Multi-Source Transfer Learning for Chemical Process Fault Diagnosis with Multi-Channel Feature Extraction

Ruoshi Qina, Jinsong Zhaoa,b,\*

a State Key Laboratory of Chemical Engineering, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China

b Beijing Key Laboratory of Industrial Big Data System and Application, Tsinghua University, Beijing 100084, China

jinsongzhao@tsinghua.edu.cn

Abstract

In recent times, there has been a rising preference for employing deep learning models in intelligent chemical process fault diagnosis. However, a considerable portion of the established methodologies operate on the premise that both training and testing data stem from identical feature distributions, which proves inaccurate in real-world scenarios characterized by multiple working conditions. In order to facilitate the preservation of domain-specific characteristics and the extraction of common features across both domains simultaneously, a novel domain adaptation deep network with multi-channel feature extraction is proposed in this research. The model employing a Transformer-CNN-based feature extractor and a domain adaptation module with polynomial kernel-induced maximum mean discrepancy is expected to achieve accurate anomaly diagnosis from diverse working conditions in one process and similar processes. Experiments on the Tennessee Eastman process and an industrial case of fluid catalytic cracking prove the effectiveness and advancement of the proposed method.

**Keywords**: Fault diagnosis, Transfer learning, Domain adaptation, Multi-channel feature extraction, Tennessee Eastman process, Fluid catalytic cracking.

* 1. Introduction

Intelligent fault diagnosis techniques assume a pivotal role in process monitoring to ensure manufacturing safety. In recent years, many scholars have made great achievements in the field of fault diagnosis application with the development of deep learning. However, the current application of deep learning in chemical process fault diagnosis relies on two critical assumptions that sufficient abnormal data is labeled for model training and the test dataset is identically distributed as the training dataset. In most practical situations, faulty samples in chemical processes are scarce and the distributions of data from different working conditions or multiple similar facilities are distinct (Bi et al., 2022). While transfer learning presents a promising avenue for the transfer of knowledge across diverse domains (Qin et al., 2022), it is worth noting that published studies that focused on the scenario of multimode and multi-source transfer learning are still rare. Merely transferring all source samples to a shared feature space and utilizing the cross-domain common features for fault diagnosis cannot guarantee high classification accuracy (Xiao et al., 2022).

The latest studies reveal that a domain adaptation deep network can be specially designed in a multi-channel form to learn the domain-invariant features more effectively. Lu (2023) designed a multi-view and multi-level network (MMNet) for rotating machinery fault diagnosis. Zhu (2023) launched the Transformer-convolutional neural network (TrCNN) based multi-scale distribution alignment network to extract deep diagnostic information and align the characteristics from various aspects. But they are all limited to the simulation situation and hard to apply in real industry.

In this paper, a novel multi-source domain adaptation deep network with multi-channel feature extraction is proposed to increase the generalization ability of the fault diagnostic model. The multi-source domain adaptation network (MSDAN) employs a Transformer-CNN-based multi-channel architecture to achieve the extraction of the common features across domains and the specific features in respective domains. The domain-specific information inherent within samples that have been deemed unsuitable for transfer is proficiently attenuated to facilitate exacting classification. A multi-source domain adaptation module based on polynomial kernel-induced maximum mean discrepancy (PK-MMD) is introduced to accomplish distribution adaptation among the source and target domains. The domain-invariant learned features are utilized for fault classification evaluated by a relation score from the query sample and the template ones.

The remaining parts of the article are arranged as follows. In the upcoming section, the proposed method MSDAN is established with the basic theory of Transformer-CNN model and PK-MMD algorithm. Section 3 elaborates on the experiments on the lab simulation and the real plant and compares the results among different popular transfer learning models. The conclusion and outlook of this article are drawn in Section 4.

* 1. Multi-Source Domain Adaptation Network
     1. Transformer-CNN Model

A standard Transformer consists of an encoder and a decoder, both sharing a similar design featuring multi-headed attention layers, feed-forward network layers, residual connections, and normalization layers. The pivotal multi-head attention mechanism facilitates effective interaction between the dual parts. CNN operates as a multi-layer feed-forward neural network, systematically extracting features through the arrangement of convolutional layers and pooling layers. It culminates with a fully connected layer to amalgamate local information and with a classifier like Softmax to accomplish categorization. By connecting Transformer in tandem with CNN to form the backbone network of the feature extractor, the diagnostic model can better extract the multiscale features of time-series data. The overall architecture borrows the module arrangement from relation network, and each module contains two branches indicated as source branch and cross-domain branch, which process the input normal and faulty samples respectively.

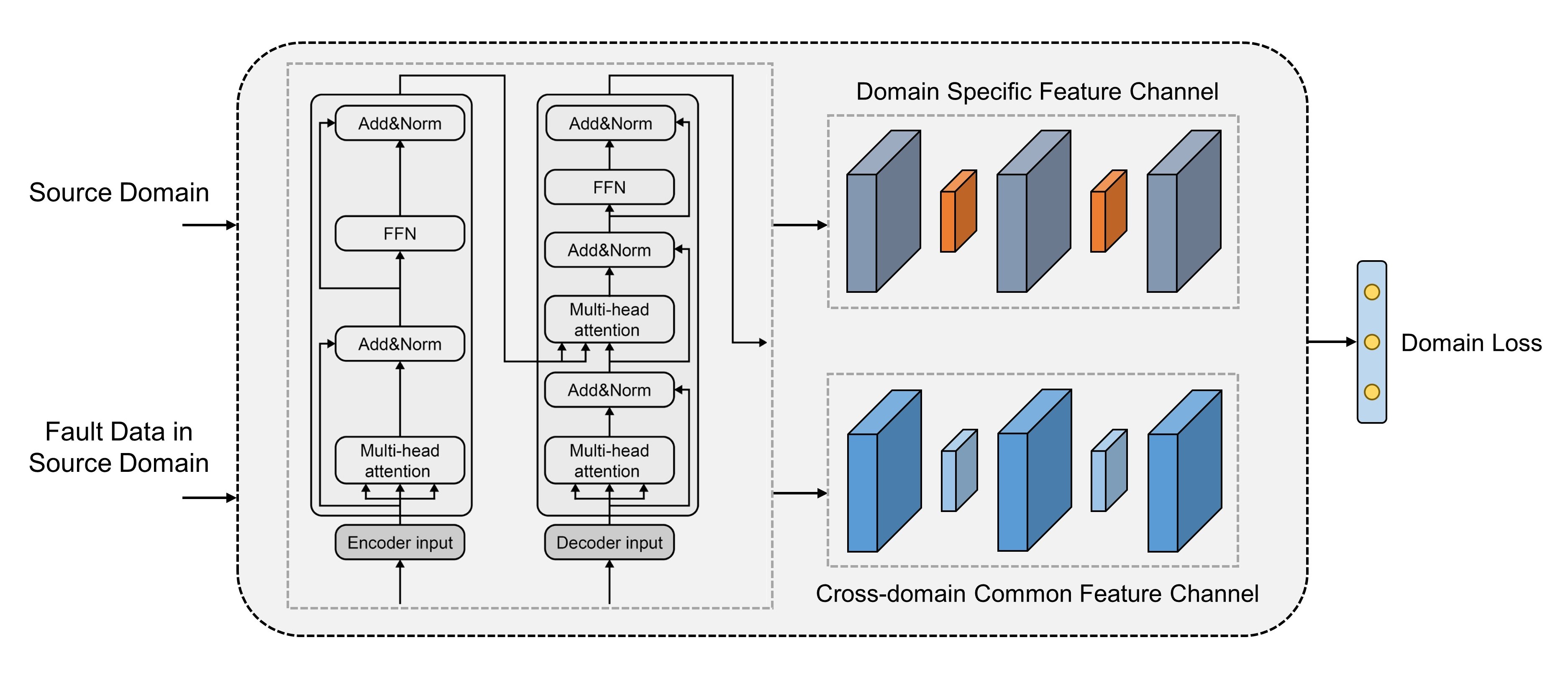


Fig.1. Transformer-CNN model structure for source domain.

* + 1. Polynomial Kernel-induced Maximum Mean Discrepancy

Maximum mean discrepancy (MMD) is a widely used method to assess how transferable features differ in their distribution. Specifically, these features are initially transformed into a Reproduced Kernel Hilbert Space (RKHS), within which the average distance between these features is considered as the metric indicating their distribution discrepancy. Gaussian kernels are commonly utilized to induce RKHS for estimating MMD. However, there are notable shortcomings associated with diagnosis models that employ Gaussian kernel-induced MMD (GK-MMD). Firstly, GK-MMD relies on the mean distance but ignores high-order moments. This limitation hinders the accurate assessment of distribution discrepancy in transferable features. Secondly, the high time complexity of GK-MMD demands significant computation resources during the model training. Lastly, the transfer performance of GK-MMD-based models is highly sensitive to the Gaussian kernel parameters, impeding convergence to an optimal point during the distribution adaptation process.

In order to overcome the identified weaknesses, a refined distance metric that uses polynomial kernels to induce MMD instead of Gaussian kernels is expressly crafted. Polynomial kernel-induced MMD (PK-MMD) not only computes the weighted sum of order-wise moment distances but also possesses the ability to adjust kernel parameters, offering potential solutions to address a range of diverse transfer learning tasks.

Give the function of polynomial kernels as

|  |  |
| --- | --- |
|  | (1) |

where the output of polynomial kernels is determined by the slope , the intercept , and the order .

By means of the binomial theorem, the empirical estimation of PK-MMD can be calculated by

|  |  |
| --- | --- |
|  | (2) |
|  | (3) |

Under the condition , the expression given in (2) symbolizes the distance between the means, explicitly denoting the first moment of the provided datasets and . And when , the formula is construed as a weighted summation of distances, encompassing the order-wise central moments within these datasets. The coefficients, represented by the slope and the intercept , serve as weights to balance the influences of both low-order and high-order moments in dataset pairs. A higher value for the slope diminishes the impact of high-order moments on the distribution discrepancy observed between paired datasets.

* + 1. MSDAN-based fault diagnosis framework

The framework of the proposed MSDAN-based fault diagnosis method is shown in Fig.2. In this domain adaptation deep network, the adjustment of network weights aims to enhance the network's classification performance. Consequently, samples unsuitable for domain adaptation exert reduced influence. Effective feature extraction occurs only from samples conducive to aligning the source and target domains. To extract both cross-domain common features and domain-specific features, multiple similar isolated network channels are designed in MSDAN. In each source channel, the dual branches share the same weights. The cross-domain common feature channel aims at extracting the common features via domain adaptation, while the domain-specific feature channel extracts the discriminant features facilitating both fault classification and domain classification. PK-MMD-based domain adaptation is utilized to acquire shared features across both the source and target domains. The model design utilizes historical data from chemical processes. Leveraging transferred knowledge, the diagnosis network oversees the online process and provides the fault classification outcome.

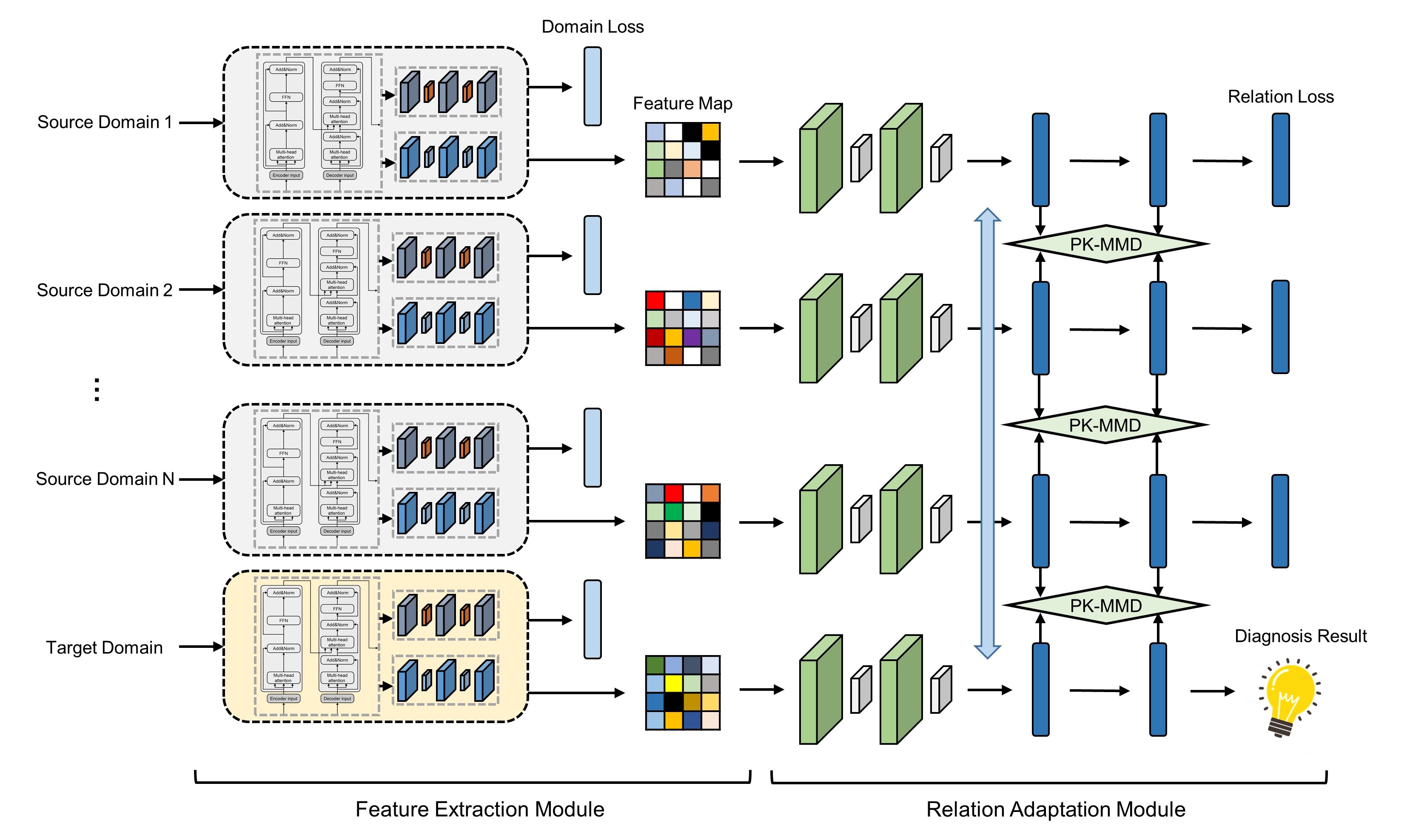


Fig.2. Architecture of MSDAN.

* 1. Case Studies

In this section, the benchmark Tennessee Eastman process (TEP) and the industrial fluid catalytic cracking (FCC) are applied to evaluate the performance of the proposed model.

|  |  |
| --- | --- |
| C:\Users\qrsh1\AppData\Local\Microsoft\Windows\INetCache\Content.Word\5-P&ID of Tennessee Eastman process.jpg | G:\Tencent\WeChat\WeChatFiles\WeChat Files\wxid_z1oclgj5qhwy12\FileStorage\Temp\2d5985650703aae32eb5a2a7526baf1.jpg |
| Fig.3. P&ID of TEP. | Fig.4. Scene of FCC case. |

* + 1. TEP

The TEP holds widespread usage in tasks related to monitoring chemical processes. Its unit operations encompass a reactor, condenser, recycle compressor, vapor-liquid separator, and stripper. This study builds upon the revised version of TEP as outlined in Bathelt (2015) at http://depts.washington.edu/control/LARRY/TE/download.html. The simulation involves 12 process-manipulated variables and 30 continuous process measurements. 19 fault types and 5 steady-state operating modes are introduced in this research to explore the multimode fault diagnosis performance. The selection of fault types and operational conditions mirrors prior published work, and the datasets are prepared in alignment with Wu (2020).

To explore the generalizability of deep transfer learning methods, the model trained for the source mode has to be applied to the new target mode. Note that there is no label information in all target domain data. Compared with the other transfer learning models (Qin et al., 2023), such as joint adaptation networks along with CNN (JAN-CNN) and dynamic adversarial adaptation network (DAAN), the proposed method achieves better domain alignment and higher classification accuracy in most scenarios. In particular, the more information from different source domains is imported for training, the more accurate unsupervised fault diagnosis is achieved.

Table 1. Unsupervised fault diagnosis performance of TEP.

1. JAN-CNN

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | + | ++ | +++ |
|  | - | 96.0% | 96.4% | 91.1% | 97.6% | - | - | - |
|  | 95.6% | - | **95.9%** | 90.8% | 96.4% | - | - | - |
|  | 93.5% | 95.6% | - | 89.6% | **97.3%** | 95.8% | - | - |
|  | **88.8%** | 89.7% | 88.2% | - | 90.6% | **91.2%** | 92.4% | - |
|  | 93.9% | 95.8% | 97.3% | 86.9% | - | 95.9% | 97.8% | 96.5% |

1. DAAN

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | + | ++ | +++ |
|  | - | **97.2%** | 96.8% | 92.5% | 97.3% | - | - | - |
|  | 96.6% | - | **95.9%** | 93.9% | 96.7% | - | - | - |
|  | 95.5% | 96.1% | - | **92.0%** | 97.2% | 96.9% | - | - |
|  | 87.2% | **91.7%** | **90.4%** | - | 91.1% | 90.8% | 93.0% | - |
|  | 94.7% | 96.0% | 97.3% | 88.4% | - | 96.2% | 97.8% | 97.2% |

1. MSDAN

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  | + | ++ | +++ |
|  | - | 96.7% | **97.1%** | **93.8%** | **98.2%** | - | - | - |
|  | **97.1%** | - | 95.6% | **94.0%** | **97.6%** | - | - | - |
|  | **96.6%** | **96.8%** | - | 91.5% | **97.3%** | **98.1%** | - | - |
|  | 86.4% | 91.5% | 90.3% | - | **93.3%** | 90.0% | **93.5%** | - |
|  | **95.2%** | **96.9%** | **97.7%** | **91.0%** | - | **97.2%** | **98.3%** | **98.4%** |

Table 2. Unsupervised fault diagnosis summary of TEP.

|  |  |  |  |
| --- | --- | --- | --- |
|  | JAN-CNN | DAAN | MSDAN |
| Average accuracy rate | 93.7% | 94.5% | **95.0%** |

* + 1. FCC

Fluid catalytic cracking (FCC) is a pivotal process within the petroleum refining industry used to convert high-molecular-weight hydrocarbons into valuable gasoline, diesel, and other lighter fractions. Key components of the fluid catalytic cracking process include feedstock pre-treatment, reaction in the fluidized bed, mixture separation, catalyst regeneration, and product recovery. In this study, the process data from a large chemical plant in southeastern China is collected for nearly two years. The plant owns two FCC units with similar but not identical processes. The separation sections in these two sets are highlighted in this experiment because they have over eighty key alarm variables.

On the basis of the experience in the above tests, the two sets of devices are respectively set up as source or target domains for transfer learning. A total of ten common faults are selected as classification criteria, including high-temperature alarm at the top of the tower, low-level alarm at the bottom of the tower, and abnormal feed flow rate. The training dataset includes a handful of labeled fault data from the source mode and solely one batch of unlabeled fault data from the target mode. This configuration is utilized to conduct unsupervised anomaly diagnosis tasks, aimed at validating the effectiveness of the model. The brief results are shown below which confirm the advantages of MSDAN again.

Table 3. Unsupervised fault diagnosis performance of FCC.

|  |  |  |  |
| --- | --- | --- | --- |
|  | JAN-CNN | DAAN | MSDAN |
|  | 76.4% | 83.5% | **88.2%** |
|  | 69.8% | 72.9% | **81.0%** |

* 1. Conclusions

In the present work, a multi-source domain adaptation network is proposed for cross-domain fault diagnosis in chemical processes. This advanced transfer learning method applies a more efficient multi-channel feature extractor which combines the strengths of Transformer and CNN. The multi-source domain adaptation with PK-MMD is superior to other widely-used adaptation techniques. The fault diagnosis results presented in this paper prove that this model can apply the diagnosis knowledge learned from different working conditions of one process and even similar processes in the real industry. This exploration will promote the practical application of deep transfer learning in chemical process fault diagnosis. Ongoing efforts are underway to improve this model, focusing on reducing reliance on labeled data from the source domain and tapping its potential for more industrial process tasks.

Acknowledgments

The authors gratefully acknowledge support from the National Science and Technology Innovation 2030 Major Project of the Ministry of Science and Technology of China (2018AAA0101605) and the National Natural Science Foundation of China (21878171, 62003004).

References

A. Bathelt, N.L. Ricker, M. Jelali, 2015, Revision of the Tennessee Eastman Process Model, *IFAC-PapersOnLine*, 48, 8, 309-314.

X. Bi, R. Qin, D. Wu, S. Zheng, J. Zhao, 2022, One step forward for smart chemical process fault detection and diagnosis, *Computers and Chemical Engineering*, 164, 107884.

N. Lu, Z. Cui, H. Hu, T. Yin, 2023, Multi-view and Multi-level network for fault diagnosis accommodating feature transferability, *Expert Systems with Applications*, 213, 119057.

R. Qin, J. Zhao, 2022, Adaptive multiscale convolutional neural network model for chemical process fault diagnosis, *Chinese Journal of Chemical Engineering*, 50, 398-411.

R. Qin, J. Zhao, 2023, Cross-domain Fault Diagnosis for Chemical Processes through Dynamic Adversarial Adaptation Network, *Computer Aided Chemical Engineering*, 52, 867-873.

H. Wu, J. Zhao, 2020, Fault detection and diagnosis based on transfer learning for multimode chemical processes, *Computers and Chemical Engineering*, 135, 106731.

H. Xiao, H. Ogai, W. Wang, 2022, Multi-Channel Domain Adaptation Deep Transfer Learning for Bridge Structure Damage Diagnosis, *IEEJ Transactions on Electrical and Electronic Engineering*, 17, 11, 1637-1647.

B. Yang, Y. Lei, F. Jia, N. Li, Z. Du, 2020, A Polynomial Kernel Induced Distance Metric to Improve Deep Transfer Learning for Fault Diagnosis of Machines, *IEEE Transactions on Industrial Electronics*, 67, 11, 9747-9757.

Q. Zhu, Y. Qian, N. Zhang, Y. He, Y. Xu, 2023, Multi-scale Transformer-CNN domain adaptation network for complex processes fault diagnosis, *Journal of Process Control*, 130, 103069.