Adversarial Autoencoder based fault diagnosis model for complex chemical processes

Kyojin Jang[1], Jonggeol Na[2], Minsu Kim[1], Seokyoung Hong[1] and Il Moon[1,]*

[1]Department of Chemical and Biomolecular Engineering, Yonsei University, Seoul 03722, Republic of Korea

[2]Clean Energy Research Center, KIST, Seoul 02791, Republic of Korea

Extended Abstract :

The scale of modern industrial systems, especially chemical processes, has become more and more complex. Timely detection of faults in these systems is critical to ensuring the safety of people and property and ensuring product quality. However, currently these methods are relying upon human operators who need to be aware of abnormal situations and make corrective decisions in a timely manner. Different operators may have different levels of safety awareness. Moreover, operators should handle a large amount of alarms. Therefore, there has been a need in the industry to develop an intelligent automated fault detection and diagnosis (FDD) system to assist operators in handling abnormal situations.

To this end, various methods have been developed over three decades. Among those methods, data-driven methods possess great potential to be applied to chemical processes since a large amount of data is collected and stored. Initially, principal component analysis (PCA), the most popular feature extraction method, has been widely used for monitoring linear processes. However, the data characteristics of many modern industrial processes are complicated, and the relationships among different variables are highly nonlinear, limiting the applicability of PCA. To overcome this, several nonlinear methods have been proposed. One of them is KPCA which involves data transformation from the low-dimensional nonlinear observation space into the high-dimensional linear feature space. However, standard kernel functions does not always guarantee good results and the performance is also very sensitive to some hyper-parameters.

Recently, deep learning (DL) has received a lot of attention, especially in process systems engineering, because of its high model flexibility. Among various DL-based methods, autoencoder provides good performance by achieving automated key features (latent variables) extraction. In particular, a method based on the stacked denoising auto-

encoder (SDAE) and variational autoencoder (VAE) were utilized to model the process data. However, the SDAE-based method lacked good model interpretability while the VAE-based model is not expressive enough to capture the true posterior distribution.

To address these issues, an algorithm for process monitoring based on adversarial autoencoder (AAE) has been developed. AAE is a generative autoencoder that uses generative adversarial networks (GAN) to impose an arbitrary prior distribution on the latent code. It uses the adversarial training to perform variational inference by matching the aggregated posterior of the hidden code vector of the autoencoder with an arbitrary prior distribution. Matching the aggregated posterior to the prior samples ensures that generating ones from any part of prior space results in meaningful samples. As a result, the decoder of the AAE learns a deep generative model that maps the imposed prior ones to the data distribution.

In VAE, loss consists of reconstruction error and KL regularization term and the algorithm updates the KL divergence term to impose the prior samples on the latent space. To backpropagate through the KL divergence, the form of the prior distribution is assumed to be Gaussian distribution. However, in AAE, KL term in the loss is replaced by an adversarial loss. Therefore, only sampling from the prior distribution is needed to induce the latent distribution and the performance is not affected by whatever the probability distribution function is.

Training an AAE has two phases as follows :

1. Reconstruction phase – update the encoder and decoder to minimize reconstruction error.

2. Regularization phase – update the discriminator to distinguish true prior sample from generated samples and update the generator to fool the discriminator.

After the AAE-based process model is developed, a latent code can be obtained and the key features can be extracted. To detect fault, the k-Nearest Neighbor(kNN) rule is applied to construct the monitoring statistics in a feature space. The control threshold is determined by calculating the kNN squared distance of normal data. When a new sample arrives, the sample data is mapped into a feature vector and the kNN distance is calculated. If it is smaller than the threshold, the sample is normal, otherwise, it is faulty.

The proposed AAE based fault diagnosis method including offline and online stages is described as follows :

1. Normalize process variables and divide the dataset into training set and validation set.

2. Design the architecture of AAE and train the model.

3. Finetuning some hyperparameters by minimizing the reconstruction error and GAN loss.

4. Obtain the feature spaces from the trained model and determine the control limit.

5. Online data is collected and preprocessed.

6. A new sample is input to the trained model and compare the index with the corresponding control limit.

7. Detect whether it is normal or faulty.

To demonstrate the feasibility and advantages of the proposed method, the benchmark Tennessee Eastman (TE) process is employed. TE process is based on simulation that contains 5 major unit operations and 53 variables for process monitoring. There are 21 different process faults including 16 known faults and 5 unknown faults. In this case study, the proposed method, and several comparative methods (PCA, KPCA, AE), are applied to detecting the faults of the process. In order to show the experiment result of the fault diagnosis, define the confusion matrix and fault diagnosis rate (FDR). Fault alarm rate (FAR) is also calculated by applying to fault free case.

For large magnitude faults 1, 2, 4, 6, 7, 8, 12, 13, 14, and 17, the performance of the AAE method equivalent to the other methods. All methods cannot detect some small fault types, such as 3, 9, and 15. The FDR of proposed method outperform that of the other three methods in faults 5, 10, 11, 16, 18, 20, and 21. Thus overall, the FDR of the proposed method is higher on average than that of the other three methods. In FAR, the proposed method gives significantly better results. Comparing computing time for each method, the time required to train the AEE model is a bit longer than others, which is cause by the iterative optimization process. Nonetheless, it can be acceptable because once the model is trained, model can detect fault for new samples in a very short time.

To extract fault features from practical industrial process, a novel AAE based fault detection algorithm is proposed. The AAE based method consists of VAE and GAN. The two algorithms complement each other and therefore the benefits are maximized. For detecting faults, the kNN rule is applied to feature spaces to construct the monitoring statistics. The application of the method on TE process shows outstanding performance, especially at the average fault diagnosis rate. The AAE model is particularly good for nonlinear, non-Gaussian systems. It automatically extracts features through a deep neural network, so can be applied to large-scale data. Unlike the past VAE model, it can handle prior samples which have non-Gaussian distribution.