



# A one-class feature extraction method based on space decomposition

Guangzao Huang<sup>1</sup> · Xiaojing Chen<sup>1</sup> · Xi Chen<sup>1</sup> · Xiao Chen<sup>2</sup> · Wen Shi<sup>1</sup>

Accepted: 25 February 2022 / Published online: 19 April 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

One-class classification is an important branch of machine learning. Feature extraction is an important means to improve the performance of one-class classifiers, but there is no generalized method yet reported to solve this problem. In this paper, a framework is proposed for one-class feature extraction. The proposed framework divides the original feature space into two orthogonal spaces, namely the principal space and the complementary space. The principal space is used to learn the features of the target class, and the complementary space is used to learn the features of the abnormal class. The features extracted from the two spaces are fused as the final one-class feature vector of the original feature space. Furthermore, a specific implementation method, complete principal component analysis (CPCA), is proposed. First, CPCA conducts principal component analysis to calculate the projection scores of the target class samples in the principal space. Then, according to the projection vectors of the principal components (obtained in the principal space), the corresponding complementary space is constructed. The projection of the sample in the complementary space is calculated and transformed into the first-order norm as the extracted feature in the complementary space. Several datasets are used to verify the effect of this proposed method. The experimental results show that the proposed CPCA has good universality for one-class feature extraction problems.

**Keywords** One-class classification · Feature extraction · Principal space · Complementary space · Principal component analysis

## 1 Introduction

The one-class classification (OCC) is a fundamental problem in the field of machine learning (Kemmler et al. 2013; Xu et al. 2013; Xiao et al. 2015; 2014; Burnaev and Smolyakov 2017; Huang et al. 2017; De Santana et al. 2019). For the OCC problem, only positive samples exist, while negative samples do not exist or are very less in number when training the classifier. This situation is common in practice, such as in fault diagnosis (Fernández-Francos et al. 2013; Bing et al. 2018) or target recognition (Fei et al. 2018; Yu and Xiao 2018). For fault diagnosis, normal samples are easily obtained, while fault samples are

often rare or unavailable. For the target recognition problem, many non-target samples exist, which results in negative samples that are available for learning not being representative. Since there are no suitable negative class samples for learning, the classifier can only learn from the positive class samples. Such problems are collectively referred to as OCC problems and are also known as concept learning, single-class classification, data description, and anomaly detection.

For the OCC problem, feature dimensionality reduction (feature selection or extraction) (Liu et al. 2017; Jia et al. 2018; Kim 2018) has a significant impact on the performance of a one-class classifier. Especially for high-dimensional data, the distribution of samples in the feature space is sparse. Some conventional one-class classifier design methods cannot be directly applied to high-dimensional data. For example, the one-class classifier design methods that are based on probability density estimations require the samples to be densely distributed in the feature space. Compared with the supervised classification problem, the OCC problem lacks the information of negative samples. Moreover, compared with the unsupervised

✉ Xiaojing Chen  
chenxj@wzu.edu.cn

<sup>1</sup> College of Physical and Electronic Information Engineering, Wenzhou University, Wenzhou 325035, People's Republic of China

<sup>2</sup> Wenzhou Institute of Industry&Science, Wenzhou 325035, People's Republic of China

classification problem, the OCC problem has additional information that the training samples belong to the same class. So, the problem of one-class feature reduction has unique characteristics and is difficult to solve.

Some feature dimension reduction methods for OCC problems have been proposed. For instance, Lorena et al. (2014) considered that the discriminative features should be structured, and put forward some indexes (information score, Pearson correlation, intra-class distance, and interquartile range) for filtering one-class features. Jeong et al. (2012) supposed that the features which have less influence on the volume of one-class data are more discriminatory and proposed a feature recursive elimination method based on the support vector data description for OCC problems. Tax (2003) proposed that the compression direction of the target class data with less variance often contains more discriminative feature information for OCC problems. Lian (2012) pointed out that selecting some principal components with less information is often more useful for distinguishing between target samples and abnormal samples when using the principal component analysis (PCA) method to extract one-class features. However, compared with traditional multi-classification problems, there are few studies on the feature dimensionality reduction of one-class classification problems. Moreover, these studies mainly focus on feature selection, rather than feature extraction. At present, the feature extraction methods of one-class classification problems are mainly based on traditional unsupervised methods, such as PCA. However, such practices cannot well match the one-class classification problems as they ignore the fact that one-class classification problems are between supervised and unsupervised.

In this paper, the information that the “training data belong to the same class” is integrated into unsupervised learning to propose a suitable feature extraction strategy for OCC problems. The proposed strategy divides the original feature space into two parts, the main space, and the complementary space, which are used to extract the feature information of the target class and the abnormal class, respectively. According to this strategy, this paper puts forward a specific implementation method, which is named complete principal component analysis (CPCA). CPCA extracts the information of the target class using PCA in the main space, extracts the information of the abnormal class using the first-order norm in the complement space, and finally combines the two as the extracted one-class feature vector of the original feature space.

The rest of this paper is organized as follows. Section 2 introduces the proposed one-class feature extraction method. The experiment is demonstrated in Sect. 3. Section 4 describes the applications and results. Section 5

presents the discussion. Section 6 addresses the paper conclusion and future work.

## 2 The proposed feature extraction method for one-class classification

How to extract discriminative features using only the target class samples is the primary problem that needs to be solved. The following one-class feature extraction strategy is designed. First, the target class samples are used to find a projection space called the main space, which can describe the distribution characteristics of the target class. The projection score in the main space can be used to measure the similarity between an unknown sample and the target class samples. Then, the remaining space is used to construct the complement space. The intensity of the projection vector of a sample in the complement space can indicate its difference from the target class. The main space and complement space are orthogonal. If the main space cannot fully reflect the characteristics of the target class, then the remaining characteristic information of the target class leak into the complement space, which decreases the recognition effect of abnormal samples.

Figure 1 demonstrates the basic idea of the proposed one-class feature extraction strategy. The abnormal class samples green circle, green square and green pentagon can be identified in the main space and complement space, respectively. Therefore, when using the ordinary PCA method to extract one-class features, some abnormal samples fail to be detected. Therefore, it is necessary to retain the features that are extracted in the complement space because there is no prior knowledge to distinguish the discriminative information of abnormal samples (either in the main space or in the complement space).

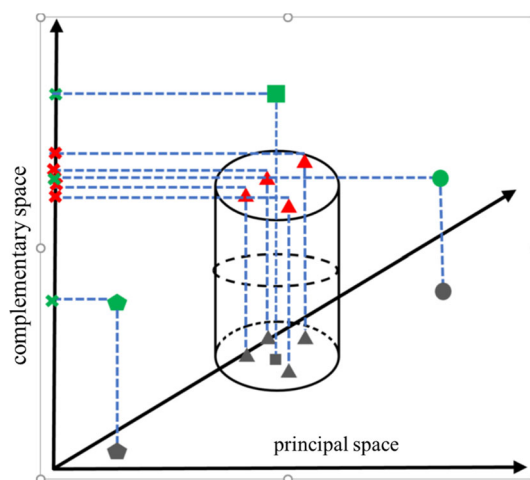


Fig. 1 Red triangle indicates the target class sample, and green circle, green pentagon and green square indicate the abnormal class samples

The following implementation according to the above OCC feature extraction strategy was proposed. First, the PCA (Wold et al. 1987; Koltchinskii and Lounici 2016) is performed on the target class samples to find a suitable principal component projection space as the main space that can describe the principal distribution characteristics of the target class. Then, according to the obtained main space, the corresponding complement space is constructed. The projection score in the main space and the first-order norm of the projection in the complement space are combined as the final extracted one-class features. The specific implementation steps are as follows.

Suppose the training target class data are  $\mathbf{X}_{n \times p}$ , where  $n$  is the number of samples and  $p$  is the feature dimension. First, PCA is performed on  $\mathbf{X}_{n \times p}$  and  $v\%$  of the variation information of the training target class data is retained to construct the principal component projection space  $\mathbf{W}_{p \times l}$  as the main space  $S$ , where  $l$  is the number of retained principal components. The projection score  $\mathbf{t}$  of an unknown sample  $\mathbf{x}_{1 \times p}$  in the main space  $S$  is

$$\mathbf{t} = \mathbf{x}\mathbf{W} \tag{1}$$

Then, the complement space  $S^\perp$  is

$$S^\perp = \mathbf{I} - \mathbf{W}^T\mathbf{W} \tag{2}$$

where  $\mathbf{I}$  is the identity matrix and symbols  $^T$  and  $^+$  indicate the transpose and Pseudo-inverse. The projection vector  $\mathbf{x}^\perp$  in the complement space  $S^\perp$  is

$$\mathbf{x}^\perp = \mathbf{x}S^\perp \tag{3}$$

Finally,  $\mathbf{t}$  and  $\|\mathbf{x}^\perp\|$  are combined as the final extracted one-class feature vector.

### 3 Experiments

#### 3.1 Evaluation of the proposed method by using one-class classifiers

A one-class classifier measures the similarity of a sample to the target class. In this paper, the Mahalanobis distance (Galeano et al. 2013; Washizawa and Hotta 2017) and one-class support vector machine (OC-SVM) (Schölkopf et al. 2000; Guerbai et al. 2018) were used to verify the effect of the proposed one-class feature extraction method.

The Mahalanobis distance is a generalized distance that fully considers the covariance between variables. Compared with the common Euclidean distance, the Mahalanobis distance can eliminate the influence of the dimension and the correlation between variables. The Mahalanobis distance is calculated as

$$\|\mathbf{x} - \mathbf{m}\|_M = \sqrt{(\mathbf{x} - \mathbf{m}) \sum^{-1} (\mathbf{x} - \mathbf{m})^T} \tag{4}$$

where  $\mathbf{m}$  is the center vector of the sample set and  $\sum$  is the estimated covariance matrix. By using the Mahalanobis distance, it is considered that the distribution of the target class obeys the following multi-dimensional Gaussian distribution:

$$P(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \sqrt{\sum}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}) \sum^{-1} (\mathbf{x}-\mathbf{m})^T} \tag{5}$$

where  $n$  is the number of samples. Therefore, the Mahalanobis distance can be used as a simple one-class classifier based on the probability density estimation. However, Mahalanobis distance would not be better for high-dimensional data because  $\sum^{-1}$  does not exist. Therefore, using Mahalanobis distance to classify high-dimensional data, feature dimensionality reduction is a necessary pre-processing means.

The OC-SVM is a derivative of the traditional two-class support vector machine in the OCC field. The abnormal class is set as the origin. The basic idea of the OC-SVM is to locate most of the target class samples on one side of the hyperplane when maximizing the distance between the hyperplane and the origin. The hyperplane can be obtained by solving the following objective function:

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 + c \sum_{i=1}^n \zeta_i - \rho \\ \text{s.t.} & \begin{cases} w^T \phi(x_i) \geq \rho - \zeta_i \\ \zeta_i > 0 \end{cases} \end{aligned} \tag{6}$$

where  $\phi(\cdot)$  is a feature projection function that maps an input vector into a higher dimensional feature space,  $w$  is a decision hyperplane normal vector which is perpendicular to the hyperplane,  $\rho$  is an intercept term, and  $\zeta_i$  are nonzero slack variables for penalizing the outliers.

By using Lagrangian techniques and a kernel function for the dot-product calculations, the output function becomes:

$$f(\mathbf{x}) = \sum_{i=1}^n a_i K(\mathbf{x}_i, \mathbf{x}) - \rho \tag{7}$$

where  $a_i$  is a Lagrange multiplier and  $k(x_i, x) = \phi(x_i)^T \phi(x)$  is a kernel function. A radial basic function (RBF) kernel is employed in our experiment:

$$k(x_i, x) = e^{-\gamma \|x_i - x\|^2}, \gamma > 0 \tag{8}$$

The OC-SVM has been widely used in OCC problems because of its advantages of not requiring prior knowledge and minimal structural risk.

**Table 1** Summary of the experimental datasets

Dataset		No. of classes	No. of target class samples/Total sample size	Feature dimension
Low dimension	Iris	3	50/150	4
	Wine	3	59/178	13
	bupa	2	139/345	6
	spectheart	2	55/264	44
	thyroid	3	166/7200	21
	segment	7	330/2310	19
	High dimension	USPS	10	1553/ 9298
Alphadigits		36	39/1404	320
warpAR10P		13	13/130	2400
warpPIE10p		10	21/210	2420
milk powder		2	90/360	1762
pixraw10P		10	10/100	10,000

### 3.2 Experimental datasets

The effectiveness of the proposed CPCA was verified with 30 datasets, the characteristics of which are shown in Table 1. Due to the limited space of the paper, the relevant tables in the main text only contain the information and experimental results of some datasets, and the others are in Appendix. According to the feature dimension, these datasets were divided into two types: low-dimensional data (UCI machine learning repository) and high-dimensional data.<sup>1</sup> There were a total of 12 low-dimensional datasets, and the rest were high-dimensional datasets, including handwritten digital data, face recognition data, gene expression data, etc. These data are usually used for multi-class experiments and need to be reorganized for one-class problems. The first class in the data was used as the target class, and the remaining classes were used as the abnormal class. Two-thirds of the target class samples were randomly selected as training set data, and the remaining samples made the test set. It should be noted that the milk powder data are infrared spectral data, which was specifically used for rapid adulteration detection of milk powder with one-class classification experiments (Huang et al. 2022). Infrared spectroscopy is a fast and nondestructive testing technology, which is necessary to construct an appropriate recognition model. For milk powder adulteration detection, the adulteration information is uncertain, which makes it difficult for traditional discriminant models to deal with it. However, the one-class classification method is suitable for this problem because it only needs to learn from target samples (pure milk powder) to identify abnormal samples

(different forms of adulterated milk powder). All the code ran on MATLAB 2012.

### 4 Applications and results

First, a low-dimensional dataset (iris) and a high-dimensional dataset (warpPIE10p) were used to visually demonstrate the performance of CPCA and PCA in one random experiment. Figure 2 shows that the target class and abnormal class in the iris dataset were well separated with the second principal component (PC2). Figure 4 shows that the different classes in the warpPIE10p dataset seriously overlap in the PC1 × PC2 space. To visually display the discriminant effect of the features extracted from the main space and the complementary space, the Mahalanobis distance of the PCA scores in the main space was applied as the abscissa and the feature value in the complementary space as the ordinate to show the distribution effect of different classes in the constructed low-dimensional space. The left subgraphs in Figs. 3 and 5 correspond to iris dataset and warpPIE10p dataset, respectively (Fig. 4).

Figure 3 demonstrates the influence of the number of principal components that are extracted. CPCA worked well when extracting 1, 2, 3, and 4 principal components while the traditional PCA method only worked well when extracting 2, 3, and 4 components. Figure 5 shows that the target class and the abnormal class were well distinguished in the complement space when the number of extracted components was greater than 1, but they were completely inseparable in the main space. Therefore, the features that were extracted from the complement space well reflected the characteristics of the anomaly class and completed the

<sup>1</sup> <http://featureselection.asu.edu/datasets.php>.

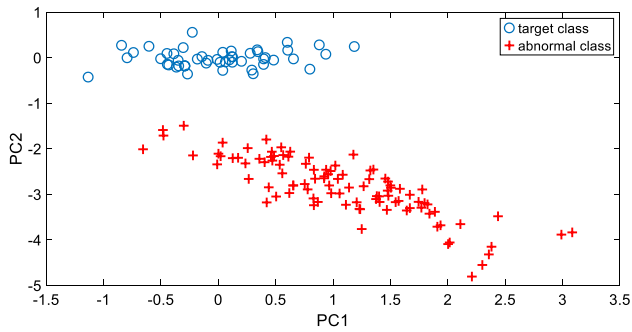


Fig. 2 Score plots of the iris dataset in the PC1 × PC2 space

deficiency of the features that were extracted in the main space.

Next, the effect of the proposed method using statistical analysis was analyzed. The dimension of the main space was determined by retaining  $\nu\%$  of the variation information of the target class samples. The raw features and one-class features (extracted by PCA) were used for comparative analysis. Each dataset was randomly tested 50 times, and the average area under curve (AUC) was used as the evaluation indicator.

Table 2 shows the dimensions of the main space when 70%, 85%, 95%, and 99% of the variation information of the target classes samples are retained, respectively. One can see that the main information of the target class data often exists in a low-dimensional space, especially for some high-dimensional small sample data. For example, for face recognition datasets, such as warpAR10P,

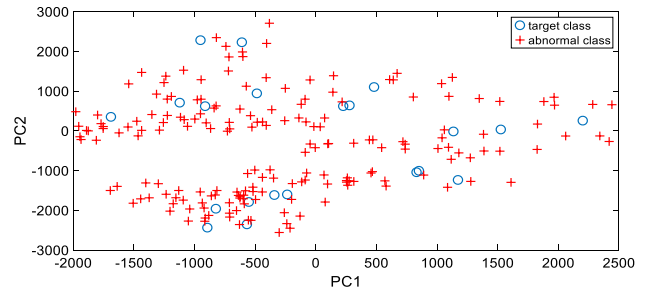


Fig. 4 Score plots of the warpPIE10p dataset in the PC1 × PC2 space

warpPIE10p, and pixraw10P, the dimension of the main space was less than 2% of the original feature dimension. For high-dimensional data, the dimension of the main space was much smaller than the dimension of the original feature space. Therefore, the discriminative information that was useful to detect the abnormal class more easily appeared in the complement space.

The classification results using the Mahalanobis distance are shown in Table 3. For some low-dimensional datasets such as iris, wine, and bupa, CPCA is generally equivalent to using raw features and PCA. And for thyroid and segment, CPCA is better than PCA while using raw features fails. For the spectheart data, CPCA is comparable to the PCA method. For high-dimensional data such as Alphadigits, warpAR10P and milk powder, using raw features fails and CPCA is significantly better than PCA. Therefore, CPCA is significantly better than using raw

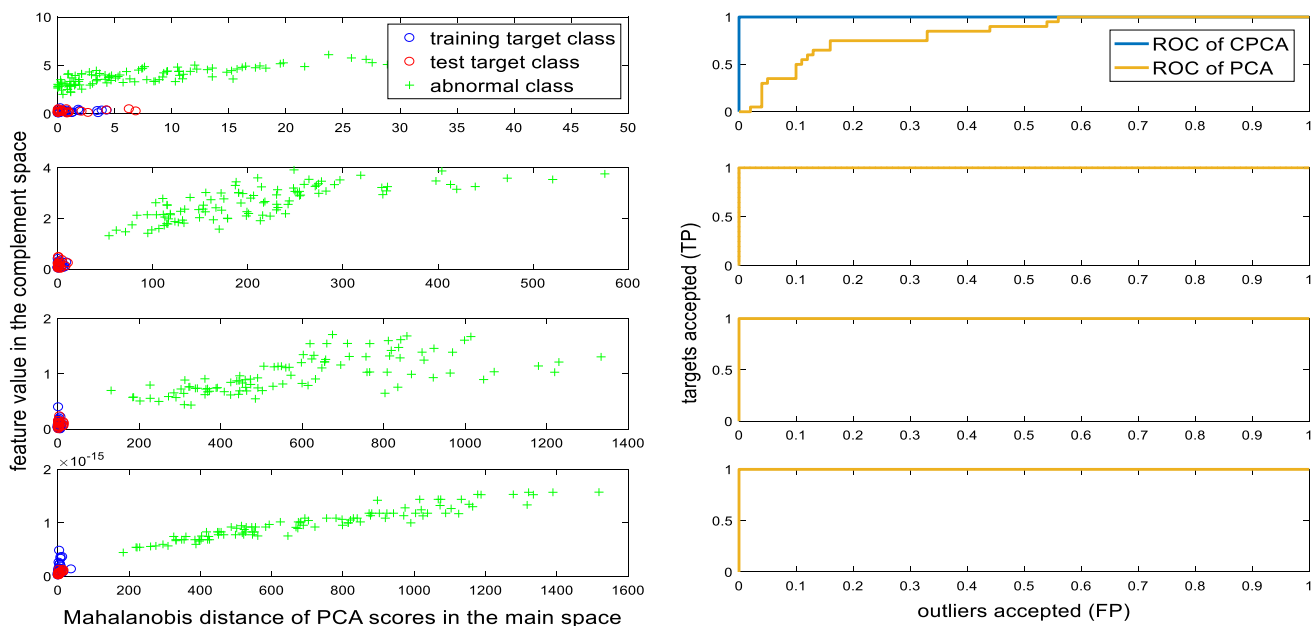
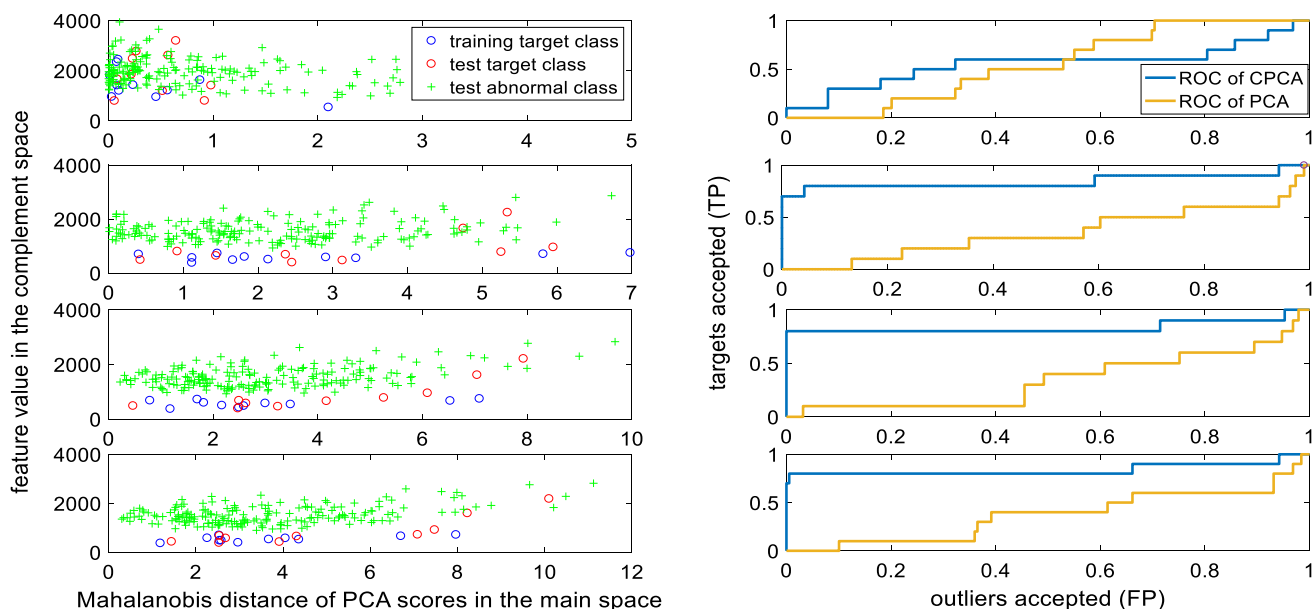


Fig. 3 The feature extraction effects of CPCA and PCA on the iris dataset. The subgraphs from top to bottom correspond to the cases where the number of principal components is 1, 2, 3, and 4, respectively



**Fig. 5** The feature extraction effect of CPCA and PCA on the warpPIE10p dataset. The subgraphs from top to bottom correspond to the cases where the number of principal components is 1, 2, 3, and 4, respectively

**Table 2** Average no. of extracted principal components

Dataset	70% var	85% var	95% var	99% var	
Low dimension	Iris	1.00	2.00	3.02	4.00
	Wine	1.00	1.00	1.00	1.00
	bupa	1.86	2.18	3.00	5.00
	spectheart	7.28	12.00	19.18	27.18
	thyroid	3.86	5.92	8.38	11.08
	segment	1.00	2.00	2.00	3.00
High dimension	USPS	4.76	9.82	22.64	48.44
	Alphadigits	8.22	13.82	20.06	23.50
	warpAR10P	2.38	3.42	5.06	7.22
	warpPIE10p	1.92	2.04	4.18	9.52
	milk powder	2.30	3.21	3.43	4.20
	pixraw10P	2.82	4.00	5.14	6.00

features and PCA when the Mahalanobis distance is used as a one-class classifier.

The one-class classifier with Mahalanobis distance is based on probability density estimation, which is sensitivity to data dimension. To further verify the universality of the proposed method and the support vector machine-based one-class classifier, OC-SVM was applied to these datasets and the results are shown in Table 4. The OC-SVM can work for all the datasets because it introduces kernel functions to avoid ill-posed problems that are caused by high dimensions. For low-dimensional datasets, CPCA, PCA, and using raw features worked well. For most of the high-dimensional datasets, CPCA and using raw features were closer and better than PCA. For some datasets, such

as warpAR10P, pixraw10P, and milk powder, CPCA was significantly better than using raw features. In general, CPCA improved the classification effect of the OC-SVM on high-dimensional data to a certain extent.

### 5 Discussion

CPCA compresses all the feature information of the complement space into one dimension to achieve feature reduction. For the OCC problem, feature reduction has two effects. First, feature reduction leads to the lack of some information, which reduces the recognition effect of some abnormal class samples. Second, feature reduction can



**Table 3** AUC of classification with Mahalanobis distance (/ indicates modeling failure)

Dataset		Raw features	PCA				CPCA			
			70% var	85% var	95% var	99% var	70% var	85% var	95% var	99% var
Low dimension	Iris	<b>1.00</b>	<b>0.78</b>	<b>0.98</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
	Wine	<b>0.98</b>	0.95	0.96	0.96	0.96	0.95	0.96	0.95	0.95
	bupa	0.59	0.56	0.57	0.59	0.58	0.55	0.57	<b>0.60</b>	0.59
	spectheart	/	0.80	0.82	<b>0.84</b>	0.82	0.82	0.81	0.80	0.77
	thyroid	/	0.51	0.51	0.53	0.57	0.52	0.55	0.60	<b>0.89</b>
	segment	/	0.60	0.76	0.76	0.92	0.87	0.96	0.96	<b>1.00</b>
High dimension	USPS	0.97	0.78	0.91	0.98	<b>0.99</b>	0.96	0.97	0.98	<b>0.99</b>
	Alphadigits	/	0.67	0.74	0.85	0.86	<b>0.97</b>	0.96	0.96	0.96
	warpAR10P	/	0.53	0.48	0.43	0.44	0.88	0.93	0.94	<b>0.95</b>
	warpPIE10p	/	0.45	0.43	0.44	0.40	0.91	0.95	<b>0.99</b>	0.98
	milk powder	/	0.61	0.68	0.64	0.69	0.93	0.95	0.97	<b>0.92</b>
	ORL	/	0.30	0.37	0.38	0.47	0.96	0.96	0.96	<b>0.98</b>

**Table 4** AUC of classification with OC-SVM

Dataset		Raw features	PCA				CPCA			
			70% var	85% var	95% var	99% var	70% var	85% var	95% var	99% var
Low dimension	Iris	<b>1.00</b>	0.73	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>
	Wine	<b>1.00</b>	0.92	0.93	0.92	0.92	<b>0.96</b>	0.95	0.95	<b>0.96</b>
	bupa	0.58	0.56	0.57	0.59	0.58	0.56	0.58	<b>0.60</b>	0.58
	spectheart	0.84	0.80	0.82	<b>0.83</b>	0.81	0.82	0.81	0.80	0.78
	thyroid	<b>0.94</b>	0.60	0.58	0.65	0.67	0.63	0.75	0.79	0.90
	segment	<b>1.00</b>	0.62	0.80	0.80	0.95	0.91	0.98	0.98	<b>1.00</b>
High dimension	USPS	0.97	0.76	0.91	0.98	0.99	0.96	0.97	0.98	<b>0.99</b>
	Alphadigits	0.95	0.50	0.63	0.79	0.86	<b>0.97</b>	<b>0.97</b>	0.96	0.96
	warpAR10P	0.83	0.51	0.53	0.48	0.49	0.89	0.92	<b>0.97</b>	0.89
	warpPIE10p	<b>1.00</b>	0.49	0.47	0.43	0.44	0.95	0.94	0.99	0.97
	milk powder	0.90	0.62	0.63	0.68	0.72	0.95	0.97	0.97	<b>0.99</b>
	pixraw10P	0.71	0.43	0.21	0.29	0.36	<b>1.00</b>	<b>1.00</b>	0.93	0.50

avoid the influence of over-fitting caused by high-dimensional features, which improves the recognition effect of some abnormal class samples. CPCA extracts feature information that is different from the target class in the complement space. If the feature information of the target class that is extracted in the main space is too little, then the remaining target class feature information will leak into the complement space, thereby reducing the recognition effect of the abnormal class. This is also the principle by which the main space and the complementary space are constructed. To avoid over-fitting, it is necessary to optimize the number of extracted principal components for the proposed CPCA. According to the experimental results shown in Tables 3 and 4, the CPCA is robust to the number

of extracted principal components. When the number of extracted principal components can retain 85–95% variation information, the CPCA can meet the needs of practical problems. In summary, CPCA introduces a new feature compression method in the complement space to successfully extend the ordinary PCA to the one-class feature extraction field. The CPCA method has a good generalization effect in extracting features for OCC problems and can be used in practical problems, especially for high-dimensional small sample data.

## 6 Conclusion and future work

This paper proposed a new feature extraction strategy for OCC problems. The strategy divided the original feature space into two parts: the main space and the complement space. The main space was used to learn the feature information of the target class, and the complement space was used to learn the feature information of the exception class. By extracting the features of the main space and the complement space separately, the one-class features of the whole space were obtained. According to the proposed one-class feature extraction strategy, a specific implementation, CPCA, was also provided. In CPCA, the features of the main space were compressed by PCA, and the features of the complement space were compressed by the first-order norm. Then, the two types of features were combined as the extracted one-class features of the original space. Several different types of actual datasets were used to verify the effect of the proposed method. The experimental results show that CPCA has good generalization for one-class data, especially for the high-dimensional small sample data. In summary, the proposed method is a feature extraction method that is truly oriented to OCC problems. It also provides a good reference for how to transform other unsupervised feature extraction methods for one-class classification problems. In the following research, we will study the feature space decomposition in the nonlinear space, such as kernel mapping space. The feature compression of manifold learning, deep learning on the main space and subspace will also be explored to further improve the feature extraction effect of OCC problems.

**Funding** The authors would like to acknowledge the financial support provided by the Natural Science Foundation of Zhejiang (LY21C200001 and LQ20F030059) and the National Natural Science Foundation of China (62105245 and 61805180) and the Wenzhou science and technology bureau general project (S2020011 and G20200044).

**Data availability** Available upon request.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Human or animal rights** This article does not contain any studies with human participants or animals performed by any of the authors.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s00500-022-07067-y>.

## References

- Bing L, Liu M, Guo Z, Ji Y (2018) Mechanical fault diagnosis of high voltage circuit breakers utilizing EWT-improved time frequency entropy and optimal GRNN classifier. *Entropy* 20:448–459
- Burnaev E, Smolyakov D (2017) One-Class SVM with Privileged Information and Its Application to Malware Detection. In: *IEEE International Conference on Data Mining Workshops*
- De Santana FB, Neto WB, Poppi RJ (2019) Random forest as one-class classifier and infrared spectroscopy for food adulteration detection. *Food Chem* 293:323–332
- Fei G, Teng H, Sun J, et al (2018) A new algorithm of sar image target recognition based on improved deep convolutional neural network. *Cogn Comput* 1–16
- Fernández-Francos D, MartíNez-Rego D, Fontenla-Romero O, Alonso-Betanzos A (2013) Automatic bearing fault diagnosis based on one-class v-SVM. *Comput Ind Eng* 64:357–365
- Galeano P, Joseph E, Lillo RE (2013) The mahalanobis distance for functional data with applications to classification. *Technometrics* 57:281–291
- Guerbai Y, Chibani Y, Hadjadj B (2018) Handwriting gender recognition system based on the one-class support vector machines. In: *Seventh International Conference on Image Processing Theory*
- Huang G, Yang Z, Chen X, Ji G (2017) An innovative one-class least squares support vector machine model based on continuous cognition. *Knowl-Based Syst* 123:217–228
- Huang G, Yuan L, Shi W et al (2022) Using one-class autoencoder for adulteration detection of milk powder by infrared spectrum. *Food Chem* 372:131219
- Jeong YS, Kang IH, Jeong MK, Kong D (2012) A new feature selection method for one-class classification problems. *IEEE Trans Syst Man Cybern Part C Appl Rev* 42:1500–1509
- Jia F, Yan Y, Zhang J (2018) K-means based feature reduction for network anomaly detection. *J Tsinghua Univ* 58:137–142
- Kemmler M, Rodner E, Wacker ES, Denzler J (2013) One-class classification with Gaussian processes. *Pattern Recognit* 46:3507–3518
- Kim K (2018) An improved semi-supervised dimensionality reduction using feature weighting: application to sentiment analysis. *Expert Syst Appl* 109:49–65
- Koltchinskii V, Lounici K (2016) New asymptotic results in principal component analysis. *Sankhya A* 79:1–44
- Lian H (2012) On feature selection with principal component analysis for one-class SVM. *Pattern Recognit Lett* 33:1027–1031
- Liu C, Wang W, Konan M et al (2017) A new validity index of feature subset for evaluating the dimensionality reduction algorithms. *Knowl-Based Syst* 121:83–98
- Lorena LHN, Carvalho ACPLF, Lorena AC (2014) Filter feature selection for one-class classification. *J Intell Robot Syst* 80:1–17
- Schölkopf B, Smola A, Williamson R, Bartlett P (2000) New support vector algorithms. *Neural Comput* 12:1207–1245
- Tax DMJ (2003) Feature extraction for one-class classification. In: *Joint International conference on artificial neural networks and neural information processing*. pp 342–349
- Washizawa Y, Hotta S (2017) Mahalanobis distance on extended grassmann manifolds for variational pattern analysis. *IEEE Trans Neural Netw Learn Syst* 25:1980–1990
- Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemom Intell Lab Syst* 2:37–52
- Xiao Y, Wang H, Zhang L, Xu W (2014) Two methods of selecting Gaussian kernel parameters for one-class SVM and their application to fault detection. *Knowl-Based Syst* 59:75–84



- Xiao Y, Wang H, Xu W (2015) Hyperparameter Selection for Gaussian Process One-Class Classification. *IEEE Trans Neural Netw Learn Syst* 26:2182–2187
- Xu L, Yan SM, Cai CB, Yu XP (2013) One-class partial least squares (OCPLS) classifier. *Chemom Intell Lab Syst* 126:1–5
- Yu G, Xiao H (2018) Genetic algorithm-tuned adaptive pruning SVDD method for HRRP-based radar target recognition. *Int J Remote Sens* 39:3407–3428

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.