SCIENCE CHINA
Information Sciences

• LETTER •

May 2022, Vol. 65 159203:1–159203:3
https://doi.org/10.1007/s11432-020-2964-7

A novel deep quality-supervised regularized autoencoder model for quality-relevant fault detection

Zhichao LI, Li TIAN & Xuefeng YAN*

Key Laboratory of Advanced Control and Optimization for Chemical Processes of Ministry of Education, East China University of Science and Technology, Shanghai 200237*, China*

Received 29 February 2020/Revised 18 April 2020/Accepted 29 May 2020/Published online 15 March 2021

Citation Li Z C, Tian L, Yan X F. A novel deep quality-supervised regularized autoencoder model for quality-relevant fault detection. Sci China Inf Sci, 2022, 65(5): 159203, https://doi.org/10.1007/s11432-020-2964-7

Dear editor,

In the industrial processes, timely detection of key quality variables is very important for tracking the product quality, monitoring the process status, and achieving stable and reliable control. However, the key quality variables are difficult to measure or have obvious time delay. The process variables that are easy to measure are often used to establish a monitoring model to ensure the safety of production process and the stability of product quality. As reviewed by Ge [1] and Jiang et al. [2], traditional quality-relevant monitoring technologies mainly include methods based on principal component regression (PCR) and partial least squares, which usually perform linear transformations between variables. Therefore, the nonlinear correlation between the variables cannot be described.

Recently, deep learning technology has become the focus of academia and industry to effectively dealing with the complex non-linearity in industrial data [3]. Deep learning has a deeper network structure than shallow learning, so that more effective and deeper features can be obtained from the original features. Process monitoring algorithms based on various types of autoencoders (AEs) have been proposed [4–6]. However, although these algorithms can effectively detect abnormal conditions, it is unable to identify routine process faults and those that seriously affect product quality.

Actually, the features learned by traditional unsupervised models are good representations of the original input data. However, they may not be good representations for quality-relevant process monitoring. Good representations should be guided by measured and quality variables. Features learned from the measured variables should be highly correlated with the quality variables and can reconstruct the original input data well. To increase the generalization ability of the model, the learned features should be as irrelevant as possible. Besides, in order to prevent quality-relevant information from being lost, we also constrain the residuals to make them independent of quality. Therefore, this study proposes a novel deep quality-supervised regularized autoencoder (QS-RAE) model to further improve the quality-relevant monitoring performance.

QS-RAE. Each QS-RAE contains a non-linear encoding layer and a linear decoding layer. QS-RAE adds three new regularization terms to the loss function (mean squared error (MSE)) of traditional autoencoders. Therefore, the optimization objectives of the QS-RAE are as follows.

Objective 1. To maintain the global structure of the input data, QS-RAE must be able to reconstruct the input data well. Therefore, the first objective is to minimize the MSE as

$$L_1 = \min\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{m}(x_{ik}-\tilde{x}_{ik})\right), \tag{1}$$

where n is the number of samples and m is the dimension.

Objective 2. The features extracted from the input data should contain more quality-relevant information. Therefore, quality data are used to supervise the features extracted by the model during training phase. Assuming the extracted features are $\boldsymbol{H}=[\boldsymbol{h}_1,\ldots,\boldsymbol{h}_l]$, the absolute value of the correlation coefficient between the pth feature $\boldsymbol{h}_p$ and the quality variable is calculated as follows:

$$\begin{aligned}\rho_{\boldsymbol{h}_p,\boldsymbol{y}} &= \left|\frac{\mathrm{cov}(\boldsymbol{h}_p,\boldsymbol{y})}{\sigma_{\boldsymbol{h}_p}\sigma_{\boldsymbol{y}}}\right| \\ &= \left|\frac{E(\boldsymbol{h}_p\boldsymbol{y})-E(\boldsymbol{h}_p)E(\boldsymbol{y})}{\sqrt{E(\boldsymbol{h}_p^2)-E^2(\boldsymbol{h}_p)}\sqrt{E(\boldsymbol{y}^2)-E^2(\boldsymbol{y})}}\right|,\end{aligned} \tag{2}$$

where $E(\cdot)$ represents mean. Then, the correlation coefficient vector can be obtained as $\rho_{\boldsymbol{H},\boldsymbol{y}}=[\rho_{\boldsymbol{h}_1,\boldsymbol{y}},\ldots,\rho_{\boldsymbol{h}_l,\boldsymbol{y}}]^{\mathrm{T}}$. In order to ensure that each feature has a large correlation with the quality variable, this study achieves it by maximizing the minimum value in $\rho_{\boldsymbol{H},\boldsymbol{y}}$. Then the second optimization objective is expressed as

$$L_2 = \max(\min(\rho_{\boldsymbol{H},\boldsymbol{y}})) = \min(1-\min(\rho_{\boldsymbol{H},\boldsymbol{y}})). \tag{3}$$

* Corresponding author (email: xfyan@ecust.edu.cn)

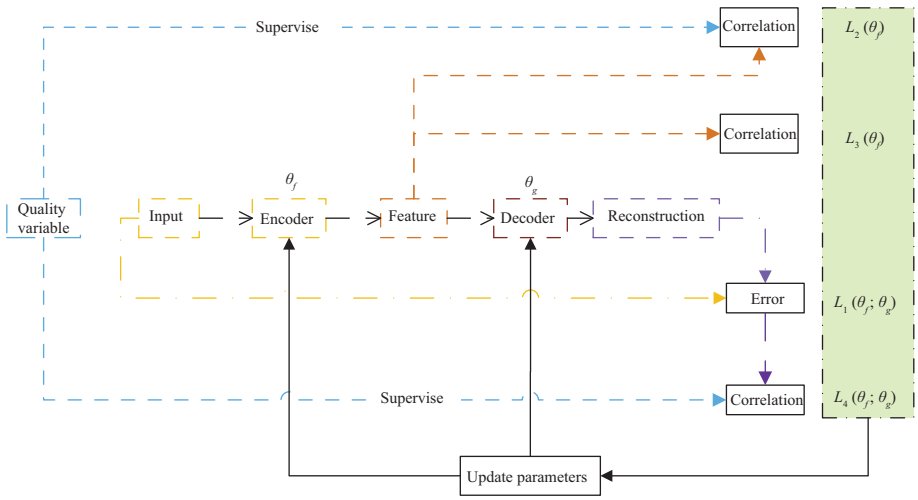

Figure 1 (Color online) Illustration of QS-RAE.

Objective 3. To make the model more robust, the correlation between the extracted features should be as small as possible. Similar to (2), the correlation coefficient between the pth feature and the qth feature, $\rho_{\boldsymbol{h}_p,\boldsymbol{h}_q}$, can be calculated. Then, the correlation coefficient matrix can be obtained as follows:

$$\boldsymbol{\rho}_{\text{within}H} = \begin{bmatrix} \rho_{\boldsymbol{h}_1,\boldsymbol{h}_1} & \cdots & \rho_{\boldsymbol{h}_1,\boldsymbol{h}_l} \\ \vdots & \ddots & \vdots \\ \rho_{\boldsymbol{h}_l,\boldsymbol{h}_1} & \cdots & \rho_{\boldsymbol{h}_l,\boldsymbol{h}_l} \end{bmatrix}. \tag{4}$$

The third optimization objective can be achieved by minimizing the maximum value in $\boldsymbol{\rho}_{\text{within}H}$ (except for elements in the main diagonal), which is expressed as

$$L_3 = \min(\max(\rho_{\boldsymbol{h}_p,\boldsymbol{h}_q})), \quad p \neq q. \tag{5}$$

Objective 4. To avoid the loss of quality-relevant information, this study also requires that the residuals are not correlated or have a small correlation with the quality variable. Assuming that the residual is $\text{Res} = \boldsymbol{X} - \tilde{\boldsymbol{X}}$, the correlation coefficient between the reconstruction error of the kth variable and the quality variable, $\rho_{\boldsymbol{r}_k,\boldsymbol{y}}$, is calculated similar to (2). Then, the correlation coefficient vector can be obtained as $\rho_{\text{Res},\boldsymbol{y}} = [\rho_{\boldsymbol{r}_1,\boldsymbol{y}},\ldots,\rho_{\boldsymbol{r}_m,\boldsymbol{y}}]^{\text{T}}$. The fourth optimization objective can be achieved by minimizing the maximum value in $\rho_{\text{Res},\boldsymbol{y}}$, which is expressed as

$$L_4 = \min(\max(\rho_{\text{Res},\boldsymbol{y}})). \tag{6}$$

Taking these optimization objectives into consideration, the loss function of QS-RAE is defined as

$$\begin{aligned} &L(\theta_f^*;\theta_g^*) \\ &= \min L_1(\theta_f,\theta_g) + \alpha L_2(\theta_f) + \beta L_3(\theta_f) + \gamma L_4(\theta_f,\theta_g), \end{aligned} \tag{7}$$

where α,β,γ are hyperparameters used to determine the importance of each optimization objective. Afterwards, the proposed QS-RAE model can be trained by a stochastic gradient descent algorithm, as shown in Figure 1. θ_f is the parameter of the encoder; θ_g is the parameter of the decoder.

By stacking multiple QS-RAEs and training the deep QS-RAE model using a layer-by-layer greedy training method, quality-irrelevant information is gradually reduced, while quality-relevant information is more concentrated in the extracted features.

Quality-relevant fault detection. Although the features extracted by deep QS-RAE contain much quality-relevant information, there is also interference from quality-irrelevant information. Therefore, this study further decomposes the feature space into quality-relevant subspace and quality-irrelevant subspace by orthogonal decomposition. According to the extracted features by deep QS-RAE and the quality variable, the following formula can be obtained based on the idea of PCR:

$$\begin{cases} \boldsymbol{y}_{\text{pre}} = \boldsymbol{H}\boldsymbol{P}, \\ \boldsymbol{P} = (\boldsymbol{H}^{\text{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\text{T}}\boldsymbol{y}. \end{cases} \tag{8}$$

Perform singular value decomposition on $\boldsymbol{P}\boldsymbol{P}^{\text{T}}$:

$$\boldsymbol{P}\boldsymbol{P}^{\text{T}} = \begin{bmatrix} \boldsymbol{P}_y & \boldsymbol{P}_o \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_y & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{P}_y^{\text{T}} \\ \boldsymbol{P}_o^{\text{T}} \end{bmatrix}. \tag{9}$$

Then, the extracted features $\boldsymbol{H}$ can be decomposed into a quality-relevant subspace $\boldsymbol{H}_y = \boldsymbol{H}\boldsymbol{P}_y\boldsymbol{P}_y^{\text{T}}$ and a quality-irrelevant subspace $\boldsymbol{H}_o = \boldsymbol{H}\boldsymbol{P}_o\boldsymbol{P}_o^{\text{T}}$, respectively.

Considering quality-related subspace, quality-irrelevant subspace and residual space, quality-relevant and quality-irrelevant fault detection can be achieved by constructing three monitoring statistics, quality-relevant statisitc ($\boldsymbol{T}_{y,\text{new}}^2$) and quality-irrelevant statistics ($\boldsymbol{T}_{o,\text{new}}^2$ and SPE_{new}). The details can be found in Appendix A. Kernel density estimation is used to calculate the control limits for each monitoring statistic [4].

For online monitoring, if all the statistics are below the corresponding control limits, no fault occurs; if $\boldsymbol{T}_{y,\text{new}}^2$ is less than the control limit, and the other two statistics exceed the control limits, a quality-irrelevant fault occurs; once $\boldsymbol{T}_{y,\text{new}}^2$ exceeds the control limit, a quality-relevant fault occurs.

Case study. Experiments on the Tennessee-Eastman process [7] confirm the presented model's validity and superiority. The details of the experiment results on TE process and an ablation study are shown in Appendixes B and C, respectively.

Conclusion. The experimental results on TE process indicate that, compared with several state-of-the-art methods, the proposed deep QS-RAE not only performs better

in the detection of process faults, but also has higher reliability in the detection of quality-relevant faults. The ablation study indicates that the lack of any of the three regularization terms will have an impact on the reliability of quality-relevant fault detection, thus verifying the necessity for each regularization term.

Acknowledgements This work was supported by National Key Research and Development Program of China (Grant No. 2020YFA0908303) and National Natural Science Foundation of China (Grant No. 21878081).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Ge Z Q. Review on data-driven modeling and monitoring for plant-wide industrial processes. Chemometr Intell Lab Syst, 2017, 171: 16–25
2 Jiang Q C, Yan X F, Huang B. Review and perspectives of data-driven distributed monitoring for industrial plant-wide processes. Ind Eng Chem Res, 2019, 58: 12899–12912
3 Du B, Xiong W, Wu J, et al. Stacked convolutional denoising auto-encoders for feature representation. IEEE Trans Cybern, 2017, 47: 1017–1027
4 Zhang Z, Jiang T, Li S, et al. Automated feature learning for nonlinear process monitoring — an approach using stacked denoising autoencoder and k-nearest neighbor rule. J Process Control, 2018, 64: 49–61
5 Wang K, Forbes M G, Gopaluni B, et al. Systematic development of a new variational autoencoder model based on uncertain data for monitoring nonlinear processes. IEEE Access, 2019, 7: 22554–22565
6 Lv F, Wen C, Liu M. Representation learning based adaptive multimode process monitoring. Chemometr Intell Lab Syst, 2018, 181: 95–104
7 Downs J J, Vogel E F. A plant-wide industrial process control problem. Comput Chem Eng, 1993, 17: 245–255