEEE. https://doi.org/10.1109/CCDC.2019.8832908

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Soft Sensor for Ball Mill Load Based on Multi-view Domain Adaptation Learning

Xuqi Guo[a], Fei Yan[b], Yusong Pang[c], Gaowei Yan[d]

a. College of electrical and power engineering, Taiyuan University of Technology, Taiyuan, 030600
E-mail: guoxuqi0330@link.tyut.edu.cn

b. College of electrical and power engineering, Taiyuan University of Technology, Taiyuan, 030600
E-mail: yanfei140222@163.com

c. Section Transport Engineering and Logistics, Delft University of Technology, Netherlands, 2628CD
E-mail: Y.Pang@tudelft.nl

d. College of electrical and power engineering, Taiyuan University of Technology, Taiyuan, 030600
E-mail: yangaowei@tyut.edu.cn

Abstract: In the operation process of wet ball mill, there are often multi-modal and multi-condition problems. In this paper, a multi-view based domain adaptive extreme learning machine (MVDAELM) was used to measure the mill load. Firstly, the correlation relationship between the load parameters and the two views (vibration and acoustic signals of the ball mill) was obtained by Canonical Correlation Analysis (CCA) respectively. Secondly, a small number of labeled data from the target domain were introduced to construct a Domain Adaptation Extreme Learning Machine (DAELM) model under manifold constraints, which solve the mismatch problem caused by the change of working conditions in the multi-condition grinding process. Finally, based on the correlation coefficient obtained before, the two views domain adaptive load parameter soft sensor model was integrated to solve the uncertainty problem in single-modal data modeling. The experimental results show that the proposed method can effectively improve the learning accuracy of the soft sensor model under multi-modal conditions.

Key Words: transfer learning; domain adaptation; multi-view; mill load; soft sensor

1   INTRODUCTION

Mill load (ML) refers to the general name of the important parameter in the grinding process of the mineral processing industry[1]. Accurate detection of ML plays a key role in achieving optimal control of grinding process, energy saving and improvement of grinding efficiency and quality of grinding products[2].

Wet ball mill load detection is mainly for the internal parameters of the mill that can accurately characterize ML (Material to ball volume ratio, MBVR, Pulp dnsity, PD, Charge volume ratio, CVR). In [3] a soft sensor model for ML parameters is established by extracting and selecting various spectral characteristics of the cylinder vibration signal. But it has poor generalization and low precision due to the single cylinder vibration signal. Multi-view learning (MVL)[4] is a semi-supervised learning that uses a variety of information derived from the same research object to improve learning performance. It can alleviate over-learning problems and effectively process heterogeneous data. It is combined with integrated learning[5] in the literature[6]. The proposed multi-view integration learning has achieved better model performance than traditional integrated learning, but still does not solve the problem of reduced prediction accuracy caused by

changes in operating conditions. In [7], an adaptive soft sensor method based on just-in-time learning (JITL) is proposed for the multi-condition system, where the prediction accuracy degradation due to the change of working conditions, so that the soft sensor model can adapt to the change of working conditions to a certain extent.

However, JITL is still based on the premise that the modeling data meets the requirements of independent and identical distribution. In the actual industrial process, the data distribution is inconsistent between the operating conditions due to the sudden change of the system operating conditions. Modeling on such a multi-condition system needs to solve the problem of data difference between different working conditions. Therefore, this paper introduces the transfer learning [8] strategy to solve the problem that test conditions only a small number of labeled samples, and the distribution is different between the test and model working conditions.

In [9], the DAELM was proposed and applied to the problem of classifier performance deterioration caused by sensor characteristic drift, which solves the model mismatch problem caused by data distribution difference. But the weight random initialization setting of ELM network[14] determines that it will result in a nonlinear structure of the data. Therefore, we use the manifold regularization[10] there to constrain the data geometry in the DAELM network structure, thereby improving the stability of the model.

In this paper, the MVL strategy[11] is used to extract the correlation between the two-view data (bearing vibration and acoustic signal) and the output load parameter by

CCA[12]; then introduce a small amount of labeled data of the target condition in the two views with the original working condition data to construct the DAELM model Respectively. Integrating[13] the two-view model through the correlation under the constraints of the manifold framework to achieve the measurement of the load parameters of the test condition.

In summary, this paper adopts MVDAELM to solve the problem of model misalignment caused by changing working conditions of multimodal system. Experiments on laboratory ball mill data verify that the soft sensor model established in this paper has good performance.

2 RELATED WORK

2.1 Canonical Correlation Analysis (CCA)

The canonical correlation analysis is a multivariate statistical method that studies the overall linear correlation between two sets of variables.

Suppose that the two sets of experimental samples as $\mathbf{X}=[x_1,x_2,...x_p]^{\mathrm{T}}$ and $\mathbf{Y}=[y_1,y_2,...,y_q]^{\mathrm{T}}$, p and q denote the dimensions of the samples X and Y, respectively. The basic idea of CCA is to find a pair of projection vectors and satisfy the maximum correlation coefficient between the sums. Then by the typical variables u1 and the correlation study of v1 replaces the correlation between the original two sets of variables. The correlation coefficient of u1 and v1can be deduced as :

$$\rho_1(\mathbf{u}_1,\mathbf{v}_1)=\frac{\mathbf{a}_1^{\mathrm{T}}\Sigma_{\mathbf{XY}}\mathbf{b}_1}{\sqrt{(\mathbf{a}_1^{\mathrm{T}}\Sigma_{\mathbf{XX}}\mathbf{a}_1)(\mathbf{b}_1^{\mathrm{T}}\Sigma_{\mathbf{YY}}\mathbf{b}_1)}} \qquad (1)$$

Where $\Sigma_{\mathbf{XX}}$ and $\Sigma_{\mathbf{YY}}$ denote the covariance matrix of feature sets X and Y, respectively, and $\Sigma_{\mathbf{XY}}$ is the cross-covariance matrix of X and Y. If there is no constraint on a1 and b1 in the above formula, it is clear that there are infinitely many solutions satisfying the correlation coefficient. In order to make the equation have a unique solution, add the constraint as：

$$\mathbf{a}_1^{\mathrm{T}}\Sigma_{\mathbf{XX}}\mathbf{a}_1=\mathbf{b}_1^{\mathrm{T}}\Sigma_{\mathbf{YY}}\mathbf{b}_1=1 \qquad (2)$$

The objective function can be expressed as:

$$\max \rho_1(\mathbf{u}_1,\mathbf{v}_1)=\mathbf{a}_1^{\mathrm{T}}\Sigma_{\mathbf{XY}}\mathbf{b}_1 \qquad (3)$$

$$\begin{cases}\mathbf{a}_1^{\mathrm{T}}\Sigma_{\mathbf{XX}}\mathbf{a}_1=\mathbf{b}_1^{\mathrm{T}}\Sigma_{\mathbf{YY}}\mathbf{b}_1=1\\ \mathbf{u}_1=\mathbf{a}_1^{\mathrm{T}}\mathbf{X}\\ \mathbf{v}_1=\mathbf{b}_1^{\mathrm{T}}\mathbf{Y}\end{cases} \qquad (4)$$

Solving the equation according to the Lagrangian multiplier method, assume that the non-zero eigenvalue of $\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}$ as $\lambda_1^2\geq\lambda_2^2\geq\cdots\geq\lambda_p^2$, and find the optimal solution $(\rho_k,\mathbf{a}_k,\mathbf{b}_k)$, $k=1,2,\cdots,p$. The corresponding *k-th* pair of typical correlation variables of X and Y is $(\mathbf{u}_k,\mathbf{v}_k)=(\mathbf{a}_k^{\mathrm{T}}\mathbf{X},\mathbf{b}_k^{\mathrm{T}}\mathbf{Y})$, and the typical correlation coefficient is :

$$\rho(\mathbf{u}_k,\mathbf{v}_k)=\sqrt{\lambda_k^2} \qquad (5)$$

2.2 Domain Adaptive Extreme Learning Machine (DAELM)

In the algorithm, the DAELM is based on the ELM combines the idea of transfer learning.

Suppose that the source domain and target domain are represented by *Ds* and *Dt*, in this paper, we assume that all the samples in the source domain are labeled, and has a small number of tags in the target domain, where the source domain and the target domain represent source modeling conditions and unmodeled conditions, respectively. DAELM aims to train the learner parameters through all the labeled samples of the source domain and a small number of labeled samples of the target domain. The objective function is as follow:

$$\min_{\beta,\xi_S^i,\xi_T^i}\frac{1}{2}\|\beta\|^2+C_S\frac{1}{2}\sum_{i=1}^{N_S}\|\xi_S^i\|^2+C_T\frac{1}{2}\sum_{i=1}^{N_T}\|\xi_T^i\|^2 \qquad (6)$$

$$s.t.\begin{cases}H_S^i\beta=t_S^i-\xi_S^i,\quad i=1,...,N_S\\ H_T^j\beta=t_T^j-\xi_T^j,\quad j=1,...,N_T\end{cases} \qquad (7)$$

where $H_S^i\in R^{1\times L},\xi_S^i\in R^{1\times m},t_S^i\in R^{1\times m}$ denote the output of hidden layer, the prediction error, and the label with respect to the *i-th* training instance x_s^i from the source domain; $H_T^j\in R^{1\times L},\xi_T^j\in R^{1\times m},t_T^j\in R^{1\times m}$ denote the output of hidden layer, the prediction error, and the label vector with respect to the *j-th* guide samples x_{T}^j from the target domain; $\beta\in R^{L\times m}$ is the output weights being solved; N_S and N_T denote the number of training instances and guide samples from the source domain and target domain, respectively; and C_S and C_T are the penalty coefficients on the prediction errors of the labeled training data from source domain and target domain, respectively. From(6), we find that very few labeled guide samples from target domain can make the learning of β transferable, and realize the knowledge transfer between source domain and target domain by introducing the third term as regularization coupling with the second constraint in(7).

Since the random initialization of the weight of the DAELM causes the data to produce a nonlinear structure, the manifold constraint is introduced to maintain the data structure. The objective function is as follows:

$$\min_{\beta_s,\xi_S^i,\xi_T^i}\frac{1}{2}\|\beta_s\|^2+C_S\frac{1}{2}\sum_{i=1}^{N_S}\|\xi_S^i\|^2$$
$$+C_T\frac{1}{2}\sum_{i=1}^{N_T}\|\xi_T^i\|^2+\frac{C_g}{2}tr(Y^TLY) \qquad (8)$$

Where the matrix L represents a manifold regulation Laplacian matrix, the specific solution formula is detailed in Document [10], C_g represents the penalty coefficient of the manifold constraint.

3 SOFT SENSOR STRATEGY

In summary, multi-view learning strategies can be introduced for various modal signals such as vibration, acoustic and current pressure difference in the working

process of the ball mill, so that the correlation between multi-modal signals can be fully considered to solve the loss of soft measurement model caused by missing signal or noise interference in single-mode signal. In the experiment, the multi-view learning method is introduced with bearing vibration and acoustic signal. At the same time, the domain adaptation algorithm is used to solve the model mismatch problem caused by the change of working conditions.

Therefore, this paper adopts the MVDAELM to construct a soft sensor model, and integrates the domain adaptive learning network established on the two views of the vibration and acoustic to solve the model adaptation problem caused by the change of working conditions in the multi-case task, and overcome the model misalignment caused by the uncertainty of data single-modal data. The soft sensor strategy is shown in Figure 1.
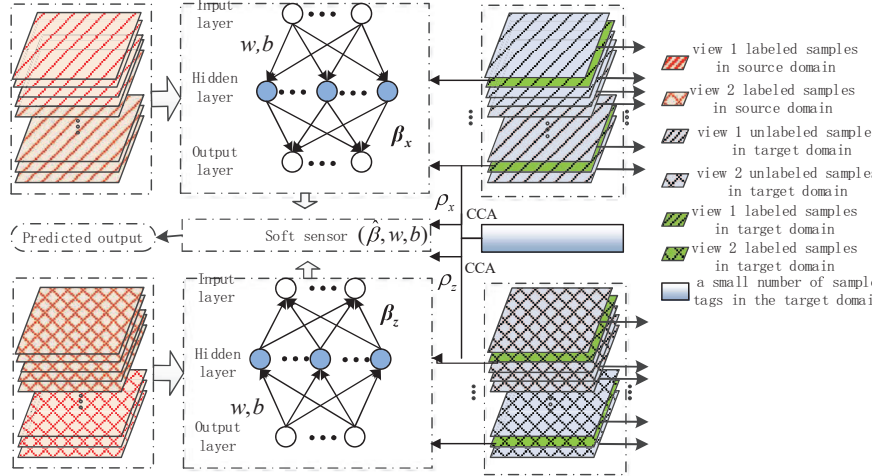

Fig 1. Soft sensor strategy.

3.1 Problem Statement

In order to make the model universal, assumed that a source domain with all labeled sample and a target domain with only a small number of labeled sample are included in one experiment.

The purpose of the model is to estimate the tags of unlabeled samples in the target domain from the source domain samples in the historical database and a small number of tagged samples in the target domain. The relationship between input and output of the model is shown in Table 1.

Table 1 Relationship between input and output of the prediction model

Input	Output
$D_S = \{(x_i, z_i, y_i)\}_{i=1}^n \cup D_t = \{D_{ti}^l, Dl_{tj}^u\}$	y_j

3.2 Proposed method

Suppose that n represents the number of source domain samples, m represents the number of tag samples in the target domain, r represents the number of unlabeled samples in the target domain. $D_S = \{(x_i, z_i, y_i)\}_{i=1}^n$ refers to the source domain data set, and x_i is the column vector of the i-th sample in the first view (i.e. bearing vibration signal samples), z_i is the column vector of the i-th sample of the second view (i.e. the acoustic signal sample), y_i is the corresponding label; $D_t^l = \{(x_i, z_i, y_i)\}_{i=1}^m$ indicates the labeled samples set in target domain, and $D_t^u = \{(x_i, z_i, y_i)\}_{i=1}^r$ denotes the unlabeled data set of

the target domain, the overall objective function can be described as follows:

$$
\begin{aligned}
\hat{\beta} &= \rho_x \beta_x + \rho_z \beta_z \\
&= \rho_x (I + C_{Sx} H_{Sx}^T H_{Sx} + C_{Tx} H_{Tx}^T H_{Tx} + C_g H_{Sx}^T L H_{Sx})^{-1} \\
&\quad (C_{Sx} H_{Sx}^T t_{Sx} + C_{Tx} H_{Tx}^T t_{Tx}) \\
&\quad + \rho_z (I + C_{Sz} H_{Sz}^T H_{Sz} + C_{Tz} H_{Tz}^T H_{Tz} + C_g H_{Sz}^T L H_{Sz})^{-1} \\
&\quad (C_{Sz} H_{Sz}^T t_{Sz} + C_{Tz} H_{Tz}^T t_{Tz})
\end{aligned} \tag{9}
$$

Where ρ_x and ρ_z denote the correlation coefficient between the two-view signal and the load parameter to be tested, respectively. And β_x, β_z are the output weights of the DAELM model established in the two views respectively. The algorithm process as is shown in **Algorithm 1**.

4 EXPERIMENTAL RESULTS

4.1 Data preprocessing

The experiment was based on a laboratory small wet ball mill. Since the medium filling rate (MFR) varies between 0.3 and 0.5 in the actual industrial process, in order to simulate the change of working conditions in the actual industrial process, the experiment fixed medium filling separately. The rate of 5 sets of experimental program data, that is, each set of experiments fixed the mass of the steel ball and water in the corresponding mill drum, and changed the mass of the material to obtain different load parameters.

Input: $D_S = \{(x_i, z_i, y_i)\}_{i=1}^{n}$, $D_t = \{D_t^l, D_t^u\}$,

Number of hidden layer nodes of ELM, penalty

coefficient C_S, C_T, C_g

Output: Output weight $\hat{\beta}$, Forecast label y_j

Step1: Data preprocessing

Step2: Model parameter initialization

For i=1 to n

According to the ELM algorithm, the source domain

model is separately trained in two views, and the output

weight β_{x0}, β_{z0} is obtained according to formula (8).

  For j=1 to m

    Calculate the correlation coefficient ρ_x, ρ_y between

    the two-view sample and the output load parameter

    according to equation (5)

    Use the target domain label data to migrate the

    source domain model and calculate the output

    weight β_x, β_z according to equation (11)

    For k=1 to r

      For the unlabeled data in target domain, calculated

      $\hat{\beta}$ according to equation (12)

      And calculate the predicted output $\hat{y}_i$

    End For

  End For

End For

Simultaneously collect the vibration and acoustic signal of the ball mill under different working conditions. The number of experiments under various working conditions is shown in Table 2.

Table 2　Number of signal acquisition times

MFR	0.3	0.35	0.4	0.45	0.5
Times	139	103	88	95	102

On this basis, the bearing vibration signals collected in each experiment are equally divided into 28 samples (each sample covers the data of the ball mill rotating for more than 1 week), and each sample is non-overlapping sliding with a window of $l = 1024$, performs welch transformation on each window, and the average value of each window is averaged to complete the vibration signal feature extraction process. For the experimental acquisition of the vibration signal, the fourier transform (FFT) is performed in the window, and the rest is the same as the vibration signal processing.

4.2　Experimental results and analysis

In order to verify the effectiveness of CCA-based MVDAELM, extreme learning machine (ELM), Bagging, JITL and DAELM are also used as comparison methods,. The prediction experiments were carried out on the three mill load parameters of MBVR, PD and CVR.

In the soft sensor modeling of mill load parameters, for the convenience of comparison, the working condition 1 is taken as the source domain, and the other four working conditions are taken as the target domains respectively. The experimental results are expressed as 1-2, 1-3, and 1-4, 1-5, respectively.

Table 3 is the root mean square error (RMSE) value of the soft sensor results of different prediction models. In which 1-1 indicates that the training set and test set data are both derived from the working condition 1. It can be seen that ELM, Bagging and JITL satisfy the independent and identical distribution in the data. At the same time, the ideal prediction accuracy can be obtained. Compared with other prediction models, MVDAELM can achieve the lowest prediction error for the three mill load parameters when the working conditions is changing, and compared with the single-mode DAELM experimental results. It can also be seen that the domain adaptation learning of the integrated two views can obtain higher prediction accuracy when predicting load parameters.

Figure 2 shows the prediction results of the load parameters of each mill when the target domain is working condition 2 with ELM and MVDAELM respectively, it can be seen that transfer learning can effectively solve the problem in soft sensor modeling that the data distribution mismatch caused by the change of working conditions.

Figure 3 shows the prediction results comparison of PD over the three algorithm Bagging, JITL, and DAELM under target domain is working condition2, combined with the corresponding prediction results in Figure 2. We can find
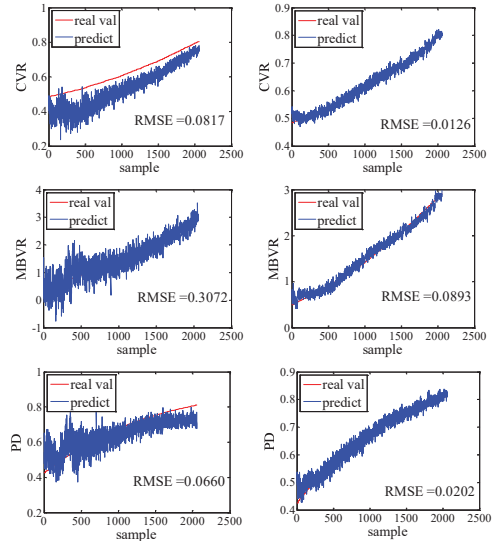

Fig 2. Prediction result when the target domain is condition 2

Table 3 RMSE comparison of mill load parameter prediction results

Algorithm	Load parameter	1-1	1-2	1-3	1-4	1-5
ELM	MBVR	0.1819	0.3072	1.0044	1.2182	1.1312
	PD	0.0315	0.0660	0.0656	0.1763	0.2541
	CVR	0.0231	0.0817	0.3063	0.2516	0.3607
Bagging	MBVR	0.0923	0.1618	0.3379	0.3919	0.2346
	PD	0.0157	0.0470	0.0897	0.0880	0.1263
	CVR	0.0086	0.0808	0.1270	0.1395	0.2012
JITL	MBVR	0.0829	0.3688	0.4490	0.7043	0.9147
	PD	0.0138	0.0695	0.1149	0.3406	0.3467
	CVR	0.0094	0.0908	0.1428	0.1777	0.2471
DAELM	MBVR	-	0.1349	0.0854	0.0843	0.0709
	PD	-	0.027	0.017	0.0153	0.0178
	CVR	-	0.0159	0.0114	0.0084	0.0081
MVDAELM	MBVR	-	0.0893	0.0766	0.0526	0.0457
	PD	-	0.0202	0.0145	0.0123	0.0145
	CVR	-	0.0126	0.0102	0.0068	0.0070
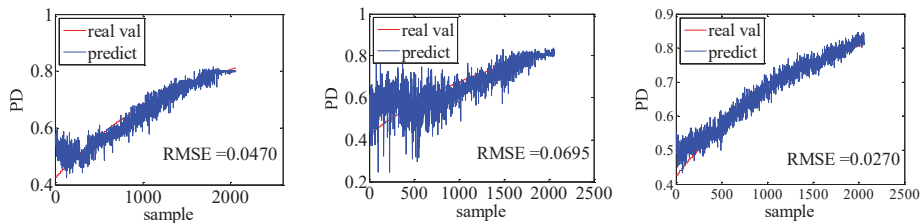
Fig 3. PD prediction results comparison over the three algorithm under working condition 2
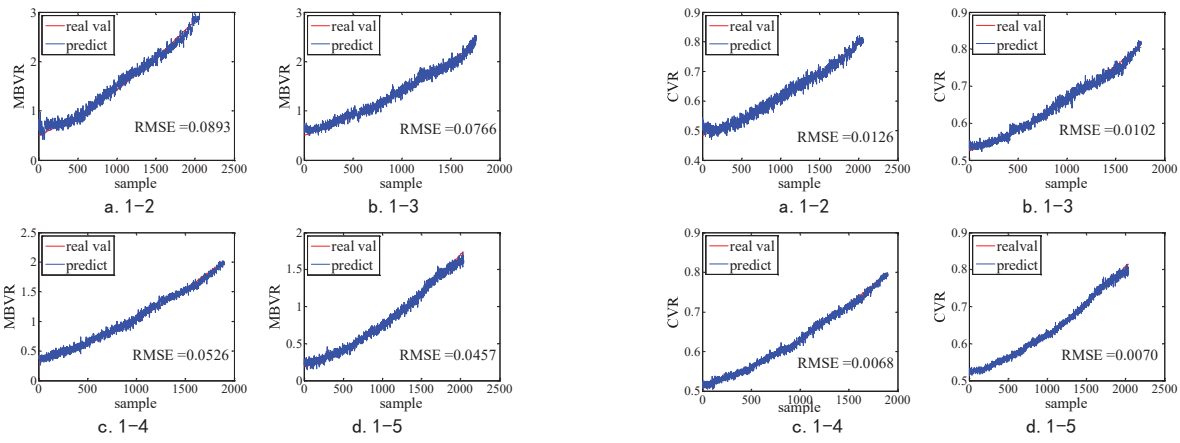
a. 1-2 b. 1-3 c. 1-4 d. 1-5
(a) Prediction results of MBVR

a. 1-2 b. 1-3 c. 1-4 d. 1-5
(c) Prediction results of CVR
Fig 4. Prediction results of MVDAELM
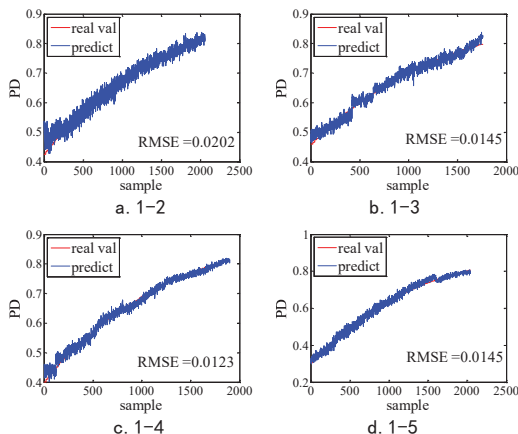
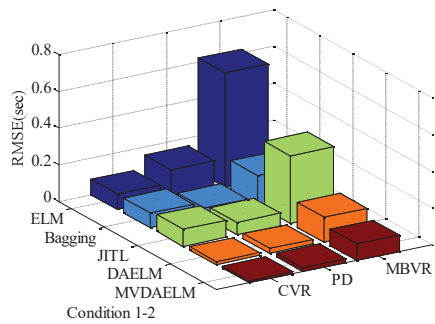a. 1-2 b. 1-3 c. 1-4 d. 1-5
(b) Prediction results of PD

Fig 5. Comparison of modeling methods when the target domain is condition 2
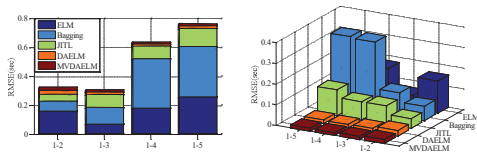
Fig 6. Comparison of modeling prediction results for PD

that the proposed algorithm MVDAELM can always obtain the best prediction results in the soft sensor experiment. The comparison between the DAELM prediction results and the MVDAELM prediction results shows that after the introduction of multi-view learning, the prediction results are more stable and the error is reduced. Figure 4 shows the prediction results of three mill load parameters for domain adaptive learning with the operating condition 1 as the source domain.

Figure 5 shows the comparison of the prediction errors of the three types of mill load parameters with the condition 1 as the source domain and the condition 2 as the target domain. Figure 6 shows the prediction error comparison of the PD over different method. It is straightforward to see that the MVDAELM can obtain the lowest prediction error by the stacking histogram.

The experiment proves that when the working condition of the multi-mode wet ball mill changes, the performance of the soft sensor model established by historical data decreases rapidly. To solve the misalignment of the soft sensor model caused by the change of working conditions, the domain adaptive learning can effectively improve the adaptability of the model.

It can be seen from the comparison results that after introducing multi-view learning to fully consider the multi-modal signals generated during the working process of the mill, the generalization performance and prediction accuracy of the MVDAELM model are improved compared with the DAELM model, so that the effectiveness of the method can be verified.

5   CONCLUSION

In order to make full use of the multi-modal signals collected during the operation of the wet ball mill, and solve the problem of model misalignment caused by the mismatch between real-time data and historical data when the working conditions during the working process of the wet ball mill suddenly changed. This paper introduces multi-view learning ideas, the two signals of bearing vibration and acoustic are preprocessed by welch and FFT method respectively as two-view data, and the transfer learning strategy is introduced, so that a multi-view domain adaptive extreme learning machine model based on canonical correlation analysis is established.

The experimental results show that when the mill operating conditions changed, the multi-view and transfer learning strategies are introduced, which can fully utilize the multi-modal data generated during the operation of the wet mill to reduce the uncertainty of the prediction model, and further the transfer learning through the domain adaptation can reduce the cost of data collection, solve the problem of multiple working conditions in the working process of the

ball mill, effectively improve the generalization ability of the model, and improve the accuracy and credibility of the prediction results.

REFERENCES

[1] Tang J, Zhao L, Chai T, Yue H. On-line Soft-sensing Modelling of Mill Load Based on Vibration Spectrum[J]. Information & Control, 2012(in Chinese)

[2] Yang Z G, Zhang Jie, Li Y J. Analysis of factors affecting grinding noise and summary of mill load detection methods [J].Metal Mine, 2015,22(2): 139-144 (in Chinese)

[3] Tang J, Chai T Y, Yu W, Zhao L J. Feature extraction and selection based on vibration spectrum with application to estimating the load parameters of ball mill in grinding process. Control Engineering Practice, 2012, 20(10): 991-1004

[4] Shijagurumayum E S, Chang F C. Preliminary Survey of Multiview Synthesis Technology[C]// International Conference on Intelligent Information Hiding & Multimedia Signal Processing. IEEE, 2010:163-166.

[5] Tang Jian，Chai Tian-You，Yu Wen，Zhao Li-Jie. On-lineKPLS algorithm with application to ensemble modeling pa-rameters of mill load. Acta Automatica Sinica，2013，39(5): 471-486

[6] Guo Yun, Zhang Daoqiang, Song Tong. Constraint-based ensemble learning algorithm for canonical correlation analysis [J]. Pattern recognition and artificial intelligence, 2012, 25 (5): 851-858

[7] Shao W，Tian X，Wang P. Supervised local and non-local structure preserving projections with application to just-in-time learning for adaptive soft sensor ☆[J]. Chinese Journal of Chemical Engineering (English Edition)，2015，23(12):1925-1934.

[8] Pan S J，Yang Q. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering，2010，22:1345‐1359.

[9] Zhang L，Zhang D. Domain Adaptation Extreme Learning Machines for Drift Compensation in E-Nose Systems[J]. IEEE Transactions on Instrumentation & Measurement，2015，64(7):1790-1801.

[10] X Jia-Ming, ZHANG, Wei-Qiang, YANG Deng-Zhou, Liu Jia. Manifold Regularized Extreme Learning Machine for Language Recognition[J]. Acta Automatica Sinica, 2015, 41(9): 1680-1685(in Chinese)

[11] Shi jagurumayum E S, Chang F C. Preliminary Survey of Multiview Synthesis Technology[C]// International Conference on Intelligent Information Hiding & Multimedia Signal Processing. IEEE, 2010:163-166.

[12] Hardoon D R, Szedmak S, Shawe-Taylor J. Canonical Correlation Analysis: An Overview with Application to Learning Methods[J]. Neural Computation, 2014, 16(12):2639-2664.

[13] Fern A, Givan R. Online Ensemble Learning: An Empirical Study[J]. Machine Learning, 2003, 53(1-2):71-109.

[14] Liu X，Lin S，Fang J，et al. Is Extreme Learning Machine Feasible A Theoretical Assessment (Part I)[J]. IEEE Transactions on Neural Networks & Learning Systems，2014，26(1):7-20.