Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

FeO Content Prediction for an Industrial Sintering Process based on Supervised Deep Belief Network

Xiaofeng Yuan, Yongjie Gu, Yalin Wang*, Zhiwen Chen, Bei Sun, Chunhua Yang

School of Automation, Central South University, Changsha, 410083, China
(E-mail: yuanxf@csu.edu.cn, gyj184611076@csu.edu.cn, ylwang@csu.edu.cn, czhiwen@csu.edu.cn, 215098@csu.edu.cn, ychh@csu.edu.cn)

Abstract: In industrial sintering processes, it is very important to monitor and control key quality indicators, which are often difficult to measure online. Soft sensor technology is a good solution for online prediction of quality indicators. Nowadays, deep learning is widely used in soft sensors due to its powerful ability in processing nonlinear data. In this paper, a supervised deep belief network (SDBN) is proposed by introducing quality variable into the input variables at each restricted Boltzmann machine to extract quality-related features for soft sensor. With case study on an actual industrial sintering process, SDBN shows much better prediction performance than the original deep belief network and stacked autoencoder.

Keywords: Soft sensor, Deep learning, Quality prediction, DBN, SDBN

1. INTRODUCTION

With the development of science and technology, the automation level of various industries has been significantly improved. Modern industrial plants are developing towards the direction of digitalization, automation, networking and intelligence. However, there are still some problems in the development of industrial intelligence, especially in the process industries represented by the pharmaceutical, chemical, petrochemical, and steel manufacturing industries. For example, it often faces multi-objective and multi-scale dynamic optimization problems in the process operation and schedule due to the multi-association and strong coupling relationship in the manufacturing process. At the same time, due to unpredictable raw material properties, the complex physical mechanism and chemical reactions, it is difficult to establish accurate mathematical models for the real-time monitoring and optimization of the production processes.

In order to conduct complex dynamic optimization and real-time monitoring, it is often necessary to timely measure the key quality variables. However, most of these key variables are difficult to measure online due to technical and economical limitations. Thus, soft sensor technology [Brosilow et al. (1978) ;Joseph et al. (1978) ;Yuan et al. (2019d) ;Yuan et al. (2020c)] has been widely developed and applied in industrial processes to tackle this problem. The basic idea of soft sensor technology is to establish mathematical models to estimate the difficult-to-measure variables by using the variables that are easy to measure in the production processes. For example, the sintering process is one important part of the metallurgical industry. In the sintering process, the artificial enrichment of ore is sintered to reduce the smelting cost of the blast furnace, and improve the smelting efficiency and quality of the blast furnace. In order to timely optimize the production process and improve product quality, quality variables in the sintering process must be monitored in real time. However, due to the harsh production environment and expensive measurement cost, some key quality variables like the total Fe content and the ferrous oxide (FeO) content in the production process cannot be measured directly [Zhang et al. (2007) ;Wan et al. (2009) ;Chen et al. (2010) ;Wang et al. (2014)]. Therefore, it is necessary to establish soft sensor models to predict these key quality variables [Wang (2006)] in the sintering process online.

In general, soft sensor methods are divided into mechanism modelling methods and data-driven modelling methods. Mechanism modelling [Kadlec et al. (2009) ;Dai et al. (2020)] requires soft sensor designers to be fully familiar with the physicochemical knowledge of the production process. In most cases, it is very difficult to obtain accurate mechanism models for real industrial processes. With the advancement of technology, enough data and efficient computing devices can provide the basis for the development of data-driven models [Yuan et al. (2019b)]. Thus, data-driven methods have become the most popular soft sensing approaches for its outstanding performance in high-dimensional nonlinear complex data modelling [Chen et al. (2018) ;Yuan et al. (2018b) ;Yuan et al. (2020a) ;Yuan et al. (2020b)].

For a long time, principal component regression (PCR) [Yuan et al. (2017a) ;Yuan et al. (2017b)] and partial least squares regression (PLSR) [Ge et al. (2014)] are the most commonly used linear data-driven models. However, they do not perform well in complex nonlinear industrial processes. Thence, kernel PCR [Ge et al. (2009)], kernel PLSR [Rosipal et al. (2002)] and artificial neural network (ANN) [Wang et al. (2019)] have been developed to build nonlinear models to deal with the nonlinear relationship in industrial processes. For example, neural networks [Chen et al. (2013b)] were used to predict the quality of the sintering process. Also,

the backpropagation (BP)-based neural networks [Zhang et al. (2002) ;Chen et al. (2013a)] were constructed to predict the FeO content in the sintered ore.

Although traditional neural networks had achieved good prediction performance in some applications, the problems of expensive training time, gradient vanishing and gradient explosion have limited their applications to large datasets and complex processes in the era of big data. To handle these problems, deep learning has been proposed for feature learning and network training with the unsupervised layer-wise pretraining and supervised fine-tuning techniques, which were first used in the deep belief networks (DBN) [Hinton et al. (2006)]. Then, deep learning-based method has been widely used for quality prediction and fault monitoring in various industrial processes [Yuan et al. (2019c) ;Wang et al. (2020)]. As can be seen, DBN and SAE are pretrained in an unsupervised way. They cannot guarantee that the features extracted by the hidden layer are related to the quality variables. However, there are strong correlations between the input variables and quality variable in actual industrial processes Hence, it is a good practice to introduce quality information into the procedure of pre-training for better feature learning [Yuan et al. (2018a) ;Yuan et al. (2019a) ;Yuan et al. (2020d)]. In this paper, a supervised DBN is proposed for soft sensor modelling. First, supervised restricted Boltzmann machine (SRBM) is designed by introducing the quality variables into the visible layer of RBM to guide the learning direction of the model for quality-related features. Then, multiple SRBMs are hierarchically stacked to construct a deep SDBN, which can realize quality prediction by learning deep quality-related features layer by layer. In this way, SDBN-based soft sensor model can largely improve the prediction performance since the quality information is layer-wise utilized in the pre-training of each layer. The proposed SDBN model is applied for online prediction of the sinter quality in an industrial sintering process.

The rest of the paper is structured as follows. In Section 2, an overview of RBM and DBN is briefly introduced. Section 3 presents the proposed SRBM and SDBN in detail. Then, the predictive performance of SDBN is validated in a practical industrial process in Section 4. Finally, summaries are given in Section 5.

2. PRELIMINARIES

Deep belief network [Hinton et al. (2006)] is a probabilistic generative model proposed by Hinton in 2006. DBN is a stack of multiple restricted Boltzmann machines (RBM), which establishes a joint distribution between observation data and hidden feature data. With a unique training technique of pre-training and fine-tuning, DBN can effectively model high dimensional nonlinear data.

2.1 Restricted Boltzmann Machine

The structure of RBM is shown in Fig. 1. Assume $v^t=\{v_1,v_2,v_3,\cdots,v_{n_v}\}$ is the input variable vector at the visible layer at time t. $h^t=\{h_1,h_2,\cdots,h_{n_h}\}$ is the feature variable vector at the hidden layer. n_v and n_h represent the dimensions of the visible layer and the hidden layer, respectively.

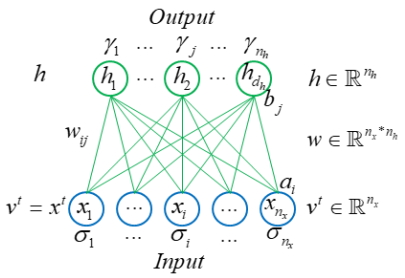

Fig. 1. Network structure of the RBM

RBM is an energy-based model. When used in quality prediction, the visible and hidden layer units of RBM are usually selected as Gaussian units to represent real-valued variables. The energy function $E(v,h)$ is defined as follows

$$E(v,h)=\sum_{i\in v}\frac{(a_i-v_i)^2}{2\sigma_i^2}+\sum_{j\in h}\frac{(b_j-h_j)^2}{2\gamma_j^2}-\sum_{i\in v}\sum_{j\in h}\frac{v_i}{\sigma_i}\frac{h_j}{\gamma_j}w_{ij} \quad (1)$$

where σ_i , γ_j are the standard deviations of the Gaussian noise of the visible unit i and the hidden unit j, respectively. $a^t=\{a_1,a_2,a_3,\cdots,a_{n_v}\}$ and $b^t=\{b_1,b_2,\cdots,b_{n_h}\}$ are the bias terms at the visible and hidden layers, respectively. w_{ij} is the weight between the visible layer unit i and the hidden layer unit j. To train the RBM model, it needs to learn model parameters that can maximize the loglikelihood of the probability distribution of the training data, which can be expressed equivalently as

$$\begin{aligned}\ln(P(D))&=\ln\prod_{t=1}^{T}P(v^t)=\sum_{t=1}^{T}\ln P(v^t)\\&=\sum_{t=1}^{T}[\ln\int_h e^{-E(v^t,h)}-\ln\int_{(v,h)}e^{-E(v,h)}]\end{aligned} \quad (2)$$

where T represents the total sample number in the training dataset D, in which $D=\{v^1,v^2,....,v^T\}$, $t=1,2,...,T$. Here, v^t represents the t^{th} training sample, and $v^t=(v_1^t,v_2^t,...,v_{n_v}^t)$. Then, the gradient of parameters for a sample v^t is computed as follows

$$\frac{\partial\ln(P(v^t))}{\partial w_{ij}}=\Delta w_{ij}=<\frac{v_i^t}{\sigma_i}\frac{h_j^t}{\gamma_j}>_{P(h|v^t)}-<\frac{v_i}{\sigma_i}\frac{h_j}{\gamma_j}>_{P(v,h)} \quad (3)$$

$$\frac{\partial\ln(P(v^t))}{\partial a_i}=\Delta a_i=<\frac{v_i^t}{\sigma_i^2}>_{P(h|v^t)}-<\frac{v_i}{\sigma_i^2}>_{P(v,h)} \quad (4)$$

$$\frac{\partial\ln(P(v^t))}{\partial b_j}=\Delta b_j=<\frac{h_j^t}{\gamma_j^2}>_{P(h|v^t)}-<\frac{h_j}{\gamma_j^2}>_{P(v,h)} \quad (5)$$

where $<formula>$ indicates the mathematical expectation under the subscript P distribution. Since it is difficult to directly calculate the second term in the above equations, the CD algorithm [Hinton (2002)] is used for approximate

calculation to simplify the computational complexity. The calculation results are as follows

$$\frac{\partial \ln(P(v^t))}{\partial w_{ij}} = \Delta w_{ij} \approx \frac{v_i^{t(0)}}{\sigma_i}\frac{h_j^{(0)}}{\gamma_j} - \frac{v_i^{t(k)}}{\sigma_i}\frac{h_j^{(k)}}{\gamma_j} \tag{6}$$

$$\frac{\partial \ln(P(v^t))}{\partial a_i} = \Delta a_i \approx \frac{v_i^{t(0)}}{\sigma_i^2} - \frac{v_i^{t(k)}}{\sigma_i^2} \tag{7}$$

$$\frac{\partial \ln(P(v^t))}{\partial b_j} = \Delta b_j \approx \frac{h_j^{(0)}}{\gamma_j^2} - \frac{h_j^{(k)}}{\gamma_j^2} \tag{8}$$

where the superscript (k) indicates the k^{th} sampling result of a certain term.

2.2 Deep Belief Networks

DBN is constructed with a hierarchical stack of multiple RBMs. The detailed structure of DBN is shown in Fig. 2. First, the unsupervised pre-training is performed for the first RBM. Then, the hidden layer of the first RBM is used as the visible layer for the next RBM, and the unsupervised pre-training procedure is repeated for the second RBM. In this way, the unsupervised pre-training is completed step by step for the whole DBN. After that, an output layer is added to the top hidden layer for specific tasks like classification or regression. Finally, the supervised fine-tuning is carried out by using the BP algorithm to train the whole network.

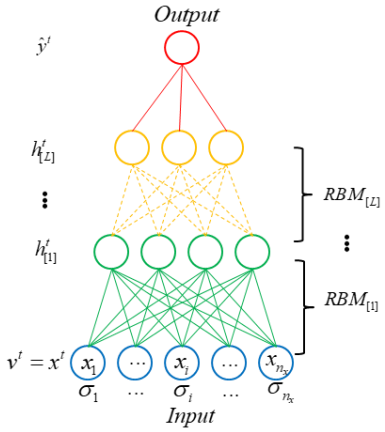

Fig. 2. Network structure of the DBN

3. Supervised Deep Belief Networks

3.1 Supervised Restricted Boltzmann Machine

Since the pre-training process of the DBN only learns the nonlinear features of the visible layer through the hidden unit of the RBM, it cannot guarantee that the learned features are associated with quality variables. To solve this problem, supervised RBM is proposed in this paper, which introduces quality variables at the visible layer during pre-training. The structure of SRBM is shown in Fig. 3. Assume $y = [y_1, \cdots y_I, \cdots y_{n_y}]^T \in R^{n_y}$ ($I = 1, 2, ..., n_y$) is the quality variable vector. In most cases, n_y is equal to 1. With the introduction of the quality variables, the energy function of SRBM can be expressed as

$$E(x,y,h) = \sum_{i\in n_x}\frac{(a_i - x_i)^2}{2\sigma_i^2} + \sum_{j\in n_h}\frac{(b_j - h_j)^2}{2\gamma_j^2} - \sum_{i\in n_x}\sum_{j\in n_h}\frac{x_i}{\sigma_i}\frac{h_j}{\gamma_j}w_{ij} + \sum_{I\in n_y}\frac{(c_I - y_I)^2}{2\beta_I^2} - \sum_{I\in n_y}\sum_{j\in n_h}\frac{y_I}{\beta_I}\frac{h_j}{\gamma_j}w_{Ij} \tag{9}$$

where c_I and β_I is the bias and standard deviations of quality unit y_I, respectively. w_{Ij} is the connection weight between quality variable unit y_I and the hidden layer unit h_j .Then, the CD algorithm is used to calculated the gradient of the parameters in each SRBM as follows

$$\frac{\partial \ln(P(v^t))}{\partial w_{ij}} = \Delta w_{ij} \approx \frac{v_m^{t(0)}}{\delta_m}\frac{h_j^{(0)}}{\gamma_j} - \frac{v_m^{t(k)}}{\delta_m}\frac{h_j^{(k)}}{\gamma_j} \tag{10}$$

$$\frac{\partial \ln(P(v^t))}{\partial w_{Ij}} = \Delta w_{Ij} \approx \frac{y_I^{t(0)}}{\beta_I}\frac{h_j^{(0)}}{\gamma_j} - \frac{y_I^{t(k)}}{\beta_I}\frac{h_j^{(k)}}{\gamma_j} \tag{11}$$

$$\frac{\partial \ln(P(v^t))}{\partial a_i} = \Delta a_i \approx \frac{v_i^{t(0)}}{\sigma_i^2} - \frac{v_i^{t(k)}}{\sigma_i^2} \tag{12}$$

$$\frac{\partial \ln(P(v^t))}{\partial c_I} = \Delta c_I \approx \frac{y_I^{t(0)}}{\beta_I^2} - \frac{y_I^{t(k)}}{\beta_I^2} \tag{13}$$

$$\frac{\partial \ln(P(v^t))}{\partial b_j} = \Delta b_j \approx \frac{h_j^{(0)}}{\gamma_j^2} - \frac{h_j^{(k)}}{\gamma_j^2} \tag{14}$$

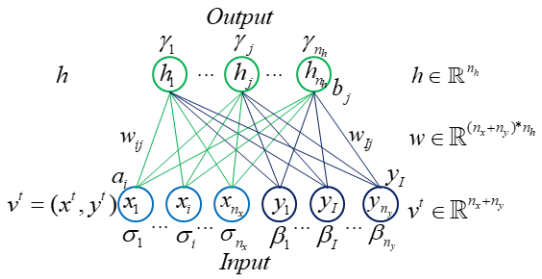

Fig. 3. The detailed structure of SRBM

3.2 Supervised Deep Belief Network

The SDBN is established by hierarchically stacking multiple SRBMs and then adding an output layer to the top hidden layer for the quality variables. The detailed structure of SDBN is shown in Fig. 4. Assume there are a total of L SRBM modules in the SDBN. The l^{th} SRBM is denoted as $\text{SRBM}_{[l]}$. Here, subscript $[l]$ is used to identify the terms with regard to the l^{th} SRBM. The SDBN model completes training through pre-training and fine-tuning processes. After pre-training, an output layer is added to the top hidden layer as shown in red circles in Fig. 4. Assume $w_{[L+1]}$ is the connecting weights from the top hidden layer to the output layer, and $b_{[L+1]}$ is the corresponding bias term. The predicted $\hat{y}^t$ of the quality value y^t can thus be calculated as follows

$$\hat{y}^t = Affine(w_{[L+1]}h_{[L]} + b_{[L+1]}) \tag{15}$$

where *Affine* is a mapping function that generates the predicted quality variable value in output layer. To obtain a good prediction model, the whole network can be fine-tuned by the BP algorithm.

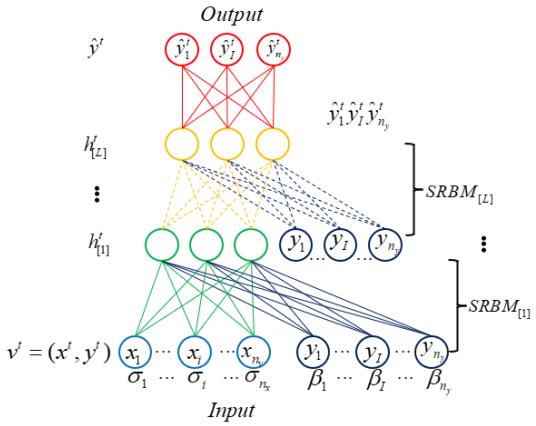

Fig. 4. The detailed structure of the SDBN

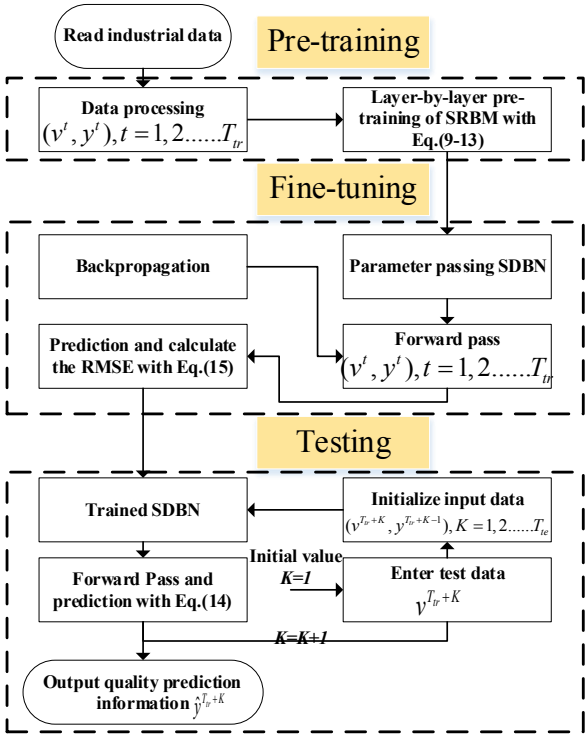

Fig. 5. The flowchart of the SDBN-based soft sensor model

By introducing the quality information to the visible layer in each SRBM, the proposed SDBN can learn quality-related features in the hidden layer, which can achieve an improvement in soft sensing. The SDBN-based soft sensor framework mainly consists of the pre-training, fine-tuning and testing steps, which are shown in Fig. 5.

The root mean squared error (RMSE) indicator is used to evaluate the performance of the proposed SDBN. The RMSE is calculated as follows

$$RMSE_{tr} = \sqrt{\sum_{t=1}^{T_{tr}} (y^t - \hat{y}^t)^2 / T_{tr}} \tag{16}$$

$$RMSE_{te} = \sqrt{\sum_{t=T_{tr}+1}^{T_{tr}+T_{te}} (y^t - \hat{y}^t)^2 / T_{te}} \tag{17}$$

where T_{tr} , T_{te} is the number of samples for the training and testing datasets, respectively.

4. CASE STUDIES

In this section, the proposed SDBN is applied to the sintering process of a steel manufacturing enterprise to validate its performance.

4.1 Sintering Process

Sintering is an important technique in the artificial iron-rich process for the blast furnace in modern steel-making industry. The quality of the sinter directly affects the production cost and steel production capacity since it is the main raw material for ironmaking. Therefore, real-time monitoring of sinter quality is of great significance. The sintering process is shown in Fig. 6 for its flowchart. The raw materials such as laterite ore and lime are first mixed and then fed into the sintering trolley for sintering, crushing, and cooling. Finally, the sintering products are sent to the blast furnace for smelting. In order to produce high quality molten iron in the blast furnace, it is important to control and monitor the chemical composition of the sinter. The FeO content is an important quality indicator in the sinter, which can reflect the production status of the sintering process.

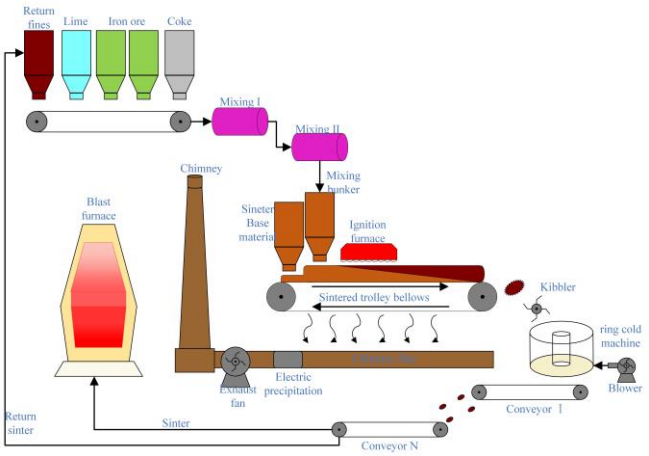

Fig.6. The flowchart of the sinter process

At present, the content of FeO is measured by offline lab analysis with very low sampling frequency. There are problems like large time lag and low sampling frequency in this manner. To alleviate these problems, soft sensor technology is used to estimate the FeO content in the sinter. Thus, the temperature and pressure of the main pipe and the branch pipe that are related to the quality variable are selected as the input variables for the soft sensor model. The data was collected from an industrial sintering process in China. The data samples were from January to October in 2018 with sampling frequency of 4 hours per sample. There were a total of 1400 data samples during this period. For model training and testing, the first 1000 samples are used as

the training dataset, and the last 400 samples are used as the testing dataset.

4.2 Results

To construct the soft sensor models, some parameters should be first determined. By trial-and-error technique, the SDBN is determined with a network structure of two hidden layers. Moreover, the number of hidden neurons is 72 for each hidden layer. Also, the iteration of model training is 100. In order to verify the performance of the proposed model, the original SAE and DBN-based models with the same structure are also established for comparison. The RMSE results of the training and testing datasets are shown in Table 1 for the three models. As can be seen, both SAE and DBN have similar prediction performance with large RMSE. SDBN can provide much better prediction accuracy than SAE and DBN. This is because both SAE and DBN are pretrained unsupervised without any reference to the quality data. However, SDBN uses the quality variable to guide quality-relevant feature learning in the pretraining of each SRBM. In this way, the prediction performance of the SDBN is greatly improved.

Table 1. RMSE of SAE, DBN and SDBN Model

RMSE	SAE	DBN	SDBN
Training dataset	0.1286	0.1204	0.0038
Testing dataset	0.1399	0.1381	0.0090

After pre-training, Fig. 7 shows the RMSE trend with the iteration in the fine-tuning procedure. Here, the number of iterations is set to 100 for model training. As can be seen, SAE and DBN have a fine-tuning convergent state in the range of [0.14, 0.15]. However, SDBN can achieve convergence with fine-tuning loss much smaller than 0.05. Furthermore, the detailed prediction results are shown for the training and testing datasets with the three models, which are shown in Figs. 8-10.

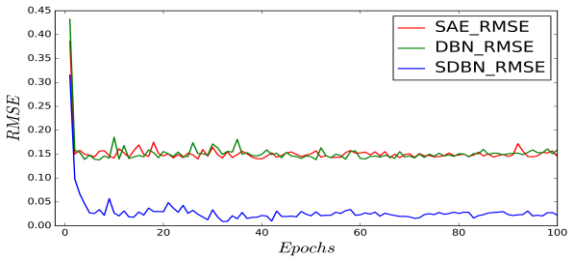

Fig. 7. The RMSE curve in the fine tuning

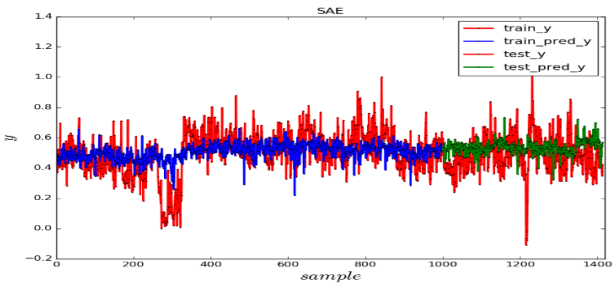

Fig. 8. The detailed predictions of SAE model

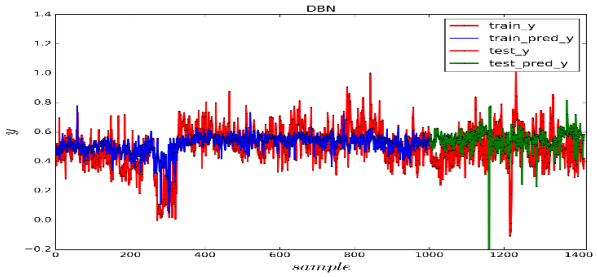

Fig. 9. The detailed predictions of DBN model

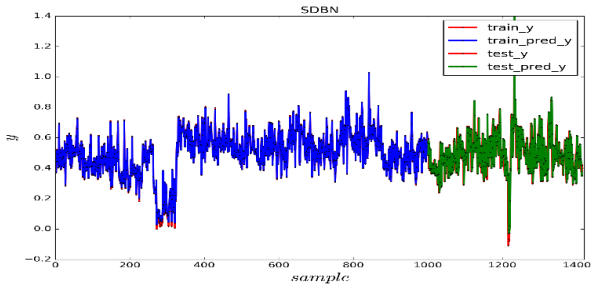

Fig. 10. The detailed predictions of SDBN model

It can be seen from the prediction curves of Fig. 8 and Fig. 9 that the predicted output values with SAE and DBN cannot track the true values well on both the training and testing datasets. The predicted errors are very large for most data points. After the quality variable is introduced to the visible layer of the RBM modules, the SDBN has excellent prediction performance on both datasets for the FeO content in sintered ore. It can be clearly seen from Fig. 10 that the predicted output values of SDBN can track the true quality values very well compared with SAE and DBN. There are only small deviations between the real and predicted quality value. Thus, SDBN is well suited for soft sensor modelling of complex industrial processes.

5. CONCLUSIONS

This paper proposes a new SDBN model for quality prediction, which is constructed with a hierarchical stack of multiple SRBMs. By introducing the quality variables to the visible layer of the original RBM module, the hidden layer can learn the features related to quality variables in SRBM. By applying the SDBN model for the prediction of FeO content to the sintering process, the SDBN model shows a significant improvement in prediction performance compared to the original DBN and SAE models. It provides a new solution for practical industrial soft sensor modelling.

REFERENCES

Brosilow, C. and M. Tong (1978). Inferential control of processes: Part II. The structure and dynamics of inferential control systems. *AIChE Journal*, 24, 492-500.

Chen, N., J. Dai, X. Yuan, W. Gui, W. Ren and H. N. Koivo (2018). Temperature Prediction Model for Roller Kiln by ALD-Based Double Locally Weighted Kernel Principal Component Regression. *IEEE Transactions on Instrumentation and Measurement*, 67, 2001-2010.

Chen, W., B. X. Wang, Y. Chen, H. J. Zhang and X. Li (2013a). Using BP Neural Network to Predict the Sinter

Comprehensive Performance: FeO and Sinter Yield. *Advanced Materials Research*, 771, 209-212.
Chen, W., H. J. Zhang, B. X. Wang, Y. Chen and X. Li (2013b). Prediction Model of the Sinter Comprehensive Performance Based on Neural Network. *Advanced Materials Research*, 753-755, 62-65.
Chen, X., X. Fan and J. Tao (2010). Operation Guidance System for Iron Ore Sintering Process. International Conference on Intelligent System Design & Engineering Application.
Dai, J., N. Chen, X. Yuan, W. Gui and L. Luo (2020). Temperature prediction for roller kiln based on hybrid first-principle model and data-driven MW-DLWKPCR model. *ISA Transactions*, 98, 403-417.
Ge, Z., Z. Song, L. Zhao and F. Gao (2014). Two-level PLS model for quality prediction of multiphase batch processes. *Chemometrics and Intelligent Laboratory Systems*, 130, 29-36.
Ge, Z., C. Yang and Z. Song (2009). Improved kernel PCA-based monitoring approach for nonlinear processes. *Chemical Engineering Science*, 64, 2245-2255.
Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural Comput.*, 14, 1771-1800.
Hinton, G. E. and R. R. Salakhutdinov (2006). Reducing the dimensionality of data with neural networks. *Science*, 313, 504-507.
Joseph, B. and C. B. Brosilow (1978). Inferential control of processes: Part I. Steady state analysis and design. *AIChE Journal*, 24, 485-492.
Kadlec, P., B. Gabrys and S. Strandt (2009). Data-driven soft sensors in the process industry. *Computers & Chemical Engineering*, 33, 795-814.
Rosipal, R. and L. J. Trejo (2002). Kernel partial least squares regression in reproducing kernel hilbert space. *The Journal of Machine Learning Research*, 2, 97-123.
Wan, Z., F. Z. Meng, A. Y. Hao and Y. L. Wang (2009). Optimization of the mixture design for alumina sintering with fuzzy ingredients. *Hunan Daxue Xuebao/journal of Hunan University Natural Sciences*, 36, 55-58.
Wang, J. and F. Qiao (2014). Cost and energy consumption collaborative optimization for sintering burdening in iron and steel enterprise. 2014 IEEE International Conference on Automation Science and Engineering (CASE).
Wang, P. (2006). Design and Application of Soft Sensor Object in Process Control. Sixth International Conference on Intelligent Systems Design and Applications.
Wang, Y., Z. Pan, X. Yuan, C. Yang and W. Gui (2020). A novel deep learning based fault diagnosis approach for chemical process with extended deep belief network. *ISA Transactions*, 96, 457-467.
Wang, Y., D. Wu and X. Yuan (2019). A two-layer ensemble learning framework for data-driven soft sensor of the diesel attributes in an industrial hydrocracking process. *Journal of Chemometrics*, 33, e3185.
Yuan, X., Z. Ge, B. Huang and Z. Song (2017a). A probabilistic just-in-time learning framework for soft sensor development with missing data. *IEEE Transactions on Control Systems Technology*, 25, 1124-1132.
Yuan, X., Z. Ge, B. Huang, Z. Song and Y. Wang (2017b). Semisupervised JITL Framework for Nonlinear Industrial Soft Sensing Based on Locally Semisupervised Weighted PCR. *IEEE Transactions on Industrial Informatics*, 13, 532-541.
Yuan, X., Y. Gu, Y. Wang, C. Yang and W. Gui (2019a). A deep supervised learning framework for data-driven soft sensor modeling of industrial processes. *IEEE Transactions on Neural Networks and Learning Systems*, DOI: 10.1109/TNNLS.2019.2957366.
Yuan, X., B. Huang, Y. Wang, C. Yang and W. Gui (2018a). Deep learning based feature representation and its application for soft sensor modeling with variable-wise weighted SAE. *IEEE Transactions on Industrial Informatics*, 14, 3235-3243.
Yuan, X., L. Li, Y. A. W. Shardt, Y. Wang and C. Yang (2020a). Deep learning with spatiotemporal attention-based LSTM for industrial soft sensor model development. *IEEE Transactions on Industrial Electronics*, DOI: 10.1109/TIE.2020.2984443.
Yuan, X., L. Li and Y. Wang (2020b). Nonlinear dynamic soft sensor modeling with supervised long short-term memory network. *IEEE Transactions on Industrial Informatics* 16, 3168-3176.
Yuan, X., L. Li, Y. Wang, C. Yang and W. Gui (2019b). Deep learning for quality prediction of nonlinear dynamic process with variable attention-based long short-term memory network. *The Canadian Journal of Chemical Engineering*, DOI: 10.1002/cjce.23665.
Yuan, X., C. Ou, Y. Wang, C. Yang and W. Gui (2019c). Deep quality-related feature extraction for soft sensing modeling: A deep learning approach with hybrid VW-SAE. *Neurocomputing*.
Yuan, X., C. Ou, Y. Wang, C. Yang and W. Gui (2019d). A layer-wise data augmentation strategy for deep learning networks and its soft sensor application in an industrial hydrocracking process. *IEEE Transactions on Neural Networks and Learning Systems*, DOI: 10.1109/TNNLS.2019.2951708.
Yuan, X., S. Qi and Y. Wang (2020c). Stacked Enhanced Auto-encoder for Data-driven Soft Sensing of Quality Variable. *IEEE Transactions on Instrumentation and Measurement*, DOI: 10.1109/TIM.2020.2985614
Yuan, X., Y. Wang, C. Yang, Z. Ge, Z. Song and W. Gui (2018b). Weighted linear dynamic system for feature representation and soft sensor application in nonlinear dynamic industrial processes. *IEEE Transactions on Industrial Electronics*, 65, 1508-1517.
Yuan, X., J. Zhou, B. Huang, Y. Wang, C. Yang and W. Gui (2020d). Hierarchical quality-relevant feature representation for soft sensor modeling: a novel deep learning strategy. *IEEE Transactions on Industrial Informatics*, 16, 3721-3730.
Zhang, J.-h., A.-g. Xie and F.-m. Shen (2007). Multi-Objective Optimization and Analysis Model of Sintering Process Based on BP Neural Network. *Journal of Iron and Steel Research, International*, 14, 1-5.
Zhang, J. H., F. M. Shen and A. G. Xie (2002). Application of G-BP in FeO content prediction during sintering. *Journal of Northeastern University*, 23, 1073-1075.