

A Multi-model Approach for Soft Sensor Development Based on Feature Extraction Using Weighted Kernel Fisher Criterion^{*}

LÜ Ye (吕业) and YANG Huizhong (杨慧中)**

Key Laboratory of Advanced Process Control for Light Industry of Jiangnan University, Wuxi 214122, China

Abstract Multi-model approach can significantly improve the prediction performance of soft sensors in the process with multiple operational conditions. However, traditional clustering algorithms may result in overlapping phenomenon in subclasses, so that edge classes and outliers cannot be effectively dealt with and the modeling result is not satisfactory. In order to solve these problems, a new feature extraction method based on weighted kernel Fisher criterion is presented to improve the clustering accuracy, in which feature mapping is adopted to bring the edge classes and outliers closer to other normal subclasses. Furthermore, the classified data are used to develop a multiple model based on support vector machine. The proposed method is applied to a bisphenol A production process for prediction of the quality index. The simulation results demonstrate its ability in improving the data classification and the prediction performance of the soft sensor.

Keywords feature extraction, weighted kernel Fisher criterion, classification, soft sensor

1 INTRODUCTION

Petrochemical process is a typical continuous industrial process which is multivariable, nonlinear and uncertain in nature. The process operational data is featured with multi-modes, high disturbance and time-dependent dynamics [1]. Therefore, using a single model to describe a process may result in poor prediction performance and generalization ability [2, 3]. A multi-model approach with object properties considered can solve such problems. Unfortunately, traditional clustering algorithms may lead to overlapping phenomenon in subclasses so that edge classes and outliers cannot be effectively handled. Linear discriminant analysis (LDA) enables different subclasses to have maximum discriminant by searching for an optimal projection vector. This technology is mainly applied to image processing and face recognition for improving classification accuracy [4–8]. Kernel Fisher discriminant analysis (KFDA) is a nonlinear extension of LDA, which is more efficient in solution of nonlinear classification problems [9–14]. However, due to the structural limitation of KFDA, *i.e.*, the definition of between-class scatter matrix is less than optimal, it is still difficult to accurately classify the data when edge classes and outliers coexist [15].

In order to further improve the prediction performance of the multi-model based on traditional clustering methods, a multi-model approach based on feature extraction using weighted kernel Fisher discriminant analysis (WKFDA) is introduced in this paper. First of all, based on the Fisher criterion, a new weighted function with variable weight is proposed to rectify the within-class scatter matrix and between-class scatter matrix. Secondly, the category information measurement is put forward to describe the scatter

degree of within-class and between-class for each dimensional component. This index is applied to determine the number of feature vectors that should be included in the optimal projection matrix. Furthermore, the original data is successfully classified into multiple sub-classes based on the extracted projection feature. Then, the support vector machine (SVM) is employed to establish the sub-models using all the samples belonging to that sub-class [16–20]. Finally, a multi-model based soft sensor is obtained by using a switcher to combine the SVM sub-models. The proposed method is evaluated on a bisphenol A production process to indicate the real-time phenol composition at the dissolution tank outlet of a crystallization unit. The simulation results demonstrate the effectiveness of the proposed method.

2 WEIGHTED KERNEL FISHER DISCRIMINANT ANALYSIS

2.1 Kernel Fisher criterion

LDA searches for an optimal discriminant vector in a lower-dimension feature space to maximize the ratio of between-class scatter matrix over the within-class scatter matrix. After projection of the original high dimensional data to the feature space, it becomes much easier to classify because the samples belonging to different sub-classes are more clearly separated, while the samples within-class are more compact.

Assume $\mathbf{X}$ is an n -dimensional data set which consists of N samples. It can be divided into C sub-classes, represented by $\mathbf{X}_i$ ($i = 1, 2, \dots, C$), which is composed of N_i samples. The between-class scatter matrix $\mathbf{S}_b$ and the within-class scatter matrix $\mathbf{S}_w$ are given by [21, 22]

Received 2012-08-10, accepted 2012-10-06.

^{*} Supported by the National Natural Science Foundation of China (61273070) and the Foundation of Priority Academic Program Development of Jiangsu Higher Education Institutions.

^{**} To whom correspondence should be addressed. E-mail: yhz@jiangnan.edu.cn

$$\mathbf{S}_b = \sum_{i=1}^C P_i (\mathbf{m}_i - \mathbf{m}_0)(\mathbf{m}_i - \mathbf{m}_0)^T \quad (1)$$

$$\mathbf{S}_w = \sum_{i=1}^C \sum_{k=1}^{N_i} P_i E[(\mathbf{x}_{ik} - \mathbf{m}_i)(\mathbf{x}_{ik} - \mathbf{m}_i)^T | \mathbf{x}_{ik} \in \mathbf{X}_i] \quad (2)$$

where $\mathbf{m}_0$ is the overall mean of $\mathbf{X}$, $\mathbf{m}_i$ and P_i are the mean and prior probability of the subclass $\mathbf{X}_i$, respectively.

The objective function based on Fisher criterion can be expressed as

$$J(\mathbf{w}_{\text{opt}}) = \arg \max \frac{|\mathbf{w}^T \mathbf{S}_b \mathbf{w}|}{|\mathbf{w}^T \mathbf{S}_w \mathbf{w}|} = [w_1, w_2, \dots, w_n]^T \quad (3)$$

where $\mathbf{w}_{\text{opt}}$ can be obtained by solving a generalized eigenproblem specified by

$$\mathbf{S}_b \mathbf{w} = \lambda \mathbf{S}_w \mathbf{w} \quad (4)$$

and λ is a generalized eigenvalue.

KFDA can extract diagnostic characteristics through mapping the original data into a high-dimensional feature space, in which the diagnostic characteristics are essentially nonlinear discriminant features in the original data space. Similar to SVM and kernel principle component analysis (KPCA), KFDD also adopts the kernel trick [23]. Its basic idea is to project an original dataset $\mathbf{X}$ into a high-dimensional feature space F using a nonlinear mapping $f, F: \mathbf{x} \mapsto f(\mathbf{x})$. The input space is implied to give a strong nonlinear discriminant with the linear discriminant so that each category is separable. Suppose the mean value of the total samples is zero after projection, the objective function based on Fisher criterion is modified to

$$J(\mathbf{w}_{\text{opt}}^f) = \arg \max \frac{|\mathbf{w}^f{}^T \mathbf{S}_b^f \mathbf{w}^f|}{|\mathbf{w}^f{}^T \mathbf{S}_w^f \mathbf{w}^f|} = [w_1^f, w_2^f, \dots, w_n^f]^T \quad (5)$$

where the between-class scatter matrix is

$$\mathbf{S}_b^f = \sum_{i=1}^C \frac{N_i}{N} (\mathbf{m}_i^f)(\mathbf{m}_i^f)^T \quad (6)$$

and the within-class scatter matrix is

$$\begin{aligned} \mathbf{S}_w^f &= \sum_{i=1}^C \frac{N_i}{N} \frac{1}{N_i} \sum_{k=1}^{N_i} [f(\mathbf{x}_{ik}) - \mathbf{m}_i^f][f(\mathbf{x}_{ik}) - \mathbf{m}_i^f]^T \\ &= \frac{1}{N} \sum_{i=1}^C \sum_{k=1}^{N_i} f(\mathbf{x}_{ik}) f(\mathbf{x}_{ik})^T - \sum_{i=1}^C \frac{N_i}{N} (\mathbf{m}_i^f)(\mathbf{m}_i^f)^T \end{aligned} \quad (7)$$

where

$$\mathbf{m}_i^f = \frac{1}{N} \sum_{k=1}^{N_i} f(\mathbf{x}_{ik}), \quad i = 1, \dots, C$$

is the mean value of the i th sub-class. Because mapping into the high-dimensional space will lead to increased computational complexity, an inner product kernel function is introduced for implicit calculation so as to extract the feature in the high-dimensional

space. The inner product kernel function is defined as

$$K(\mathbf{x}_i, \mathbf{x}_j) = f^T(\mathbf{x}_i) f(\mathbf{x}_j) \quad (8)$$

According to reproducing kernel theory, the solution of F space can be obtained by the samples in F space. Therefore, $\mathbf{w}^f$ can be determined by

$$\mathbf{w}^f = \sum_{l=1}^N \alpha_l f(\mathbf{x}_l) \quad (9)$$

where α_l ($l = 1, 2, \dots, N$) is the corresponding coefficients.

Using the inner product to replace the kernel function, we have

$$\begin{aligned} (\mathbf{w}^f)^T \mathbf{m}_i^f &= \sum_{l=1}^N \alpha_l \frac{1}{N} \sum_{k=1}^{N_i} f(\mathbf{x}_l)^T f(\mathbf{x}_{ik}) \\ &= \sum_{l=1}^N \alpha_l \frac{1}{N_i} \sum_{k=1}^{N_i} K(\mathbf{x}_l, \mathbf{x}_{ik}) = \boldsymbol{\alpha}^T \mathbf{U}_i \end{aligned} \quad (10)$$

where

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T,$$

$$\mathbf{U}_i = \left[\frac{1}{N_i} \sum_{k=1}^{N_i} K(\mathbf{x}_1, \mathbf{x}_{ik}), \dots, \frac{1}{N_i} \sum_{k=1}^{N_i} K(\mathbf{x}_N, \mathbf{x}_{ik}) \right]^T.$$

Using Eqs. (6) and (10), the following equation can be obtained

$$(\mathbf{w}^f)^T \mathbf{S}_b^f \mathbf{w}^f = \boldsymbol{\alpha}^T \mathbf{K}_b \boldsymbol{\alpha} \quad (11)$$

where $\mathbf{K}_b$ is calculated by

$$\mathbf{K}_b = \sum_{i=1}^C \frac{N_i}{N} (\mathbf{U}_i)(\mathbf{U}_i)^T = \frac{1}{N} \mathbf{K} \mathbf{A} \mathbf{K}^T$$

and $\mathbf{K}$ is an $N \times N$ kernel function matrix from a kernel mapping and $\mathbf{A} = \text{diag}(\mathbf{A}_i)_{i=1,2,\dots,C}$ is a $N \times N$ block diagonal matrix and each element in the matrix is $1/N_i$.

Since

$$\begin{aligned} (\mathbf{w}^f)^T f(\mathbf{x}_{ik}) &= \sum_{l=1}^N \alpha_l f(\mathbf{x}_l)^T f(\mathbf{x}_{ik}) \\ &= \sum_{l=1}^N \alpha_l \sum_{k=1}^{N_i} K(\mathbf{x}_l, \mathbf{x}_{ik}) \end{aligned} \quad (12)$$

combining Eqs. (9), (10) and (12) yields

$$(\mathbf{w}^f)^T \mathbf{S}_w^f \mathbf{w}^f = \boldsymbol{\alpha}^T \mathbf{K}_w \boldsymbol{\alpha} \quad (13)$$

where $\mathbf{K}_w = (\mathbf{K} \mathbf{K}^T - \mathbf{K} \mathbf{A} \mathbf{K}^T) / N$.

The Fisher linear discriminant in F can be obtained through maximizing

$$\begin{aligned} J(\mathbf{w}_{\text{opt}}^f) &= \arg \max \frac{|\mathbf{w}^f{}^T \mathbf{S}_b^f \mathbf{w}^f|}{|\mathbf{w}^f{}^T \mathbf{S}_w^f \mathbf{w}^f|} \\ &= \frac{|\boldsymbol{\alpha}^T \mathbf{K}_b \boldsymbol{\alpha}|}{\boldsymbol{\alpha}^T \mathbf{K}_w \boldsymbol{\alpha}} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T \end{aligned} \quad (14)$$

The solution can be obtained by solving the eigenvector of $\lambda \mathbf{K}_\omega \boldsymbol{\alpha} = \mathbf{K}_b \boldsymbol{\alpha}$.

2.2 Improved algorithm of KFDA

The definition of between-class scatter matrix is less than optimal based on traditional Fisher criteria. According to Eq. (1), calculating the distance between the mean of the total mapped samples and the mean of the i th sub-class mapped samples, each subclass may easily deviate from the center of the mapped sample set and result in sub-classes overlap. Hence, the between-class scatter matrix in Eq. (1) is redefined by calculating the distances between each subclass as

$$\tilde{\mathbf{S}}_b = \sum_{i=1}^{C-1} \sum_{j=i+1}^C P_i P_j (\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T \quad (15)$$

Considering the example shown in Fig. 1, the sub-class 3 is defined as an edge class if it is far away from the sub-class 1 and sub-class 2. The projection axis A shown in Fig. 1 is obtained based on traditional feature extraction method. Apparently, because the edge sub-class 3 is excessively emphasized by its maximization between-class distance, the adjacent sub-classes overlap is resulted, as the case of sub-classes 1 and 2 in Fig. 1.

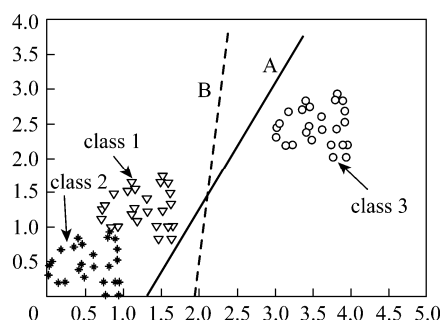


Figure 1 Different projection axes of different definition of $\mathbf{S}_b$

Accordingly, the between-class scatter matrix and the within-class scatter matrix are redefined to

$$\tilde{\mathbf{S}}_b = \sum_{i=1}^{C-1} \sum_{j=i+1}^C P_i P_j W(\Delta_{ij}) (\mathbf{m}_i - \mathbf{m}_j)(\mathbf{m}_i - \mathbf{m}_j)^T \quad (16)$$

$$\tilde{\mathbf{S}}_\omega = \sum_{i=1}^C \sum_{k=1}^{N_i} P_i \frac{1}{N_i} W(\Delta_{ik}) (\mathbf{x}_{ik} - \mathbf{m}_i)(\mathbf{x}_{ik} - \mathbf{m}_i)^T \quad (17)$$

where $W(\Delta_{ij})$ and $W(\Delta_{ik})$ are the weighted functions for between-class and within-class scatter, respectively.

The weighted scatter matrices of within-class and between-class defined in Eqs. (16) and (17) can decrease effectively the impacts of edge classes and outliers. Projection axis B is obtained by the modified scatter matrices, which has better discrimination abil-

ity than projection axis A. Similarly, the nonlinear form for the novel method is further derived using Eqs. (16) and (10). We have

$$(\mathbf{w}^f)^T \tilde{\mathbf{S}}_b^f \mathbf{w}^f = \boldsymbol{\alpha}^T \mathbf{K}'_b \boldsymbol{\alpha} \quad (18)$$

$$\text{where } \mathbf{K}'_b = \sum_{i=1}^{C-1} \sum_{j=i+1}^C P_i P_j W(\Delta_{ij}) (\mathbf{U}_i - \mathbf{U}_j)(\mathbf{U}_i - \mathbf{U}_j)^T.$$

Combining Eqs. (9), (10) and (12) yields

$$(\mathbf{w}^f)^T \tilde{\mathbf{S}}_\omega^f \mathbf{w}^f = \boldsymbol{\alpha}^T \mathbf{K}'_\omega \boldsymbol{\alpha} \quad (19)$$

where $\mathbf{K}'_\omega = (\mathbf{K} \mathbf{K}'^T - \mathbf{K}' \mathbf{A} \mathbf{K}'^T) / N$, and $\mathbf{K}'$ is a $N \times N$ weighted kernel matrix obtained by the kernel function and weighted function, and its each element $k_{ik} = W(\Delta_{ik}) k(\mathbf{x}_i, \mathbf{x}_{ik})$ is called the weighted inner product. Searching for the optimal projection vector is converted to solving the corresponding eigenvector of

$$\lambda \mathbf{K}'_\omega \boldsymbol{\alpha} = \mathbf{K}'_b \boldsymbol{\alpha} \quad (20)$$

A Gaussian radial basis function is used as the inner kernel product function, defined as

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{\sigma^2}\right) \quad (21)$$

This function is widely used for its good learning ability. Furthermore, since the corresponding eigenspace is infinite, the finite sampled data must be linearly separable in this eigenspace. The weighted function is similarly defined as

$$W(\Delta_{ik}) = \exp\left(-\frac{\|\mathbf{x}_{ik} - \mathbf{m}_i\|^2}{\sigma^2}\right) \quad (22)$$

The weight is transformed by the same nonlinear transformation of the sampled data. Therefore, it can reflect the weighted relationship of the sampled data in the feature space.

2.3 Category information measure and kernel parameters selection

The projection matrix $[a_1, a_2, \dots, a_d]$ based on the kernel Fisher criterion is an eigenvector matrix corresponding to the first d maximum eigenvalues in the general eigen equation $\lambda \mathbf{K}'_\omega \boldsymbol{\alpha} = \mathbf{K}'_b \boldsymbol{\alpha}$.

Generally, d is artificially determined. The larger d is, the more information is contained. But for the classification, remaining valid category information is the most important. After eigen transformation of the original data, the aggregation degree of the within-classes and the dispersion degree of the between-classes can be evaluated by its each dimension.

The aggregation degree of the within-classes for the k th dimensional component can be described as

$$\sigma_k^2 = \sum_{i=1}^C S_{ik}^2 \quad (23)$$

where S_{ik}^2 is the standard deviation of the k th dimensional component in the i th class.

The dispersion degree of the between-classes for the k th dimensional component can be described as

$$D_k^2 = \sum_{i=1}^{C-1} \sum_{j=i+1}^C \|m_{ik} - m_{jk}\|^2 \quad (24)$$

where m_{ik} and m_{jk} are the mean value of the k th dimensional component in the i th class and the j th class, respectively.

Combining Eqs. (23) and (24), an evaluation index is presented by the category information measurement, and defined as

$$J_k = \frac{\sigma_k^2}{D_k^2} \quad (25)$$

It is obvious that the smaller J_k is, the higher classification accuracy will be obtained.

In order to reserve the most effective category information, the d characteristic components corresponding to the minimum J_k have to be found out. Define the accumulated contribution rate of category information as

$$\lambda_d = \sum_{k=1}^d \frac{1}{J_k} / \sum_{k=1}^L \frac{1}{J_k} \quad (26)$$

where L is the number of selected initial eigenvectors, satisfying $\lambda_L > 10\lambda_{L+1}$. Then the first d characteristic components satisfying $\lambda_d \geq \bar{\lambda}$ are selected where $\bar{\lambda}$ is the set value.

3 A MULTI-MODEL METHOD BASED ON FEATURE EXTRACTION USING WEIGHTED KERNEL FISHER CRITERION

The weighted between-class scatter matrix and weighted within-class scatter matrix can decrease the impacts of edge classes and outliers. The category information measurement is applied to determine the number of eigenvectors and further determine the optimal feature projection matrix. The subclasses with obvious characters can be obtained by mapping the original data into the feature space. Then, the sub-models corresponding to the subclasses are established based on support vector machine (SVM), and a switcher is used to combine all SVM sub-models.

3.1 Calculation Procedure

A multiple model is established by combining WKFDA feature extraction and the SVM algorithm. The specific calculation procedure is summarized as follows.

Step 1: Set up the initial parameters, such as the clustering number C and accumulated contribution rate for category information $\bar{\lambda}$. The training data are

initially classified by the fuzzy C-means clustering algorithm and the clustering number can be determined by the subtractive clustering [24].

Step 2: The inner product kernel parameters σ^2 in the l group are selected within a certain range. The eigenvectors and corresponding eigenvalues are calculated based on Eqs. (18)–(20). The eigenvectors corresponding to the first maximum M_i ($i=1, 2, \dots, l$) eigenvalues are selected as the initial projection matrix.

Step 3: After the eigen transformation, the category information measurement J_k ($k=1, \dots, M_i$) is calculated by Eq. (25) and effective eigenvectors d_i ($i=1, 2, \dots, l$) corresponding to l group kernel parameters are determined by Eq. (26) and set value $\bar{\lambda}$. Then the Fisher criterion values corresponding to l group effective eigenvectors are calculated and the maximum Fisher value corresponding to σ^2 is selected as the optimal kernel parameter. The corresponding optimal projection matrix is constituted by d_i eigenvectors.

Step 4: The projection transformation for training data is made by the optimal projection matrix. As to each subclass, the SVM sub-model is established by the transformed data separately.

Step 5: A switcher is used to combine SVM sub-models and then a multi-model based soft sensor is obtained.

After projection transformation of the original test sample, which subclass it belongs to is determined by the WKFDA classification algorithm. And then the test sample after transformation is input into the corresponding sub-model to calculate the output prediction. Fig. 2 shows the schematic of the proposed soft sensor, where S represents the switcher and Y is the predicted output.

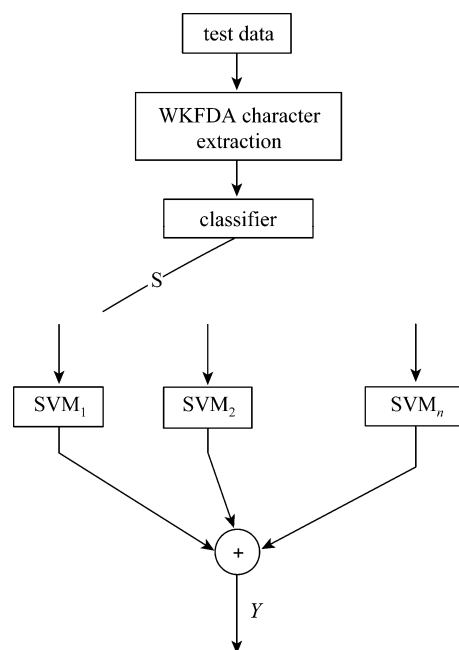


Figure 2 Schematic of the proposed soft sensor

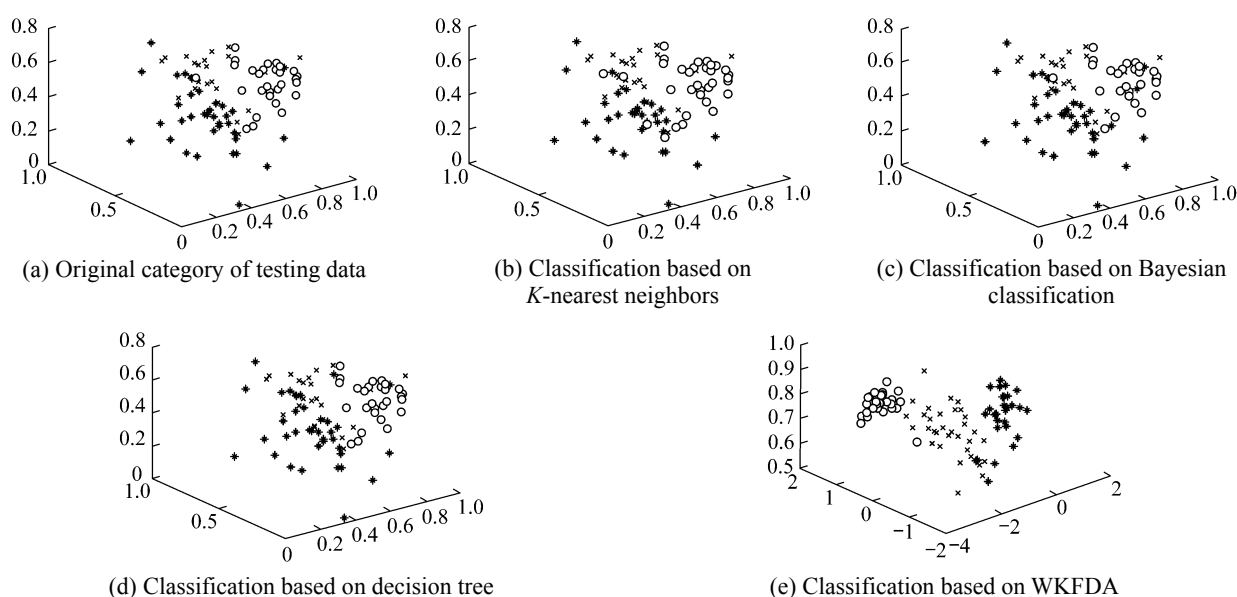


Figure 3 Classification of the testing data

3.2 Example analysis

To verify the classification accuracy of the proposed method, a normal dataset of wine from UCI machine learning database [25] is used for experimental analysis. The wine dataset has 178 group sample data, including 13 input attributes and 1 category attribute. One half of the wine dataset is used as training data and the rest is used as testing data. The classification results are compared with *K*-nearest neighbors algorithm, Bayesian classification algorithm and decision tree algorithm. The experiment is repeated for several times and their average values are taken as the experimental results, as shown in Table 1.

Table 1 Classification results

Algorithm	Error number	Classification accuracy/%
WKFDA	5.500	93.75
<i>K</i> -nearest neighbors	5.667	93.56
Bayesian	5.837	93.4
decision tree	7.16	91.93

It is obvious that WKFDA has the highest classification accuracy. Furthermore, the sample dataset is easier to be classified after eigen transformation. The first 3 dimensions of the eigenmatrix are selected to be compared with the same dimensions of the normalized original dataset.

It is clear from Fig. 3 that the subclasses of original dataset are overlapped and the existing methods cannot separate them clearly. The proposed method can be used for eigen transformation of the original dataset so as to find out the optimal classified projection

matrix. The simulation results demonstrate that the transformed data using WKFDA is easier to distinguish and classify.

4 INDUSTRIAL CASE STUDY

Crystallization process is an important step in production of bisphenol A. The online phenol content, a key quality index of a dissolution tank outlet in the crystallizer, is estimated by the proposed method. The online phenol content is subject to the following 7 main operation variables: dissolving tank temperature, tank level, vacuum filter flow rate, and the contents of 4 components from the upstream unit exiting stream. There are 284 group data collected from a production site, 4/5 of which are used for training and 1/5 for testing.

The parameters of inner kernel function are set as $\sigma^2 = 2.5$, $\bar{\lambda} = 0.8$. According to the analysis for the operating condition, the number of category is $C = 4$. KFDA and WKFDA are applied for classification and the classification results are shown in Figs. 4 and 5, respectively. From these two figures, it is obvious that the WKFDA is more effective in clustering and has better performance.

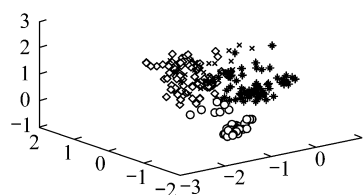


Figure 4 Classification based on KFDA extraction of training data

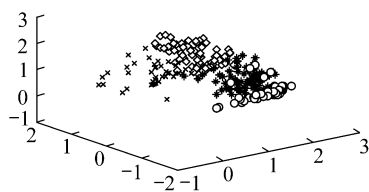


Figure 5 Classification based on WKFDA extraction of training data

In order to verify the effectiveness of the proposed method, a single SVM model, a multiple SVM model based on KFDA and a multiple SVM model based on WKFDA are built separately so as to do simulation comparison. The kernel parameters and punishing coefficients of single SVM model, multiple SVM model based on KFDA and multiple SVM model based on WKFDA are specified as follows:

(1) $r = 0.26$, $\rho = 100$.

(2) $r_1 = 0.001$, $\rho_1 = 16$; $r_2 = 0.36$, $\rho_2 = 31$; $r_3 = 0.51$, $\rho_3 = 76$; $r_4 = 1.46$, $\rho_4 = 16$.

(3) $r_1 = 0.005$, $\rho_1 = 50$; $r_2 = 1.5$, $\rho_2 = 150$; $r_3 = 1$, $\rho_3 = 75$; $r_4 = 0.01$, $\rho_4 = 31$.

The test errors of simulating results are shown in Table 2.

Table 2 Test errors

Model	Maximal relative error/%	Average relative error	Squared error
single SVM	4.24	0.0104	1.032
KFDA multi-model	2.79	0.0076	0.7319
WKFDA multi-model	2.27	0.0069	0.6539

Table 2 demonstrates that the multiple SVM model based on WKFDA is better than the other two models. The corresponding output prediction results of testing data are shown in Fig. 6.

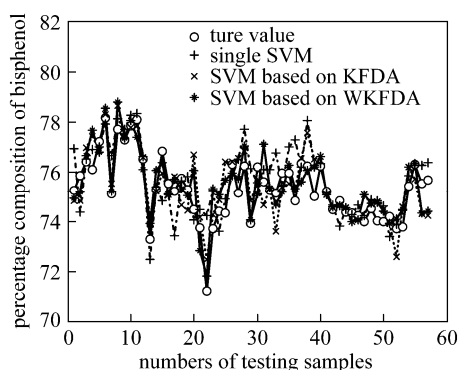


Figure 6 Testing curves

5 CONCLUSIONS

A new feature extraction based on a weighted

kernel Fisher criterion is proposed by modifying the definition of within-class scatter matrix and between-class scatter matrix. The impacts of edge classes and outliers are decreased by weighting so as to avoid subclasses overlap. The weighting method is derived with a nonlinear form so that the nonlinear problems are more effectively handled. The optimal number of projection vectors is determined by category information measurement, and the better classification results are obtained. In order to further validate the effectiveness of the proposed method, an experiment is demonstrated by a normal dataset of wine from UCI machine learning database. The method is applied to a bisphenol A production process to estimate an online quality index. The simulation results validate that the classification accuracy and estimation precision are well improved by the proposed method.

REFERENCES

- Li, X.L., Su, H.Y., Chu, J., "Multiple models soft-sensing technique based on online clustering arithmetic", *Journal of Chemical Industry and Engineering*, **58** (11), 2835–2839 (2007). (in Chinese)
- Liu, L.L., Zhou, L.F., Xie, S.G., "A novel supervised multi-model modeling method based on k -means clustering", In: *Control and Decision Conference (CCDC)*, Shenyang, 684–689 (2010). (in Chinese)
- Li, N., Li, S.Y., Xi, Y.G., "Multi-model modeling method based on satisfactory clustering", *Control Theory & Applications*, **20** (5), 783–788 (2003). (in Chinese)
- Li, W.B., Sun, L., Zhang, D.K., "Text classification based on labeled-LDA model", *Chinese Journal of Computer*, **31** (4), 621–626 (2008). (in Chinese)
- Yang, Q., Ding, X., "Discriminant local feature analysis with applications to face recognition", *Journal of Tsinghua University (Science and Technology)*, **44** (4), 530–533 (2004). (in Chinese)
- Huerta, E.B., Duval, B., Hao, J., "A hybrid LDA and genetic algorithm for gene selection and classification of microarray data", *Neurocomputing*, **73** (13–15), 2375–2383 (2010).
- Sabatier, R., Reynès, C., "Extensions of simple component analysis and simple linear discriminant analysis", *Computational Statistics and Data Analysis*, **52** (10), 4779–4789 (2008).
- Mohammadi, M., Raahemi, B., Akbari, A., Nassersharif, B., Moeinzadeh, H., "Improving linear discriminant analysis with artificial immune system-based evolutionary algorithms", *Information Sciences*, **189**, 219–232 (2012).
- Chen, C.K., Yang, J.Y., Yang, J., "Fusion of PCA and KFDA for face recognition", *Control and Decision*, **19** (10), 1147–1154 (2004). (in Chinese)
- Sun, D.R., Wu, L.N., "Face recognition based on nonlinear feature extraction and SVM", *Journal of Electronics & Information Technology*, **26** (2), 308–312 (2004). (in Chinese)
- Li, Y., Jiao, L.C., "Target recognition based on kernel Fisher discriminant", *Journal of Xidian University*, **30** (4), 179–183 (2003). (in Chinese)
- Zheng, Y.J., Jian, Y.G., Yang, J.Y., Wu, X.J., "A reformative kernel Fisher discriminant algorithm and its application to face recognition", *Neurocomputing*, **69** (13–15), 1806–1810 (2006).
- Sun, J.C., Li, X.H., Yang, Y., "Scaling the kernel function based on the separating boundary in input space: A data-dependent way for improving the performance of kernel methods", *Information Sciences*, **184** (1), 140–154 (2012).
- Maldonado, S., Weber, R., Basak, J., "Simultaneous feature selection and classification using kernel-penalized support vector machines", *Information Sciences*, **181** (2), 115–128 (2011).

- 15 Wen, Z.M., Cairing, Z., Zhao, L., "Weighted maximum margin discriminant analysis with kernels", *Neurocomputing*, **67**, 357–362 (2005).
- 16 Peng, X.J., "A v-twin support vector machine (v-TSVM) classifier and its geometric algorithms", *Information Sciences*, **180** (2), 3863–3875 (2010).
- 17 Amari, S., Wu, S., "Improving support vector machine classifiers by modifying kernel functions", *Neural Networks*, **12** (6), 783–789 (1999).
- 18 Kumar, M.A., Khemchandani, R., Gopal, M., Chandra, S., "Knowledge based least squares twin support vector machines", *Information Sciences*, **180** (23), 4606–4618 (2010).
- 19 Wu, Q., "Fault diagnosis model based on Gaussian support vector classifier machine", *Expert Systems with Applications*, **37** (9), 6251–6256 (2010).
- 20 He, Q., Wu, C.X., "Separating theorem of samples in Banach space for support vector machine learning", *International Journal of Machine Learning and Cybernetics*, **184** (1), 49–54 (2011).
- 21 Loog, M., Duin, R.P.W., Haeb-Umbach, R., "Multiclass linear dimension reduction by weighted pairwise Fisher criteria", *IEEE Transactions on Pattern Analysis And Machine Intelligence*, **23** (7), 762–766 (2001).
- 22 Li, X.H., Li, X., Guo, G.R., "Feature extraction of HRRP based on LDA algorithm", *Journal of National University of Defense Technology*, **27** (5), 72–77 (2005). (in Chinese)
- 23 Du, S.Q., "Methods of face recognition based on kernel Fisher discriminant", Master Thesis, Shanxi Normal University, China (2007). (in Chinese)
- 24 Zhang, X., "Research and application of FCM initialization method", Master Thesis, Southwest University, China (2006). (in Chinese)
- 25 Frank, A., Asuncion, A., "UCI machine learning repository", 2010, <http://archive.ics.uci.edu/ml>.