Fault Detection and Diagnosis in Tennessee Eastman Process with Deep Autoencoder

Zhongying Xiao[1], Arthur Kordon[2] and Subrata Sen[3]

[1]*Microsoft, Redmond, WA 98052*
johnnyxiao@microsoft.com

[2]*Kordon Consulting, Fort Lauderdale, Florida 33308*
arthur@kordon-consulting.com

[3]*greyBOX Innovation LLC, Atlanta, GA 30152*
subrata@gbinnov.com

ABSTRACT

Data-driven modeling has been considered as an attractive approach for fault detection in chemical processes. Of special interest to industry are methods that represent nonlinear phenomena and detect complex faults. In this paper, a semi-supervised deep learning method - deep autoencoder for fault detection in Tennessee Eastman Process (TEP) is proposed. The TEP process is a simulated benchmark for evaluating process control and monitoring methods. The performance of the proposed method is evaluated and compared to Principal Component Analysis (PCA). The experimental results demonstrate that the proposed optimized five-layers DAE model for fault detection outperforms the standard PCA. Of special importance to real-world applications is its capability for automatic variable selection. In comparison to PCA it demonstrated higher prediction accuracy for most of the generated faults. Deep autoencoder has the potential to become an excellent approach for process monitoring and fault detection in chemical processes.

1. INTRODUCTION

Chemical processes have become more and more automated after deployment of advanced process control systems during the last three decades. Despite their benefits during production, there are still many tragic chemical process accidents, resulting in assets loss and environmental damages. Due to highly dynamic process and high-frequency records in these industrial systems, fault detection is far from being the default state in large-scale application. Data-driven process monitoring and fault detection is becoming one of the most active field in chemical process control (Chiang, Russell, & Braatz, 2000a; Qin & Chiang, 2019; Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003). Among them, multivariate statistical methods, such as Principal Component Analysis (PCA) (Kresta, Macgregor, & Marlin, 1991; Wise, Ricker, Veltkamp, & Kowalski, 1990), Partial Least Squares (PLS) (Khan, Moyne, & Tilbury, 2008; Kresta et al., 1991; Kruger & Dimitriadis, 2008; MacGregor, Jaeckle, Kiparissides, & Koutoudi, 1994),
Fisher Discriminant Analysis (FDA) (Chiang, Kotanchek, & Kordon, 2004; Chiang, Russell, & Braatz, 2000b; He, Qin, & Wang, 2005; Zhu & Song, 2011) have been extensively studied during last decades (Yin, Ding, Haghani, Hao, & Zhang, 2012). These statistical methods provide promising ways to detect faults at early stages of abnormality. Most of these methods, however, are limited by the assumption that fault data could be distinguished with linear transformations. Therefore, some non-linear relationships between variables and outcome cannot be well captured by these linear methods.

For statistical process monitoring, including fault detection, PCA is a widely used method in the chemical and petrochemical industry due to its simplicity, popularity and effectiveness (He & Wang, 2011; Joe Qin, 2003; Yin et al., 2012). PCA can be viewed as the linear projection of a data set to maximize the variance in the projected space. It can handle high dimensional, noisy, and highly correlated data generated from chemical processes and reduce dimensionality to a small number of principal components. In addition, PCA only requires the historical data of normal operation to build the fault detection model.

Although the PCA-based monitoring methods have been successfully applied in many applications, they have their limitations. For instance, PCA does not consider the probability density of the observed data. Also, the PCA-based process monitoring scheme assumes that the process behaves linearly, which limits its applicability for monitoring nonlinear processes. Although a special version of PCA - kernel PCA (KPCA) can deal with nonlinearity, it is difficult or even impossible for KPCA to find an inverse mapping function from the feature space to the original space (Lee, Yoo, Choi, Vanrolleghem, & Lee, 2004).

PLS is another popular multivariate statistical method and extensively used for model building, fault detection, and diagnosis. It uses an off-line trained correlation model and online process measurements to predict online key performance indicators of an industrial process. For the purpose of process monitoring, PLS can detect the faults which occurred in the process input by the use of the information contained in the input–output correlation. PLS extracts the correlation model from the process inputs and outputs for further prediction and fault diagnosis purposes. Unlike PCA, PLS inclines to discover the faults that occurred in process inputs, which might influence the key performance indicators. Recently, the applicability of PLS and its variants for process monitoring and fault detection have been comprehensively studied (Yin et al., 2012).

FDA is a linear dimensionality reduction technique, which is optimal in terms of maximizing the separation between several classes. It determines a set of projection vectors, ordered in terms of maximizing the scatter between the classes while minimizing the scatter within each class. When an additional class of data represents the normal operating conditions, FDA can also be applied on industrial processes for fault detection (Chiang et al., 2004).

An inherent limitation of traditional approaches is the assumption of Gaussian distribution of process data. Additional basic limitation of these methods is that the developed statistical models are based on a single layer of features and may not achieve the best monitoring and fault detection performance. Another class of fault detection method is based on non-linearity of features in data. For instance, Support Vector Machine (SVM) was applied to fault detection in industrial systems (Chiang et al., 2004; Kulkarni, Jayaraman, & Kulkarni, 2005; Mahadevan & Shah, 2009; Yélamos, Escudero, Graells, & Puigjaner, 2009). It can capture nonlinear features embedded in the data and detect complex faults which are similar to normal data, even though there are subtle nuances between the two classes.

Artificial neural network-based approach is another option for fault detection of nonlinear features, and different architectures of ANN have been explored. Recently, a classical neural net for classification has been successfully used for fault detection (Heo & Lee, 2018). A nonlinear autoregressive with exogenous input (NARX) neural network has been implemented for the detection of both internal and external faults in the distillation column for dynamic system monitoring and to predict the probability of failure (Taqvi, Tufa, Zabiri, Maulud, & Uddin, 2018). A different architecture, auto-associative neural network which is trained in an unsupervised fashion, is used in (Heo & Lee, 2019). It overcomes one of the key limitations in fault detection applications; that the neural networks are trained in a supervised manner assuming that the normal/fault labels were available.

Recently, deep learning has shown significant progress in its capabilities and has been utilized in diverse application areas such as, image and natural language processing (Goodfellow, Bengio, & Courville, 2016). Deep learning is an algorithm containing stacked neural network layers with linear transformation and non-linear activation, including restricted Boltzmann machines (RBM), convolutional, recursive, and pooling layers. In deep learning method, low level features such as edges are emphasized and transformed to a higher and more abstract level features (Goodfellow et al., 2016). With sufficient transformation and activation, giant functions aiming at specific tasks are learned and optimized based on backpropagation. The key advantage of this method is that it automatically discovers features with gradually increased complexity. With the rapid development of powerful graphics cards and deep learning frameworks, deep learning has become a viable alternative for potential industrial applications. Recently, there is a growing interest in exploring deep learning for fault detection and diagnosis of chemical processes. A hierarchical deep neural network (HDNN) (Xie & Bai, 2016), a deep belief network (DBN) (Zhang & Zhao, 2017), a Deep convolutional neural network (CNN) model were proposed for diagnosing the faults on the TE process (Cheng, He, & Zhao, 2019; Wu & Zhao, 2018). However, these methods still require tedious variable selection and models with complex architecture, which will constrain their application in real-time process monitoring.

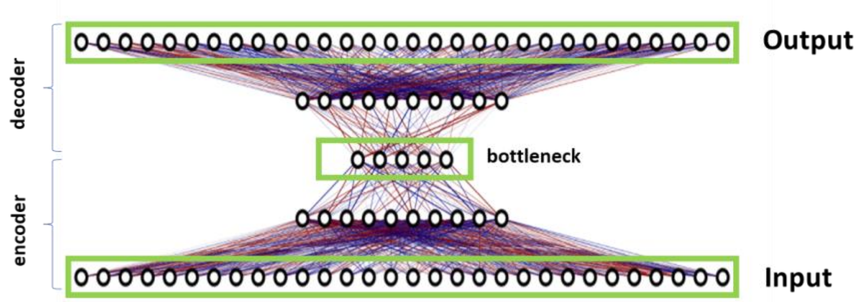

Fig. 1: Structure of a Deep Autoencoder

Despite the progress of the above two categories of data-driven methods, fault detection is still far from widely used in practical applications due to three major issues. First, these methods require large amount of labelled data to train a well-performed model. Second, these methods often require significant amount of domain expertise for variable selection and model validation. Third, the imbalance between normal and fault data makes model development process a real challenge.

This paper proposes a deep learning neural network structure, called Deep Autoencoder (DAE) algorithm, to detect faults without tedious feature selection. The proposed DAE framework is trained based on time series data in normal process condition without manual variable selection. The article demonstrates the model performance of the DAE through testing it for detecting different types of faults in Tennessee Eastman Process (TEP). To compare DAE with traditional statistical models, PCA method is used as a benchmark method.

2. DEEP AUTOENCODER

Autoencoder is a type of neural network which is adopted to copy significant information of its input to its output (Fig.1). The idea of autoencoders has been a vital part of neural networks for decades (Kramer, 1991). Historically, autoencoders have been used to de-noise signals, extract features and reduce dimensionality (Goodfellow et al., 2016; Hinton & Salakhutdinov, 2006). DAE has been deployed as anomaly detection method, such as monitoring vibration data (Qi et al., 2017; Qu, He, Deutsch, & He, 2017; Reddy, Sarkar, Venugopalan, & Giering, 2016) and telemetry data (Sakurada & Yairi, 2014; Zhao, Meng, Zeng, & Qi, 2017). There are several autoencoder applications to classify faults in chemical process as well (Cheng et al., 2019; Jiang, Ge, & Song, 2017). As an unsupervised learning method, DAE consists of three components: an input layer, single or multiple hidden layers, and an output layer. At the middle of the structure is a bottleneck layer where the information of data is most concentrated and represented. Each layer can have different number of neurons. In DAE, the input vector $x \in R^d$ is mapped into a hidden layer h, by a linear transformation $z = Wx + b$, followed by a nonlinear activation $h = f(z)$. W is the weight matrix and b is the bias, $f(.)$ is the activation function. Some of the common activation functions include sigmoid function, tanh function, Rectified Linear Units (ReLU) and their derivatives. The encoder is mapped reversely to reconstruct the input vector x by another process, with $y = f(W^T h + b^T)$, where W^T, b^T stand for transposed matrix of W and b, respectively. We use the same weight to encode the input vector and decode the hidden representation. The learning process is to minimize the loss function

$$J(W, b) = \frac{1}{N} \sum_{n=1}^{N} \|x - y\|^2$$

The parameters W, b are optimized via backpropagation to minimize the loss function. Gradient descent optimization algorithms are the most common ways to optimize neural networks. In this paper, Adaptive Moment Estimation

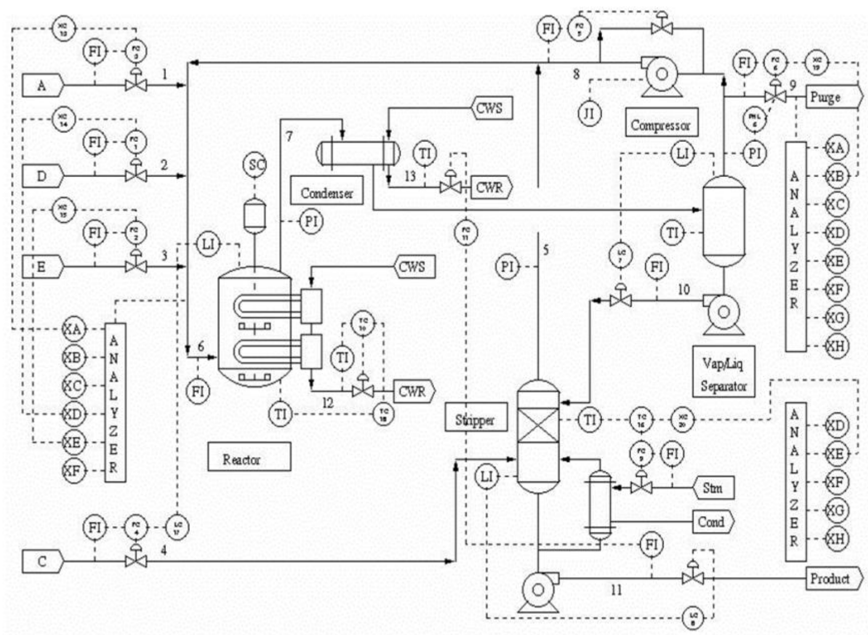

Fig. 2. The Tennessee Eastman process diagram

(Adam) gradient descent optimization algorithm - is used to optimize the deep neural network.

In complex industrial systems, the relationships between predicting variables and outcome are intended to be nonlinear. The statistical methods, such as PCA, PLS, etc., can only transform raw signals linearly but cannot capture nonlinear relationships. In this case, nonlinear features must be approximated by linear methods, which could result in inaccurate feature selections, especially in difficult fault detection scenario, such as, Fault 5 in TEP. However, nonlinear neural networks can overcome these difficulties. In DAE, there are two steps of transformation between two layers. The first step is linear multiplication, which is very similar to PCA and PLS methods. The second step is nonlinear activation, with sigmoid function, PReLU or ReLU, to generate nonlinear features in deeper layer, which are optimized by back-propagation.

3. TENNESSEE EASTMAN PROCESS

TEP model is a realistic simulation program of a chemical plant which is recognized as a benchmark for process control and fault detection studies. The process is described in (Downs & Vogel, 1993) and the MATLAB code for process simulation is available over the website (https://depts.washington.edu/control/LARRY/TE/download.html). The system, shown in Fig. 2, consists of five major units, i.e., reactor, condenser, compressor, separator and stripper. The process generates two products from four reactants. In addition, an inert and a by-product are also present making a total of 8 components denoted as A, B, C, D, E, F, G and H. The gaseous reactants A, C, D, and E and the inert B are fed to the reactor where the liquid products G and H are formed. The reactions in the reactor are irreversible, exothermic, and approximately first-order with respect to the reactant concentrations. The reactor product stream is cooled through a condenser and then fed to a vapor-liquid separator. The vapor exiting the separator is recycled to the reactor feed through a compressor. A portion of the recycle stream is purged to keep the inert and by-products from accumulating in the process. The condensed components from the separator (Stream 10) are pumped to the stripper. Stream 4 is used to strip the remaining reactants in Stream 10 and is combined with the recycle stream. The products G and H exiting the base of the stripper are sent to a downstream process which is not included in this process.

To investigate the ability of DAE for fault detection in this chemical process, the TEP simulator was used to generate 21 classes of faulty data, which correspond to Faults **1-21** specified by the TEP (Table 1). For each faulty case, two sets of data were generated. The training data containing only normal operation data were used to build the models and the test data containing both normal and faulty operations data were used for model validation. Both training and test data contain 960 observations. In test data, the first 160 observations were based on normal operation and the corresponding faults occurred after the 161st observation. Each dataset contains 52 process variables.

Table 1. Process faults for the Tennessee Eastman process

Fault Number	Description	Type
1	A/C feed ratio, B composition constant	Step
2	B composition, A/C ration constant	Step
3	D feed temperature	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	A feed loss	Step
7	C header pressure loss-reduced availability	Step
8	A, B, and C feed composition	Random variation
9	D feed temperature	Random variation
10	C feed temperature	Random variation
11	Reactor cooling water inlet temperature	Random variation
12	Condenser cooling water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16	Unknown	Unknown
17	Unknown	Unknown
18	Unknown	Unknown
19	Unknown	Unknown
20	Unknown	Unknown
21	The valve fixed at steady state position	Constant position

4. RESULTS AND DISCUSSION

4.1 DAE Model Architectures

It is a real challenge to find an optimal architecture for the deep autoencoder. Most architecture is problem-dependent and based on the data structure. To find a proper architecture, we have tuned several models with various number of layers, neurons, and different activation functions. Several activation functions were tested in this study with the best performance of Parametric Rectified Linear Units (PReLU). On top of the selected activation functions, a series of architectures with the most outstanding fault detection performance are displayed in Fig. 4.

In process data streaming, the sampled data point is highly correlated with nearby data point, therefore, the temporal relationship and variations should not be neglected. Considering time relationship between the data points of

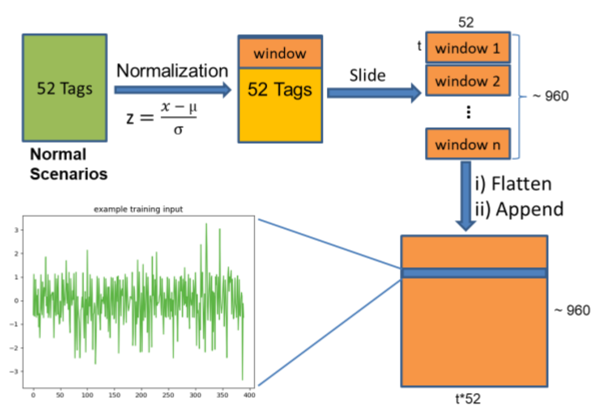

Fig. 3. Data Preprocessing

process data, dynamic deep autoencoder model was introduced by using dynamic time-variable matrix with t*m dimensions (t is time span, m is the number of variables). After concatenating time span from all sensors, the length of a single input vector is t*m. The total training dataset contains 960 data points. Dynamic deep autoencoder is a great way to extract the features of process data from both spatial and temporal domains. (Fig. 3)

Several architectures were explored and evaluated with the prepared dataset. With PReLU as activation function and MSE as loss function, model performance was evaluated by changing number of layers and number of moving windows. For small-sized dataset, complex neural networks with very deep layers and large number of neurons will cause severe over-fitting issue. The validation error is significantly higher than training error. The best way to narrow the gap is to reduce number of layers and neurons in each layers. The optimized architecture has 5 neural layers and slide window with 3 data points, resulting in 156 neurons at the input layer. This architecture generated an excellent model with very low training and test errors, which shows a low bias and variance. Therefore, this DAE structure was selected to train and test the explored datasets.

4.2 Automatic Variable Selection

Unlike other machine learning methods, the explored deep autoencoder does not need additional variable selection based on domain knowledge or statistical methods, such as stepwise regression, ridge regression, and mutual information. For regression, some of the popular variable selection methods include, forward selection, backward selection, PLS, mutual information, etc. All of them require tedious work and detailed statistical knowledge to select a set of good predictors.

However, deep autoencoder was trained by normal operation scenarios, the output is trying to preserve the information of the input, by minimizing the reconstruction error during model training. Individual-variable reconstruction errors at the output layer are also minimized in normal operation scenarios. In a faulty process, variables leading to or affected by faults would show huge differences compared with normal scenarios. When trained DAE model was mapped into data with faulty scenarios, these highly related variables would show large reconstruction errors relative to the other unrelated variables.

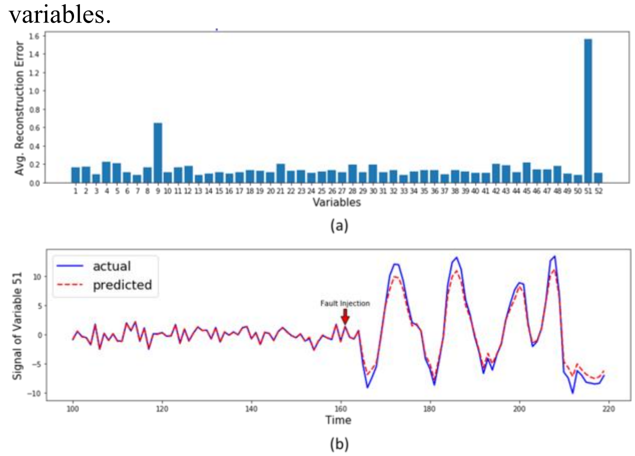

Fig. 4. Automatic Variable Selection of Fault 11. (a) Important Variable Selection; (b) Comparison of actual data and predicted data for Variable 51

An example of automatic variable selection for Fault 11, based on the reconstruction errors of all input variables for Fault 11 is shown in Fig. 4(a). Clearly, two spikes of high reconstruction errors are displayed for Variable 9 and 51, while the other variables have relatively small reconstruction errors. The signal of these two variables

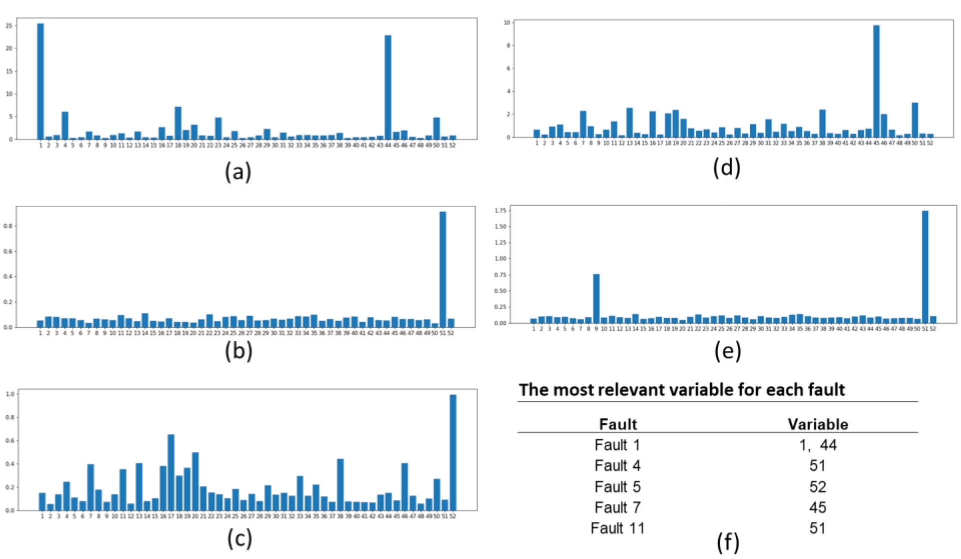

Fig 5. Reconstruction Errors of 52 variables for 5 Faults: (a) Fault 1; (b) Fault 4; (c) Fault 5; (d) Fault 7; (e) Fault 11; (f) most relevant variable for each fault

demonstrated significant changes after Fault 11 has occurred. Trained DAE could not capture enough features from Variable 51, therefore, the predicted values (grey line) have large differences with actual values (green line) after Fault 11 was injected, resulting in a large reconstruction error for Variable 51 (Fig. 4(b)). Fig. 5 shows important variables for Fault 1, 4, 5, 7 and 11 selected by DAE, which is consistent with published literatures (Chiang et al., 2004, 2000a; Downs & Vogel, 1993). Automatic variable selection, based on reconstruction errors, is a major advantage of DAE compared with the other methods for fault detection. First, all variables can be used in the training and test stages to generate a robust model. Hand-crafting variable selection process is not needed. Second, selected important variables with DAE model provide very useful information for root-cause analysis of the faults, especially in real-time process analytics.

Another advantage of DAE is the non-linear relationships between predictors and outcome represented by this method, which is aligned with the reality in complex industrial systems. As a result, it is assumed that the DAE can detect differences between normal and fault scenarios with much higher accuracy as compared to corresponding linear approaches. The results from a performance comparison between DAE and PCA for all 21 faults are given in this section. Two generally used metrics, fault detection rate (FDR) and false alarm rate (FAR), are evaluated here for fault detection performance. High FDR and low FAR are two pre-requisites for fault detection methods. For PCA, 9 PCs were selected. Based on PCA, loading matrix of normal scenarios was obtained. Applying loading matrix to test dataset, *Hotelling's* T^2 and Squared Prediction Error (SPE) were calculated as benchmarks for fault detection.

With the same training and test dataset, the accuracy of

Table 2. Fault Detection Rates of The Three Methods (%)

Type	**Controllable Faults**			**Back to control Faults**			**Uncontrollable Faults**														
Faults	**3**	**9**	**15**	**4**	**5**	**7**	**1**	**2**	**6**	**8**	**10**	**11**	**12**	**13**	**14**	**16**	**17**	**18**	**19**	**20**	**21**
DAE	3.6	3.5	7.9	100	100	100	100	99	100	98	78	79	99	96	100	77	98	91	76	68	43
T^2	5.9	5.6	5.8	18	29	45	99	99	99	97	42	33	98	94	85	24	78	90	3.8	39	38
SPE	7.6	5.6	5.9	100	31	100	100	99	100	98	37	77	98	96	100	32	94	91	35	53	50

Table 3. False Alarm Rates of The Three Methods (%)

Type	**Controllable Faults**			**Back to control faults**			**Uncontrollable faults**														
Faults	**3**	**9**	**15**	**4**	**5**	**7**	**1**	**2**	**6**	**8**	**10**	**11**	**12**	**13**	**14**	**16**	**17**	**18**	**19**	**20**	**21**
DAE	1.3	5	1.9	3.1	3.1	1.3	2.5	3.8	0.6	0.6	1.9	5	6.9	0.6	0.6	7.5	1.3	0.6	0.6	1.9	2.5
T^2	2.5	7.5	0.6	1.9	1.9	0.6	3.1	1.3	1.3	1.9	1.9	3.8	3.1	0	1.3	14	1.9	3.1	1.3	0	1.3
SPE	5	5.6	1.9	5	6.3	3.1	6.9	5.6	0.6	3.8	1.9	5.6	5.6	5.6	5	6.9	3.8	7.5	4.4	6.9	8.1

Table 4. Fault Detection Delays of The Three Methods (min)

Type	**Controllable Faults**			**Back to control faults**			**Uncontrollable faults**														
Faults	**3**	**9**	**15**	**4**	**5**	**7**	**1**	**2**	**6**	**8**	**10**	**11**	**12**	**13**	**14**	**16**	**17**	**18**	**19**	**20**	**21**
DAE	159	15	213	0	0	0	0	12	0	24	51	15	6	27	0	27	51	69	0	216	819
T^2	42	6	279	0	0	0	21	33	24	27	15	18	6	138	3	3	0	42	30	21	63
SPE	60	0	231	0	0	0	6	36	0	51	54	15	6	108	0	33	72	0	30	33	36

4.3 Higher Prediction Accuracy

fault detection of the DAE with the optimized architecture is evaluated. Table 2 shows the Fault Detection Rate (FDR) of the three different methods. Table 3 shows the False Alarm Rate (FAR) of the three different methods. Apparently, DAE based method generated better results with much higher FDR and lower FAR for most faults. For controllable faults which are hard-to-detect (i.e. Fault 3, 9 and 15), none of these three methods can produce satisfactory results. For back-to-control faults (i.e. Fault 4, 5 and 7), DAE based method can generate 100% FDR along with low FARs, which performs obviously better than T^2 and SPE. For the rest of the uncontrollable faults, DAE generates higher FDRs and lower FARs than traditional linear methods. Among them, DAE based methods to detect Fault 1, 2, 6, 8, 12, 13, 14, 17 and 18 produce 90% or higher FDR with 3% or lower FAR, which overwhelmingly outperforms T^2 and SPE. For Fault 10, 11, 16, 19, 20 and 21, none of the methods can produce good results, even though DAE based method has significant improvement of FDR and FAR, compared with T^2 and SPE.

In industrial practice, fault detection delay is an important issue that we need to consider. Fewer delay means faster fault detection once the fault has happened, which could save significant time to proactively fix the faults and therefore, prevent system failure. Table 4 listed the corresponding Fault Detection Delays (FDD) of these three methods. For controllable faults (Fault 3, 9, 15), DAE has longer FDD compared with T^2 and SPE. It is likely that DAE is insensitive towards signals of controllable faults and cannot detect faults at initial stage. For back to control faults (Fault 4, 5 ,7), all the three methods can detect fault signal immediately when fault occurs and have no fault detection delay issue. For most of uncontrollable faults, DAE has much shorter FDD time than T^2 and SPE.

$$FDR = \frac{No.of\ True\ Faults}{No.of\ Total\ Faults} \times 100$$

$$FAR = \frac{No.of\ False\ Alarms}{No.of\ Total\ Normality} \times 100$$

In Fig. 6, three methods were utilized to monitor the process of Fault 5 which was injected at sample 161. Both Hotelling's T^2 and SPE could only detect errors at early stages and their statistics became similar with normal scenarios at later stages after sample 350-400. Their FDRs are 26.6% and 31.0%, respectively. Most important variables behaved similarly to those of normal scenarios-they returned to their set-points at latter stage of the fault. With DAE, however, we can conduct much more accurate process monitoring for Fault 5. Due to non-linear transformation of deep neural network, DAE can preserve more detailed features from data and detect deviations from trained data with higher sensitivity when fault occurs. For Fault 5, the misclassification rate is 0. Area under curve of receiver operating characteristic (ROC) test is very close to 1, indicating very strong robustness of the fault detection.

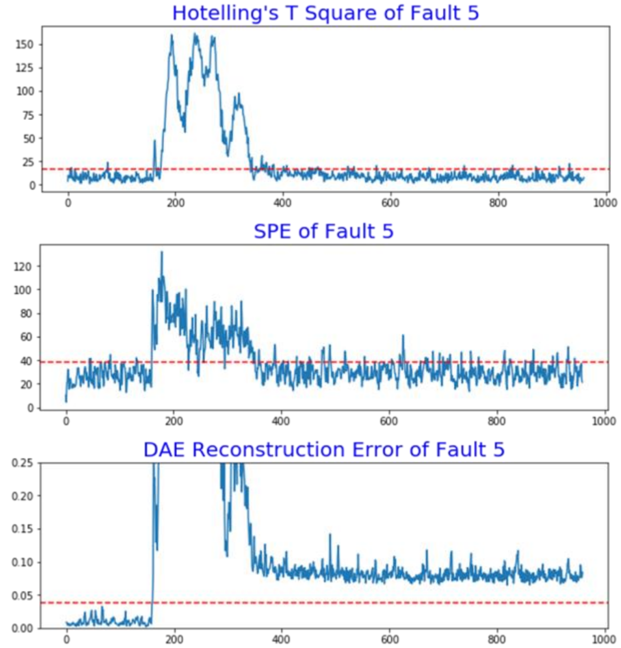

Fig. 6. Process Monitoring with Hotelling's T^2, SPE and DAE in case of Fault 5

5. CONCLUSIONS

An important branch of deep learning neural networks - deep autoencoder has been studied for fault detection in Tennessee Eastman Process benchmark. The performance of an optimized five-layers DAE model for fault detection of all TEP-generated faults is compared with an established linear method, PCA. A big advantage of the proposed DAE is the automatic variable selection it provides, based on reconstruction errors. The method provides superior results with automatic variable selection and higher fault detection rate. Without tedious variable selection before training process, DAE simplified the modeling procedures by detecting all the variables, which is suitable for monitoring large industrial systems. Furthermore, important variables, selected by DAE algorithm, is a vital information for root-cause analysis of the faults by engineers and data analysts. Compared with linear PCA method, nonlinear transformation of features embedded in the dataset by DAE can capture more useful information when fault occurs, resulting in a higher fault detection rate. The higher rates have been demonstrated for most of the explored faults. Despite the advantages of our proposed method, DAE can only be applied to steady processes. To adopt DAE into dynamic industrial systems with thousands of variables is a formidable challenge. The next step will be focusing on designing a proper DAE architecture for a real-world application.

ACKNOWLEDGMENTS

Sincere acknowledgements go to Dr. N. Lawrence Ricker in making the dataset available for research, Shakir Ali with meaningful discussion of TEP dataset, Dr. Guoyi Li with manuscript revision suggestions.

REFERENCES

Cheng, F., He, Q. P., & Zhao, J. (2019). *A novel process monitoring approach based on variational recurrent autoencoder*. *129*. https://doi.org/10.1016/j.compchemeng.2019.106515

Chiang, L. H., Kotanchek, M. E., & Kordon, A. K. (2004). Fault diagnosis based on Fisher discriminant analysis and support vector machines. *Computers and Chemical Engineering*. https://doi.org/10.1016/j.compchemeng.2003.10.002

Chiang, L. H., Russell, E. L., & Braatz, R. D. (2000a). *Fault detection and diagnosis in industrial systems*. Springer Science & Business Media.

Chiang, L. H., Russell, E. L., & Braatz, R. D. (2000b). Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *50*(2), 243–252. https://doi.org/https://doi.org/10.1016/S0169-7439(99)00061-1

Downs, J. J., & Vogel, E. F. (1993). A plant-wide industrial process control problem. *Computers & Chemical Engineering*, *17*(3), 245–255. https://doi.org/https://doi.org/10.1016/0098-1354(93)80018-I

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

He, Q. P., Qin, S. J., & Wang, J. (2005). A new fault diagnosis method using fault directions in Fisher discriminant analysis. *AIChE Journal*, *51*(2), 555–571. https://doi.org/10.1002/aic.10325

He, Q. P., & Wang, J. (2011). Statistics pattern analysis: A new process monitoring framework and its application to semiconductor batch processes. *AIChE Journal*, *57*(1), 107–121. https://doi.org/10.1002/aic.12247

Heo, S., & Lee, J. H. (2018). Fault detection and classification using artificial neural networks. *IFAC-PapersOnLine*, *51*(18), 470–475. https://doi.org/https://doi.org/10.1016/j.ifacol.2018.09.380

Heo, S., & Lee, J. H. (2019). Statistical Process Monitoring of the Tennessee Eastman Process Using Parallel Autoassociative Neural Networks and a Large Dataset. *Processes*, *7*(7). https://doi.org/10.3390/pr7070411

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science*, *313*(5786), 504 LP – 507. https://doi.org/10.1126/science.1127647

Jiang, L., Ge, Z., & Song, Z. (2017). Semi-supervised fault classi fi cation based on dynamic Sparse Stacked auto-encoders model. *Chemometrics and Intelligent Laboratory Systems*, *168*(June), 72–83. https://doi.org/10.1016/j.chemolab.2017.06.010

Joe Qin, S. (2003). Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, *17*(8-9), 480–502. https://doi.org/doi:10.1002/cem.800

Khan, A. A., Moyne, J. R., & Tilbury, D. M. (2008). Virtual metrology and feedback control for semiconductor manufacturing processes using recursive partial least squares. *Journal of Process Control*, *18*(10), 961–974. https://doi.org/https://doi.org/10.1016/j.jprocont.2008.04.014

Kramer, M. A. (1991). *Nonlinear Principal Component Analysis Using Autoassociative Neural Networks*. *37*(2), 233–243.

Kresta, J. V, Macgregor, J. F., & Marlin, T. E. (1991). Multivariate statistical monitoring of process operating performance. *The Canadian Journal of Chemical Engineering*, *69*(1), 35–47. https://doi.org/10.1002/cjce.5450690105

Kruger, U., & Dimitriadis, G. (2008). Diagnosis of process faults in chemical systems using a local partial least squares approach. *AIChE Journal*, *54*(10), 2581–2596. https://doi.org/10.1002/aic.11576

Kulkarni, A., Jayaraman, V. K., & Kulkarni, B. D. (2005). Knowledge incorporated support vector machines to detect faults in Tennessee Eastman Process. *Computers & Chemical Engineering*, *29*(10), 2128–2133. https://doi.org/https://doi.org/10.1016/j.compchemeng.2005.06.006

Lee, J.-M., Yoo, C., Choi, S. W., Vanrolleghem, P. A., & Lee, I.-B. (2004). Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, *59*(1), 223–234. https://doi.org/https://doi.org/10.1016/j.ces.2003.09.012

MacGregor, J. F., Jaeckle, C., Kiparissides, C., & Koutoudi, M. (1994). Process monitoring and diagnosis by multiblock PLS methods. *AIChE Journal*, *40*(5), 826–838. https://doi.org/10.1002/aic.690400509

Mahadevan, S., & Shah, S. L. (2009). Fault detection and diagnosis in process data using one-class support vector machines. *Journal of Process Control*, *19*(10), 1627–1639. https://doi.org/https://doi.org/10.1016/j.jprocont.2009.07.011

Qi, Y., Shen, C., Wang, D., Shi, J., Jiang, X., & Zhu, Z. (2017). Stacked Sparse Autoencoder-Based Deep

Network for Fault Diagnosis of Rotating Machinery. *IEEE Access*, *5*, 15066–15079. https://doi.org/10.1109/ACCESS.2017.2728010

Qin, S. J., & Chiang, L. H. (2019). *Advances and opportunities in machine learning for process data analytics*. *126*, 465–473. https://doi.org/10.1016/j.compchemeng.2019.04.003

Qu, Y., He, M., Deutsch, J., & He, D. (2017). Detection of Pitting in Gears Using a Deep Sparse Autoencoder. *Applied Sciences*, *7*(5). https://doi.org/10.3390/app7050515

Reddy, K. K., Sarkar, S., Venugopalan, V., & Giering, M. (2016). Anomaly Detection and Fault Disambiguation in Large Flight Data: A Multi-modal Deep Auto-encoder Approach. *Phm*, (i), 1–8. https://doi.org/10.1039/c0Ob00047g

Sakurada, M., & Yairi, T. (2014). Anomaly Detection Using Autoencoders with Nonlinear Dimensionality Reduction. *Proceedings of the MLSDA 2014 2Nd Workshop on Machine Learning for Sensory Data Analysis*, 4:4--4:11. https://doi.org/10.1145/2689746.2689747

Taqvi, S. A., Tufa, L. D., Zabiri, H., Maulud, A. S., & Uddin, F. (2018). Fault detection in distillation column using NARX neural network. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-018-3658-z

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., & Yin, K. (2003). A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering*, *27*(3), 327–346. https://doi.org/https://doi.org/10.1016/S0098-1354(02)00162-X

Wise, B. M., Ricker, N. L., Veltkamp, D. F., & Kowalski, B. R. (1990). Theoretical basis for the use of principal component models for monitoring multivariate processes. *Process Control and Quality*, *1*(1), 41–51.

Wu, H., & Zhao, J. (2018). Deep convolutional neural network model based chemical process fault diagnosis. *Computers and Chemical Engineering*, *115*, 185–197. https://doi.org/10.1016/j.compchemeng.2018.04.009

Xie, D., & Bai, L. (2016). A hierarchical deep neural network for fault diagnosis on Tennessee-Eastman process. *Proceedings - 2015 IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015*, 745–748. https://doi.org/10.1109/ICMLA.2015.208

Yélamos, I., Escudero, G., Graells, M., & Puigjaner, L. (2009). Performance assessment of a novel fault diagnosis system based on support vector machines. *Computers & Chemical Engineering*, *33*(1), 244–255. https://doi.org/https://doi.org/10.1016/j.compchemeng.2008.08.008

Yin, S., Ding, S. X., Haghani, A., Hao, H., & Zhang, P. (2012). A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *Journal of Process Control*, *22*(9), 1567–1581. https://doi.org/10.1016/j.jprocont.2012.06.009

Zhang, Z., & Zhao, J. (2017). A deep belief network based fault diagnosis model for complex chemical processes. *Computers and Chemical Engineering*, *107*, 395–407. https://doi.org/10.1016/j.compchemeng.2017.02.041

Zhao, W., Meng, Q. H., Zeng, M., & Qi, P. F. (2017). Stacked sparse auto-encoders (SSAE) based electronic nose for chinese liquors classification. *Sensors (Switzerland)*, *17*(12). https://doi.org/10.3390/s17122855

Zhu, Z.-B., & Song, Z.-H. (2011). A novel fault diagnosis system using pattern classification on kernel FDA subspace. *Expert Systems with Applications*, *38*(6), 6895–6905. https://doi.org/https://doi.org/10.1016/j.eswa.2010.12.034

Zhongying Xiao is a Data & Applied Scientist at Microsoft with a Ph.D. in chemistry and a Master in Machine Learning. He has 5 years of experience in manufacturing, healthcare, and technology industries. He specializes in Machine Learning and Deep Learning for anomaly detection, medical claim automation, and natural language processing. He previously worked at Anthem and Georgia-Pacific as a data scientist, deploying machine learning models and developing ML/DL-based anomality detection methods. He published several research papers in chemistry and anomaly detection field.

Arthur Kordon is a CEO of Kordon Consulting LLC and an internationally recognized pioneer in applying advanced analytics and artificial intelligence in the industry with more than 30 years of experience. He is currently consulting several global corporations to introduce and use analytical solutions and artificial intelligence in their business. His current projects include developing cognitive models of the Enterprise, preventive maintenance, energy cost reduction analysis, office space analytics, etc. In his previous position as Advanced Analytics Leader at Dow Chemical, He has successfully applied advanced analytics solutions to various business problems in forecasting, business cycle analysis, price elasticity analysis, etc. He introduced several novel technologies for improved manufacturing and new product design based on artificial intelligence, such as robust inferential sensors, operating discipline, and accelerated fundamental model building. He was granted a US patent and has published more than 70 papers and 16 book chapters in applied artificial intelligence and advanced

analytics. He is the author of two books Applying Data Science and Applying Computational Intelligence, published recently by Springer. He is a co-author of the book Applied Data Mining for Forecasting, published by SAS Press. He is a respected member of the international scientific community in the field of applied AI systems as a participant in the program committees of the most recognized conferences in this field and industrial committees in the Computational Intelligence Systems society of IEEE and IFAC (International Federation on Automatic Control.)

Subrata Sen is the Principal of greyBOX Innovation LLC. He provides advisory to senior corporate leadership on the appropriate use of AI and analytics tools and techniques in critical application for business success. His prior roles include Senior Director of Data Science and Analytics at Georgia-Pacific (a subsidiary of Koch Industries), Principal Scientist and Program Leader of new product development in the corporate R&D of Dow Inc. In his professional career, he has utilized his formal training in using first principles, statistical and machine learning technics in solving high value industrial problems. Subrata received BS, MS and PhD in Chemical Engineering, with specialization in coupling multiphase flow physics and chemistry. He has published numerous technical articles and has three granted patents.