ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Visual Process Monitoring by Data-Dependent Kernel Discriminant Analysis with t-Distributed Similarities

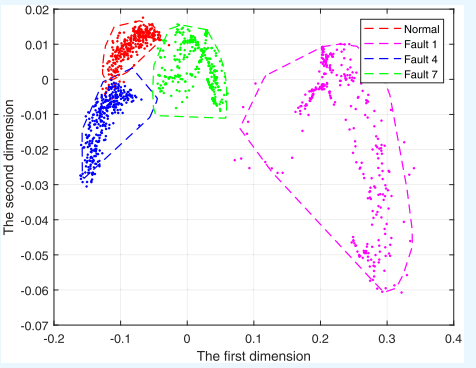
Chihang Wei,* Chenglin Wen, Jieguang He, and Zhihuan Song*

Cite This: *ACS Omega* 2023, 8, 38013−38024

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Visual process monitoring would provide more directly appreciable and more easily comprehensible information about the process operating status as well as clear depictions of the occurrence path of faults; however, as a more challenging task, it has been sporadically discussed in the research literature on conventional process monitoring. In this paper, the Data-Dependent Kernel Discriminant Analysis (D^2K-DA) model is proposed. A special data-dependent kernel function is constructed and learned from the measured data, so that the low-dimensional visualizations are guaranteed, combined with intraclass compactness, interclass separability, local geometry preservation, and global geometry preservation. The new optimization is innovatively designed by exploiting both discriminative information and t-distributed geometric similarities. On the construction of novel indexes for visualization, experiments of visual monitoring tasks on simulated and real-life industrial processes illustrate the merits of the proposed method.



1. INTRODUCTION

Process monitoring is a kind of technique to ensure process safety, improve production efficiency, and reduce energy consumption and pollution.[1−3] A large amount of data can be expediently collected due to the improvement of measurement and information technology. Data-driven methods, featured by easy implementation, nice generalization, and less dependence on process mechanisms, are attracting more and more attention.[4−10] For example, the topology-guided graph learning fault diagnosis framework was developed that combined the concept of graphs with process physics to focus on the intrinsic relationships between inputs and outputs, particularly the physical consistency of model prediction logic.[11] The novel graph convolutional network-based soft sensor utilizing localized spatial-temporal correlations, aiding in comprehending the intricate interactions among the included variables, was proposed with high model transparency.[12] The novel dynamic latent variable (DLV)-based transfer learning approach, called transfer DLV regression (TDLVR), for quality prediction of multimode processes with dynamics was developed; this model can overcome data marginal distribution discrepancy and enrich the information on the new mode.[13]

Human is naturally and constantly exposed to a world full of visual stimuli. Visualization of the actual process status would provide more directly appreciable and easily comprehensible information about the process operating status for enhancing engineers' and operators' understanding. What's more, it would be able to clearly depict the occurrence path of faults. However, due to respective technical limits, this kind of technique has been sporadically reported in conventional process monitoring, which mainly focuses on fault detection, fault diagnosis, fault isolation, etc. Note that the "visualization" in this paper refers to not only the exhibition of mapped training data in a comprehensible low-dimensional (like two/three-dimensional) space but also the path of the mapped novel samples; the former is consensual in the conventional regime of pattern recognition, but the latter is of more significance in the regime of process monitoring.

In the past, discriminant analysis was developed to classify observations and find combinations of features that characterize or separate two or more classes of objects or events.[14] It is used as a tool for classification and dimension-reduction, and can also be employed in the data visualization task to some extent.[15,16] Representative algorithms include linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and kernel discriminant analysis (KDA). However, the conventional discriminant analysis algorithms usually ignore the geometry structures of data; thus, they may not work well for visualization where both local geometry structure and global geometry structure should be simultaneously preserved in the low-dimensional space. Besides, the traditional kernel

Received: May 19, 2023
Accepted: September 21, 2023
Published: October 5, 2023

function for discriminant analysis, implicitly defining the kernel space, is artificially determined, which only aims at solving the nonlinearities considering neither the discriminative information nor geometry structures. It is desirable to learn the kernel function from specific data that adapt better to visualization.

Manifold learning is a kind of technique for nonlinear dimensionality reduction and visualization, aiming at recovering the low-dimensional embeddings from the original data through exploiting the geometry of data distribution,[17−23] that is, the local linearity of manifolds which locally resembles a Euclidean space, since each local neighborhood may have its unique spatial geometric distribution characteristics. Ideally, given the data with nonlinear and high-dimensional characteristics, the real manifold structure of the nonlinear data in high dimensions would be revealed from the geometric information, and the boundary of the spatial distribution of the samples in the input space would be preserved by exploiting the intrinsic local geometry structure indicating the local relationships among the data samples.[24] Stochastic neighbor embedding (SNE) and t-distributed stochastic neighbor embedding (t-SNE)[25] are two representative algorithms for visualization. In spite of the advantageous features of manifold learning techniques, most of them are nonparametric in nature, resulting in the failure to find explicit mapping functions from high-dimensional original data to low-dimensional embeddings, suffering from the generalization problem for newly collected samples. It is reported that a regression algorithm can be successively adopted to build the explicit relationship between the original space and the visualization space of t-SNE,[26] but the separate procedures may be challenged by reliability and scalability in practice.

Based on the above-mentioned aspects, it is an obvious choice to make a compact cooperative between discriminant analysis (classification) and manifold learning (dimension-reduction) to learn from each other, which is however not an easy task. In this study, the D^{2}K-DA model is proposed as a consistent framework for visual monitoring. Instead of an artificially determined kernel function, a unique data-dependent kernel (DDK) function is constructed and directly learned from specific data through one compact step of optimization with a scalable quantity of parameters, so that the low-dimensional visualizations can be pursued. The objective function is integrated into two parts: one stands for class discrimination by exploiting discriminative information and the other stands for dimension-reduction by manifold learning solely derived from the process variables, as they contain the manifold structure of data. The pairwise similarities between samples to quantify the manifold structure of data are measured with the heavy-tailed Student's t distribution, which can be seen as an infinite mixture of Gaussians and helps to solve the crowding problem of local techniques for multidimensional scaling. From the view of discriminant analysis, the proposed D^{2}K-DA exploits discriminative information to pursue both intraclass compactness and interclass separability to group samples from the same class and separate samples from different classes. From the view of manifold learning, local geometry structure and global geometry structure would be faithfully and synchronously preserved in the low-dimensional visualizations, to compress significant information to the leading two-/three-dimensional visualization space and avoid the destruction of geometry structure. All of the above aspects are further described. Table 1 lists the important variables and notations of this article.

Table 1. Nomenclature of Important Variables and Notations

symbol	description
$\boldsymbol{X}$	$\boldsymbol{X} = \{(\boldsymbol{x}_i, c_i)\}_{i=1}^{N}$, a set of N-labeled samples.
c_i	the class information on $\boldsymbol{x}_i$; $c_i \in \{1,\cdots,T\}$.
$\boldsymbol{Y}$	$\boldsymbol{Y} = \{\boldsymbol{y}_i\}_{i=1}^{N}$, a set of N low-dimensional visualizations.
$\boldsymbol{y}_i$	the low-dimensional visualization of $\boldsymbol{x}_i$, $\boldsymbol{y}_i = f(\boldsymbol{x}_i)$.
$\kappa(\cdot,\cdot)$	the data-dependent kernel.
$\kappa_0(\cdot,\cdot)$	the basic kernel of the data-dependent kernel.
$r(\cdot,\cdot)$	the factor function.
$\boldsymbol{e}_h$	the empirical cores.
$\boldsymbol{K}$	the kernel matrix of training samples with $\kappa(\cdot,\cdot)$.
$\boldsymbol{K}_0$	the kernel matrix of training samples with $\kappa_0(\cdot,\cdot)$.
$\boldsymbol{R}$	the diagonal matrix with elements $r(\boldsymbol{x}_1)$, ⋯, $r(\boldsymbol{x}_N)$.
k	the number of nearest neighbors.
L^{Dis}	the subobjective function for class discrimination.
L^{Geo}	the subobjective function for dimension-reduction.
S^W	the measurement of intraclass compactness.
S^B	the measurement of interclass separability.

2. PRELIMINARY STUDY

In spite of the advantageous features of the kernel learning methods in aspects of theory, application performance, and flexibility, conventional kernel learning techniques construct a kernel matrix that is effective for training samples only; they cannot deal with new samples.[22] In this paper, to tackle this problem without repeating the whole training procedures and additional extrapolating extensions, the flexible DDK is creatively adopted for good generalization.[22,23]

In general, the data-dependent kernel function is defined as

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = r(\boldsymbol{x}_i) r(\boldsymbol{x}_j) \kappa_0(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{1}$$

where $\kappa_0(\cdot,\cdot)$ is the basic kernel. The factor function $r(\cdot)$

$$r(\boldsymbol{x}_i) = \alpha_0 + \sum_{h=1}^{H} \alpha_h \kappa_1(\boldsymbol{x}_i, \boldsymbol{e}_h) \tag{2}$$

where α_h is to be optimized. The empirical cores $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_H$ are specifically selected from the training data.

For easy derivations, the inner product matrix (Gram matrix) is defined as $\boldsymbol{K} \in \mathbb{R}^{N\times N}$, where the ijth element $\boldsymbol{K}_{ij} = \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = f(\boldsymbol{x}_i) \cdot f(\boldsymbol{x}_j) = f(\boldsymbol{x}_i)^{\mathrm{T}} f(\boldsymbol{x}_j)$. It is easy to get

$$\boldsymbol{K} = \boldsymbol{R}\boldsymbol{K}_0\boldsymbol{R} \tag{3}$$

$$\boldsymbol{r} = \begin{bmatrix} 1 & \kappa_1(\boldsymbol{x}_1, \boldsymbol{e}_1) & \cdots & \kappa_1(\boldsymbol{x}_1, \boldsymbol{e}_H) \\ \vdots & \vdots & \ddots & \vdots \\ & \kappa_1(\boldsymbol{x}_N, \boldsymbol{e}_1) & \cdots & \kappa_1(\boldsymbol{x}_N, \boldsymbol{e}_H) \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_H \end{bmatrix} = \boldsymbol{K}_1\boldsymbol{\alpha} \tag{4}$$

$$\begin{aligned} \boldsymbol{K}_{ij} &= \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = r(\boldsymbol{x}_i) r(\boldsymbol{x}_j) \kappa_0(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ &= [(\boldsymbol{K}_1)_{i.}\boldsymbol{\alpha}] \times [(\boldsymbol{K}_1)_{j.}\boldsymbol{\alpha}] \times [(\boldsymbol{K}_0)_{ij}] \\ &= \boldsymbol{\alpha}^T (\boldsymbol{K}_1)_{i.}^T (\boldsymbol{K}_1)_{j.} (\boldsymbol{K}_0)_{ij} \boldsymbol{\alpha} \end{aligned} \tag{5}$$

where $(\boldsymbol{K}_1)_{i.}$ denotes the ith row of $\boldsymbol{K}_1$.

3. D^{2}K-DA MODEL

It is both theoretically and practically significant to focus on visual monitoring to obtain more directly appreciable and more easily comprehensible information about the process operating status and clear depictions of the occurrence path of

https://doi.org/10.1021/acsomega.3c03496
ACS Omega 2023, 8, 38013−38024

faults. However, the existing approaches would not be competent due to the neglect of discriminative information or geometry structures. Taking advantage of both discriminant analysis and manifold learning, this paper proposes the D^{2}K-DA model as a consistent framework to learn a unique DDK function for visual monitoring. The compact optimization is formulated in terms of the combination coefficient $\boldsymbol{\alpha}$ of DDK to facilitate calculation. The objective function is integrated into two parts: (1) $\mathcal{L}^{\mathrm{Dis}}$ represents class discrimination (classification) by exploiting discriminative information to guarantee intraclass compactness and interclass separability; (2) $\mathcal{L}^{\mathrm{Geo}}$ represents dimension-reduction by manifold learning to guarantee local and global structure preservation, where pairwise similarities between samples are measured with the heavy-tailed Student's t distribution.

3.1. Construction of $\mathcal{L}^{\mathrm{Dis}}$. In the proposed D^{2}K-DA model, samples from the same class should be grouped and samples from different classes should be separated in the low-dimensional space. To realize this, the measurements of intraclass compactness and interclass separability, respectively expressed as S^W and S^B, are constructed.

Specifically, let us suppose the set of k-nearest neighbors of each sample $\boldsymbol{x}_i$ to be $knn(\boldsymbol{x}_i)$, and $knn^W(\boldsymbol{x}_i) = \{\boldsymbol{x}_j \mid \boldsymbol{x}_j \in knn(\boldsymbol{x}_i)$ and $c_i = c_j\}$ $knn^B(\boldsymbol{x}_i) = \{\boldsymbol{x}_j \mid \boldsymbol{x}_j \in knn(\boldsymbol{x}_i)$ and $c_i \neq c_j\}$. Then the two weight matrices are respectively defined as

$$\boldsymbol{W}_{ij}^W = \begin{cases} 1 & \text{if} \boldsymbol{x}_i \in knn^W(\boldsymbol{x}_j) \text{ or} \boldsymbol{x}_j \in knn^W(\boldsymbol{x}_i) \\ 0 & \text{otherwise} \end{cases}$$

$$\boldsymbol{W}_{ij}^B = \begin{cases} 1 & \text{if } \boldsymbol{x}_i \in knn^B(\boldsymbol{x}_j) \text{ or} \boldsymbol{x}_j \in knn^B(\boldsymbol{x}_i) \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The measurements of intraclass compactness and interclass separability are, respectively, calculated as

$$S^W = \sum_{i=1}^{N}\sum_{j=1}^{N} \left\| f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j) \right\|^2 \boldsymbol{W}_{ij}^W$$

$$S^B = \sum_{i=1}^{N}\sum_{j=1}^{N} \left\| f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j) \right\|^2 \boldsymbol{W}_{ij}^B \tag{7}$$

With the introduction of DDK, the optimization should be formulated in terms of $\boldsymbol{\alpha}$. S^W in (eq 7) can be reformulated as

$$\begin{aligned} & S^W \\ &= \sum_{i,j=1}^{N} [f(\boldsymbol{x}_i)^{\mathrm{T}} f(\boldsymbol{x}_i) - 2f(\boldsymbol{x}_i)^{\mathrm{T}} f(\boldsymbol{x}_j) + f(\boldsymbol{x}_j)^{\mathrm{T}} f(\boldsymbol{x}_j)] \boldsymbol{W}_{ij}^W \\ &= \sum_{i,j=1}^{N} [\boldsymbol{K}_{ii} - 2\boldsymbol{K}_{ij} + \boldsymbol{K}_{jj}] \boldsymbol{W}_{ij}^W \\ &= \boldsymbol{\alpha}^{\mathrm{T}} \sum_{i,j=1}^{N} [\boldsymbol{\Psi}^{ii} - 2\boldsymbol{\Psi}^{ij} + \boldsymbol{\Psi}^{jj}] \boldsymbol{W}_{ij}^W = \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Phi}^W \boldsymbol{\alpha} \end{aligned} \tag{8}$$

where $\boldsymbol{\Psi}^{ij} = (\boldsymbol{K}_1)_{i\cdot}^{T} (\boldsymbol{K}_1)_{j\cdot} (\boldsymbol{K}_0)_{ij}$.

Similarly

$$S^B = \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Phi}^B \boldsymbol{\alpha} \tag{9}$$

The objective to minimize S^W and maximize S^B is given as

$$\mathcal{L}^{Dis} = \frac{\eta}{2} S^W - \left(1 - \frac{\eta}{2}\right) S^B = \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{\Phi} \boldsymbol{\alpha} \tag{10}$$

where $\boldsymbol{\Phi} = \frac{\eta}{2}\boldsymbol{\Phi}^W - \left(1 - \frac{\eta}{2}\boldsymbol{\Phi}^B\right)$. $\eta \in [0,1]$ is the parameter to regulate the relative significance.

3.2. Construction of $\mathcal{L}^{\mathrm{Geo}}$. $\mathcal{L}^{\mathrm{Geo}}$ aims to preserve as much of both the local geometry and global geometry structure of the original data (high-dimensional) as possible in the low-dimensional space.

Inspired by the classical algorithm t-SNE,[25] the similarity between pairwise data points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ in the original space is measured by $p_{ij} = (p_{i|j} + p_{j|i})/2N$, where $p_{i|j}$ is the conditional probability. Mathematically, $p_{i|j}$ is given as

$$p_{i|j} = \frac{\exp\left(-\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2 / 2\sigma_i^2\right)}{\sum_{z \neq i} \exp\left(-\|\boldsymbol{x}_i - \boldsymbol{x}_z\|^2 / 2\sigma_i^2\right)} \tag{11}$$

that $\boldsymbol{x}_i$ would pick $\boldsymbol{x}_j$ as its neighbor; neighbors were proportionally picked to the probability density under a Gaussian centered at $\boldsymbol{x}_i$; $p_{i|i} \equiv 0$. σ_i is the variance of the Gaussian centered to $\boldsymbol{x}_i$, controlling the perplexity of $p_{i|j}$. Besides, in the low-dimensional space, the similarity q_{ij} between pairwise data points $f(\boldsymbol{x}_i)$ and $f(\boldsymbol{x}_j)$ is measured by probability with the heavy-tailed Student's t distribution to solve the crowding problem

$$\begin{aligned} q_{ij} &= \frac{[1 + \left\|f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)\right\|^2]^{-1}}{\sum_{z \neq l} [1 + \|f(\boldsymbol{x}_z) - f(\boldsymbol{x}_l)\|^2]^{-1}} \\ &= \frac{[1 + (\boldsymbol{K}_{ii} - 2\boldsymbol{K}_{ij} + \boldsymbol{K}_{jj})]^{-1}}{\sum_{z \neq l} [1 + (\boldsymbol{K}_{zz} - 2\boldsymbol{K}_{zl} + \boldsymbol{K}_{ll})]^{-1}} \\ &= \frac{[1 + \boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{\Psi}^{ii} - 2\boldsymbol{\Psi}^{ij} + \boldsymbol{\Psi}^{jj})\boldsymbol{\alpha}]^{-1}}{\sum_{z \neq l} [1 + \boldsymbol{\alpha}^{\mathrm{T}}(\boldsymbol{\Psi}^{zz} - 2\boldsymbol{\Psi}^{zl} + \boldsymbol{\Psi}^{ll})\boldsymbol{\alpha}]^{-1}} \end{aligned} \tag{12}$$

Define $q_{ii} \equiv 0$. Note that the variance of the Gaussian is set to $\frac{1}{\sqrt{2}}$ in (12); other values only result in a rescaled version.

For dimension-reduction, the construction of $\mathcal{L}^{\mathrm{Geo}}$ is based on the Kullback–Leibler divergence between P and Q

$$\mathcal{L}^{\mathrm{Geo}} = KL(P\|Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{13}$$

3.3. Consistent Framework. Note that both $\mathcal{L}^{\mathrm{Dis}}$ and $\mathcal{L}^{\mathrm{Geo}}$ should take minimum, such that the compact objective function $\mathcal{L}$ is calculated as

$$\mathcal{L} = \mathcal{L}^{\mathrm{Dis}} + \xi \mathcal{L}^{\mathrm{Geo}} \tag{14}$$

where ξ is a supplement to adjust the order of magnitudes. In eq 14, the objective for class discrimination $\mathcal{L}^{\mathrm{Dis}}$ by exploiting discriminative information to guarantee intraclass compactness and interclass separability and the objective for dimension-reduction $\mathcal{L}^{\mathrm{Geo}}$ by manifold learning to guarantee local and global structure preservation are combined to construct the whole objective function $\mathcal{L}$ by additive operation with a weight factor; a consistent framework for visual monitoring would be obtained, containing the concept of contribution of the proposed method. Note that the proposed construction of the objective function in eq 14 realizes the main contribution

https://doi.org/10.1021/acsomega.3c03496
ACS Omega 2023, 8, 38013−38024

of the proposed D^{2}K-DA method to make compact cooperatives between discriminant analysis (classification) and manifold learning (dimension-reduction) to learn from each other. Strictly speaking, the optimization here should be a bi-objective optimization; however, to be convenient to calculate in practice, the two objectives, both taking the minimum, are simply added. Besides, ξ should be supplemented for a desired optimization as there may be an order of magnitude difference between $\mathcal{L}^{\mathrm{Dis}}$ and $\mathcal{L}$. Through optimization with a proper ξ, both objectives would take the balance, such that both intraclass compactness and interclass separability to group samples from the same class and separate samples from different classes would be pursued, while at the same time, significant information would be compressed to the leading two/three dimensions for visualization without obvious destruction of the geometry structure.

The solution of (eq 14) may not be globally optimal. To find a refined solution with both time and space efficiency, interior-point methods and sequential quadratic programming are successively employed, while the former rapidly converges to a rough range around the optimum and the latter helps to find the precise solution; it empirically works well for repeated simulations.

4. D^{2}K-DA-BASED VISUAL PROCESS MONITORING

4.1. Parameter Optimization. After constructing the consistent framework of D^{2}K-DA, there are several parameters and hyper-parameters to be tuned.

4.1.1. Construction of the DDK.

- The value of H and the selection criterion of empirical cores $\boldsymbol{e}_h$: As recommended by ref 23, one-third of the training data are chosen as the empirical core in this paper, according to the spatial distribution of the data. Specifically, the empirical cores are chosen step by step: First, one training sample is randomly chosen as a new empirical core from the training data. Second, two training samples nearest to the chosen sample are selected. Third, the training sample in the first procedure is chosen and its two neighbors are removed from the training data. The three procedures are repeated until there are no samples left in the training data. In this way, the selected empirical cores would roughly depict the spatial structure of the data.
- The kernel width τ_0 of kernel function $\kappa_0(\cdot,\cdot)$ and the kernel width τ_1 of kernel function $\kappa_1(\cdot,\cdot)$: In practice, τ_0 is empirically set to $\tau_0 = 10\zeta\iota^2$ according to ref 23, where ζ and ι^2 are, respectively, the dimension of process variables and the variance of the data. τ_1 is set to $\tau_1 = \tau_0/5$, recommended by ref 23.

4.1.2. Construction of $\mathcal{L}^{\mathrm{Dis}}$.

- The number of nearest neighbors k: As recommended by ref 27, the initial value would be roughly determined according to an empirical equation, and then the final value would be thoroughly adjusted around the initial value.
- The parameter η regulates the relative significance between S^W and S^B: Generally, η is larger than 0.5 because the information on two similar samples is more important than that of two dissimilar samples. The specific value would be chosen by grid search from small to large. Because the labeled data are usually quite precious and scarce, it is recommended to employ k-fold cross-validation to avoid splitting an independent validation data set.

4.1.3. Construction of $\mathcal{L}^{\mathrm{Geo}}$.

- σ_i for each $\boldsymbol{x}_i$ to calculate $p_{i|j}$: In practice, the value of σ_i is usually determined by a binary search with a fixed perplexity recommended by a highly cited paper.[25] The perplexity $Perp(P_i)$ is defined as

$$Perp(P_i) = 2^{E(P_i)} \tag{15}$$

- where $E(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i}$ is the Shannon entropy of P_i. According to ref 25, the performance would be fairly robust to changes in the perplexity, and typical values are between 5 and 50.

4.1.4. Construction of $\mathcal{L}$.

- The parameter ξ adjusts the order of magnitudes between $\mathcal{L}^{\mathrm{Dis}}$ and $\mathcal{L}^{\mathrm{Geo}}$. Too small ξ implies insufficient geometry structure information is preserved which leads to the destruction of the geometry structure in low-dimensional visualization space, whereas too large ξ results in inadequate intraclass compactness and interclass separability by losing the role of discriminative information. Practically, ξ is roughly chosen from the exponential candidate set $\{10^e | e = -8, -7, \cdots, 7, 8\}$ by seriate search, and then the final value would be adjusted around the rough value.

4.2. Visualization Based on D^{2}K-DA. Once the combination coefficient vector $\boldsymbol{\alpha}$ is obtained, the low-dimensional visualizations can be obtained where as much information as possible has been compressed to the leading dimensions. Given the eigenvalue decomposition of $\boldsymbol{K}$,

$$\lambda\boldsymbol{\beta} = \frac{1}{N}\boldsymbol{K}\boldsymbol{\beta} \tag{16}$$

Solving (eq 16) yields the orthonormal eigenvectors $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_N$ and the associated corresponding eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$.

To make a visualization $\boldsymbol{y}_{\mathrm{new}}$ at a new observation $\boldsymbol{x}_{\mathrm{new}}$, the visualization system can be expressed as

$$\boldsymbol{y}_{\mathrm{new},l} = \frac{1}{\sqrt{\lambda_l}} \sum_{n=1}^{N} \boldsymbol{\beta}_n^l \kappa(\boldsymbol{x}_n, \boldsymbol{x}_{\mathrm{new}}) \tag{17}$$

where l is the serial of dimension. The status and occurrence path of $\boldsymbol{y}_{\mathrm{new}}$ would provide more directly appreciable and more easily comprehensible information about the process operating status and can be used to forecast the tendency of the process.

4.3. D^{2}K-DA-Based Visual Process Monitoring. A good visualization for multiclasses usually means there are as few overlaps as possible between different classes. To quantify the performance of visualization, this work introduces the Delaunay triangulation-based correct visualization rate (DT-CVR). Specifically, the boundary for each operation status, actually the convex envelope line, would be obtained by the Delaunay triangulation[28] based on low-dimensional visualizations of training samples of each class; the low-dimensional visualization $\boldsymbol{y}_{\mathrm{new}}$ of the query sample is checked if it locates inside of the polygonal region defined by the boundary; only if $\boldsymbol{y}_{\mathrm{new}}$ locates in the correct region and it does not locate in any other region is the query sample determined as correctly visualized. Finally, DT-CVR is obtained. A higher DT-CVR

https://doi.org/10.1021/acsomega.3c03496
ACS Omega 2023, 8, 38013−38024

denotes fewer overlaps among different classes. The decision regions usually do not fill up the whole space.

DT-CVR can also help to quantify the performance of fault detection and classification. Note that this kind of metrics is more rigorous and conservative than traditional ones. A high DT-CVR denotes a good performance of detection and classification; however, a low DT-CVR may not denote a poor performance of detection and classification. (1) If the mapped query sample still locates near but outside the boundary, it is not determined as being correctly visualized; however, it would be correctly classified by common classifiers, like k-nearest neighbors. (2) If the mapped query sample locates in the overlaps, it is also not determined as correctly visualized; however, it may be correctly classified by common classifiers; some traditional classifiers are able to perform well with interlaced boundaries, which would not be desired in visualization tasks. To quantify the performance of detection and classification as conventional, the k-nearest-neighbor-based correct classification rate (KNN-CCR) is introduced. To illustrate the difference between DT-CVR and KNN-CCR, Figure 1 gives a sketch map.

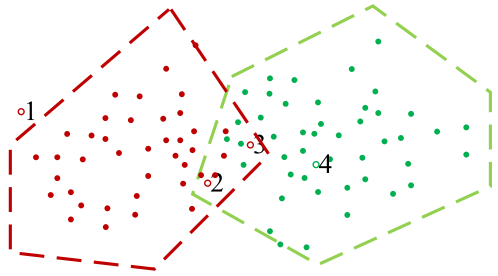


Figure 1. Sketch map of DT-CVR and KNN-CCR. Legends: Imaginary lines: decision boundaries of DT-CVR; solid dots: training samples; hollow dots: query samples, marked as the correct class. Results: (1) Hollow dot "1": DT-CVR (×), KNN-CCR (√); (2) Hollow dot "2": DT-CVR (×), KNN-CCR (√); (3) Hollow dot "3": DT-CVR (×), KNN-CCR (×); (4) Hollow dot "4": DT-CVR (√), KNN-CCR (√).

In a nutshell, Figure 2 presents the flowchart of D^{2}K-DA-based visual process monitoring.

5. RESULTS AND DISCUSSION

In this section, two industrial processes, the Tennessee Eastman (TE) process and a real-world polyethylene process, are employed to evaluate the feasibility and efficiency of the proposed D^{2}K-DA model, respectively, for generalization and pragmaticism.

To better show the performance of the proposed method, several state-of-the-art data-driven classification and visualization methods are employed: (1) Fisher linear discriminant analysis (FDA); (2) self-organizing map (SOM), the classical method for visual monitoring[29]; (3) supervised maximum variance unfolding (SMVU),[30] a kernel-learning-based classification method; (4) supervised autoencoder (SAE)[31] as a representative of supervised neural network; specifically, a multilayer form is adopted to enhance the ability to extract information; (5) t-SNE-BP, where t-SNE helps to find the visualization of data, and successively back-propagation (BP) helps to build the explicit relationship between the original space and the visualization space.[25]

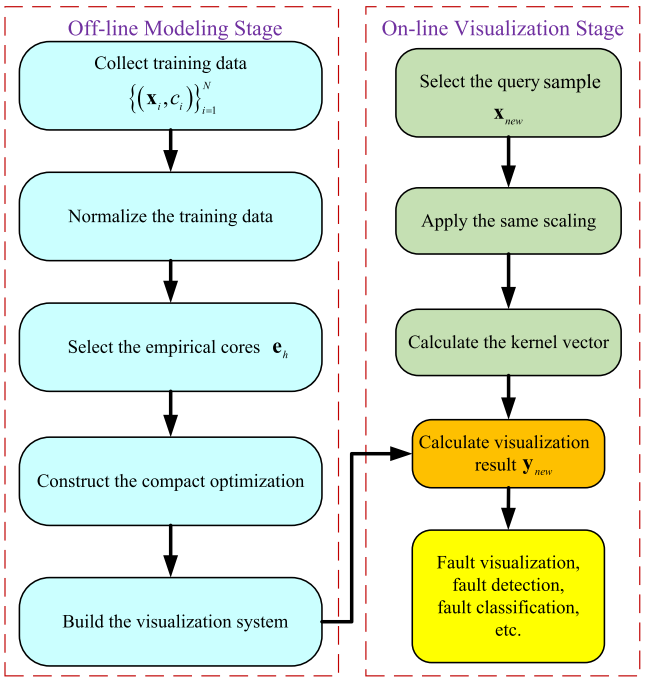


Figure 2. Flowchart of D^{2}K-DA-based visual process monitoring.

5.1. Tennessee Eastman Process. TE process has been a widely adopted benchmark in the research fields of process system engineering to evaluate the performance of various models on fault detection, fault classification, etc.

The process is designed with several operating states; among them, the normal operating state and three faulty operating states (conventionally denoted as faults 1, 4, and 7) are chosen to evaluate the monitoring performance of the proposed method and other related methods. Besides, a total of 16 easily measured process variables are chosen for modeling, including A feed, D feed, E feed, A and C feed, recycle flow, reactor feed rate, reactor temperature, purge rate, product separator temperature, product separator pressure, product separator underflow, stripper pressure, stripper temperature, stripper steam flow, reactor cooling water outlet temperature, and separator cooling water outlet temperature. For offline training, each operating state is simulated for 48 h while the fault is introduced at the beginning; the sampling interval is set to 3 min, such that 480 samples are collected for each operating state. For online testing, 480 samples are collected for each operating state, while the fault is introduced at the beginning of the ninth hour (161st sample).

5.1.1. Quantified Result of Visualization and Classification. For D^{2}K-DA, the kernel widths of $\kappa_0(\cdot,\cdot)$ and $\kappa_1(\cdot,\cdot)$ are, respectively, set to $\tau_0 = 160$ and $\tau_1 = 32$;[23] the number of nearest neighbors is set to $k = 3$;[27] the parameter to regulate the relative significance between S^W and S^B is set to $\eta = 0.6$; the perplexity $Perp(P_i) = 34$;[25] the supplement between $\mathcal{L}^{\mathrm{Dis}}$ and $\mathcal{L}^{\mathrm{Geo}}$ is set to $\xi = 10$. For the SOM, the structure is set to 9-13. For SMVU, the number of nearest neighbors is also set to $k = 3$ as D^{2}K-DA, while $\rho = 10$. For SAE, the structure is set to 16-5-2-5-16, while other hyper-parameters are set to 0.001, 128, and 100. For t-SNE-BP, the perplexity is set to 34, the same as D^{2}K-DA; the hidden layer of BP is set to 50. For KNN-CCR, the number of nearest neighbors is set to 3.

Detailed DT-CVRs and KNN-CCRs are listed in Tables 2 and 3. The results show that the proposed D^{2}K-DA performs

https://doi.org/10.1021/acsomega.3c03496
ACS Omega 2023, 8, 38013−38024

Table 2. DT-CVRs of Different Methods in the TE Process

type	FDA	SOM	SMVU	SAE	t-SNE-BP	D^{2}K-DA
normal	0.0000	0.0000	0.0000	0.0000	0.0000	0.9313
fault 1	0.9625	0.5313	0.0000	0.1688	0.8219	0.9469
fault 4	0.0250	0.0000	0.1719	0.0000	0.0031	0.9406
fault 7	0.2875	0.4125	0.1000	0.1875	0.0969	0.9844
average	0.3125	0.2359	0.0680	0.0891	0.2305	0.9508

Table 3. KNN-CCRs of Different Methods in the TE Process

type	FDA	SOM	SMVU	SAE	t-SNE-BP	D^{2}K-DA
normal	0.5156	0.9969	0.4313	0.5406	0.5469	0.9406
fault 1	0.9844	0.9656	0.0094	0.6844	0.9344	0.9969
fault 4	0.3094	0.0000	0.3406	0.2781	0.3312	1.0000
fault 7	0.5000	0.6781	0.9563	0.4406	0.4594	0.9969
average	0.5773	0.6602	0.4344	0.4859	0.5680	0.9836

the best in both visualization and classification. FDA cannot handle the nonlinearity of data, which is quite common in real practice. The structure of SOM would not be capable of describing complex characteristics with shallow layers and limited nodes. SMVU learns a data-dependent kernel from specific data with sectional discriminative information; however, separate training procedures may not guarantee a good generalization performance, while the between-class structure is factitiously determined. SAE has a deeper multilayer structure to extract much information from the original space to the leading dimensionalities. It still gives unsatisfactory results without explicit consideration of geometry structure. t-SNE-BP enjoys the advantages of t-SNE to capture much of the local structure of the high-dimensional data and geometry preservation to promote visualization and classification performance, as well as the generalization by BP. However, the neglect of discriminative information and separate modeling procedures may be challenged by reliability and scalability in practice. The proposed consistent framework, D^{2}K-DA, integrates discriminative information (classification) and geometry information (dimension-reduction by manifold learning) together to directly learn a unique DDK function from specific data. It compresses significant information to the leading dimensions for visualization with both intraclass compactness and interclass separability with the restraint of crowding problem to guarantee visualization performance.

Note that the results show that DT-CVR is usually smaller than KNN-CCR because the former is more rigorous and conservative than the latter. The decision boundary of DT-CVR is the convex envelope of mapped samples, and the decision regions usually do not fill up the whole space. A high DT-CVR denotes few overlaps between classes, but KNN-CCR may still work when the mapped query sample locates outside the convex boundary of DT-CVR or in the overlaps.

5.1.2. Visualized Result of Operating Status. To visualize process monitoring, Figure 3a−f gives the two-dimensional visualizations of each testing data set as well as the boundaries of DT-CVR (imaginary lines), where dots are colored in the corresponding correct class. Only 161−480 samples are included because the fault is introduced in the 161st sample. It is obvious that low-dimensional mapped samples are mixed in Figure 3a−e but are separated for different classes and grouped for the same class in Figure 3f. The results further show that D^{2}K-DA outperforms the other models.

To better analyze the operating status, Figure 4a−d illustrates the trajectories under D^{2}K-DA, where dots are colored in the corresponding class and the black lines denote the sampling order. In the beginning, the trajectories show the TE process operates in the normal status in the first 8 h. Then as the fault is introduced, the trajectories directly move to the corresponding region, almost without delay, meaning a specific fault has happened. Finally, the process operates in the fault status. Engineers and operators visually perceive and access this information with D^{2}K-DA.

5.1.3. Sensitivity Test of Parameter Values and Random Starts. To test the sensitivity of the parameter ξ, various experiments are conducted to show the effects of different values of ξ. According to eq 14, ξ adjusts the order of magnitudes between $\mathcal{L}^{\mathrm{Dis}}$ and $\mathcal{L}^{\mathrm{Geo}}$. Too small ξ implies insufficient geometry structure information is preserved which leads to the destruction of the geometry structure in low-dimensional visualization space, whereas too large ξ results in inadequate intraclass compactness and interclass separability by losing the role of discriminative information. Means of CVR and CCR are shown in Figure 5a,b under different values of ξ; 20 repeated experiments are conducted. It can be seen from Figure 5 that model performances are basically robust to the fluctuation of ξ. These experiments demonstrate that the performance of the proposed D^{2}K-DA will be a certain improvement when the parameter ξ varies in a wide range. In practice, this parameter is optimized by Section 4.1.4.

The proposed D^{2}K-DA is partially derived from t-SNE, which is stochastic. Different random starts would lead to different results. To test the robustness for random starts, the proposed model is repeatedly trained and tested 20 times under the same (hyper)parameters. The averages of DT-CVR and KNN-CCR are 0.9472 and 0.9816, respectively, while the standard deviations are 0.0066 and 0.0025, repetitively. This result shows that the proposed method is robust despite the randomness of starts. Additionally, experiment results on different starts are presented in Part I of the Supporting Information (Figures S1−S8), where detailed DT-CVRs and KNN-CCRs are synchronously given.

To test the proposed D^{2}K-DA upon more faults in the TE process, more experiments are further conducted while visualizations and trajectories are presented in Part II of the Supporting Information (Figures S9−S14). The results show that the low-dimensional visualizations of a fraction of faults are more or less overlapped by the low-dimensional visual-

https://doi.org/10.1021/acsomega.3c03496
ACS Omega 2023, 8, 38013−38024

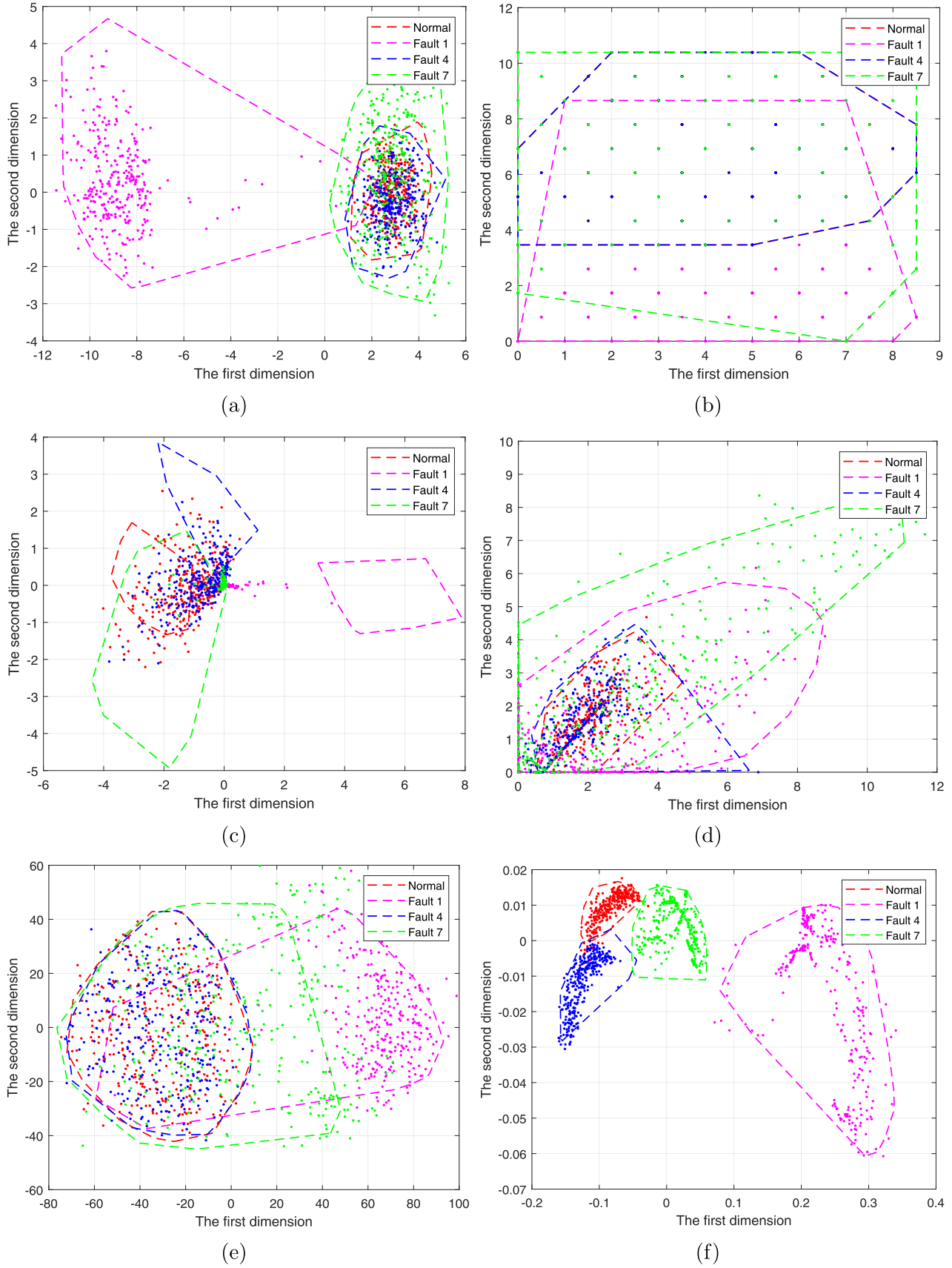


Figure 3. Visualizations of the 161−480 samples of each testing data set and the boundaries of DT-CVR in the TE process. Methods: (a) FDA, (b) SOM, (c) SMVU, (d) SAE, (e) t-SNE-BP, and (f) D^2K-DA.

https://doi.org/10.1021/acsomega.3c03496
ACS Omega 2023, 8, 38013−38024

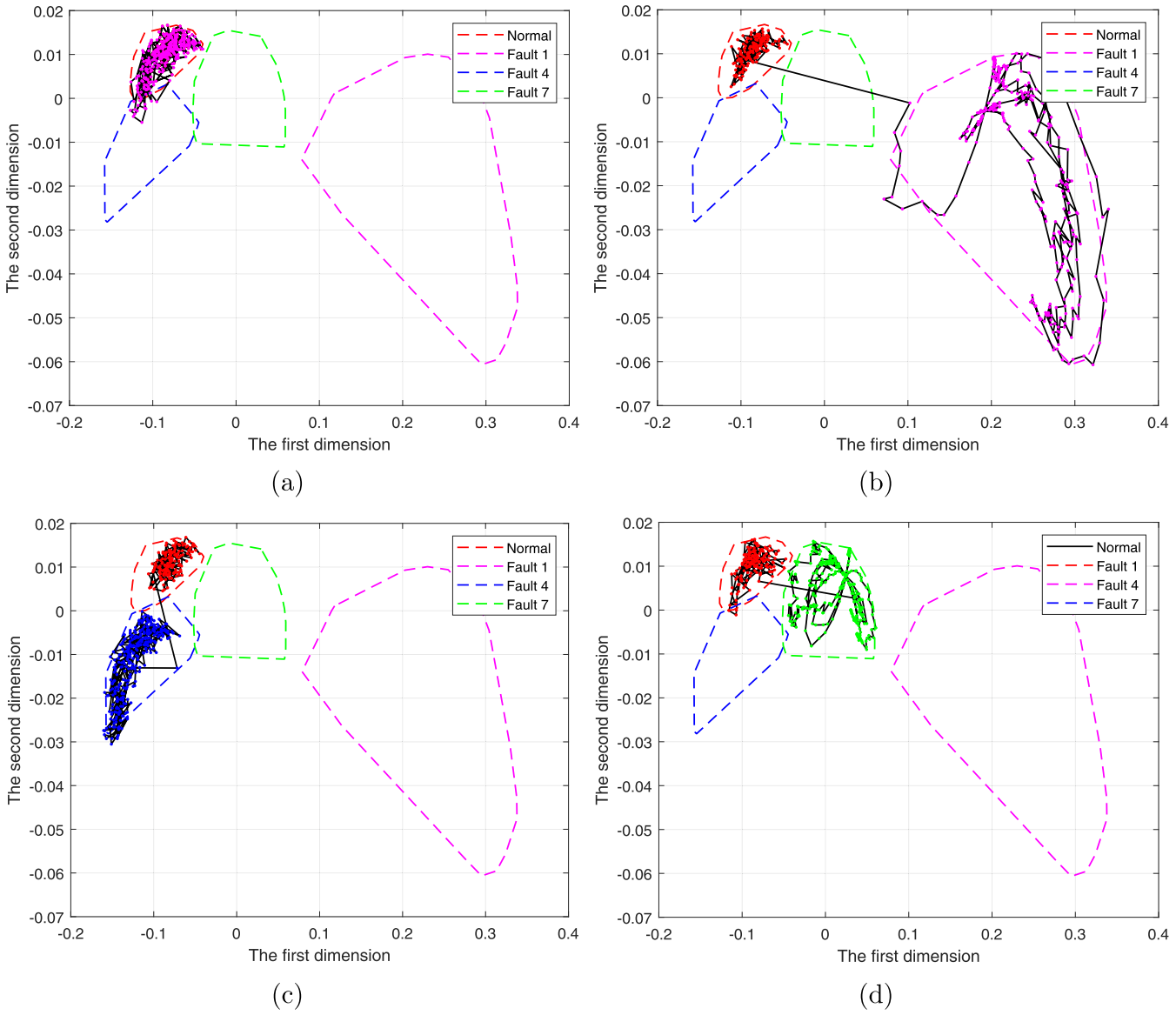


Figure 4. Trajectories of the operating status in the TE process under D^{2}K-DA. Testing data sets: (a) Normal, (b) Fault 1, (c) Fault 4, and (d) Fault 7.

izations of the normal operating state. The reason would be that some pieces of operating data after faults have occurred may not be obviously discriminative with normal operating state, and even may not be obviously discriminative with other faults. This situation is common in random faults and ramp faults. Besides, the visualization system forces the dimensions of data to be reduced to two for human cognition. For complex faults, more dimensions would be needed to discriminate differences between faults and normal operating state, even though significative information has been compressed to the leading two dimensions; complex faults are not two-dimensional divisible.

5.2. Polyethylene Process. The proposed method is further validated by the monitoring and visualization task of the melting index of a real-life industrial polyethylene process. All of the data sets are collected from the routine process records and the corresponding laboratory analysis. As the product quality of polyethylene production cannot be directly measured online, the operating variables have to be maintained the same until the analysis result is obtained; the product quality is analyzed and recorded in the laboratory once per day. For confidential reasons, the other details of this process will be concealed.

According to the operating experience, 11 process variables that are closely correlated with the product quality are selected for modeling. Four working conditions are concerned, including a normal working condition and three faulty conditions (denoted as faults 1, 2, and 3). In this paper, 480 samples of four working conditions in a product line are collected. The number of training samples for each working condition is 50 (a total of 200 samples). The number of testing samples for each working condition is 70 (a total of 280 samples), while the fault is introduced in the 21st sample for faulty working conditions.

5.2.1. Quantified Result of Visualization and Classification. For D^{2}K-DA, $\tau_0 = 110$ and $\tau_1 = 22$;[23] nearest neighbors $k = 3$;[27] $\eta = 0.7$; $Perp(P_i) = 30$;[25] $\xi = 1$. For the SOM, the structure is 8-10. For SMVU, nearest neighbors are also set to $k = 3$ as D^{2}K-DA, while $\rho = 5$. For SAE, the structure is set to 11-4-2-4-11, and other hyper-parameters are set to 0.001, 128,

https://doi.org/10.1021/acsomega.3c03496
ACS Omega 2023, 8, 38013−38024

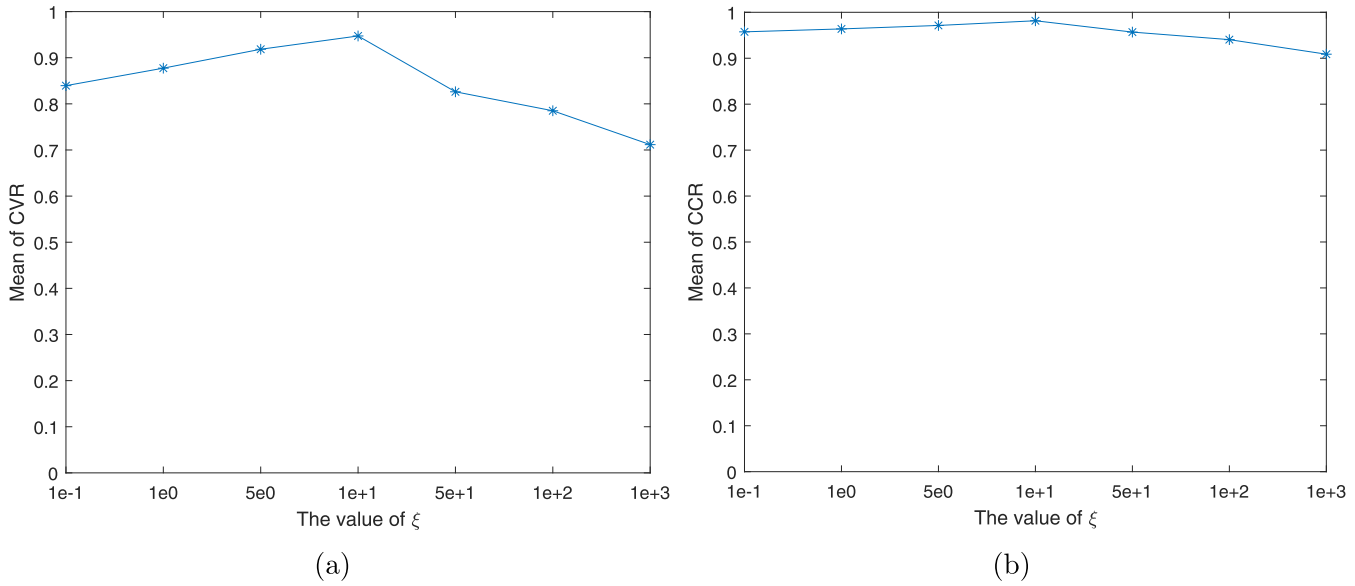


Figure 5. Sensitivity test about the parameter ξ: (a) Means of CVR versus the value of ξ, and (b) Means of CCR versus the value of ξ.

Table 4. DT-CVRs of Different Methods in the Polyethylene Process

type	FDA	SOM	SMVU	SAE	t-SNE-BP	D^{2}K-DA
normal	0.0600	0.1000	0.0000	0.4600	0.2000	0.9400
fault 1	0.0200	0.0000	0.0000	0.0000	0.2000	0.9800
fault 4	1.0000	0.7800	0.7600	0.9200	0.7400	0.8000
fault 7	0.1000	0.0400	0.0800	0.0000	0.0200	1.0000
average	0.2950	0.2300	0.2100	0.3450	0.2900	0.9300

Table 5. KNN-CCRs of Different Methods in the Polyethylene Process

type	FDA	SOM	SMVU	SAE	t-SNE-BP	D^{2}K-DA
normal	0.6800	0.8800	0.6000	0.9200	0.8200	1.0000
fault 1	0.4400	0.8400	0.3000	0.0000	0.8200	1.0000
fault 4	1.0000	0.9000	0.9000	0.9800	0.9800	1.0000
fault 7	0.2800	0.4200	0.3200	0.0200	0.5200	1.0000
average	0.6000	0.7600	0.5300	0.4800	0.7850	1.0000

and 100. For t-SNE-BP, the perplexity is set to 30; the hidden layer of BP is set to 50. For KNN-CCR, the number of nearest neighbors is set to 3.

Tables 4 and 5, respectively, tabulate detailed DT-CVRs and KNN-CCRs. Similar to the first case, D^{2}K-DA performs the best in both visualization and classification tasks. It is emphasized that this is a result of a compact cooperative between discriminative information and geometry information to reveal ideal low-dimensional visualizations.

5.2.2. Visualized Result of Operating Status. Figure 6a−f presents the visualizations of 21−70 samples of each testing data set. The results of D^{2}K-DA (Figure 6f) clearly show that the mapped query samples are well-separated for different classes and well-grouped for the same class. Other methods have obviously worse performance. Figure 7a−d illustrates the trajectories of the operating status under D^{2}K-DA. When the process operates in the normal status and in the fault status, the trajectories are mainly located in the corresponding convex boundaries (envelop lines); when a fault just happens, the trajectories directly move to the corresponding region with little delay. This low-dimensional visual information helps engineers and operators to timely and accurately monitor the process operating status and acquire depictions of the occurrence path of faults.

All of the results have shown the superiority of D^{2}K-DA.

6. CONCLUSIONS

Visual monitoring helps engineers and operators to perceive and access visualized information about process operating status, depictions of the occurrence path of faults, etc. In this study, the D^{2}K-DA framework is proposed. The features of the proposed method are summarized as follows.

- The proposed D^{2}K-DA novelly integrates discriminative information and geometry information as a consistent framework. In virtue of kernel learning, a unique DDK kernel is directly learned from specific data through one compact step of optimization, of which the flexible kernel space has enough discriminant and fewer leading degrees.
- Inheriting the class discrimination scenario, both intraclass compactness and interclass separability are guaranteed by exploiting label information to group samples from the same class and separate samples from different classes.

https://doi.org/10.1021/acsomega.3c03496
ACS Omega 2023, 8, 38013−38024

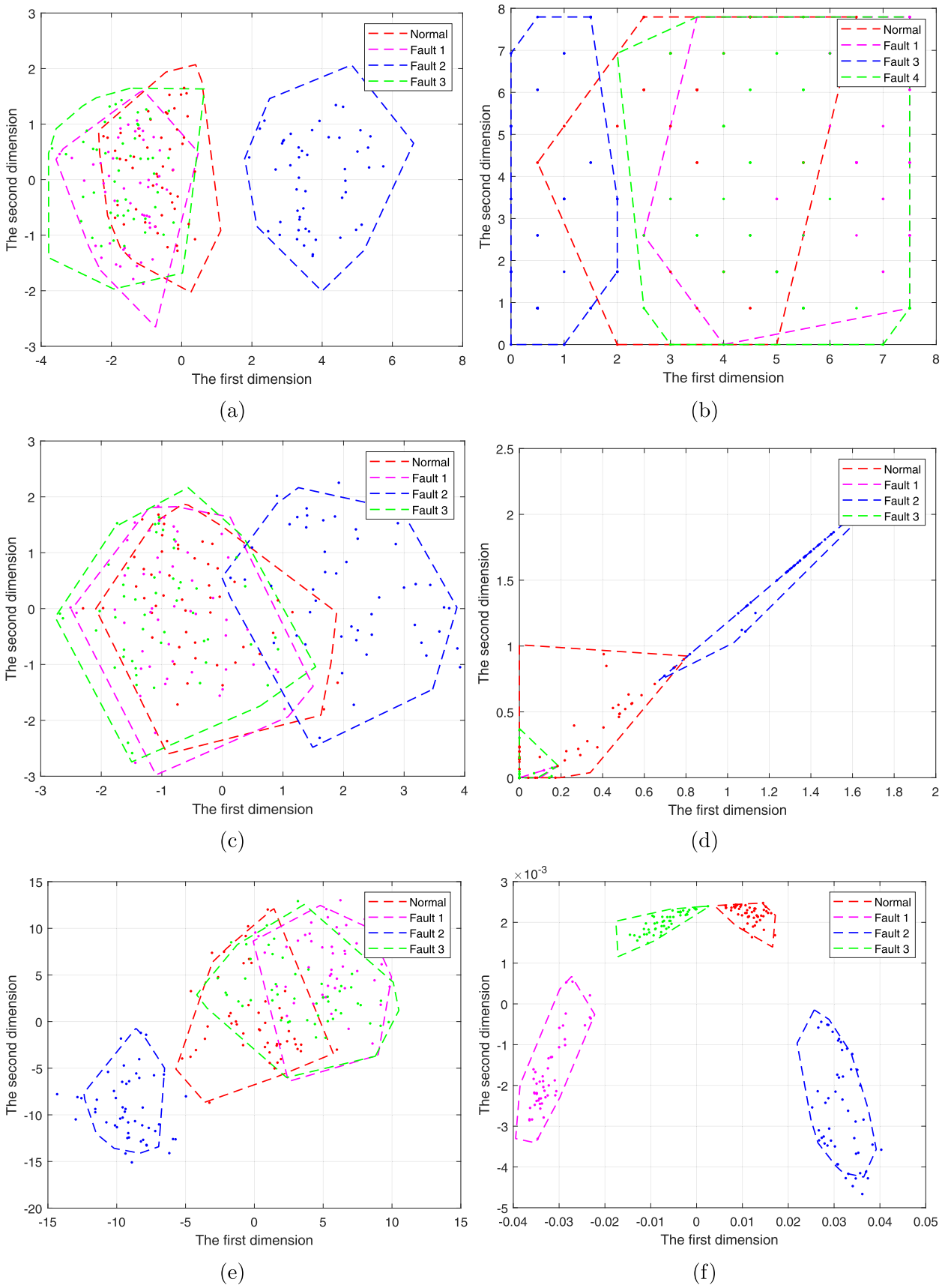


Figure 6. Visualizations of the 21−70 samples of each testing data set and the boundaries of DT-CVR in the polyethylene process. Methods: (a) FDA, (b) SOM, (c) SMVU, (d) SAE, (e) t-SNE-BP, and (f) D^2K-DA.

https://doi.org/10.1021/acsomega.3c03496
ACS Omega 2023, 8, 38013−38024

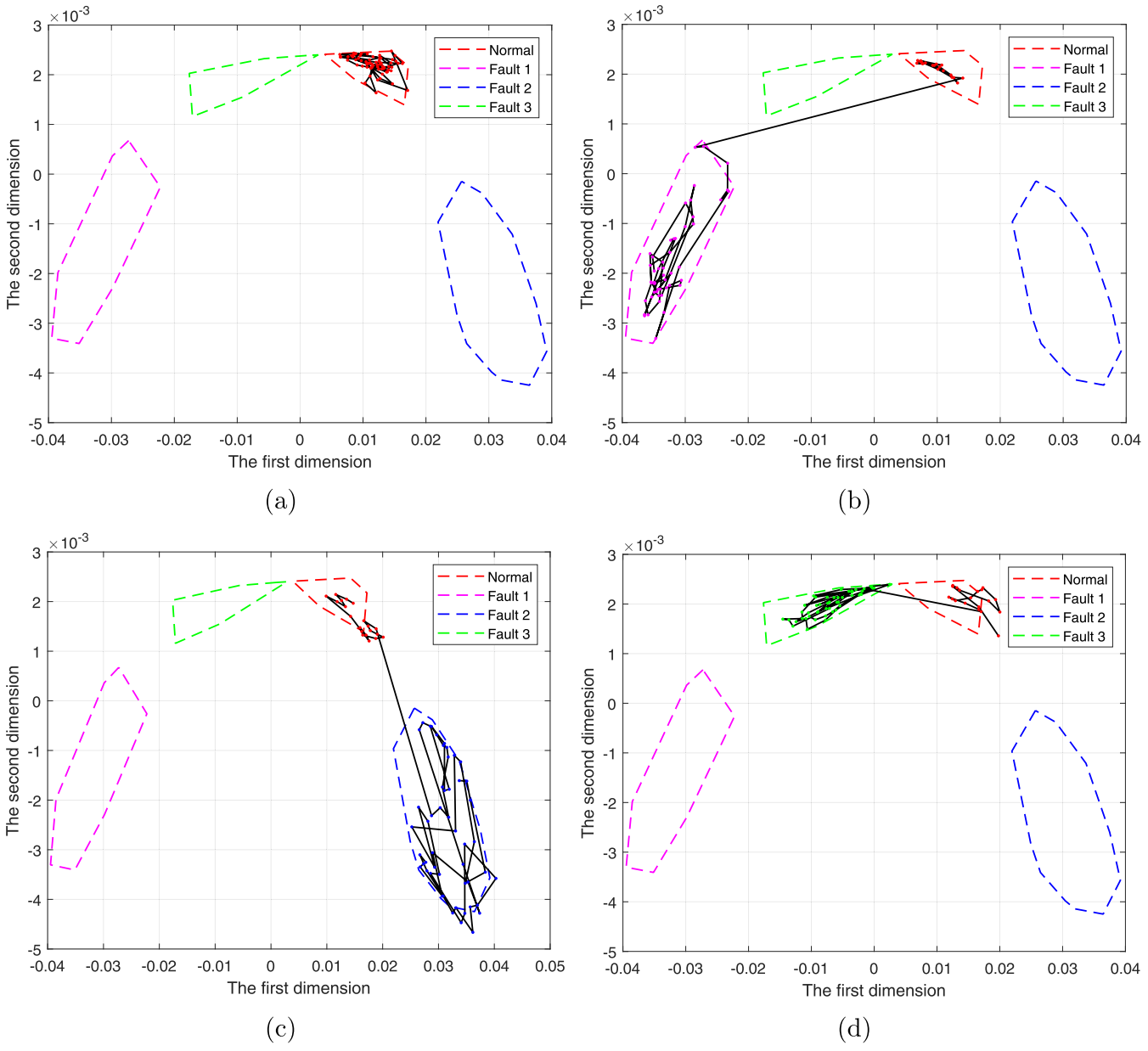


Figure 7. Trajectories of the operating status in the polyethylene process under D^{2}K-DA. Testing data sets: (a) Normal, (b) Fault 1, (c) Fault 2, and (d) Fault 3.

- Inheriting the manifold learning scenario, both local structure and global structure preservation are conducted upon Student's *t* distribution-based similarities to compress as much significative information as possible to the leading degrees of kernel space to pursue low-dimensional visualizations.

This article mainly focuses on the model construction of the proposed method and provides some preliminary analysis and application of the low-dimensional visualizations. The future work will focus on further exploiting the low-dimensional visualizations, i.e., trajectories of operating status, occurrence paths of faults, etc., and exploring potential applications on fault forecast, fault recovery, and process optimization. It is noted that the proposed D^{2}K-DA framework is currently designed in a fully supervised form in this paper, but can be easily extended to a more generalized form for partially labeled data. Also note that some regularization terms can be integrated into the proposed framework, which would extend the scope of application and future study.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c03496.

Part I: Visualizations of samples, the boundaries of DT-CVR, trajectories of the operating status, and detailed DT-CVRs and KNN-CCRs in the TE process with different random starts.; Part II: Visualizations of samples, the boundaries of DT-CVR, trajectories of the operating status, and detailed DT-CVRs and KNN-CCRs on several different faults in the TE process (PDF)

https://doi.org/10.1021/acsomega.3c03496
ACS Omega 2023, 8, 38013−38024

AUTHOR INFORMATION

Corresponding Authors

Chihang Wei − *Guangdong Key Laboratory of Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China; School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China;* orcid.org/0000-0003-1267-2763; Email: weichihang@hznu.edu.cn

Zhihuan Song − *State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China*; Email: songzhihuan@zju.edu.cn

Authors

Chenglin Wen − *Guangdong Key Laboratory of Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China*

Jieguang He − *Guangdong Key Laboratory of Petrochemical Equipment Fault Diagnosis, Guangdong University of Petrochemical Technology, Maoming 525000, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c03496

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 61933013 and 62103364).

REFERENCES

(1) Yin, S.; Gao, H.; Qiu, J.; Kaynak, O. Fault detection for nonlinear process with deterministic disturbances: A just-in-time learning based data driven method. *IEEE Trans. Cybern.* **2017**, *47*, 3649−3657.

(2) Yin, S.; Xie, X.; Lam, J.; Cheung, K. C.; Gao, H. An improved incremental learning approach for KPI prognosis of dynamic fuel cell system. *IEEE Trans. Cybern.* **2016**, *46*, 3135−3144.

(3) Zhang, J.; Chen, H.; Chen, S.; Hong, X. An improved mixture of probabilistic PCA for nonlinear data-driven process monitoring. *IEEE Trans. Cybern.* **2019**, *49*, 198−210.

(4) Shao, W.; Ge, Z.; Song, Z. Semisupervised Bayesian Gaussian Mixture Models for Non-Gaussian Soft Sensor. *IEEE Trans. Cybern.* **2021**, *51* (7), 3455−3468, DOI: 10.1109/TCYB.2019.2947622.

(5) Qin, S. J. Process data analytics in the era of big data. *AIChE J.* **2014**, *60*, 3092−3100.

(6) Ge, Z.; Song, Z.; Ding, S. X.; Huang, B. Data mining and analytics in the process industry: The role of machine learning. *IEEE Access* **2017**, *5*, 20590−20616.

(7) Ge, Z. Review on data-driven modeling and monitoring for plant-wide industrial processes. *Chemom. Intell. Lab. Syst.* **2017**, *171*, 16−25.

(8) Yin, S.; Rodriguez-Andina, J. J.; Jiang, Y. Real-Time Monitoring and Control of Industrial Cyberphysical Systems: With Integrated Plant-Wide Monitoring and Control Framework. *IEEE Ind. Electron. Mag.* **2019**, *13*, 38−47.

(9) Jiang, Y.; Yin, S.; Kaynak, O. Data-Driven Monitoring and Safety Control of Industrial Cyber-Physical Systems: Basics and Beyond. *IEEE Access* **2018**, *6*, 47374−47384.

(10) Colombo, A. W.; Karnouskos, S.; Kaynak, O.; Shi, Y.; Yin, S. Industrial Cyberphysical Systems: A Backbone of the Fourth Industrial Revolution. *IEEE Ind. Electron. Mag.* **2017**, *11*, 6−16.

(11) Jia, M.; Hu, J.; Liu, Y.; Gao, Z.; Yao, Y. Topology-guided graph learning for process fault diagnosis. *Ind. Eng. Chem. Res.* **2023**, *62*, 3238−3248.

(12) Jia, M.; Xu, D.; Yang, T.; Liu, Y.; Yao, Y. Graph convolutional network soft sensor for process quality prediction. *J. Process Control* **2023**, *123*, 12−25.

(13) Yang, C.; Liu, Q.; Liu, Y.; Cheung, Y.-M. Transfer Dynamic Latent Variable Modeling for Quality Prediction of Multimode Processes. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, 1−14, DOI: 10.1109/TNNLS.2023.3265762.

(14) Franklin, J.; Tibshirani, R.; Friedman, J. H. The elements of statistical learning: data mining, inference, and prediction. *Math. Intell.* **2005**, *27*, 83−85.

(15) Ge, Z.; Zhong, S.; Zhang, Y. Semisupervised kernel learning for FDA model and its application for fault classification in industrial processes. *IEEE Trans. Ind. Inf.* **2016**, *12*, 1403−1411.

(16) Liu, J.; Song, C.; Zhao, J.; Ji, P. Online learning based Fisher discriminant analysis and its application for fault classification in industrial processes. *Chemom. Intell. Lab. Syst.* **2019**, *191*, 30−41.

(17) Lawrence, N. D. A Unifying Probabilistic Perspective for Spectral Dimensionality Reduction: Insights and New Models. *J. Mach. Learn. Res.* **2012**, *13*, 1609−1638.

(18) Van Der Maaten, L.; Postma, E.; Van den Herik, J. Dimensionality reduction: a comparative. *J. Mach. Learn. Res.* **2009**, *10*, 66.

(19) Zhou, J. L.; Ren, Y.; Wang, J. Quality-relevant fault monitoring based on locally linear embedding orthogonal projection to latent structure. *Ind. Eng. Chem. Res.* **2019**, *58*, 1262−1272.

(20) Yu, J. Process monitoring through manifold regularization-based GMM with global/local information. *J. Process Control* **2016**, *45*, 84−99.

(21) Su, Z.; Tang, B.; Liu, Z.; Qin, Y. Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine. *Neurocomputing* **2015**, *157*, 208−222.

(22) Wei, C.; Song, Z. Generalized Semi-Supervised Self-Optimizing Kernel Model for Quality Related Industrial Process Monitoring. *IEEE Trans. Ind. Electron.* **2020**, *67*, 10876−10886.

(23) Wei, C.; Chen, J.; Song, Z.; Chen, C.-I. Development of self-learning kernel regression models for virtual sensors on nonlinear processes. *IEEE Trans. Autom. Sci. Eng.* **2019**, *16*, 286−297.

(24) Wei, C.; Zuo, L.; Zhang, X.; Song, Z. Hessian Semisupervised Scatter Regularized Classification Model With Geometric and Discriminative Information for Nonlinear Process. *IEEE Trans. Cybern.* **2022**, *52*, 8862−8875.

(25) van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579−2605.

(26) Tang, J.; Yan, X. Neural network modeling relationship between inputs and state mapping plane obtained by FDA-t-SNE for visual industrial process monitoring. *Appl. Soft Comput.* **2017**, *60*, 577−590.

(27) Kouropteva, O.; Okun, O.; Pietikäinen, M. Selection of the optimal parameter value for the locally linear embedding algorithm. *FSKD* **2002**, *2*, 359−363.

(28) De Loera, J.; Rambau, J.; Santos, F. *Triangulations: Structures for Algorithms and Applications*; Springer, 2010; Vol. *25*.

(29) Brito da Silva, L. E.; Wunsch, D. C. An Information-Theoretic-Cluster Visualization for Self-Organizing Maps. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 2595−2613.

(30) Wei, C.; Chen, J.; Song, Z. Developments of two supervised maximum variance unfolding algorithms for process classification. *Chemom. Intell. Lab. Syst.* **2016**, *159*, 31−44.

(31) Hinton, G. E.; Salakhutdinov, R. R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504−507.