Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 183 (2021) 100–106

Procedia Computer Science

www.elsevier.com/locate/procedia

10th International Conference of Information and Communication Technology (ICICT-2020)

Research on equipment health prediction technology based on edge computing and VAE-TCN

Yanfang Yang[a,b,*], Bo Yu[a,b], Wei Wang[c]

[a]Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China
[b]University of Chinese Academy of Sciences, Beijing 100049, China
[c] School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China

Abstract

The Internet of Things technology is developing rapidly, and the data generated has also exploded. Traditional cloud computing technology can no longer meet the demand for efficient processing of massive data. Edge computing technology can move the amount of calculation down to the edge of the network, which can greatly improve computing efficiency. Applying edge computing to the field of equipment health prediction, the combination of strong responsiveness and computing capabilities of edge computing and high-precision prediction technology makes production operation and maintenance more reliable and efficient. At the same time, a neural network prediction model combining Variational Auto-Encoder (VAE) and Time Convolutional Network (TCN) is proposed to improve the accuracy of equipment health prediction. This model uses VAE for dimensionality reduction, extracts the hidden information in the original data, reconstructs high-quality sample data, and then uses TCN to mine the internal connection between the features and the target in the long sequence information. Compared with five benchmark prediction models on the C-MAPSS dataset, experiments show that the proposed model has higher prediction accuracy.

Keywords: edge computing;equipment health prediction;Variational Auto-Encoder;Time Convolution Network;

* Corresponding author. Tel.: +86-157-3851-9266
E-mail address: yangyanfang18@mails.ucas.ac.cn

1877-0509 © 2021 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of the scientific committee of the 10th International Conference of Information and Communication Technology.
10.1016/j.procs.2021.02.036

1. Introduction

As massive devices are connected to the Internet of Things, the acquisition of condition monitoring data has become more and more convenient, and the era of industrial big data is coming. Traditional cloud computing uses the way of centralized data processing, which can not process massive data efficiently. To solve this problem, edge computing has gradually become a research hotspot[1,2,3]. In edge computing, edge devices undertake some or all data processing tasks, and send the results to the cloud. Compared with traditional cloud computing, the data processing near the edge computing reduces the network data transmission and cloud load. Accordingly, the combination of edge computing and equipment health prediction will greatly enhance the level of industrial intelligent operation and maintenance.

Equipment health prediction technology[4] exists in many fields such as manufacturing, aerospace, and energy production[5,6,7]. As the core of equipment health prediction technology, remaining life prediction technology directly determines the accuracy and timeliness of system failure warning and maintenance decisions, and has important research value.

This paper uses data-based prediction methods, which can learn the trend of equipment degradation behavior and extract key information through the collected equipment operation data. It is easier to deal with complex modeling problems by constructing the mapping relationship between equipment operating data and remaining life through methods such as machine learning and mathematical statistics learning[8].

Equipment health prediction is a sequential problem. Recurrent Neural Network (RNN) is a classical model, but it has gradient vanishing problem. Later, the Long Short-Term Memory (LSTM) network was proposed, in which a gating mechanism was added to the RNN to realize the learning of long-term dependence. This method has achieved great success and has become the mainstream method to solve the sequential problem. With the passage of time, the Long Short-Term Memory network also exposed some shortcomings, such as the difficulty of processing long-term sequence and the large amount of calculation when the network is deep. In 2018, the Time Convolution Network (TCN)[9] was proposed to solve the concurrency problem of LSTM.

To improve the ability of mining and extracting effective information from monitoring data, this paper proposes a prediction model based on VAE-TCN. VAE[10] is used to reduce the dimension and extract the hidden layer information. The extracted feature information is transferred into TCN to mine the mapping relationship between the degradation information and the remaining life in the long sequence. Through research and comparison, the model improves the prediction accuracy.

2. Edge computing architecture

The equipment health prediction relies on the massive data generated by the equipment. If all the collected data is uploaded to the cloud for processing, it will inevitably cause huge network load. Edge computing does not require a response from the cloud through the network, which can greatly reduce system latency. Therefore, the Internet of Things based on edge computing can effectively construct equipment predictive maintenance programs.

Using edge computing technology to predict the health of equipment, it is first necessary to connect sensors at the industrial production site to collect the data required for equipment health prediction, and then transmit the collected data to the edge computing platform. The data is analyzed and preprocessed at the edge side, the invalid data is discarded, and the valid data is forwarded to the cloud. In the cloud, the received data is input into the equipment health prediction model for prediction. The architecture of the equipment health prediction system based on edge computing as shown in Fig. 1.

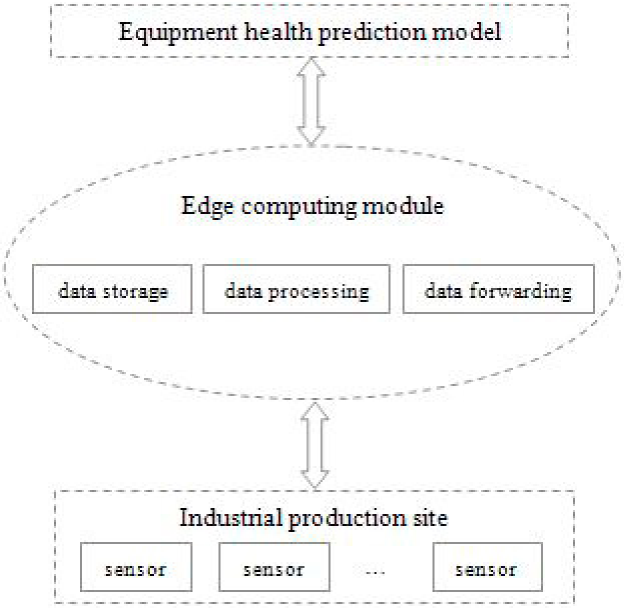

Fig. 1. Architecture of edge computing.

3. Equipment health prediction model

3.1. Variational Auto-Encoder

VAE is a variant of Auto-Encoder (AE), which is a semi-supervised model. The encoding network encodes the original data x to generate a hidden variable z, and the decoding network learns the mapping relationship between the hidden variable z and the data distribution p(x). The latent variable z and the data sample x can form a joint probability density distribution P(z,x). The training target of VAE can be regarded as the minimum KL divergence between the posterior distribution and the prior distribution. The objective function of VAE is shown in formula (1).

$$\mathrm{L}(\theta,\phi)=KL\big(p_\theta(z,x)\,\|\,q_\phi(z,x)\big)=\int p_\theta(z,x)\ln\frac{p_\theta(z,x)}{q_\phi(z,x)}dzdx \tag{1}$$

3.2. Time Convolution Network

TCN is essentially a special one-dimensional convolutional network that mines long-term sequence features through a multi-layer network. TCN adopts Causal Convolutions, there will be no future input information in the output, and historical data is not missed. Moreover, TCN uses one-dimensional full convolution structure to ensure that the input and output lengths of each hidden layer are the same.

In addition, in order to make the model achieve the effect of long-term historical memory, TCN uses dilated convolution to make the network have a larger receptive field and can obtain long-term historical information. The calculation formula of dilated convolution is shown in formula (2).

$$\mathrm{F}(s)=(\mathrm{x}*_d f)(s)=\sum_{i=0}^{k-1} f(i)\cdot x_{s-d\cdot i} \tag{2}$$

In formula (2): k is the size of the convolution kernel; d is the dilated coefficient of each layer; s-d•i is the index of the previous time-frequency information.

The entire TCN network is made up of multiple residual blocks. The residual connection ensures that the input and output have the same dimension in the TCN network design principle. For input X and dilated convolution operation F, the output definition of the residual block is shown in formula (3).

$$o = Activation(X + F(X)) \tag{3}$$

3.3. VAE-TCN prediction model

The idea of the VAE-TCN prediction model is to use VAE as a feature extraction tool to reduce the dimensions of high-dimensional data and extract the hidden layer information from the original data, and then input the feature information into the TCN neural network for learning, and mine The mapping relationship between degradation information and remaining life. The overall framework of the VAE-TCN prediction model as shown in Fig. 2.

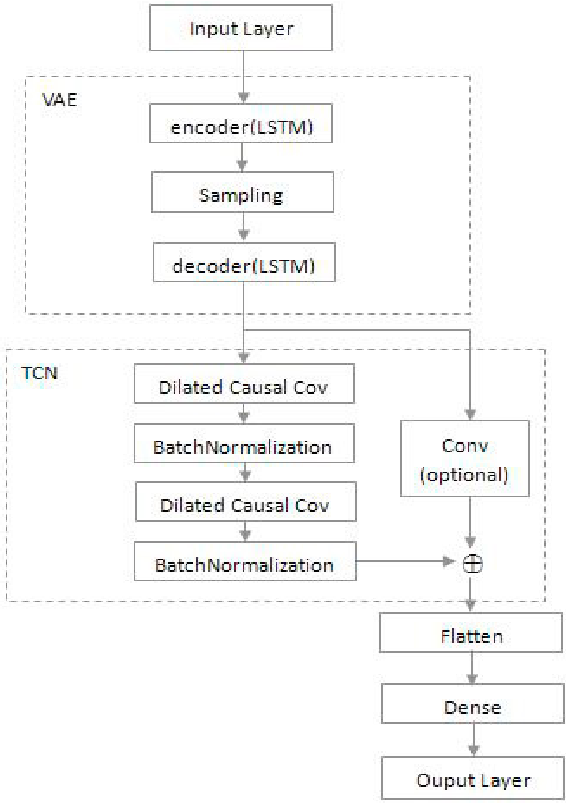

Fig. 2. Franework of VAE-TCN prediction model.

The encoder of VAE is used to learn the distribution of training data and generate compression values of the training data. The decoder of VAE reconstructs the compressed data to reconstruct high-quality data to eliminate interference samples. Both the encoder and decoder use LSTM[11].

TCN is composed of multiple residual blocks, and the number can be selected according to the dataset. Each residual block includes two BatchNormalization modules, two layers of dilated convolution, two ReLU activation functions and two Dropout layers. Then attach a one-dimensional full convolution to ensure that the input and output dimensions are equal.

4. Experiment and results analysis

4.1. Dataset description

The dataset is selected from NASA's turbofan engine dataset C-MAPSS, which is a benchmark dataset widely used in equipment remaining life prediction. There are four sub-data sets, which record the monitoring data of turbofan engine from normal operation to failure under different operating conditions and failure modes. This paper uses the first sub-data set FD001 to verify the model. The description of the FD001 as shown in Table 1.

Table 1.Information of dataset FD001.

name	training	test	condition	fault mode	sensor number	operating
FD001	100	100	1	1	21	3

In the training set of FD001, the complete process of 100 engines from normal operation to complete failure was recorded, and the process of 100 engines terminating before failure was recorded in the test set. The goal is to predict the remaining life of these 100 engines. The true remaining life has been given, which can be used to calculate the prediction accuracy of the model.

4.2. Data preprocessing

Before passing into the model, the engine data needs to be preprocessed. Since FD001 is a single operating condition, the three operating parameters may not participate in model training. Among the 21 sensor monitoring data of FD001, the sensors with sensor numbers of 1,5,6,10,16,18, and 19 are constant in the whole engine operating cycle, which make little contribution to the remaining life prediction, so they are discarded to reduce the cost of data processing.

Since each sensor returns different physical characteristic values, it is necessary to normalize the selected input variables to eliminate the influence of different dimensions on the prediction results. The MinMaxScaler method is selected to normalize the input data to [0,1].

4.3. Parameter configuration

Because the learning ability of this model is affected by parameter adjustment, it is very important to select the parameters appropriately. The main parameters involved in this model are: dilated convolution coefficient, dropout rate,Window size, etc. Find the optimal values one by one through experiments, the parameter configuration list as shown in Table 2.

Table 2.Parameter configuration list.

Parameter	Value
Optimizer	Adam
Dilated convolution cofficient	1,2,4,8
Kernel size	3
Dropout rate	0.5
Window size	30

4.4. Model training

There are 25 features in the original data, and 14 features after preprocessing. The preprocessed high-dimensional time series data is input into VAE for dimensionality reduction and feature extraction, and then input into TCN to learn the mapping relationship between feature information and remaining life. The test result as shown in Fig. 3.

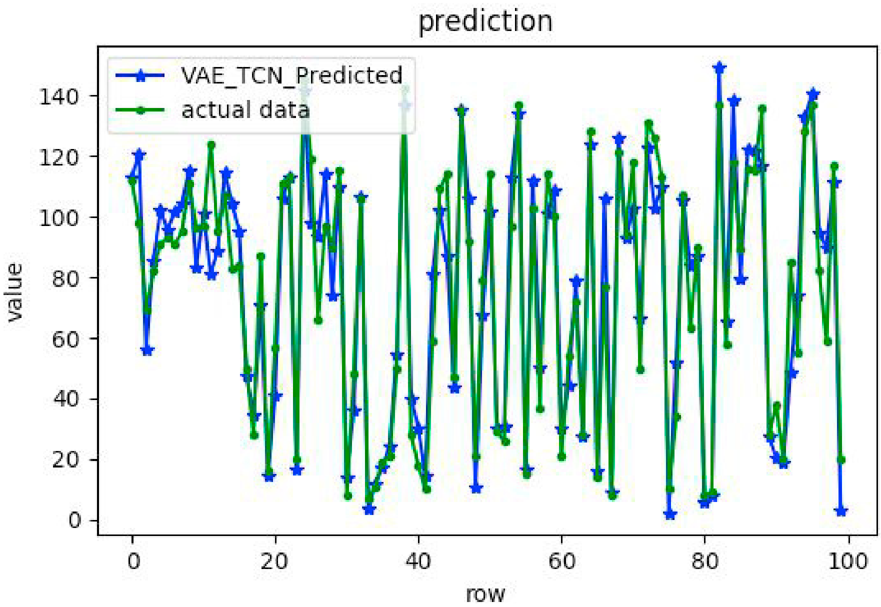

Fig. 3. RUL prediction result.

As can be seen, the RUL predicted by the model and is not much different from the real RUL. Especially the prediction effect is better when the engine life cycle is approaching the end stage. With the growth of the engine life cycle, the probability of failure gradually increases, and the prediction accuracy of the model gradually improves. In the actual working environment, engine failures mostly occur in the latter half of the life cycle, and the prediction accuracy at this stage has important guiding significance for predictive maintenance.

4.5. Analysis of experimental results

In order to quantitatively analyze the performance of the model and facilitate comparison with other algorithms, this paper selects the root mean square error (RMSE) to evaluate the prediction effect. RMSE is a common indicator for evaluating prediction errors, and the calculation formula as shown in formula (4).

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(X_i - Y_i\right)^2} \tag{4}$$

In formula (4): m is the number of turbofan engines; X_i is the actual RUL value; Y_i is the predicted RUL value.
Calculate the root mean square error of the real value and predicted value of different neural network models in the FD001, and the results of RMSE as shown in Table 3.

Table 3.RMSE of different neural network models.

Network model	SVM[12]	DCNN[13]	MLP[12]	D-LSTM[14]	MODBNE[12]	VAE-TCN
RMSE	40.72	18.45	16.78	16.14	15.04	**14.28**

As can be clearly seen from Table 3 that the error of VAE-TCN is the lowest level. Compared with the improved DCNN, D-LSTM and MODBNE, the prediction accuracy of the remaining life is further improved, and the RMSE of the prediction is reduced by 5% compared with the MODBNE prediction model.

5. Conclusion

In view of the high dimensionality and large sample size of condition monitoring data that equipment health prediction relies on, a combination of edge computing technology and VAE-TCN prediction model is proposed for analysis and prediction. Edge computing extracts effective information at the edge as the input of the prediction model, which greatly reduces the delay of network requests and the load on the cloud. VAE-TCN uses VAE for feature extraction, and TCN mines the mapping relationship between degradation information and remaining life in long sequences. It has been verified that in terms of equipment health prediction, the hybrid prediction model in this paper has higher prediction accuracy than SVM, DCNN, MLP, D-LSTM, and MODBNE. In the future, we will further study the influence of feature extraction methods and neural network architecture on the performance of prediction models, in order to obtain better prediction results.

References

1. Liu Y, Yang C, Jiang L, et al. Intelligent Edge Computing for IoT-Based Energy Management in Smart Cities. IEEE Network, 2019, 33(2):111-117.
2. Ren J, Pan Y, Goscinski A, et al. Edge Computing for the Internet of Things. IEEE Network, 2018, 32(1):6-7.
3. Pawani P, Jude O, Madhusanka L, et al. Survey on Multi-Access Edge Computing for Internet of Things Realization. IEEE Communications Surveys & Tutorials, 2018, PP:1-1.
4. Zhang LW, Lin J, Liu B, et al. A Review on Deep Learning Applications in Prognostics and Health Management. IEEE Access, 2019, 7: 162415-162438.
5. Wang D, Tsui K L, Miao Q. Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators. IEEE Access, 2018, 6(99):665-676.
6. Zhang S, Wang Y, Liu M, et al. Data-based Line Trip Fault Prediction in Power Systems Using LSTM Networks and SVM. IEEE Access, 2017:7675-7686.
7. Moleda M, Momot A, Mrozek D. Predictive Maintenance of Boiler Feed Water Pumps Using SCADA Data. Sensors, 2020, 20(2):571.
8. Yu W, Kim I Y, Mechefske C. Remaining Useful Life Estimation Using a Bidirectional Recurrent Neural Network Based Autoencoder Scheme. Mechanical Systems and Signal Processing, 2019, 129(15):764-780.
9. BAI S, KOLTER J Z, KOLTUN V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv: Learning, 2018.
10. Yoon A S, Lee T, Lim Y, et al. Semi-supervised Learning with Deep Generative Models for Asset Failure Prediction. 2017.
11. Wei W, Wu H, Ma H. An AutoEncoder and LSTM-Based Traffic Flow Prediction Method. Sensors, 2019, 19(13):2946-.
12. Zhang C, Lim P, Qin A K, et al. Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics. IEEE Trans Neural Netw Learn Syst, 2017, 28(10):2306-2318.
13. Babu G S, Zhao P, Li X L. Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life. International Conference on Database Systems for Advanced Applications. Springer, Cham, 2016.
14. Zheng S, Ristovski K, Farahat A, et al. Long Short-Term Memory Network for Remaining Useful Life estimation. 2017 IEEE International Conference on Prognostics and Health Management (ICPHM). IEEE, 2017.