International Journal of Information Technology Vol. 25 No. 2 2019

Digital Twin for Fault Detection Based on Hybrid Neural Networks

Zhiwen Pan[*†], Jiahao Li[*†], Yiqiang Chen[*†✉],
Salim Hariri[‡], Jesus Pacheco[‡] and Xiaofan Song[‡‡]

[*]Beijing Key Laboratory of Mobile Computing and Pervasive Device
[†]Institute of Computing Technology, CAS, Beijing, China
[‡]Department of Electrical and Computer Engineering, the University of Arizona, Tucson, AZ, USA
[‡‡]State Grid Henan Economic Research Institute, Zhengzhou China

pzw@ict.ac.cn, www.LiJiahao@live.cn, yqchen@ict.ac.cn,
hariri@email.arizona.edu, Jpacheco@email.arizona.edu, sxf810118@sina.com

Abstract

The Digital Twin based fault detection system becomes more and more popular nowadays for industry. It can be formalized as a behavior model for predicting the normal behaviors of monitored assets and an anomaly detector for triggering alarms based on residual between normal behavior and current behavior. However, due to the sophisticated nature of the Industrial Control Systems (ICS), the fault detection based on Digital Twin is still facing the challenges of dealing with nonlinear correlation, high dimensionality, and non-gaussian noise consisted within the physical process data.

In this paper, a Digital Twin approach based on Hybrid Neural Network is proposed to perform fault detection for Cyber Physical Systems (CPS) and their assets. The proposed behavior model is a Hybrid Neural Network consisting two components: 1) a robust convolution auto-encoder to perform noise reduction and dimensionality reduction; and 2) a LSTM neural network to predict the normal behavior. The experimental evaluation is performed on a well-recognized CPS benchmark dataset: Secure Water Treatment, and the result shows that the proposed approach outperforms both classic machine learning approaches and existing deep learning approaches.

In conclusion, the proposed approach utilizes advanced deep learning technique to automatically generate behavior model based on normal historical behavior data of monitored CPS. It is capable to describe normal behavior patterns of sophisticated systems and assets in real world. The major future research direction is to deal with concept drift phenomenon occurred during normal operation of sophisticated Cyber Physical Systems.

Keyword: Digital Twin, Fault Detection, Anomaly Detection, Deep Learning, Neural Networks

I. Introduction

The significant advancement of information technologies including Internet of Things, Cloud Computing, Big Data analytics is reforming the traditional Industrial Control Systems (IDS) as Cyber Physical Systems (CPS). By leveraging the power of data transmission and computation into physical control, Cyber Physical Systems aim to realize autonomous, intelligent, real-time and robust interactions between cyber space (e.g. cloud computing systems, edge computing systems, human machine interface) and physical space (e.g. controllers, sensors, actuators, workers, customers). Specifically, CPS collect the information which can explicitly describe the contexts of physical world from the physical space and deliver it to the cyber space. In the cyber space, the collected information is processes, managed, abstracted, analyzed and visualized so that decisions can be generated in a data-driven manner. Based on these decisions, commands and information are delivered to physical space to control the physical systems. These interactions are operated in a real-time loop of "monitor, influence and feedback".

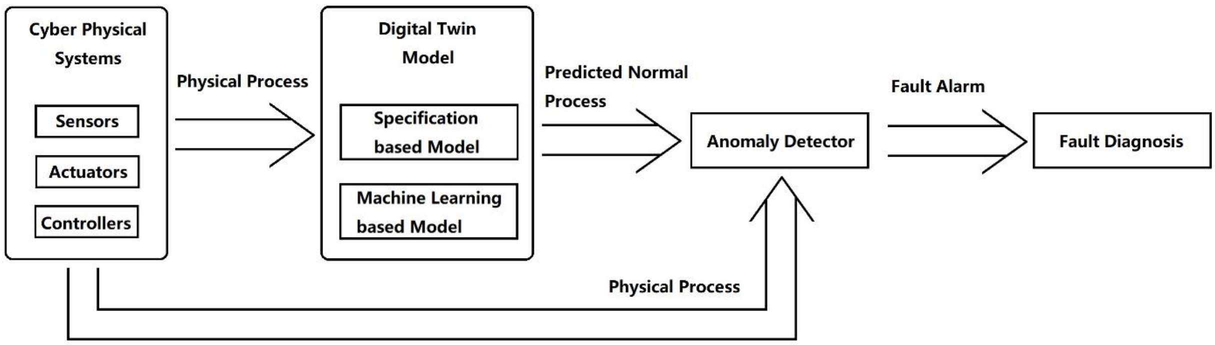


Figure 1. The architecture of Digital Twin based CPS for fault detection and diagnosis

As a core technology for CPS, Digital Twin is an organized collection of physics-based methods and advanced analytics that are used to model the operational behaviors of physical assets (e.g. controllers, sensors, actuators) [1]. The Digital Twin is proved effective since the behaviors of physical assets are inherently following the configurations, designs, and physical constraints. By utilizing the models generated by Digital Twin technology, decisions are generated to influence the physical assets.

One major application for Digital Twin is fault detection and diagnosis in ICS. Figure 1 shows the architecture of Digital Twin based CPS for fault detection and diagnosis. In the model generation phase, Digital Twin model is generated based on the historical behavioral data collected from the physical assets. In the real-time analysis phase, the generated Digital Twin model predicts the normal operational behaviors (or states) of physical assets based on their previous behaviors (or states), so that the deviation between the normal behaviors and current behaviors can be utilized to evaluate the health condition of physical assets. Currently, the Digital Twin model is generated through either specification based methods [2-3] or machine learning based methods [4-9]. However, due to the sophisticated nature of the ICS, the fault detection based on Digital Twin is still facing the following challenges:

- **Nonlinear operational patterns:** Although the traditional methods (e.g. specification based methods and machine learning based methods) perform well to model Linear Time Invariant (LTI) systems, they perform poorly on modeling sophisticated correlation patterns in nonlinear systems.
- **High dimensional "big" data:** In a medium-scale or large-scale ICS with a large amount of assets deployed, the operational data being collected are high dimensional and the data attributes are auto-correlated and self-correlated between each other. The traditional methods perform poorly for analyzing high dimensional data. Moreover, specification based methods require integrating enormous expert knowledges at modeling normal behaviors of sophisticated systems (even assets).
- **Non-gaussian noise and disturbance:** Since a majority of data collected from the physical assets are from sensor readings, they usually contain noises for many reasons. Moreover, acceptable disturbance (e.g. unstable environment, normal asset degradation) may occur during

normal operations of ICS assets. Detection performance of traditional methods are highly affected by these noise and disturbance, especially when the data bias of the noise and disturbance does not follow Gaussian distribution.

As an advanced machine learning technology, deep neural network is featured for its ability of modeling nonlinear patterns in high-dimensional data. To meet the existing challenges, a Digital Twin approach based on Hybrid Neural Network (DTHNN) is proposed in this paper to perform fault detection and diagnosis. In order to predict the future normal behavior of physical assets accurately, the proposed hybrid neural network model integrates three types of neural networks in an end-to-end fashion. It includes an autoencoder for dimension reduction, a convolution network for describing cross-correlation (namely spatial correlation) and recurrent network for describing auto-correlation (namely temporal correlation). Specifically, the proposed model contains two components: 1) an autoencoder with convolutional neural layers which can map the high dimensional physical process data into low dimensional latent vector with highly abstracted spatial-correlation features of the original data; and 2) a Long Short-Term Memory (LSTM) network which can perform accurate prediction of the assets' normal behaviors based on the latent variables. The convolutional autoencoder and the LSTM network are integrated in an end-to-end fashion, so that the parameters of the two sub-networks are trained together to reach the global optimum. Moreover, a noise reduction mechanism is integrated into the convolutional autoencoder to reduce the interference on prediction caused by sensor noise and normal disturbance, so that the noise and disturbance bias data is removed from the original data during encoding process. To summarize, the contribution of this paper are as follows:

1. A Hybrid Neural Network based Digital Twin model is proposed to predict the normal behaviors of sophisticated Industry Control Systems and their assets. The proposed model integrates autoencoder, convolution network, and recurrent network together to capture the temporal-spatial correlations within high dimensional physical process data.

2. A noise reduction mechanism is proposed to remove noise and normal disturbance from the physical process data. The mechanism is integrated into the convolutional auto-encoder to make it more robust and further improve the prediction performance.
3. A Digital Twin based fault detection framework which utilizes Digital Twin model to perform accurate fault detection, as proved in the experimental evaluation.

The rest of the paper is organized as follows. In section II, the proposed DTHNN approach is introduced in detail. In section III, an experimental evaluation on benchmark dataset is performed. The paper is concluded in section IV.

II. Digital Twin for Fault Detection Based on Hybrid Neural Networks

General Architecture

By following the framework architecture of Digital Twin based fault detection system in Figure 1, the DTHNN approach consists of two components: a neural network based Digital Twin model for predicting the current normal behavior of monitored assets and a nonparametric anomaly detector to trigger alarms based on residual of predicted behaviors and current behaviors. Specifically, the Digital Twin model consist of two components: 1) a robust convolutional auto-encoder which performs dimensionality reduction and noise reduction on collected physical process data and generates the corresponding latent vector; 2) a recurrent neural network which predicts normal behavior of monitored assets based on latent vector provided by auto-encoder. The operation of DTHNN approach consists of two phases: training phase and detection phase. During the training phase, the physical process dataset is collected to describe the normal historical behaviors of the monitored ICS assets. The dataset is then used to train the Digital Twin model to describe the normal behavior patterns of the monitored asset precisely. During the detection phase, the Digital Twin model predicts the normal current behavior of monitored assets continuously based on their previous behaviors. The anomaly detector triggers alarms based on the residual of predicted behavior and assets’ current behaviors being monitored.

Training Phase

i). Convolutional Deep Autoencoder

The convolutional deep autoencoder can well retain the temporal correlation information of multi-dimensional time series data. In the real industrial scenario, the data inevitably have large outliers and universal noise, which affects the cleanliness of the training set. We proposed a robust convolutional deep autoencoder to represent and reconstruct the input data in the real industrial environment and discover the nonlinear features of time series data effectively.

In particular, the autoencoder architecture allows the model to obtain a low-dimensional representation of the input data through reconstruction.

$$\tilde{x} = g_{\psi}^{Decode}\left(f_{\psi}^{Encode}(x)\right) \tag{1}$$

where $\boldsymbol{x}$ denotes the input data, $\boldsymbol{f_{\psi}^{Encode}}(\cdot)$ denotes an encoder, $\boldsymbol{g_{\psi}^{Decode}}(\cdot)$ denotes a decoder, $\boldsymbol{\psi}$ denotes the parameters of encoder and decoder, and $\boldsymbol{\tilde{x}}$ is reconstructed version of $\boldsymbol{x}$.

In general, an autoencoder problem can be mathematically expressed as the following optimization problem.

$$\min_{\psi} \|x - \tilde{x}\| \tag{2}$$

where $\|\cdot\|$ is usually selected as L2-norm. For computationally efficiency, the specific express is written as below.

$$\min_{\psi} \left\|\hat{x} - g_{\psi}^{Deconv}\left(f_{\psi}^{Conv}(\hat{x})\right)\right\|_2^2 \tag{3}$$

ii). Robust Autoencoder.

In order to improve the robustness of the model, the following constraint is added to the optimization training process.

$$x = \hat{x} + x_s \tag{4}$$

where $\boldsymbol{x}$ is the input data, $\hat{\boldsymbol{x}}$ denotes the part that can be reconstructed well by the autoencoder, and $\boldsymbol{x_s}$ is one that contains noise and outliers that are hard to represent.

Compared with the traditional autoencoder structure with $\boldsymbol{x}$ as input, the robust autoencoder can recover $\hat{\boldsymbol{x}}$ better by removing noise and outliers from the input data. Using $\hat{\boldsymbol{x}}$ instead of $\boldsymbol{x}$ for optimization under the robust constraint (4), we have the following specific mathematical expression.

$$\min_{\psi} \left\| \hat{x} - g_{\psi}^{Decode}\left(f_{\psi}^{Encode}(\hat{x})\right)\right\|_2 + \|x_s\|_0 \tag{5}$$

$$s.t.\ x - \hat{x} - x_s = 0 \tag{6}$$

where, $\|\cdot\|_{\mathbf{0}}$ is the number of non-zero elements in the matrix which can well reflect the sparsity of the matrix. Unfortunately, L0-norm is theoretically perfect but computationally impossible, so we use L1-norm instead of L0-norm. $\boldsymbol{\lambda}$ is a parameter that controls the sparsity of the $\boldsymbol{x_s}$. Notice that for the first item in (5), we can use the back-propagation (BP) algorithm to train by minimizing the reconstruction error.

$$\min_{\psi} \left\| \hat{x} - g_{\psi}^{Decode}\left(f_{\psi}^{Encode}(\hat{x})\right)\right\|_2 + \lambda\|x_s\|_1 \tag{7}$$

$$s.t.\ x - \hat{x} - x_s = 0 \tag{8}$$

We show the implementation of the composite architecture in Algorithm I.

Input: $\boldsymbol{x}$
Initialize $\hat{\boldsymbol{x}}$, $\boldsymbol{x_s}$ to be zero matrices, $\boldsymbol{yx_s} = \boldsymbol{x}$. Randomly initialize the $\boldsymbol{\psi}$
While (True):

1. Remove $\boldsymbol{x_s}$ from $\boldsymbol{x}$, i.e. $\hat{\boldsymbol{x}} := \boldsymbol{x} - \boldsymbol{x_s}$
2. Minimize the first term in (7) using BP algorithm.
3. Set $\hat{\boldsymbol{x}}$ to be the reconstructed data $\tilde{\boldsymbol{x}}$.
4. Set S to be the difference between $\boldsymbol{x}$ and $\hat{\boldsymbol{x}}$, i.e. $\boldsymbol{x_s} := \boldsymbol{x} - \hat{\boldsymbol{x}}$
5. Optimize $\boldsymbol{x_s}$ using $\boldsymbol{f}$ function.
   If $\frac{\|x-\hat{x}-x_s\|_2}{\|x\|_2} < \boldsymbol{\varepsilon}$ or $\frac{\|yx_s-\hat{x}-x_s\|_2}{\|x\|_2} < \boldsymbol{\varepsilon}$:
   Break
6. Update $\boldsymbol{yx_s}$, i.e. $\boldsymbol{yx_s} = \hat{\boldsymbol{x}} + \boldsymbol{x_s}$

Return $\boldsymbol{\psi}$, $\hat{\boldsymbol{x}}$, $\boldsymbol{x_s}$

Algorithm I. Implementation of Robust Autoencoder

Where, the $\boldsymbol{f}$ function is implemented as below.

$$f(x_s(i)) = \begin{cases} x_s(i) - \lambda, & x_s(i) > \lambda \\ x_s(i) + \lambda, & x_s(i) < \lambda \\ 0, & -\lambda \le x_s(i) \le \lambda \end{cases} \tag{9}$$

iii). Convolutional Long Short-Term Memory Networks with Attention Mechanism

LSTM model performs well at processing time series data. It is a kind of recurrent neural network which can ensure the persistence of previous information by continuously circulating input data.

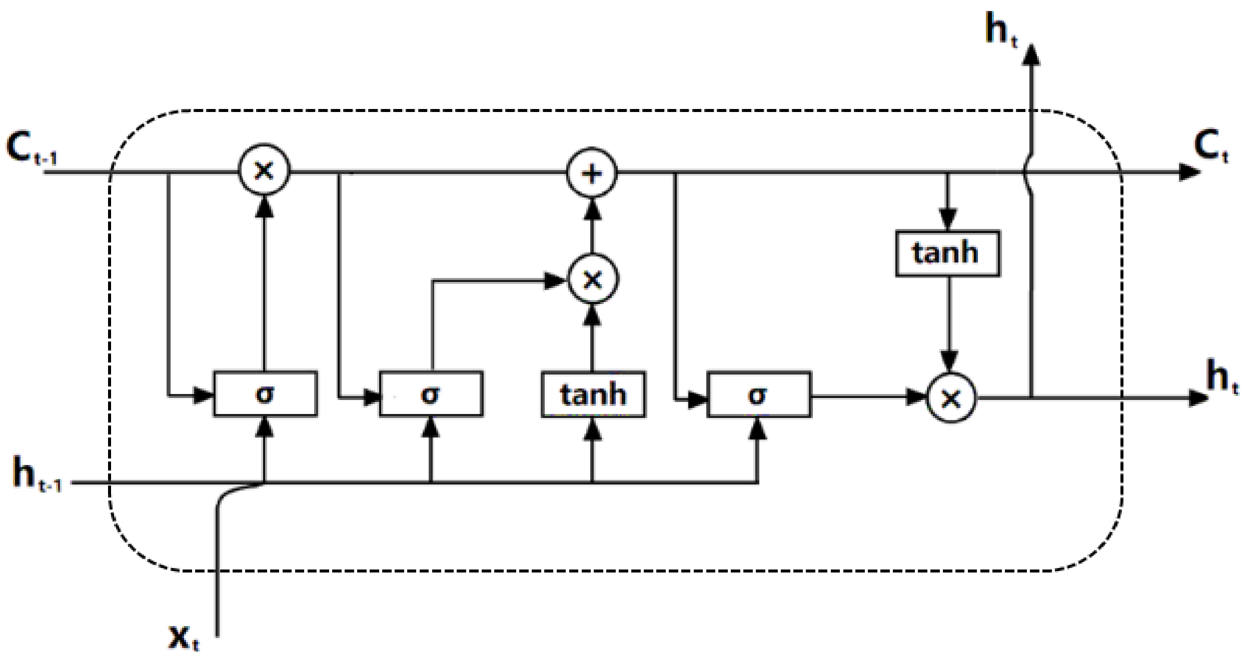


Figure 2. Cell Unit of LSTM

For the convenience of description, we make a symbolic convention as shown in Table I. The input, forget, cell, output, and hidden state of each timestep are denoted by ***I***, ***F***, ***C***, ***O*** and ***H*** respectively, the activation by $\boldsymbol{\sigma}$, and the weighted connections between states by a set of weights, ***W***.

Table 1. Symbolic Convention

Symbol	Meaning	Symbol	Meaning
I	Input Gate	W	Weights
F	Forget Gate	σ	Activate function
C	Cell State	$\tilde{C}$	Transition Cell State
O	Output Gate	$\cdot$	Normal Product
H	Hidden State	$*$	Convolution

The LSTM cell unit is mathematically described as follows.

$$I_t = \sigma(W_I \cdot [h_{t-1}, x_t] + b_I) \tag{10}$$

$$F_t = \sigma(W_F \cdot [h_{t-1}, x_t] + b_F) \tag{11}$$

$$\widetilde{C_t} = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{12}$$

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \tag{13}$$

$$C_t = F_t * C_{t-1} + I_t * \widetilde{C_t} \tag{14}$$

$$H_t = O_t * tanh(C_t) \tag{15}$$

For the traditional model, the weight of each input of the context vector is consistent, which leads to the insensitivity of the entire training model to important parts of the data, that is, the model is "distracted". Based on that, we introduced the attention mechanism into the LSTM model.

$$C_t = \sum_{j=1}^{T} w_{tj} h_j \tag{16}$$

$$w_{tj} = \frac{exp\left(h_j H_c^T \,/\, k\right)}{\Sigma_{k=1}^{T} exp(h_t H_c^T \,/\, k)} \tag{17}$$

where, $\boldsymbol{w_{tj}}$ is the attention weight.

iv). Joint Training Loss Function with L1 Regular Term

In order to realize the end-to-end training in accordance with the above composite model, we give the following specific computational loss function of its joint optimization training.

$$L(\psi) = E_P + \beta E_R + \lambda \|x - \hat{x}\|_1 + \alpha \|\psi\|_1 \tag{18}$$

where, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the weight coefficient parameters, $\boldsymbol{E_P}$ and $\boldsymbol{E_R}$ are prediction error and reconstruction error respectively.

By the joint loss function (18), we can carry out joint optimization training for the composite model mentioned in this paper, which is suitable for time series data in real industrial scenarios. The model architecture is shown in Figure 3.

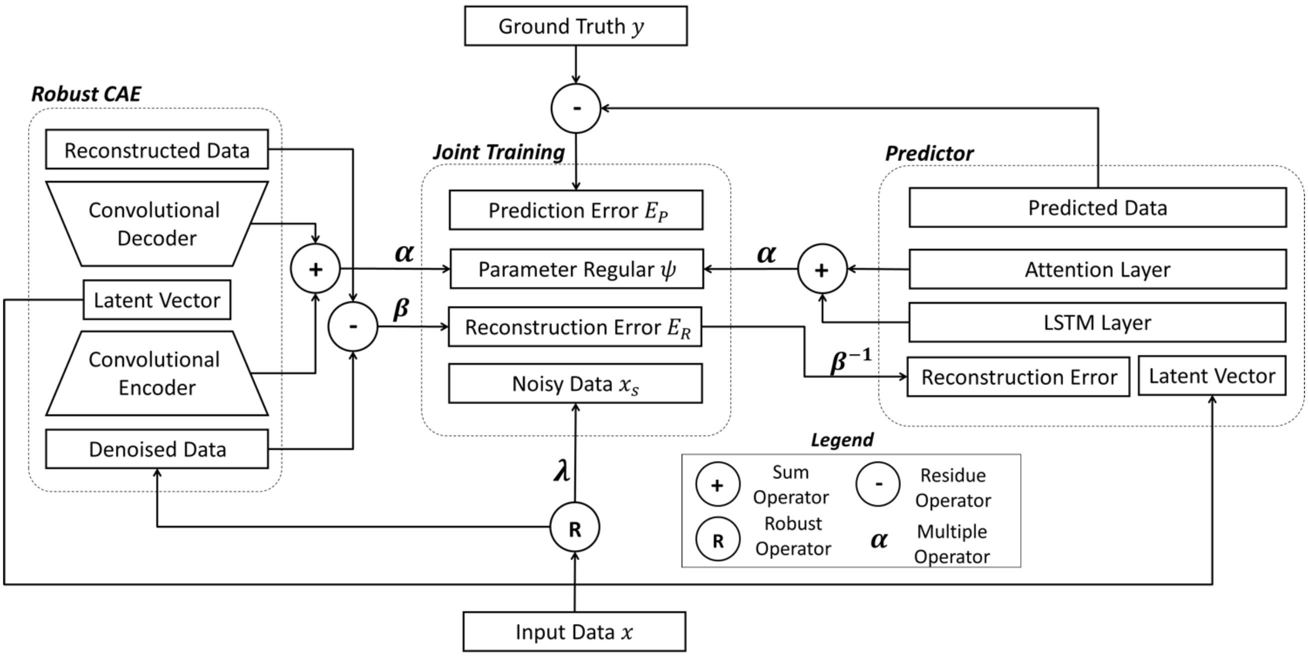


Figure 3. Composite Model

Detection Phase

During the detection phase, the physical process data describing the behaviors of monitored ICS physical assets are collected in real-time. They can be presented as $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N\} \in \mathbb{R}^{m \times N}$ where m is the number of sensors, and N is the number of time points. First, we generate sequence segments by sliding window of size $2T$. Next, we divide the $2T$ subsegments with the width of T. The first T subsegments $\boldsymbol{X}_P = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T\}$ are input into the Digital Twin model for prediction. The subsequent T subsegments $\boldsymbol{X}_C = \{\boldsymbol{x}_{T+1}, \boldsymbol{x}_{T+2}, \dots, \boldsymbol{x}_{2T}\}$ are used as the current behavior set of the monitored ICS assets.

By taking $\boldsymbol{X}_P = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T\}$ as input, the Digital Twin model outputs the predicted normal behavior $\boldsymbol{X}_C^{'} = \{\boldsymbol{x}_{T+1}^{'}, \boldsymbol{x}_{T+2}^{'}, \dots, \boldsymbol{x}_{2T}^{'}\}$. To further alleviate the predict fluctuation caused by process data noise and model bias, the Exponentially Weighted Smoothing method is applied to smooth the prediction result $\boldsymbol{X}_C^{'}$ [10]. The residual between $\boldsymbol{X}_C^{'}$ and $\boldsymbol{X}_C$ are regarded as the anomaly score: $\boldsymbol{E} = \boldsymbol{X}_C' - \boldsymbol{X}_C$. The alarm is triggered based a predefined threshold τ so that if the anomaly score exceeds the value of τ, an alarm indicating the occurrence of fault will be triggered.

III. Experimental Evaluation

A. Experimental Setup

Secure Water Treatment (SWaT) dataset is used as the experimental dataset to evaluate the performance of the DTHNN approach. It is a CPS benchmark dataset and open-accessed to research academy [11]. The dataset is collected from a realistic hardware-based CPS testbed created by Center for Research in Cyber Security of Singapore University of Technology and Design. The data collected from the testbed consists of 11 days of continuous operation. The normal data was collected under normal operation lasting for 7 days, while anomalous data was collected under 36 fault scenarios lasting for 4 days. The testbed is a water treatment system formalized as a SCADA system with 51 sensors and actuators and 3 controllers. The sensor and actuator data were collected as physical process data during data collection. The data are formatted as a matrix with 51 attributes. Unlike other simulated dataset such as dataset collected from Tennessee Eastman Process software testbed, the SWaT dataset consists of non-Gaussian noise and disturbance as well as data distribution variation (namely concept drift). Hence the dataset is more similar to the data of CPS in real world.

The performance evaluation metrics include Precision, Recall, and F1 values, which are calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{21}$$

where TP, FP, FN are true positive rate, false positive rate and false negative rate, respectively.

Experimental Results

The Digital Twin approaches based on traditional machine learning algorithm for comparisons include Principle Component Analysis (PCA) [4], One-class Support Vector Machine (OCSVM) [5], Isolation Forest (IForest) [6]. The Digital Twin approaches based on deep learning algorithm for comparisons include traditional Auto-encoder (AE) [7], One-Dimensional Convolutional Neural Network (1DCNN) [8] and LSTM based Encoder-Decoder (LSTM-ED) [9]. The experimental results shown in Table II indicate that the proposed DTHNN approach not only outperforms traditional machine learning algorithms but also outperforms existing deep learning algorithms.

Table II. Experimental Results

	SWaT Dataset		
	Pre	Rec	F1
PCA [4]	0.2492	0.2163	0.23
OCSVM [5]	0.9250	0.6990	0.7963
IForest [6]	0.1924	0.8347	0.3127
1DCNN [8]	0.867	0.754	0.7958
LSTM-ED [9]	0.9585	0.7151	0.8191
AE [7]	0.967	0.696	0.812
DTHNN	0.9406	0.7524	**0.8378**

IV. Conclusion and Future Works

In this paper, a Digital Twin approach based on Hybrid Neural Network is proposed to build the behavior model of sophisticated CPS and their assets. The robust convolutional autoencoder within the hybrid network is capable to perform dimensionality reduction and noise reduction and generate a latent vector which can abstract the cross correlations among process attributes. The LSTM network within the hybrid network is capable to capture the nonlinear temporal correlations of each process attribute, so that accurate prediction of normal behaviors can be performed based on latent vectors provided by auto-encoder. The two sub-models are integrated in an end-to-end fashion so that the parameters within the sub-models can be trained together to reach the global optimum.

One of the major future works is to deal with concept drift phenomenon which is widely occurred in CPS operation. The concept drift phenomenon is that the distribution of physical process data may vary with time due to asset degradation, change of configuration, modification of manufacturing process, etc. As a result, the Digital Twin models generated offline may not be suitable when concept drift occurs. A promising research direction is to apply incremental learning technique to the Digital Twin approach.

Acknowledgement

This work is supported by CCF-NSFOCUS KunPeng Research Fund (2018013), National Key Research & Development Program of China No. 2017YFC0803401 and the research on “Safety Protection technologies for Process Industry based on Adaptive Heterogeneous Neural Network”.

References

[1] General Electric Company, “Digital Twin for the Digital Power Plant,” in https://www.ge.com/digital/sites/default/files/download_assets/Digital-Twin-for-the-digital-power-plant-.pdf

[2] Fauri, D., Dos Santos, D. R., Costante, E., den Hartog, J., Etalle, S., & Tonetta, S. “From system specification to anomaly detection (and back),” in *Proceedings of the 2017 Workshop on Cyber-Physical Systems Security and PrivaCy*, 2017, pp. 13-24.

[3] Eckhart, Matthias, and Andreas Ekelhart. "A specification-based state replication approach for digital twins," in *Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and Privacy*. 2018, pp. 36-47.

[4] Li, Dan, et al. "Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks," in *International Conference on Artificial Neural Networks.* Springer, Cham, 2019, pp. 703-716.

[5] Inoue, Jun, et al. "Anomaly detection for a water treatment system using unsupervised machine learning," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2017, pp. 1058-1065.

[6] Liu, Fei Tony, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation forest," in *2008 Eighth IEEE International Conference on Data Mining*. IEEE, 2008, pp. 413-422.

[7] Kravchik, Moshe, and Asaf Shabtai. "Detecting cyber attacks in industrial control systems using convolutional neural networks," in *Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy*. 2018, pp. 72-83.

[8] Lai, Guokun, et al. "Modeling long-and short-term temporal patterns with deep neural networks," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 2018, pp. 95-104.

[9] Shalyga, Dmitry, Pavel Filonov, and Andrey Lavrentyev. "Anomaly detection for water treatment system based on neural network with automatic architecture optimization," in *arXiv preprint arXiv:1807.07282* ,2018.

[10] Shalyga, Dmitry, Pavel Filonov, and Andrey Lavrentyev. "Anomaly detection for water treatment system based on neural network with automatic architecture optimization," in *arXiv preprint arXiv:1807.07282* ,2018.

[11] Goh, Jonathan, et al. "A dataset to support research in the design of secure water treatment systems," in *International Conference on Critical Information Infrastructures Security.* Springer, Cham, 2016, pp. 88-99.

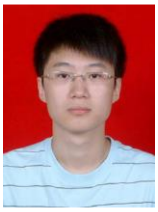
Zhiwen Pan, received the B.S. degree from the Purdue University Calumet, in 2012, and the M.S. and Ph.D. degrees from the University of Arizona, in 2014 and 2017. He is currently an Assistant Research Fellow in Research Center for Ubiquitous Computing System, Institute of Computing Technology, Chinese Academy of Science. His current research focuses on Anomaly Detection, Process Monitoring, Industrial Control System, and Deep Learning.

Jiahao Li, received the B.S. degree from the Shandong University in 2020, and is pursuing the M.S. degree from the Institute of Computing Technology, Chinese Academy of Sciences. He was algorithm engineer in Alibaba Cloud Computing Co. Ltd. His current research focuses on machine learning, anomaly detection, computer vision and ubiquitous computing.

Yiqiang Chen, received the B.S. and M.S. degrees in computer science from Xiangtan University, Xiangtan, China, in 1996 and 1999, respectively, and the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2003. In 2004, he was a Visiting Scholar Researcher with the Department of Computer Science, Hong Kong University of Science and Technology (HKUST), Hong Kong. He is currently a professor and the director of the Pervasive Computing Research Center at the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS). His research interests include artificial intelligence, pervasive computing, and hu-man-computer interaction

Salim Hariri received the M.Sc. degree from The Ohio State University, in 1982, and the Ph.D. degree in computer engineering from the University of Southern California, in 1986. He is currently a Professor with the Department of Electrical and Computer Engineering, The University of Arizona, and the Director of the NSF Center for Cloud and Autonomic Computing. His current research focuses on autonomic computing, cybersecurity, cyber resilience, secure critical infrastructures, and cloud security.

Jesus Pacheco. PhD. from the Electrical and Computer Engineering department of The University of Arizona. He is currently working as full professor at the University of Sonora in Sonora, Mexico. His research interest includes cyber security for critical infrastructures and cyber-physical.



Xiaofan Song received the B.S. and M.S. degree from the University of Manchester in 2012 and 2014. He is currently an Engineer in State Grid Henan Economic Research Institute. His current research focuses on Fault Detection for Power Systems and Anomaly Detection.