A Novel Fault Diagnosis Method based on Stacked LSTM

Qingqing Zhang, Jiyang Zhang, Jianxiao Zou, Shicai Fan

School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, China 611731 (e-mail: qqzhang121@163.com, zhangjiyang96@163.com, jxzou@uestc.edu.cn, shicaifan@uestc.edu.cn)

Abstract: Fault diagnosis is essential to ensure the operation security and economic efficiency of the chemical system. Many fault diagnosis methods have been designed for the chemical process, but most of them ignore the temporal correlation in the sequential observation signals of the chemical process. A novel deep learning method based on Stacked Long Short-Term Memory (LSTM) neural network is proposed, which can effectively model sequential data and detect the abnormal values. The proposed method is also able to fully exploit the long-term dependencies information in raw data and adaptively extract the representative features. The dataset of Tennessee Eastman (TE) process is utilized to verify the practicability and superiority of the proposed method. Extensive experimental results show that the fault detection and diagnosis model we proposed has an excellent performance when compared with several state-of-the-art baseline methods.

Keywords: Fault detection, fault diagnosis, recurrent neural network, Stacked LSTM, sequential model, process monitoring

1. INTRODUCTION

As the industrial process control systems are becoming larger and more complicated, it may cause huge damage if the control system fails. Fault diagnosis methods can be applied to ensure the reliable and efficient operation of the control system (Polverino et al. (2017)). Therefore, it is significant to develop fault diagnosis methods which can detect the fault status correctly and diagnose the type of faults immediately in production processes.

The traditional data-driven fault detection methods are mainly based on statistical methods, such as Principle Component Analysis (PCA) (Nomikos and MacGregor (1994)), Independent Component Analysis (ICA) (Lee et al. (2004)), Dynamic Principal Component Analysis (DPCA) (Chen and Liu (2002)), and Dynamic Independent Component Analysis (DICA) (Stefatos and Hamza (2010)). PCA, ICA, DPCA, and DICA are all linear dimension reduction methods while the monitoring of complex chemical processes is usually a nonlinear issue (Xu et al. (2018)). Therefore, applying these methods above for fault detection in chemical process may lose useful information and will bring misleading fault detection results.

Machine learning methods, which have excellent performance in nonlinear classification problems, are increasingly applied to fault diagnosis systems. Traditional fault diagnosis methods based on machine learning can be mainly divided into two steps: feature extraction and fault classification (Lei et al. (2019)). Fault classification methods mainly include Support Vector Machine (SVM) (Jing et al. (2014)), Random Forests (RF) (Wang et al. (2017)), Backpropagation neural networks (BP) (Yang and Wei (2011)), etc. The fault diagnosis ability of these methods depends on the quality of feature extraction. However, traditional machine learning methods have limitations on the feature representation of complex high dimensional data because of human participation. These methods have unsatisfactory performance on the fault diagnosis of industrial processes (Lv et al. (2016)).

Deep learning methods, which can automatically extract features from a large amount of data (Long et al. (2018)) and reduce the dependence of feature extraction compared to traditional models, have been applied to fault diagnosis, such as Stacked Sparse Auto Encoder (SSAE) (Lv et al. (2016)) and Convolutional Neural Network (CNN) (Chen et al. (2017)). As the observation signals in industrial process should be treated as time series signals (Liu et al. (2018)), Recurrent Neural Network (RNN) (de Bruin et al. (2017)) has achieved better feature learning and classification results in fault diagnosis. Long Short-Term Memory (LSTM) neural network (Hochreiter and Schmidhuber (1996)), which could avoid the vanishing and exploding gradient problem in RNNs, has been applied to sequential fault monitoring. Zhao et al. (2018) applied Vanilla LSTM on the Tennessee Eastman (TE) process dataset and achieved more promising fault diagnosis performance compared with other methods. However, the neural network they designed has limitations in learning deep hidden information, because Vanilla LSTM only has one hidden layer. Therefore, the fault diagnosis model has low diagnosis accuracy in some fault types when all of the 21 faults data are taken together for fault diagnosis.

In order to improve the performance of fault diagnosis model, a novel fault diagnosis method based on Stacked

LSTM is proposed in this paper. By the deeper neural network structure and the longer time series length, the proposed method can not only extract features from chemical process data automatically but also fully exploit the relationship between the previous state and the current state among sequences. Therefore, the proposed method has a state-of-the-art performance in these faults which have high false alarm rates in existing methods and tiny differences in mean and variance. Experimental results on the TE process indicated that the proposed method based on Stacked LSTM could achieve better performance compared with other method.

The remainder of this paper is divided into the following parts: (1) In the second section, the basic knowledge of the sequential model and the LSTM methods are illustrated. (2) In the third section, the process of establishing a fault diagnosis model based on Stacked LSTM are proposed. (3) In the fourth section, in order to verify the validity of the proposed method, the method is applied to the TE process fault diagnosis and compared with the state of art methods. (4) Finally, the conclusion is drawn.

2. PRELIMINARIES

2.1 Vanilla LSTM

Vanilla LSTM is an improved network structure based on the standard RNN (Saleh et al. (2017)). Compared with the standard RNN model, LSTM mainly adds three control gating units: forget gate, input gate and output gate. Therefore, Vanilla LSTM overcomes the problem of gradient vanishing which occurs in RNN (Schmidhuber (1997)).

As shown in Fig. 1, Vanilla LSTM model has one LSTM layer and the LSTM layer is based on a set of connected LSTM cells. The architecture of LSTM cell is shown in Fig. 2. $\boldsymbol{x}_t$, $\boldsymbol{H}_t$, $\boldsymbol{C}_t$, $\tilde{C}_t$, F_t, I_t, and O_t are respectively the input vector, the hidden state, the present updating memory, the new information after transformation, the forget gate, the input gate and the output gate at time t. $\boldsymbol{H}_{t-1}$ is the hidden state at time $t-1$ and $\boldsymbol{C}_{t-1}$ is the old memory of the historical information at time $t-1$. σ is the sigmoid function.

At each time step t, $\boldsymbol{H}_t$ is updated by $\boldsymbol{x}_t$, $\boldsymbol{H}_{t-1}$, $\boldsymbol{C}_{t-1}$, $\tilde{C}_t$, $\boldsymbol{C}_t$, F_t, I_t, and O_t. The following updating equations are given as follows:

$$F_t = \sigma\left(\boldsymbol{x}_t W_{xf} + \boldsymbol{H}_{t-1} W_{hf} + b_f\right) \tag{1}$$
$$I_t = \sigma\left(\boldsymbol{x}_t W_{xi} + \boldsymbol{H}_{t-1} W_{hi} + b_i\right) \tag{2}$$
$$O_t = \sigma\left(\boldsymbol{x}_t W_{xo} + \boldsymbol{H}_{t-1} W_{ho} + b_o\right) \tag{3}$$
$$\tilde{C}_t = \tanh\left(\boldsymbol{x}_t W_{xc} + \boldsymbol{H}_{t-1} W_{hc} + b_c\right) \tag{4}$$
$$\boldsymbol{C}_t = F_t \odot \boldsymbol{C}_{t-1} + I_t \odot \tilde{C}_t \tag{5}$$
$$H_t = O_t \odot \tanh\left(\boldsymbol{C}_t\right) \tag{6}$$

where W_{xi}, W_{xf}, W_{xo}, W_{xc}, W_{hi}, W_{hf}, W_{ho}, and W_{hc} are the weight parameters, b_f, b_i, b_o, and b_c are the offset parameters, these parameters are shared by all time steps and learned during model training. $\odot$ denotes the element-wise product. Because of the special architecture, Vanilla LSTM can cope with the gradient attenuation problem in the recurrent neural network and better capture the long-term dependence in the time series.

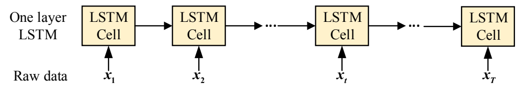

Fig. 1. The architecture of Vanilla LSTM

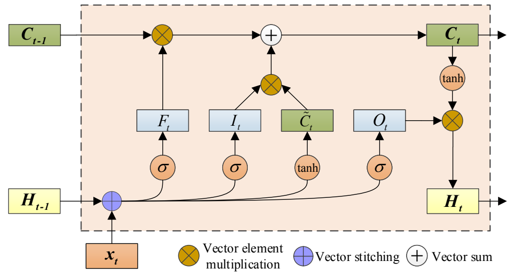

Fig. 2. The architecture of LSTM cell

2.2 Stacked LSTM

Deep network architectures have shown to be powerful in complex nonlinear feature representation (Ajami and Daneshvar (2012)). Stacked LSTM, which is a deep neural network with multiple hidden layers, becomes a stable technique for challenging sequence prediction problems (Brownlee (2017)).

The structure of Stacked LSTM is shown as Fig. 3. Stacked LSTM consists of multiple LSTM layers and each LSTM layer contains multiple connected LSTM cells. The input of LSTM layer-1 is the raw data and the input of other LSTM layers is the hidden state of the previous LSTM layer. Therefore, the hidden state in one LSTM layer is both propagated through time and passed to the next layer compared with Vanilla LSTM.

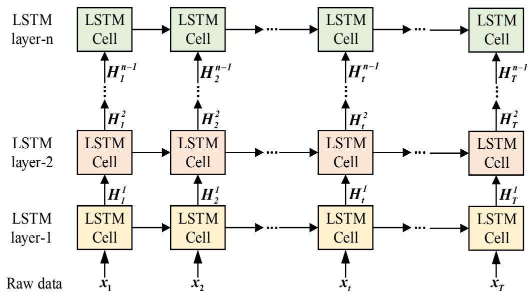

Fig. 3. The architecture of Stacked LSTM

By the Stacked LSTM, the neural network become deeper, the characteristic of input sequence can be learned better at different time scales (Zhao et al. (2016)). In order to learn the dynamic sequential information of process monitoring well, we use the Stacked LSTM network as the fault diagnosis model of the process monitoring in this paper.

3. FAULT DIAGNOSIS BASED ON STACKED LSTM

In this section, the process of applying Stacked LSTM method for fault diagnosis is described. This includes the following three parts.

3.1 Fault diagnosis model designing

The designed fault diagnosis model based on Stacked LSTM is shown in Fig. 4. By combining Stacked LSTM network and the Softmax multi-classifier, the diagnosis model can directly diagnose faults in only one shot. This Stacked LSTM-based diagnosis model can be divided into three parts: data processing layer, feature extraction layer, and fault diagnosis layer.

Data processing layer preprocesses the raw data to eliminate the impact of variables' measurement units and speed up the convergence during model training. Feature extraction layer uses multi-layer LSTM to establish a complex nonlinear mapping between the raw input sequence and the hidden status recording the historical information and implement feature extraction. Fault diagnosis layer applies the Softmax function to classify the hidden status from the last LSTM layer and outputs the fault diagnosis results. Specifically, instead of all input time steps just have one output, each step of input time has one output in our model.

3.2 Offline Modeling

Offline modeling requires collecting normal and fault raw process data for training offline fault diagnosis model.

Algorithm 1: Stacked LSTM offline modeling

1) Collect normal and fault condition raw data with time order.
2) Normalize the raw data by the formula:

$$\begin{aligned} \boldsymbol{x_i^*} &= [x_{i1}^*, x_{i2}^*, \ldots, x_{iN}^*], i = 1, 2, \ldots, M \\ x_{ij}^* &= \frac{x_{ij} - mean(\boldsymbol{x_j})}{std(\boldsymbol{x_j})}, j = 1, 2, \ldots, N \end{aligned} \tag{7}$$

where $\boldsymbol{x}^*$ is the normalized variable and i-th raw data, x_{ij} and x_{ij}^* are respectively the original and normalized result of the j-th feature in the i-th raw data, M and N are the number and feature dimension of the raw data respectively.
3) Construct sequential data of length T as training data. Under the principle of adjacent sampling, use a sliding window with length T to sample observation sequences in the raw data with time order.
4) Train the fault diagnosis model based on Stacked LSTM with Dropout (Srivastava et al. (2014)) and Adam methods (Kingma and Ba (2014)).
5) Compute the loss J by the Cross Entropy Loss Function:

$$J = -\sum_{i=1}^{N} y_i \log p_i \tag{8}$$

where y_i and p_i are respectively the true probability and the predicting probability that the sample belongs to class i, and N is the number of classes.
6) If $J > \tau$ or the number of iterations $l < N$, repeat steps 4-6. τ is a positive number and N is the predefined maximum number of iterations.
7) Save the parameters of the hidden layer and get a stable fault diagnosis model based on Stacked LSTM.

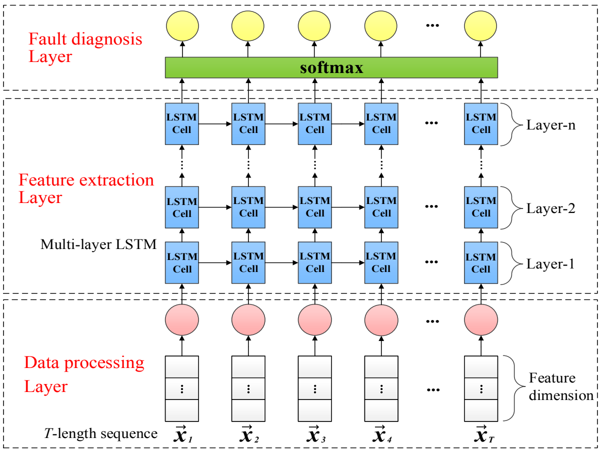

Fig. 4. Fault diagnosis model framework based on Stacked LSTM.

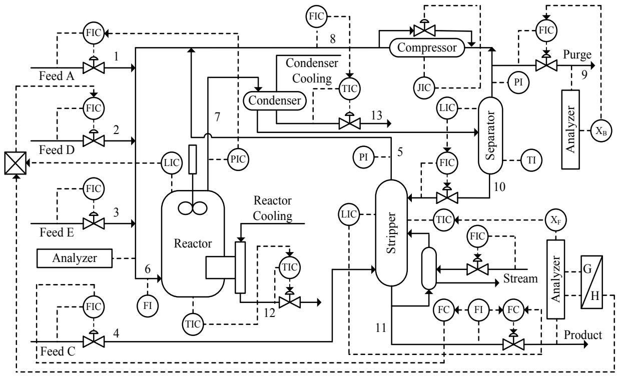

Fig. 5. Flow structure of TE process.

3.3 Online Monitoring

Online monitoring will collect new process data in real time, and then use the offline trained diagnosis model to monitor the operation status of the industrial process.

Algorithm 2: Stacked LSTM online monitoring

1) Collect new raw data with time order.
2) Normalize the raw data according to the equation (7).
3) Similar to step 3 in Algorithm 1, construct sequential data of length T, $\boldsymbol{x}^{\text{new}} = [x_1^{\text{new}}, x_2^{\text{new}}, \cdots, x_T^{\text{new}}]$ from the normalized raw data.
4) Classify the testing data x^{new} by offline trained fault diagnosis model and obtain the online monitoring result.

4. EXPERIMENTS

To verify the efficiency of the proposed method, the TE process is applied to demonstrate the advantages of this method by comparing with other methods.

The TE process was firstly created by Down and Vogel of the American chemical company Tennessee Eastman(Zhu et al. (2017)). It has been widely used to evaluate the effectiveness of some process diagnosis methods. The process flow chart is shown in Fig. 5. This process has four starting materials: A, C, D, and E, two products: G and H as well as a by-product: F. The whole reaction process contains an inert gas: B, it is used as the main catalyst for the reaction. TE process has five major operating units, they are respectively the Reactor, Condenser, Compressor, Separator, and Stripper. The TE process has 41 measurement variables and 12 operation variables. We select 52 variables (41 measurement variables and 11 operation variables) for monitoring. Besides the TE process includes 21 predefined perturbations (from fault 1 to fault 21). The experiment data can be downloaded from *http://web.mit.edu/braatzgroup/links.html*.

Therefore, the dataset contains 22 working states: 1 normal state and 21 different fault states. For model construction, the data are divided into the training part and the independent test part, which respectively has 480 samples (fault samples) and 960 samples (160 samples are normal state and 800 samples are fault state).

To verify the significant performance of proposed method, this method is compared with others. And although our Stacked LSTM diagnosis model can complete detection task and classification in one shot, for fair comparison, the fault detection rate and the fault diagnosis rate are used to evaluate the performance of proposed Stacked LSTM-based fault diagnosis method.

The fault detection and the fault diagnosis can be regarded as a binary classification problem and a multi-classification problem, respectively. And the fault detection rate and the fault diagnosis rate are respectively the accuracy in fault detection and fault diagnosis. The accuracy is defined as:

$$Accuracy = \frac{TR}{TO} \tag{9}$$

where TR and TO are respectively the number of correctly classified samples and the number of samples in the dataset.

4.1 Performances Comparison with different key parameters

In order to analyze the influence of different key parameters of LSTM network on fault diagnosis performance. We experiment with different LSTM hidden layers parameters and the length of input sequence of network. And the result is shown in Fig. 6 and Fig. 7.

In the Fig. 6, we can find that when LSTM layer=1, the network becomes a vanilla LSTM, the performance of the method is worst. When the number of layer is increased, the diagnosis rate of all faults has a significant improvement especially in the fault 3, 9, and 15. When LSTM layer=3, the diagnosis rate for all faults increases to above 97%. And our experiments show that the more than 3 layers LSTM network do not have significant effect, but increase the complexity of the network. Therefore, we stack 3 layers LSTM in the proposed model.

LSTM network is sensitive to time series data and can classify the observation signals based on the long-term dependencies information. From Fig. 7, we can find that different sequential lengths have a great impact on fault classification. Both of the fault detection rate and diagnosis rate will become higher when the input sequence length increase, but there will be no significant improvement after the sequence length is 100. Therefore, we choose the network input sequence length is 100.

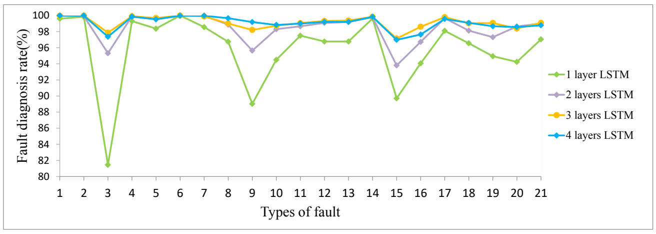

Fig. 6. The fault diagnosis rate with different LSTM layers.

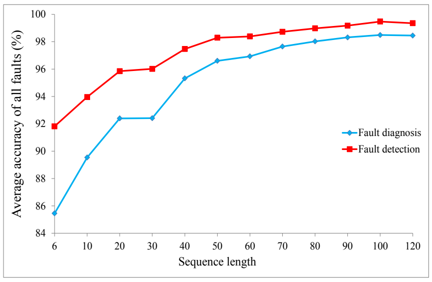

Fig. 7. The monitoring results with different sequential lengths.

4.2 Performance Comparison with other methods

Fault detection In this subsection, to show the superiority of proposed method, Stacked LSTM detection precision is compared with the traditional method and deep learning method. The detection rate of 21 faults by several monitoring methods is summarized in table 1. The average fault detection rates of Stacked LSTM, Stacked Sparse Automatic Encoder (SSAE), DICA:AO, DPCA:SPE, ICA:AO, DICA:I^2, ICA:I^2, PCA:SPE, DPCA:T^2, and PCA:T^2 are respectively 99.49%, 84.44%, 80.18%, 78.77%, 73.24%, 72.02%, 71.81%, 69.42%, 56.70%, and 51.21%.

It can be seen that the performance of SSAE and Stacked LSTM is better than that of PCA and DPCA. This shows that the deep learning method has better ability of capturing the nonlinear and complex features than the traditional method. In the traditional methods, DPCA:T^2 performs better than PCA:T^2 and DICA:AO can give better results than ICA:AO; similarly, in the deep learning algorithm, Stacked LSTM is more effective than SSAE. From the above situation, it can be seen that fully mining the temporal relationship between data has a great impact on the result of fault detection, because the data of TE process has correlative and temporal characteristics. DPCA and DICA take into account the impact of historical data information on fault detection, but do not explicitly model the temporal relationship between data. Besides, the Stacked LSTM method performs better than other fault detection methods. And there are optimal performance in fault 3, 9, and 15, which have worst detection result in other methods.

Fault diagnosis The classification accuracies of different algorithms are shown in Fig. 8, it is obvious to see that SVM and SSAE perform poor than Batch Normalization-based (BN-based) LSTM, Stacked RNN, Stacked GRU, and Stacked LSTM. This shows that considering the temporal relation is important to the fault diagnosis.

Table 1. The fault detection rates of different methods(%).

Fault	PCA:T^2	PCA:SPE	DPCA:T^2	DPCA:SPE	ICA:I^2	ICA:AO	DICA:I^2	DICA:AO	SSAE	**Stacked LSTM**
1	99.25	99.75	99.63	99.75	99.75	99.50	99.63	99.75	99.80	**100**
2	98	98.88	98.12	99.25	97.13	97.75	97.37	99	98.79	**99.99**
3	0	15.12	5.75	21.63	0.38	15.25	0	4.50	44.83	**98.47**
4	3.50	**100**	16.88	**100**	76	56.63	48.50	76.62	85.96	**100**
5	22.13	38.75	29.62	49.50	**100**	**100**	**100**	**100**	88.81	99.83
6	98.88	**100**	99	**100**	**100**	**100**	**100**	**100**	99.81	**100**
7	91.25	**100**	66.50	**100**	99.88	99.50	99.25	99.75	96.45	99.99
8	96.50	98.25	97.38	98.37	93.25	93.75	96.75	98.13	99.08	**99.91**
9	0.13	15.13	6	20.88	0.37	14.63	0	4.37	38.59	**98.64**
10	29.50	64.62	49.13	75.62	79.87	75.88	79.75	89.13	88.56	**99.45**
11	23.12	79.38	31.63	91	45.38	44.25	46.88	64.88	86.44	**99.68**
12	98	98.75	99.38	99.25	99	98.62	99.87	99.88	98.85	**99.96**
13	93.88	95.87	94.62	95.88	95.12	94.50	95.38	95.62	96.61	**99.65**
14	92.12	**100**	93.88	**100**	99.38	**100**	99.75	**100**	97.58	**100**
15	1	18.38	10.62	24	1.87	4.38	0	19.25	34.29	**97.89**
16	13.13	57.87	31.13	67.37	80	75.25	77.75	86.38	87.06	**99.28**
17	72.87	95.38	80.87	97.50	86.38	86.25	94.38	95.87	94.12	**99.86**
18	88.50	92.37	89.50	92.75	89.50	89.63	92.75	91.13	90.24	**99.66**
19	0.25	44.25	2.88	82.25	47.63	63.75	60.50	85.62	82.63	**99.24**
20	25.75	24.88	48.63	76	79.62	69.88	83.75	88.25	85.08	**98.66**
21	27.63	20.12	39.62	63.13	37.50	58.70	40.12	85.60	79.59	**99.23**

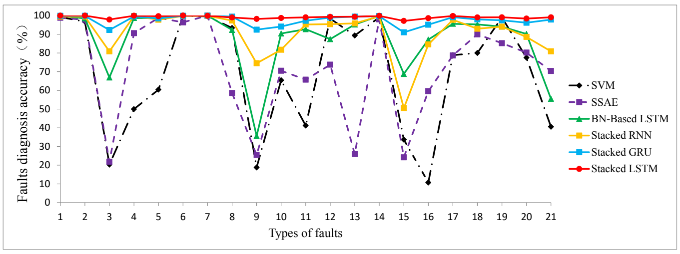

Fig. 8. The fault diagnosis rates of different classification methods.

We can also find that the methods based on stacked models outperform the BN-based LSTM, which shows that the stacked models can utilize and learn the raw data dynamic information better than BN-based one layer LSTM. Besides, The red line in the figure is the fault classification accuracy of Stacked LSTM. It is easy to find that the proposed method give optimal performance for all faults. The average classification accuracy of all faults in proposed Stacked LSTM network is 99.14%, which is 8.39% and 1.72% higher than Stacked RNN and Stacked GRU, respectively. Moreover, SVM, SSAE, and BN-based LSTM perform poor in fault 3, 9, and 15 while in our proposed method, the fault diagnosis rates of the three faults both are more than 97%.

According to the results of Stacked LSTM, we can find that the deeper of LSTM network has a strong representation capability. As shown in Fig. 9, we use t-distributed Stochastic Neighbor Embedding (t-SNE) (Hinton (2008)) to visualize the distribution of features at the different layer of the Stacked LSTM model. Because there are too many faults, we only give a two-dimensional visualization of t-SNE of fault 3, 9, and 15. In the figure different faults are represented by different colors. In each t-SNE figure, the more overlaps fault points have, the harder it is to classify these faults. Fig. 9(a). shows the original distribution of these three faults. From this figure, it can be seen that these three faults group together. It is difficult to classify if the features cannot be extracted efficiently. Fig. 9(b)-(d), show that after the features are extracted by each layer of the Stacked LSTM model, different types of fault are gradually separated, and the same faults are gradually gathered together. Therefore, the deeper LSTM network is easier to distinguish different fault and has a strong discriminant capability for fault diagnosis.

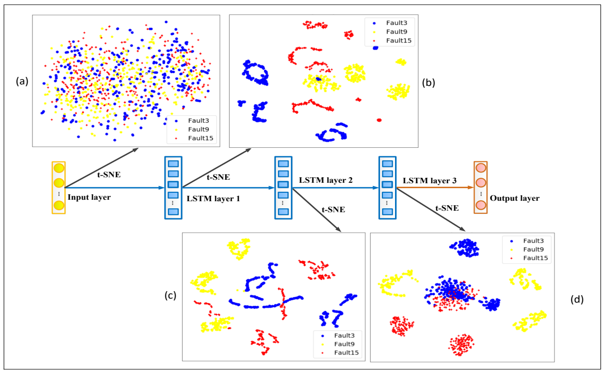

Fig. 9. The t-SNE representation result of the faults 3, 9, and 15. (a) The original faults 3, 9, and 15 in different layers.(b) The first layer of LSTM. (c) The second layer of LSTM. (d) The third layer of LSTM.

By analyzing the results of experiments, we indicate that the proposed method has good performance in fault detection and fault classification. The Stacked LSTM network greatly improves the accuracy of fault diagnosis of TE process. Obviously, the diagnosis effects of faults 3, 9, and 15 are greatly improved compared with other algorithms. Experiments have been implemented on a computer with NVIDIA GeForce RTX 2080Ti GPU and Intel Xeon CPU. The code of experiments will be available at *https://github.com/zhangqingqingq/Fault-diagnosis-based-on-LSTM.*

5. CONCLUSIONS

A fault detection and diagnosis method based on Stacked LSTM is proposed in the paper. LSTM could process sequential data of any length and encode historical information in the hidden layer. In order to fully exploit the temporal relationships among observation signals at different times and automatically extract representative features, we exhibit a model with three layers LSTM for fault diagnosis. The proposed method can implement the fault monitoring in real time, because the historical information of fault data is stored in the memory cell units. As a case study, the method is implemented on the TE process and achieves the best fault detection and diagnosis results compared with some state-of-art methods, which verifies the feasibility and effectiveness of the method. Therefore, this method can well diagnose a single fault. In future research, we will apply this method to some more complex industrial processes where multiple faults occur simultaneously.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (61872063), Fundamental Research Funds for the Central Universities (ZYGX2016J102), Sichuan Science and Technology Program (2018HH0149), and Sichuan Provincial Youth Science and Technology Innovation Team Special Projects under Grant 2015TD0018.

REFERENCES

Ajami, A. and Daneshvar, M. (2012). Data driven approach for fault detection and diagnosis of turbine in thermal power plant using independent component analysis (ICA). *International Journal of Electrical Power & Energy Systems*, 43(1), 728–735.

Brownlee, J. (2017). *Long Short-term Memory Networks with Python: Develop Sequence Prediction Models with Deep Learning*. Machine Learning Mastery.

Chen, J. and Liu, K.C. (2002). On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chemical Engineering Science*, 57(1), 63–75.

Chen, L., Wang, Z., and Zhou, B. (2017). Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. *Advanced Engineering Informatics*, 32, 139–151.

de Bruin, T., Verbert, K., and Babuška, R. (2017). Railway track circuit fault diagnosis using recurrent neural networks. *IEEE transactions on neural networks and learning systems*, 28(3), 523–533.

Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2605), 2579–2605.

Hochreiter, S. and Schmidhuber, J. (1996). LSTM can solve hard long time lag problems. In *International Conference on Neural Information Processing Systems*.

Jing, C., Gao, X., Zhu, X., and Lang, S. (2014). Fault classification on Tennessee Eastman process: PCA and SVM. In *2014 International Conference on Mechatronics and Control (ICMC)*, 2194–2197. IEEE.

Kingma, D.P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lee, J.M., Yoo, C., and Lee, I.B. (2004). Statistical process monitoring with independent component analysis. *Journal of Process Control*, 14(5), 467–485.

Lei, J., Liu, C., and Jiang, D. (2019). Fault diagnosis of wind turbine based on long short-term memory networks. *Renewable Energy*, 133, 422–432.

Liu, H., Zhou, J., Zheng, Y., Jiang, W., and Zhang, Y. (2018). Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *Isa Transactions*, S0019057818301514.

Long, W., Liang, G., and Li, X. (2018). A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on Systems Man & Cybernetics Systems*, 49(1), 136–144.

Lv, F., Wen, C., Bao, Z., and Liu, M. (2016). Fault diagnosis based on deep learning. In *2016 American Control Conference (ACC)*, 6851–6856. IEEE.

Nomikos, P. and MacGregor, J.F. (1994). Monitoring batch processes using multiway principal component analysis. *AIChE Journal*, 40(8), 1361–1375.

Polverino, P., Sorrentino, M., and Pianese, C. (2017). A model-based diagnostic technique to enhance faults isolability in solid oxide fuel cell systems. *Applied Energy*, 204, 1198–1214.

Saleh, K., Hossny, M., and Nahavandi, S. (2017). Intent prediction of vulnerable road users from motion trajectories using stacked LSTM network. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, 327–332. IEEE.

Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929–1958.

Stefatos, G. and Hamza, A.B. (2010). Dynamic independent component analysis approach for fault detection and diagnosis. *Expert Systems with Applications*, 37(12), 8606–8617.

Wang, Z., Zhang, Q., Xiong, J., Ming, X., Sun, G., and He, J. (2017). Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests. *IEEE Sensors Journal*, 17(17), 5581–5588.

Xu, Y., Shen, S.Q., He, Y.L., and Zhu, Q.X. (2018). A novel hybrid method integrating ICA-PCA with relevant vector machine for multivariate process monitoring. *IEEE Transactions on Control Systems Technology*, PP(99), 1–8.

Yang, Y. and Wei, T. (2011). Study of remote bearing fault diagnosis based on BP neural network combination. In *International Conference on Natural Computation*.

Zhao, H., Sun, S., and Jin, B. (2018). Sequential fault diagnosis based on LSTM neural network. *IEEE Access*, 6, 12929–12939.

Zhao, R., Wang, J., Yan, R., and Mao, K. (2016). Machine health monitoring with LSTM networks. In *2016 10th International Conference on Sensing Technology (ICST)*, 1–6. IEEE.

Zhu, Q.X., Meng, Q., and He, Y.L. (2017). A novel multidimensional feature pattern classification method and its application to fault diagnosis. *Industrial & Engineering Chemistry Research*, 56(31), acs.iecr.7b00027.