Letter

Process Monitoring Based on Temporal Feature Agglomeration and Enhancement

Xiao Liang, Weiwu Yan, Yusun Fu, and Huihe Shao

Dear Editor,

This letter proposes a process-monitoring method based on temporal feature agglomeration and enhancement, in which a novel feature extractor called contrastive feature extractor (CFE) extracts the temporal and relational features among process parameters. Then the feature representations are enhanced by maximizing the separation among different classes while minimizing the scatter within each class.

Process monitoring is a popular research area in process industries. With manufacturing processes becoming complex and intelligent, the demand for the safety and product quality is growing significantly. Process monitoring has played a crucial role in maintaining efficient and safe operating conditions in large-scale industrialized production. Data-driven process monitoring is one of the most fruitful areas in research and widely used in industrial applications over the last two decades [1] and [2].

Data-driven process monitoring is usually implemented using machine learning methods and multivariate statistical analyses. The key issue of data-based process-monitoring methods is to extract feature representations from industrial process data based on which the statistics of the monitoring model can be constructed in feature-representation space. Various supervised-learning methods have been introduced to learn feature representation for building process-monitoring models. PCA is the most widely used linear process-monitoring method. Nonlinear and robust methods, such as nonlinear PCA, the principal curve method, multimodal, manifold learning, and kernel-based methods (KPCA, KPLS, KICA, KFDA, etc.) are proposed to address nonlinearity and uncertainty in complex industrial processes [1]. Some methods such as manifold learning and graphical model are proposed to extract relational complicated features of industrial process data. Those process-monitoring methods extract static relationships among process variables without regard to the temporal information of process variables. Considering that the temporal variation can indicate dynamic performance changes, process-monitoring methods considering process temporal behaviors are becoming focus issues in process-monitoring researches. Dynamic latent variable (DLV) methods have been studied for time-series monitoring models by exploiting dynamic relations among process variables. To make full use of label information, supervised approaches such as Bayesian network (BN) and support vector machine (SVM) are used to coach the training of feature extraction

Corresponding author: Weiwu Yan.

Citation: X. Liang, W. W. Yan, Y. S. Fu, and H. H. Shao, "Process monitoring based on temporal feature agglomeration and enhancement," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 3, pp. 825–827, Mar. 2023.

X. Liang, Y. S. Fu, and H. H. Shao are with the Department of Automation, the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: liang_xiao@sjtu.edu.cn; fu_yusun@sina.com; hhshao@sjtu.edu.cn).

W. W. Yan is with the Department of Automation, the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, and with the Key Laboratory of System Control and Information Processing, Ministry of Education of China, and also with Shanghai Engineering Research Center of Intelligent Control and Management, Shanghai 200240, China (e-mail: yanwwsjtu@sjtu.edu.cn).

Digital Object Identifier 10.1109/JAS.2023.123114

[2]. Recently, deep learning is introduced into process monitoring for outstanding performance in extracting robust features and learning nonlinear feature representations. Deep learning-based process-monitoring models show excellent and promising performance.

Process-monitoring researches of industrial processes have obtained considerable achievements in the past few decades. However, it is extremely challenged to create process-monitoring models with high performance for large-scale, time-variant and complex industrial processes. Currently, temporal and relational features of the process data contain important information in time-variant and complex industrial production systems.

The motivation of this letter is to exploit temporal and relational information from time-variant and complex industrial process data to create process-monitoring models. Transformer-based feature extractor is introduced to abstract feature representations from time-variant process data through temporal and relational information agglomeration. Feature evaluation and feature enhancement are utilized to produce robust and interpretable feature representations through maximizing the separation among different classes while minimizing the scatter within each class.

The contributions of the letter are summarized as follows.

1) This letter proposes a process-monitoring framework based on temporal feature agglomeration and enhancement for time-variant and complex industrial processes.

2) Transformer-based feature extractor abstracts feature representations with long-term dependencies and relational information from time-variant process data.

3) The feature representations are enhanced by feature evaluation and contrastive learning to improve the robustness and generalization of the process-monitoring model.

4) The experimental results show the prominent performance of the proposed method in fault diagnosis and fault detection on the additional TE process dataset.

Related work: Some researches of process monitoring focus on extracting temporal and relational information from industrial process data.

To address the temporal feature representation and dynamic modelling issues, researchers have developed several extensions of traditional principal component methods. Ku *et al.* [3] proposed a dynamic PCA (DPCA) model, which performs classical PCA on augmented measurements with certain time lags. Li *et al.* [4] improve PCA by introducing spatiotemporal methods to integrate spatial and temporal prior into feature representations. Recurrent neural network (RNN) is used by Kiakojoori ang Khorasani [5] to represent temporal information through state inheritance. The aforementioned methods attempt to derive serial correlations between current and previous observations employing settling time. However, it is difficult for those methods to build temporal process-monitoring models with long-term dependences.

Robustness and interpretability are significant requirements for process-monitoring models. Several robust process-monitoring models employ some form of prior knowledge and expert knowledge to generate relational feature representations. Signed directed graph (SDG) is a knowledge-based fault diagnosis method, which can efficiently represent relationships among process variables and determine the fault root cause [6]. Considering that a graphical model has an easily interpreted physical meaning, it has been introduced into process monitoring of complex industrial processes. Several graphical models, such as decision trees and causal graphical models [7] and [8], have been applied in the process monitoring field. Deep
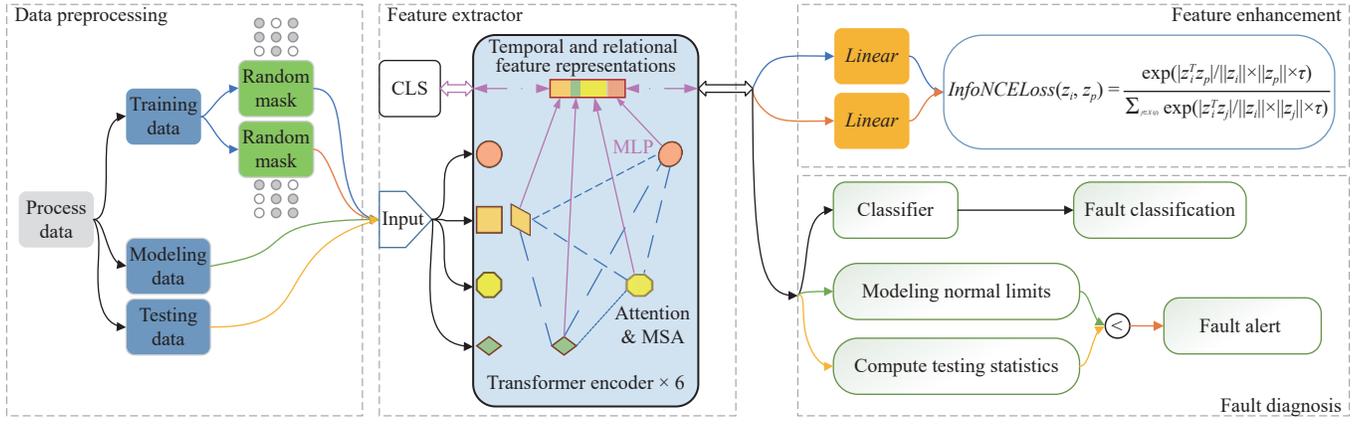
Fig. 1. The process diagram of contrastive feature extracto.

learning-based process monitoring models show well robustness and generalization [9]. However, those deep learning-based process-monitoring methods lack interpretability.

Method: Aiming for the robustness and interpretability of feature representation, the letter proposes a process-monitoring framework based on temporal feature agglomeration and enhancement. As shown in Fig. 1，the proposed framework mainly consists of four parts, which are data preprocessing, feature extractor, feature enhancement and fault diagnosis, respectively. The data preprocessing augmentations such as random masking and adding Gaussian noise generate diversified input data. The feature extractor abstracts and condenses feature representations with temporal and relational information. Then the feature representation is enhanced by implementing feature evaluation and contrastive learning. Finally, enhanced features are utilized to build the process monitoring model.

The temporal and relational information among industrial process data provides valuable information for industrial process monitoring. Transformer [10] is employed to extract temporal and relational features of process variables in process monitoring. In Transformer encoder, attention mechanisms establish the connections between timestamps and variables of diversified input data. MLP-based feed-forward layers are used to agglomerate relational information from the variable connections. Classification vector (CLS) is used to agglomerates the timestamp connections through backpropagation. The temporal and relational information agglomeration provides interpretable and comprehensive feature representations for process-monitoring modelling.

In RNN-based process monitoring, the temporal and relational information is represented by sequential hidden states. However, it is difficult for sequence hidden states to build long-term dependence because the deviations easily accumulate over time. Instead of building sequential dependencies, attention mechanisms obtain long-term dependency by calculating connections between long-term nodes. Multi-head self-attention (MSA) generates robust feature representations of process data by integrating multiple sets of attention. Attention and MSA are calculated as follows:

$$Attention_i(Q_i, K_i, V_i) = soft\max(Q_i K_i^T / \sqrt{d_K}) V_i \quad (1)$$

$$MSA = concat(Attention_1, Attention_2, \ldots, Attention_T). \quad (2)$$

For head i, $1 \le i \le T$, Q_i, K_i, $V_i Q_i$, K_i, V_i are different mappings of the input data X_i and d_K is the size of embedding for normalization.

In industrial processes, process data in the same operation mode share feature similarities, while process data in the different operation modes have significant differences. Therefore, ideal feature representations for process monitoring should maximize the separation among different classes and minimize the scatter within each class.

Feature evaluation and contrastive learning are used to enhance the discrimination and interpretability of features.

Firstly, an evaluation criterion based on similarity is introduced for feature enhancement. The similarity between features is calculated by cosine similarity as below:

$$sim(z_i, z_j) = \left| z_i^T z_j \right| / \|z_i\| \times \|z_j\| \quad (3)$$

where z_i and z_j represent two different features of the operation state. If z_i and z_j belong to the same mode, $sim(z_i, z_j)$ should be a large value. Based on the similarity between features, the evaluation criterion of the feature extractor is calculated as follows:

$$D(X) = \frac{1}{N_X} \sum_{i \in X} \frac{\sum_{j \in P(i), j \neq i} sim(z_i, z_j)}{\sum_{k \in X, k \neq i} sim(z_i, z_k)}$$

$$= \frac{1}{N_X} \sum_{i \in X} \frac{\sum_{j \in P(i), j \neq i} \left| z_i^T z_j \right| / \|z_i\| \times \|z_j\|}{\sum_{k \in X, k \neq i} \left| z_i^T z_k \right| / \|z_i\| \times \|z_k\|} \quad (4)$$

where X represents the input minibatch of the state information I_{train}, N_X is the number of the inputs. i, j and k represent vectors in the inputs, and z_i, z_j and z_k denotes their corresponding features respectively. $P(i)$ is a set of features representing the same operation mode in the industrial process.

Then, contrastive learning is used to maximize the separation among different classes and minimize the scatter within each class to enhance the discrimination of features. Contrastive learning is widely used in image classification to distil information by self-supervision [11] and [12]. The loss function of contrastive learning is derived from the evaluation criterion $D(x)$ and specially designed for different types of process data.

For unlabeled process data, the self-supervised method is utilized to generate robustness features and eliminate human cognitive bias. The loss function of self-supervised contrastive feature extractor (Self-CFE) is shown as follows:

$$InfoNCELoss(z_i, z_i') = -\log \frac{\exp(|z_i^T z_i'| / \|z_i\| \times \|z_i'\| \tau)}{\sum_{j \in X(i)} \exp(|z_i^T z_j| / \|z_i\| \times \|z_j\| \tau)} \quad (5)$$

where z_i and z_i' are features of i with different augmentations. $X(i)$ includes vectors except i in the training batch as the contrastive information. τ is the temperature to control the loss function.

For labelled process data, the label information is utilized to improve feature discrimination. The loss function of supervised contrastive feature extractor (Sup-CFE) is shown as (6).

$$Sup\_InfoNCELoss(z_i, z_i')$$

$$= -\frac{1}{N_{P(i)}} \sum_{p \in P(i)} \log \frac{\exp(|z_i^T z_i'| / \|z_i\| \times \|z_i'\| \tau)}{\sum_{j \in X(i)} \exp(|z_i^T z_j| / \|z_i\| \times \|z_j\| \tau)}. \quad (6)$$

Self-CFE and Sup-CFE enhance feature discrimination through exploiting unlabeled and labelled process data.

The complete training process of the contrastive feature extractor is shown in Algorithm 1.

Algorithm 1 The process monitoring of contrastive feature extractor

---

Input: Training input I_{train}, Batch size N, Test input I_{test};

For minibatch $X \in I_{\text{train}}$ do

For $x \in X$ do

$\quad z_i = Linear(TransformerEncoder(RandAug(x)))$

$\quad z'_i = Linear(TransformerEncoder(RandAug'(x)))$

End For

$\quad L = 0$

For all $z \in (z_i, z'_i)$ do

$\quad z_i \leftarrow z,\ z'_i \leftarrow$ Corresponding vector in Z

$\quad Loss(z_i, z'_i) = InfoNCELoss(z_i, z'_i)$ or $Sup\_InfoNCELoss(z_i, z'_i)$

$\quad L = L + \dfrac{1}{2N}\left(Loss(z_i, z'_i) + Loss(z'_i, z_i)\right)$

End For

Update L for $TransformerEncoder$ and $Linear$

End For

$Features = TransformerEncoder(I_{\text{test}})$

$Create \rightarrow$ Statistic T^2, SPE and $Classifier$

End

Experiment: Tennessee Eastman (TE) process is a well-known process dataset, which has been widely used as a benchmark in process monitoring and fault diagnosis. We use the Additional TE dataset from Harvard which has larger training and testing datasets.

An additional linear classifier is trained to classify faults with feature representations. The Self-CFE and Sup-CFE are tested on TE process data. The main parameters of process monitoring model are shown in Table 1.

Table 1. Parameters of Contrastive Feature Extractor

Names	Sample length	Process variables	Random mask rate	Encoder number	Head number	MLP layer
Value	21	52	0.2	6	4	1

Diagnosis accuracies (%) of fault 1−20 by different methods on the additional TE are shown in Table 2. It is clearly seen that CFE-based methods achieve prominent performance on most of the faults. Sup-CFE gets the best average accuracy and improves fault classification results significantly on fault 16. Self-CFE also gets good performance on faults 4, 7, 14 and 20 with unlabeled data. It should be mentioned that DCNN [13] gets well accuracy for faults 3, 9, 10 and 12.

Conclusion: The letter discusses a process-monitoring framework based on temporal feature agglomeration and enhancement. Transformer is introduced as the feature extractor to extract temporal and relational information of the variant-time process data. Feature evaluation and contrastive learning are used to enhance the discrimination of features. The proposed method provides a promising monitoring framework for large-scale, time-variant and complex modern industrial processes.

Table 2. Diagnosis Accuracy (%) of Fault 1–20 Based on Different Methods on Additional TE Dataset

Fault	PCA	IPCA	KPCA	MLP	DCNN	Self-CFE	Sup-CFE
IDV(1)	53.65	82.99	82.83	98.74	98.6	99.38	**99.83**
IDV(2)	78.00	82.17	81.90	98.16	98.5	99.75	**99.99**
IDV(3)	0.03	0.04	0.10	13.48	**91.7**	52.9	66.45
IDV(4)	83.34	42.17	9.83	100.00	97.6	**100**	**100**
IDV(5)	5.65	9.74	4.13	99.85	91.5	99.52	99.44
IDV(6)	83.50	83.50	83.50	98.91	97.5	92.45	**100**
IDV(7)	61.33	83.50	82.18	100.00	99.9	**100**	**100**
IDV(8)	65.49	79.43	75.88	51.79	92.2	66.61	**94.38**
IDV(9)	0.04	0.05	0.11	13.85	**58.4**	37.2	34.17
IDV(10)	15.26	2.90	0.90	11.61	**96.4**	61.32	90.26
IDV(11)	45.39	39.19	23.86	8.84	98.4	97.87	**99.36**
IDV(12)	74.00	79.15	71.34	22.35	**95.6**	64.319	76.94
IDV(13)	76.91	77.58	75.40	56.61	95.7	79.15	**96.09**
IDV(14)	73.50	83.43	83.36	17.30	98.7	**100**	**100**
IDV(15)	0.05	0.06	0.11	9.40	28.0	40.03	**59.78**
IDV(16)	10.47	0.64	0.29	4.79	44.2	40.47	**93.81**
IDV(17)	75.57	66.74	61.69	84.64	94.5	39.37	**97.53**
IDV(18)	77.14	77.09	76.25	87.94	93.9	90.2	**95.55**
IDV(19)	1.25	0.69	0.46	25.68	98.6	99.76	**100**
IDV(20)	40.02	23.79	17.10	71.25	93.3	**99.67**	92.44
AVG	46.03	45.74	41.56	48.63	88.2	79.21	**89.42**

References

[1] M. Nawaz, A. S. Maulud, *et al.*, "Review of multiscale methods for process monitoring with an emphasis on applications in chemical process systems," *IEEE Access*, vol. 10, pp. 49708–49724, 2022.

[2] S. A. A. Taqvi, H. Zabiri, *et al.*, "A review on data-driven learning approaches for fault detection and diagnosis in chemical processes," *ChemBioEng Reviews*, vol. 8, no. 3, pp. 239–259, 2021.

[3] W. Ku, R. H. Storer, *et al.*, "Disturbance detection and isolation by dynamic principal component analysis," *Chemometrics and Intelligent Lab. Syst.*, vol. 30, no. 1, pp. 179–196, 1995.

[4] Y. Li, X. Xiu, *et al.*, "Towards efficient process monitoring using spatiotemporal PCA," *IEEE Trans. Circuits Systems II: Express Briefs*, vol. 70, no. 2, pp. 551–555, 2023.

[5] S. Kiakojoori and K. Khorasani, "Dynamic neural networks for gas turbine engine degradation prediction, health monitoring and prognosis," *Neural Computing and Applications*, vol. 27, no. 8, pp. 2157–2192, 2016.

[6] M. A. Kramer and B. L. Palowitch Jr, "A rule-based approach to fault diagnosis using the signed directed graph," *AIChE J.*, vol. 33, no. 7, pp. 1067–1078, 1987.

[7] X. Zhou, T. Tian, *et al.*, "Research on transformer fault diagnosis technology based on adaboost-decision tree and DGA," in *Proc. 16th Annual Conf. China Electrotechnical Society*, 2022, pp. 1179–1189.

[8] C. Dong and J. Zhou, "A new algorithm of cubic dynamic uncertain causality graph for speeding up temporal causality inference in fault diagnosis," *IEEE Trans. Reliability*, 2022. DOI: 10.1109/TR.2022.3170063.

[9] L. Saidi, J. B. Ali, *et al.*, "Application of higher order spectral features and support vector machines for bearing faults classification," *ISA Trans.*, vol. 54, pp. 193–206, 2015.

[10] A. Vaswani, N. Shazeer, *et al.* "Attention is all you need," in *Proc. Advances Neural Information Processing Systems*, Long Beach, USA, 2017, pp. 5998–6008.

[11] T. Chen, S. Kornblith, *et al.*, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Machine Learning*, 2020, pp. 1597–1607.

[12] P. Khosla, P. Teterwak, *et al.*, "Supervised contrastive learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 18661–18673, 2020.

[13] H. Wu and J. Zhao, "Deep convolutional neural network model based chemical process fault diagnosis," *Computers & Chemical Engineering*, vol. 115, pp. 185–197, 2018.