Unsupervised Fault Detection for Refrigeration Showcase Systems with Kernel Principal Component Analysis based Multivariate Statistical Process Control using Feature Selection with Maximal Information Coefficient

Kiyo Arai∗ Associate, Yoshikazu Fukuyama∗∗ Senior Member
Kenya Murakami∗∗ Non-member, Tetsuro Matsui∗∗ Senior Member

This paper proposes a kernel principal component analysis (KPCA) based multivariate statistical process control (KPCA-MSPC) method for fault detection of refrigeration showcase systems using a feature selection method with maximal information coefficient (MIC). Refrigeration showcase system data include non-linear relationships among pairs of features, and only normal data can be available for training generally. KPCA-MSCP is suitable for the fault detection because it is an unsupervised method and can handle non-linear relationships. In showcase systems, a large number of measured data can be obtained and they can be utilized as features for fault detection. However, considering system costs, the number of sensors installed in the showcase systems and the amount of data stored in data centers are limited. Therefore, a feature selection method based on MIC and k-nearest neighbor algorithm (KNN) (MIC-KNN-FS) suitable for KPCA-MSPC is proposed. The effectiveness of the combination of KPCA-MSPC and the proposed MIC-KNN-FS for showcase systems is verified by comparison with the Laplacian Score feature selection method (LS-FS) and the KNN feature selection method (KNN-FS), which are typically utilized as feature selection methods, and cumulative autoencoders (CAE) and MSPC based on PCA (PCA-MSPC), which are unsupervised fault detection methods.

Keywords: showcase system, fault detection, feature selection, kernel principal component analysis, multivariate statistical process control, maximal information coefficient

1. Introduction

Refrigeration and freezing showcases (referred to as “showcase” below) are designed to maintain frozen food and beverages at low temperatures and keep the freshness of various perishable items. They are used in most supermarkets and convenience stores and have now become essential items in our daily lives. However, the internal temperature of showcases may not be kept at the preset temperature due to rare faults such as frost and refrigerant leaks. These faults may also result in disposals of displayed products and loss of sales opportunities. Therefore, to solve these kinds of issues, it is necessary to detect faults as accurately as possible.

Showcase systems are formed by multiple groups of showcases with different characteristics connected by a single refrigerator. To implement fault detection of a showcase system, it is necessary to meet the four needs below. A huge number of showcase systems of various types are installed in supermarkets and convenience stores in Japan in many different environments. Therefore, the first need is application of a method that does not require intricate parameter adjustment that can only be done by specialists. The second is that, since showcase faults are very rare events, which make it difficult to acquire abnormal value data, fault detection should be based on a method that only uses normal value data; that is, without the need for abnormal value data. The third need is that since part of the showcase data has non-linear correlations, it is necessary to apply a method that can treat the non-linear correlations. In showcase systems, various measured values are obtained as features for fault detection, but from the point of view of cost, there is a limited number of sensors that can be installed. Also, considering the data center charges, it is necessary to achieve highly accurate fault detection with a small number of sensors. Therefore, the fourth need is to only extract features that can be efficiently employed to detect faults—among a huge number of values obtained—as part of preprocessing (which is referred to as “feature selection” hereinafter).

Many showcase fault detection methods have been proposed so far, including those that use classic artificial intelligence (AI) methods such as physical models and multiple integration methods. However, all these methods must be tuned for every store, which does not meet need number 1 above. To solve this issue, we proposed a fault detection method that uses

© 2021 The Institute of Electrical Engineers of Japan.
a neural network, which is a machine learning technique. However, since this method requires fault data, it does not meet need number 2. For this reason, we proposed a fault detection method for showcase data that uses Cumulative Autoencoders (CAE) based on an ensemble approach that does not require specialist knowledge or fault data. Since this method uses a neural network, it can treat the non-linear correlations and, therefore, meets the needs number 1, 2, and 3 above. However, since this method does not consider feature selection, it cannot be easily applied to showcase systems with a high number of features and, therefore, it does not meet the need number 4. Therefore, through a combination of a feature selection method and a fault detection method that suit showcase data, it is necessary to develop a method capable of reducing the number of features and improving the fault detection accuracy as much as possible.

If we look at needs number 2 and 3, the Kernel Principal Component Analysis-based Multivariate Statistical Process Control (KPCA-MSPC) which does not require fault data and can treat non-linear correlations, may be applicable. Also, as mentioned in need number 4, it is necessary to implement a feature selection method that can filter a large set of measured values and extract only those that are effective for fault detection, as a preprocessing step. Feature selection methods are classified into three types, which are called filter methods, wrapper methods, and embedded methods. However, the wrapper and embedded methods select features using learning models, which take too much time and, therefore, are not practical. Meanwhile, the filter methods select features using only a dataset, not a learning model, which makes the feature selection a shorter and more practical process. The filter methods include, for example, an approach that ranks each feature according to an evaluation index and selects the top-ranked features, as well as an approach that selects the most representative features using the linear correlations between them. The Laplacian Score based feature selection method (referred to as “LS-FS”) and the feature selection method using the k-nearest neighbor (referred to as “KNN-FS”) are examples of these two approaches that do not require fault data. Moreover, since the fault detection using KPCA-MSPC is a method based on the non-linear correlations between features, if KNN-FS is improved so that it can be applied to the non-linear correlations, it may be developed into a feature selection method that suits KPCA-MSPC. Since the Maximal Information Coefficient (MIC) can evaluate the non-linear correlations between features, it is possible to improve KNN-FS using MIC and apply it to KPCA-MSPC.

The purpose of this paper is to improve KNN-FS using MIC that can measure the non-linear correlations between features and present it as a new feature selection method. We also propose a new showcase system fault detection method that combines MIC-KNN-FS as a feature selection method and KPCA-MSPC as a fault detection method (Fig. 1). The contributions of this paper are as follows:

1. Proposal of MIC-KNN-FS, a feature selection method that is classified as an appropriate filter method for KPCA-MSPC and only uses normal data.

2. Proposal of a new method that combines the MIC-KNN-FS above and KPCA-MSPC capable of detecting faults in showcase systems with high accuracy using the minimum amount of data (that is, without fault data) to lower the costs.

3. To verify the effectiveness of the showcase system fault detection method that combines the proposed MIC-KNN-FS and KPCA-MSPC, it is compared with two commonly used filter methods of LS-FS and KNN-FS, as well as two fault detection methods that do not require fault data—the traditional showcase fault detection method of CAE and Principal Component Analysis-based Multivariate Statistical Process Control (PCA-MSPC).

2. An Overview of Fault Detection of Showcase Systems

Showcases are installed not only in stores where customers can access but also in backyards to freeze and refrigerate products, and some convenience stores have more than ten units installed. Showcase systems are formed by multiple groups of showcases with different characteristics connected by a single refrigerator. In showcase systems, fault detection is performed using only valid measured values extracted from a large set of data, such as the individual temperature at different points and refrigerant pressure and flow rate of each showcase.

This kind of showcase system is installed in as many as 77,994 supermarkets and convenience stores in Japan (as of April 2020). Since each of these showcases has different properties and each store is in a region with a different climate, it is necessary to adjust the fault detection of the showcase systems for each store. Also, if a local server is to be installed in every store, it incurs large installation costs. Therefore, the method presented in this paper assumes a remote diagnostic system that first aggregates the measured data obtained from stores using, for example, a cloud service. Then, it automatically generates a fault detection model for each store, which is used to detect abnormalities based on online information. Therefore, if data center charges are taken into account, it is necessary to implement a highly accurate fault detection system with the lowest number of features. To this end, the first step is to select the features offline and determine the minimum required number of sensors and amount of information. Then, it automatically generates a model using past measured values and sets the control limits. Finally, it goes online and detects faults in the stores (Fig. 2). In a broad sense, the proposed method can be considered a type of artificial intelligence that can contribute to solving IoT problems using AI advocated by Society 5.0.


3. The Proposed MIC-KNN-FS

As mentioned in Chapter 1, in this paper, we propose an improved version of KNN-FS\(^{(12)}\) that can be used to analyze the non-linear correlations between features. With this, it is possible to increase the accuracy of fault detection based on KPCA-MSPC and apply it to showcase data.

The proposed MIC-KNN-FS has two procedures. The first is to evaluate the non-linear correlations between all features (Section 3.1), which is followed by the selection of the most representative features based on the calculated evaluation values of non-linear correlations between features (Section 3.2).

3.1 Evaluation of Non-linear Correlations between Features based on Maximal Information Coefficient (MIC)

The proposed MIC-KNN-FS requires an index that can evaluate the non-linear correlations between features. Since the conventional KNN-FS only treats linear correlations, the proposed feature selection method uses MIC. It is a method newly developed in 2011 that can be used to measure any correlations between features. Its effectiveness was verified by comparing it with other methods that can measure the correlations between features\(^{(13)}\).

The principle and algorithm of MIC are as follows: Fig. 3 shows the scatter plots of the data (referred to as “samples”) of two uncorrelated features \((X_1, X_2)\) and two non-linearly correlated features \((X_1, X_3)\). With MIC, a grid is set for this kind of scatter plots and, using the percentage of samples contained in each grid, the mutual information (referred to as “IC”) is calculated. Figure 3(b), which has more samples concentrated in a specific grid, has a higher IC value, while in Fig. 3(c), the samples are scattered across the grids, the IC value decreases. The IC values were calculated using Eq. (1) below:

\[
IC = \sum_{a=1}^{A} \sum_{b=1}^{B} P(a, b) \log \left( \frac{P(a, b)}{P(a)P(b)} \right)
\]

where \(A\) and \(B\): the number of regions created by vertical (A) and horizontal (B) partition lines; \(a\): \(a\)-th region in the horizontal direction created by the vertical partition lines; \(b\): \(b\)-th region in the vertical direction created by the horizontal partition lines; \(P(a, b)\): a joint probability distribution function of \(a\)-th region in the horizontal direction and \(b\)-th region in the vertical direction; \(P(a)\): a marginal probability distribution function of \(a\)-th region in the horizontal direction; \(P(b)\): a marginal probability distribution function of \(b\)-th region in the vertical direction.

Equation (1) is detailed using an example. First, the samples are divided into multiple grids using partition lines (Fig. 4(a), \((A, B) = (4, 4)\)). Then, \(P(a, b) \log(P(a, b) / (P(a)P(b)))\) is calculated for each grid. For example, in the left chart of Fig. 4(b), since one of the 20 samples is in \(a_1\) and \(b_1\) regions, \(P(a_1, b_1)/20 = 0.05\). Also, since four of all 20 samples are in \(a_1\) region, \(P(a_1)/20 = 0.2\). \(P(b_1)\) can be calculated the same way as \(P(a_1)\). Lastly, the sum of \(P(a, b) \log(P(a, b) / (P(a)P(b)))\) of all grids is the IC value (Fig. 4(c)).

As indicated in Fig. 4(b) and (c), if there are many samples in a single grid, the joint probability distribution function, \(P(a, b)\), increases, and so does the IC value. Therefore, with Eq. (1), if many samples are concentrated in a specific grid, the IC value increases, so it can detect any type of non-linear correlations regardless of the type of correlations between features\(^{(13)}\).

However, depending on how the grid lines are drawn, the number of grids where the samples are concentrated may change, and the IC value may not be calculated correctly. Figure 5 shows an example of how the IC value of two features with strong non-linear correlation varies with different...
Step 1 The total number of features before feature selection is defined as \( N_1 \), the set of all features as \( x = \{ x_i, i = 1, \ldots, N_1 \} \), and the set of features after feature selection as \( y = \{ y_i, i = 1, \ldots, N_2 \} \), \( y \leftarrow x \).

Step 2 The initial value of neighbors in the k-nearest neighbor algorithm is set as \( k_{\text{neighbors}} \), and the minimum limit value, as \( k_{\text{limit}} \). Also, \( i = 1 \).

Step 3 For features \( x_i \in y \), calculate \( \text{MIC}(x_i, x_j) \), which is the evaluation value of the non-linear correlation between features \( x_i \) and \( x_j \), the other features in set \( y \) (\( j = 1, \ldots, N_2 - 1, j \neq i \)).

Step 4 \( x_j \) when \( \text{MIC}(x_i, x_j) \) is the maximum, is defined as \( x_{j_{\text{max}}} \). Calculate \( \text{MIC}(x_{j_{\text{max}}}, x_j) \), which is the evaluation value of the non-linear correlation between \( x_{j_{\text{max}}} \) and \( x_j \) (\( j = 1, \ldots, N_2 - 1, j \neq j_{\text{max}} \)), the other features in set \( y \). Then, delete \( k_{\text{neighbors}} \) features with the largest \( \text{MIC}(x_{j_{\text{max}}}, x_j) \) from \( y (N_2 = N_2 - k_{\text{neighbors}}) \).

Step 5 If \( k_{\text{neighbors}} > N_2 - 1 \), then \( k_{\text{neighbors}} = N_2 - 1 \).

Step 6 If \( k_{\text{neighbors}} = k_{\text{limit}} \), go to Step 8; if not, go to Step 7.

Step 7 If \( i = N_1 \), go to Step 8; if not, \( i = i + 1 \) and go to Step 3.

Step 8 The set of features after feature selection is output as \( y \).

4. An Overview of KPCA-MSPC

With PCA-MSPC, when principal component analysis (referred to as “PCA”) is applied to normal data with linear correlations between features, the features of this normal data are obtained by the principal component subspace that captures the variation of the data (PCS, Fig. 7(c)) and the residual subspace that is its orthogonal complement (referred to as “RS” of Fig. 7(c)) \(^{(19)}\). For example, when \( \Phi \) PCA is applied to normal data in the feature space of Fig. 7(b), new principal component coordinate axes are created along with the data distribution (PC 1 and PC 2 of Fig. 7(c)). The space containing the principal component axis is called PCS, and in the PCS, the features of normal data can be derived like a circle in a two-dimensional space centered on PC 1 and PC 2 and controlled by Hotelling’s \( T^2 \) proposed by Jackson \(^{(19)}\). Therefore, \( T^2 \) detects fault according to the magnitude of the mapping value to the PCS of the data.

Also, since RS is the space that represents the information lost by dimensional compression (residual) when new principal component axes were created, it can express parts of the features of the normal data that cannot be expressed by PCS. Moreover, RS can be controlled by the squared prediction error (referred to as SPE) calculated from the residual (Eq. (8) to be detailed later) \(^{(19)}\). Therefore, SPE is the evaluation of data that cannot be expressed by the principal components, and the fact that SPE increases means that the parts that cannot be expressed by principal components—that is, the parts that do not suit the general properties of the process—are increasing and can be used in fault detection.

However, when PCA is used, the fault detection accuracy for data containing non-linear correlations decreases. Kernel PCA-based MSPC (KPCA-MSPC) was proposed so that it could treat the non-linear correlations \(^{(20)}\). With KPCA-MSPC, by non-linear mapping (\( \Phi \) Kernel Function in Fig. 7) the features from the existing space (Input Space in Fig. 7(a)) to a high-dimension feature space (Feature Space in Fig. 7(b)), it is possible to apply the linear method of PCA (\( \Phi \) in Fig. 7) to the mapping data and achieve a highly accurate fault diagnostic.

KPCA-MSPC is explained below using numerical formulas. With KPCA-MSPC, a matrix containing normal data with \( S \) samples for \( N \) features is defined as \( X \in \mathbb{R}^{S \times N} \) (assuming that \( X \) has already been standardized) \(^{(20)}\). Also, \( X \in \mathbb{R}^{S \times N} \) is mapped to the M-dimensional (\( \Phi(X_i, i = 1, \ldots, S) \), the matrix containing this mapping data is defined as \( z = \Phi(X) \in \mathbb{R}^{S \times M} \), and \( z \) is centralized. The inner product of this mapping data is calculated with a kernel function (Eq. (2), \( k(X_i, X_j) := \langle \Phi(X_i) \Phi(X_j) \rangle \)). With this, it is possible to apply the non-linear PCA.

\[
K_1 = x x^T = \begin{bmatrix}
\langle \Phi(X_1) \Phi(X_1) \rangle^T & \cdots & \langle \Phi(X_1) \Phi(X_S) \rangle^T \\
\vdots & \ddots & \vdots \\
\langle \Phi(X_S) \Phi(X_1) \rangle^T & \cdots & \langle \Phi(X_S) \Phi(X_S) \rangle^T
\end{bmatrix}
\]
For example, the following Gaussian kernel is applied to captures the data variation. Then, if the loading value is larger than the threshold value, it can judge it as abnormal, and if it is smaller than the threshold value, as normal.

For example, the following Gaussian kernel is applied to $k(X_i, X_j)$:

$$k(X_i, X_j) = \exp \left\{ \frac{(X_i - X_j)(X_i - X_j)^T}{C} \right\} \cdots \cdots \cdots \cdots \cdots (3)$$

where $C$: a parameter of the Gaussian kernel.

PCA is applied to the kernel matrix $K_1$, which contains the inner product of the mapping data, to calculate the eigenvalues and eigenvectors. Also, if $L$ principal components are selected using the principal component ratio set, the loading matrix $P_L \in R^{M \times L}$ is defined as Eq. (4):

$$P_L = z^T A_L \sum_{t=1}^{L-1} \cdots$$

$$= \begin{bmatrix}
  z_{s1} & \cdots & z_{sL} \\
  \vdots & \ddots & \vdots \\
  z_{M1} & \cdots & z_{ML}
\end{bmatrix} \begin{bmatrix}
  A_{s1} & \cdots & A_{sL} \\
  \vdots & \ddots & \vdots \\
  A_{M1} & \cdots & A_{ML}
\end{bmatrix} \begin{bmatrix}
  1 & \cdots & 0 \\
  0 & \ddots & \vdots \\
  0 & \cdots & 1
\end{bmatrix} \cdots \cdots \cdots \cdots \cdots (4)$$

where $A_L \in R^{S \times L}$: a matrix with $L$ eigenvectors of the kernel matrix arranged its columns, and $\Sigma_L \in R^{S \times S}$: a diagonal matrix with $L$ eigenvalues arranged in descending order on diagonal elements ($\lambda_1 > \lambda_2 > \ldots > \lambda_L$).

The column space of the loading matrix $P_L$ is a PCS that captures the data variation. Then, $\varphi(X_i) \in R^L$, the mapping data of the test data of one sample $X_i \in R^S$, is projected onto the PCS to calculate the $T^2$. For example, if the $T^2$ value of this test data of one sample is larger than the threshold value, it can judge it as abnormal, and if it is smaller than the threshold value, as normal (Fig. 7(c)). $T^2$ can be defined by the following equation:

$$T^2 = t^T \sum_{l=1}^{L-1} t_l$$

$$= \begin{bmatrix}
  t_1 & \cdots & t_L
\end{bmatrix} \begin{bmatrix}
  1 & \cdots & 0 \\
  \vdots & \ddots & \vdots \\
  0 & \cdots & 1
\end{bmatrix} \begin{bmatrix}
  t_1 \\
  \vdots \\
  t_L
\end{bmatrix} \cdots \cdots \cdots \cdots \cdots (5)$$

where $t \in R^L$: a principal component score vector for mapping data $\varphi(X_i)$.

Also, Eq. (5) can be calculated as follows:

$$t^T \sum_{l=1}^{L-1} t_l = \varphi(X_i)^T P_L \sum_{l=1}^{L-1} P_L^T \varphi(X_i)$$

$$= K_2(X_i)^T A_L \sum_{l=1}^{L-1} A_L^T K_2(X_i) \cdots \cdots \cdots \cdots \cdots (6)$$

$$K_2(X_i) = \varphi(X_i)^T (\varphi(X_i), \cdots, \varphi(X_i))^T \sum_{l=1}^{L-1} \cdots \cdots \cdots \cdots \cdots (7)$$

where $K_2(X_i)$: a vector with inner product of $\varphi(X_i)$ and $z = \varphi(X)$, the mapping data corresponding to the samples of normal data.

Next, SPE is calculated to manage RS, the residual subspace. For example, if the SPE value of this test data $X_i$ of one sample is larger than the threshold value, it can judge it as abnormal, and if it is smaller than the threshold value, as normal (Fig. 7(c)). SPE can be defined as Eq. (8):

$$Q = \| (I_M - P_L P_L^T) \varphi(X_i) \|^2 \cdots \cdots \cdots \cdots \cdots (8)$$

where $I_M \in R^{M \times M}$: the identity matrix.

Also, Eq. (8) can be calculated as follows:

$$\| (I_M - P_L P_L^T) \varphi(X_i) \|^2$$

$$= k(X_i, X_i) - K_2(X_i)^T A_L \sum_{l=1}^{L-1} A_L^T K_2(X_i) \cdots \cdots \cdots \cdots \cdots (9)$$

5. Fault Detection of Showcase Systems by KPCA-MSPC based on MIC-KNN-FS

The proposed fault detection of showcase systems by KPCA-MSPC based on MIC-KNN-FS comprises three steps: Feature selection (Step 1), generating a model that uses normal data (Step 2), and fault detection for diagnostic data (Step 3), which are detailed below:

Step 1 Feature selection is carried out with the proposed MIC-KNN-FS (chapter 3).

Step 2 Generating a model by applying KPCA-MSPC (chapter 4) to normal data, calculating $T^2_s$ and $SPE_s$ ($s = 1, \ldots, S$) of each sample under normal conditions, and determining the threshold value (the control limit) using Eqs. (10) and (11):
and $\beta$: confidence levels of $T^2$ and SPE values.

Step 3 Diagnostic data is applied to the generated model, and the $T^2_i$ and $SPE_i (i = 1, \ldots, T)$ values of the samples of diagnostic data are calculated. For example, if $SPE_i > Thre_{SPE}$, it is judged as abnormal; otherwise, as normal.

Generally, faults detected only by the $T^2$ value, for example, are caused by drastic changes in operating conditions, and faults detected by the SPE value are caused by loss correlation between features resulting from equipment fault (22). As an example, in Step 3 above, the SPE value is used, but how to evaluate the $T^2$ and SPE values is something that needs to be examined for each problem.

6. Simulation

6.1 Simulation Conditions Normal and fault time-series data obtained from actual showcase systems are used in the simulation. For the refrigerant leak data, a specialist makes a comprehensive analysis of various values measured at the initial stage of a refrigerant leak and sets a moment makes a comprehensive analysis of various values measured in the simulation. For the refrigerant leak data, a specialist

Simulation Results Table 1 shows the accuracy and specificity (average values of ten cross-validations) obtained when the showcase system fault detection methods combining each combination of feature selection and fault detection methods are applied to the test data with different numbers of selected features. The bold numbers in the table represent the best values of accuracy and specificity.

Overall, Table 1 shows that even when a limited number of features (only 10 from 110) is used, these methods can detect abnormalities very accurately, which demonstrates that it is possible to reduce the system costs.
Next, the specificity levels indicate that KPCA-MSPC and CAE, the fault detection methods that can treat non-linear correlations between features, have a higher percentage of correct fault sample detection than PCA-MSPC, which cannot treat non-linear correlations. In other words, these results confirm that fault detection methods that can treat non-linear correlations between features are more appropriate for fault detection of showcase systems.

Moreover, let us discuss which is the most appropriate for showcase systems among all methods tested (1⃝ to 9⃝ in Table 1). As indicated in Table 1, the proposed method of showcase system fault detection that combines MIC-KNN-FS and KPCA-MSPC (Table 8) has the highest accuracy with all numbers of selected features. In other words, the proposed method produces the most accurate detection results with the test data, and even with different numbers of selected features, it can consistently produce highly accurate results. This indicates that, even in cases where the number of sensors must be reduced because of cost restrictions, the proposed method can maintain a high level of detection accuracy. Also, the highest specificity value of 99.858% (when the number of selected features is 8) is obtained with the proposed method. This means that the proposed method can correctly detect fault samples with a higher percentage than the other methods. The method that combines CAE and LS-FS (Table 8) produces high specificity with 2, 4, and 10 features, but in terms of accuracy, it has a lot of room for improvement. For this reason, from the perspective of practicality, the proposed method is superior.

Lastly, let us discuss the combination of the proposed KPCA-MSPC with MIC-KNN-FS. KPCA-MSPC is combined with three feature selection methods (specificity of 1⃝ to 3⃝ in Table 1), but the results indicate that it produces the most stable accuracy and specificity values when combined with MIC-KNN-FS. This suggests that KPCA-MSPC and MIC-KNN-FS have good compatibility. With the proposed MIC-KNN-FS, the evaluation values of non-linear correlations between features are used, and representative features with non-linear correlations are selected. Since KPCA-MSPC performs fault detection based on the non-linear correlations between features, its combination with MIC-KNN-FS produces good results.

### Table 1. Average accuracy and specificity with different numbers of selected features by various fault detection methods using various features selection methods for practical showcase data

| Method | Numbers of selected features |
|--------|-----------------------------|
|        | 2  | 4  | 6  | 8  | 10 |
| ① The proposed KPCA-MSPC based method using MIC-KNN-FS | Accuracy [%] | 99.420 | 99.331 | 99.454 | 99.495 | 99.529 |
|        | Specificity [%] | 98.069 | 99.052 | 99.621 | 99.858 | 97.962 |
| ② A KPCA-MSPC based method using LS-FS | Accuracy [%] | 96.940 | 97.593 | 98.318 | 99.154 | 99.427 |
|        | Specificity [%] | 95.782 | 85.588 | 96.682 | 98.199 | 97.678 |
| ③ A KPCA-MSPC based method using KNN-FS | Accuracy [%] | 97.693 | 98.894 | 98.867 | 99.235 | 99.365 |
|        | Specificity [%] | 86.256 | 96.682 | 97.109 | 96.777 | 96.730 |
| ④ A PCA-MSPC based method using MIC-KNN-FS | Accuracy [%] | 85.590 | 85.597 | 99.140 | 99.140 | 99.440 |
|        | Specificity [%] | 0.047 | 0.142 | 94.123 | 94.123 | 96.209 |
| ⑤ A PCA-MSPC based method using LS-FS | Accuracy [%] | 86.724 | 96.505 | 85.584 | 99.406 | 99.256 |
|        | Specificity [%] | 9.289 | 75.924 | 0.000 | 96.209 | 95.166 |
| ⑥ A PCA-MSPC based method using KNN-FS | Accuracy [%] | 85.973 | 92.239 | 92.123 | 87.836 | 87.563 |
|        | Specificity [%] | 3.649 | 46.255 | 45.403 | 15.782 | 13.791 |
| ⑦ A CAE based method using MIC-KNN-FS | Accuracy [%] | 86.710 | 88.956 | 94.287 | 94.130 | 93.911 |
|        | Specificity [%] | 47.299 | 60.474 | 97.820 | 98.910 | 98.957 |
| ⑧ A CAE based method using LS-FS | Accuracy [%] | 93.474 | 93.536 | 94.150 | 94.027 | 93.706 |
|        | Specificity [%] | 98.957 | 99.100 | 98.815 | 99.052 | 99.147 |
| ⑨ A CAE based method using KNN-FS | Accuracy [%] | 94.171 | 95.065 | 95.345 | 96.990 | 99.263 |
|        | Specificity [%] | 92.749 | 98.294 | 99.573 | 80.758 | 96.628 |

1) Bold numbers mean the best values for "Accuracy" and "Specificity".

### 7. Conclusions

This paper proposes MIC-KNN-FS, a feature selection method based on the non-linear correlations between features. It also proposes a fault detection method that combines this MIC-KNN-FS with a kernel principal component analysis-based multivariate statistical process control (KPCA-MSPC), which can treat the non-linear correlations between features and can detect fault using only normal data, then applies it to showcase systems.

To verify the effectiveness of the proposed showcase system fault detection method that combines the proposed MIC-KNN-FS and KPCA-MSPC, it is compared with two feature selection methods—the Laplacian Score feature selection (LS-FS), a commonly used filter method, and KNN feature selection (KNN-FS)—and two fault detection methods that do not require fault data—the Cumulative Autoencoders (CAE), a traditional showcase fault detection method, and
principal component analysis-based multivariate statistical process control (PCA-MSPC), one of the most used methods. The simulation results indicate that the showcase system fault detection method that combines the proposed MIC-KNN-FS and KPCA-MSPC can limit the number of features required, which can help reduce the system costs. Also, using only a limited number of features, it consistently produces more accurate detection results than the other methods, as well as the highest percentage of correct detection of fault samples.

As detailed above, by limiting the number of features by feature selection, it is possible to reduce the system cost, and by improving the accuracy of fault detection, it is possible to avoid disposals of displayed products and losses of sales opportunities due to equipment faults. Our goal for the future is to develop an even more accurate feature selection method and fault detection method to further contribute to this objective.

References

(1) K. Miura, Y. Aono, T. Okhawa, N. Komoda, and F. Miyazaki: “Diagnosis of an Air Conditioning System by Stochastic Qualitative Reasoning”, Trans. of the Society of Instrument and Control Engineers, Vol.31, No.9, pp.1510–1518 (1995) (in Japanese)

(2) K. Shibata, A. Takahashi, and A. Shimizu: “Diagnostic Model of Air Conditioning System by HierarchiFrame Network”, Trans. of the Society of Heating, Air-Conditioning and Sanitary Engineers of Japan, Vol.24, No.73, pp.65–75 (1999) (in Japanese)

(3) L. Chang, H. Wang, and L. Wang: “Cloud-Based parallel implementation of an intelligent classification algorithm for fault detection and diagnosis of HVAC systems”, Proc. of the IEEE International Smart Cities Conference (2017)

(4) T. Kawai, K. Mizakami, Y. Otsawa, M. Honjo, S. Obiya, and T. Morii: “A Fault Forecast Method for Refrigerated Super-Showcases Based on Operation Data. 1st Report”, Trans. of the JSME, Vol.67, No.655, pp.744–750 (2001) (in Japanese)

(5) H. Han, Z. Cao, B. Gu, and N. Ren: “PCA-SVM-Based automated fault detection and diagnosis (AFDD) for Vapor-Compression refrigeration systems”, HVAC&R RESEARCH, Vol.16, No.3, pp.295–313 (2010)

(6) N. Otaka, Y. Fukuyama, Y. Kawamura, K. Murakami, A. Santana, T. Izaka, and T. Matsui: “Refrigerated Showcase Fault Analysis by a Correntropy-Based Artificial Neural Network Using Particle Swarm Optimization”, Proc. of SICE Annual Conference (2018)

(7) A. Santana, Y. Kawamura, K. Murakami, T. Izaka, T. Matsui, and Y. Fukuyama: “Unsupervised fault detection in refrigeration showcase with single-class data using autoencoders”, IIEE Transactions on Electronics, Information and Systems, Vol.139, No.10, pp.1191–1200 (2019)

(8) J.-M. Lee, C.K. Yoo, S.W. Choi, P.A. Vunolleghem, and I.-B. Lee: “Nonlinear process monitoring using kernel principal component analysis”, Chemical Engineering Science, Vol.59, No.1, pp.223–234 (2004)

(9) S.W. Choi, C. Lee, J.-M. Lee, J.H. Park, and I.-B. Lee: “Fault detection and identification of nonlinear processes based on kernel PCA”, Vol.75, No.1, pp.55–67 (2005)

(10) G. Chandrashekhar and F. Sahin: “A survey on feature selection methods”, Computers and Electrical Engineering, Vol.40, No.1, pp.16–28 (2014)

(11) X. He, D. Cai, and F. Nie: “Laplacian Score for feature selection”, Neural Information Processing Systems (2005)

(12) P. Mitra, C.A. Murthy, and S.K. Pal: “Unsupervised feature selection using feature similarity”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.3, pp.301–312 (2002)

(13) D.N. Reshef, Y.A. Reshef, H.K. Finucane, S.R. Grossman, G. McVean, P.J. Turnbaugh, E.S. Lander, M. Mitzenmacher, and P.C. Sabeti: “Detecting novel associations in large data sets”, Science, Vol.334, No.6062, pp.1518–1524 (2011)

(14) B.M. Wise and N.B. Gallagher: “The process chemometrics approach to process monitoring and fault detection”, Process Control, Vol.6, No.6, pp.329–348 (1996)

(15) Japan supermarket association, the number of supermarkets, http://www.j-ssom.jp/index.html

(16) Japan franchise association, statistical data of convenience stores, https://www.jfa.or.jp/particle/324.html

(17) Cabinet office, Society 5.0, https://www8.cao.go.jp/cstp/society5/4

(18) D.N. Reshef, Y.A. Reshef, H.K. Finucane, S.R. Grossman, G. McVean, P.J. Turnbaugh, E.S. Lander, M. Mitzenmacher, and P.C. Sabeti: “Supporting online material for detecting novel associations in large data sets”, https://www.sciencemag.org/content/full/334/6062/1518/DC1 (2012)

(19) J.E. Jackson: “Quality control methods for several related variables”, Technometrics, Vol.1, pp.359–377 (1959)

(20) Y. Satoyama, K. Fujiiwara, and M. Kano: “Variable elimination-based contribution method for supporting fault diagnosis”, Proc. of SICE Multi Symposium on Control Systems (2016) (in Japanese)

(21) C.F. Alcala and S.J. Qin: “Reconstruction-based contribution for process monitoring with kernel principal component analysis”, Proc. of the IEEE American Control Conference, USA (2010)

(22) M. Kano: “Statistical Process Control Using Process Chemometrics”, Trans. of the Institute of Systems, Control and Information Engineers, Vol.48, No.5, pp.165–170 (2004) (in Japanese)

(23) D.P. Kingma and J. Ba: “Adam: A Method for Stochastic Optimization”, Proc. of the 3rd International Conference for Learning Representations, San Diego (2015)