IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 14, 2019, accepted July 7, 2019, date of publication July 16, 2019, date of current version August 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929311

Feature Reduction Method for Cognition and Classification of IoT Devices Based on Artificial Intelligence

XIANG CHEN AND XIAOJUN HAO

State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System, Luoyang 471003, China

Corresponding author: Xiang Chen (cemee_xchen@163.com)

This work was supported by the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System Director Fund under Grant CEMEE2019K0101A.

ABSTRACT In recent years, the scale and application scenarios of the Internet of Things (IoT) have been expanding. Since traditional algorithms are unable to meet wireless networks computing capability requirement in the IoT, more and more research institutions and scholars have turned their eyes to artificial intelligence (AI) methods. Because the IoT device uses wireless networks to communicate in most scenarios, this paper systematically studies the method of feature dimension reduction of wireless communication signals. In this paper, we will take the power amplifier radio frequency (RF) fingerprinting as an example. Focusing on reducing the high dimensionality of RF fingerprint features and the uncorrelated or redundant features in the features space, the RF fingerprint feature dimension reduction method is mainly studied. Based on the principal component analysis (PCA), linear discriminant analysis (LDA), and auto encoder (AE) research, this paper studies the PCA–LDA method and uses the distance ratio criterion to evaluate the separability of features. The simulation results show that the classification accuracy of PCA–LDA is superior to PCA, LDA, and AE in most SNR, and the characteristics of PCA–LDA is more separable.

INDEX TERMS Radio frequency fingerprint, feature dimension reduction, principal component analysis, linear discriminant analysis, intelligent data processing.

I. INTRODUCTION

In the last two decades, the rapid development of wireless communication and hardware devices has promoted the maturity of Internet of Things (IoT) technology. Since IoT uses wireless networks as a means of communication in most scenarios, and traditional algorithms cannot meet the computing requirements of wireless networks, more and more scholars are turning their attention to the application of artificial intelligence (AI) to IoT [1]–[4]. Recently, AI concept has gradually entered people's lives, and a hot spot in the current industry. In recent years, AI has entered a stage of rapid development driven by machine learning. Whether in the fields of finance, computer networks, industrial engineering, and communications, different machine learning algorithms are emerging, and AI and machine learning are culminating again and again. In the field of communication, AI and machine learning algorithms also play a pivotal role [5]–[8]. As an important part of the machine learning method, feature dimension reduction has also received extensive attention from various institutions and scholars.

Dimensionality reduction is to project high dimensional space into low dimensional space in a certain way, so that the feature dimension after dimensionality reduction is much smaller than the dimension before dimension reduction, which realizes the compression of features and reduces the probability of dimensionality disaster.

The linear dimensionality reduction method assumes that the projection relationship from the original space to the dimensionality reduction space is linear, mainly including principal component analysis (PCA) [9] and linear discriminant analysis (LDA) [10]. These two methods are proposed earlier and the theoretical system is perfect, and the practical application effect is good. However, in many cases, nonlinear dimensionality reduction is a good way to so.

The associate editor coordinating the review of this manuscript and approving it for publication was Guan Gui.

To solve the problem. In 1998, Schölkopf *et al.* [11] ''nuclearized'' linear dimensionality reduction methods based on nuclear techniques, and proposed Kernel Principal Component Analysis (KPCA). In 1999, Mika *et al.* [12] proposed Kernel Fisher Discriminant (KFD) based on ''nucleation''. Since 2000, manifold learning has become a research hotspot. Methods such as Isometric Mapping (Isomap) [13], Locally Linear Embedding (LLE) [14], and Laplacian Eigenmaps (LE) [15] have been proposed. In 2006, Tang *et al.* [16] built an auto encoder (AE) network with dimensionality reduction capability. In 2008, van der Maaten and Geoffrey [17] proposed t-Distributed Stochastic Neighbor Embedding (t-SNE), which is an effective method for data dimensionality reduction and visualization. After that, the research methods based on fractal theory [18], [19] and correlation filtering [20], [21] have been continuously proposed by research institutions and scholars. In the last two decades, the dimensionality reduction method based on sparse representation [22]–[24], tensor [25], [26] and migration learning [27], [28] has become a research hotspot.

In the field of radio frequency fingerprinting, only some feature dimension reduction methods are applied. In 2015, Yuan *et al.* [29] used principal component analysis to complete the dimensionality reduction operation of phase space reconstruction features. In 2015, Reising *et al.* [30] extracted the RF-DNA fingerprint of the WiFi preamble and the WiMAX near-transient signal and deleted the redundant features. Only 10% of all features achieved the desired device classification effect.

Through there exists the research on the feature dimension reduction of high-dimensional data in machine learning and intelligent data processing, this paper completes some relevant theoretical research and simulation analysis in. The rest of the paper is arranged as follows: The second part introduces the basic feature dimension reduction method and the evaluation method, the third part talks about the principle of the PCA-LDA method proposed in this paper, and simulation experiment and analysis are in the fourth part, In finally part we draw conclusion.

II. FEATURE DIMENSION REDUCTION METHOD

Dimensionality reduction is to project high dimensional space into low dimensional space in a certain way, so that the feature dimension after dimensionality reduction is much smaller than the dimension before dimension reduction, which realizes the compression of features and reduces the probability of occurrence of dimensionality disaster. This section focuses on three traditional feature dimensionality reduction methods, including PCA, LDA, and auto encoder (AE).

A. PRINCIPAL COMPONENT ANALYSIS

PCA is the most commonly used method of dimensionality reduction. The principle of the method is to find a plane so that the distance between all the samples and the plane is as close as possible and the position of all the samples mapped on this plane is more dispersed. According to these two requirements, the PCA method can be realized.

Given a training set D, first normalize the data, make the mapped matrix $\{w_1, w_2, \ldots, w_d\}$, where w_i is the standard orthogonal basis vector. $\|w_i\|_2 = 1$, $w_i^T w_j = 0$ $((i \neq j))$. The sample x_i is reduced to the d' dimension and the matrix $z_i = (z_{i1}; z_{i2}; \ldots z_{id'})$ is generated. $z_{ij} = w_j^T x_i$ is the mapping of the j dimension of the mapping matrix. If x_i is reconstructed based on z_i, $\hat{x}_i = \sum_{j=1}^{d'} z_{ij} w_j$ will be obtained. In training set D, the distance between the original sample and the reconstructed sample is

$$\sum_{i=1}^{m} \left\| \sum_{j=1}^{d'} z_{ij} w_j - x_i \right\|_2^2 = \sum_{i=1}^{m} z_i^T z_i - 2 \sum_{i=1}^{m} z_i^T W^T x_i + const\ \alpha - tr\left(w^T \left(\sum_{i=1}^{m} x_i x_i^T \right) w \right) \quad (1)$$

All samples are required to be as close as possible in the plane. The optimization goal of principal component analysis is

$$\begin{aligned} &\min_W -tr\left(W^T X X^T W\right) \\ &s.t.\ W^T W = I \end{aligned} \quad (2)$$

where $\sum_i x_i x_i^T$ is the covariance matrix.

The position where all samples are mapped on this plane is required to be as scattered as possible, that is, the variance of the sample after mapping is as large as possible. Where W is the mapping matrix and satisfies.

$$\begin{aligned} &\max_W tr\left(W^T X X^T W\right) \\ &s.t.\ W^T W = I \end{aligned} \quad (3)$$

by using the Lagrangian multiplier method for equation (2) or (3), it is converted into:

$$X X^T W = \lambda W \quad (4)$$

The eigenvalue decomposition method is used to process the covariance matrices and arrange the eigenvalues from large to small $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$. All eigenvectors of the first d' eigenvalues of the covariance matrix are used to form $W = (w_1, w_2, \ldots, w_{d'})$. This is the solution to principal component analysis.

B. LINEAR DISCRIMINANT ANALYSIS

LDA belongs to linear dimensionality reduction method. This method considers the category information of the sample and is a supervised dimensionality reduction method. In practical applications, it is necessary to find a straight line so that the distance between the samples of the same category mapped onto the line is as small as possible, and the distance between the positions of the different types of sample maps is as large as possible.

Given a training set $D = \{(x_i, y_i)\}_{i=1}^{m}$, $y_i \in \{0, 1\}$, the mapping matrix is w, the sum of the covariances of the two types of data $w^T \sum_0 w + w^T \sum_1 w$ is required to be as small as possible, where $\sum_i$ is the covariances of x_i. It is required that the distance between different types of sample mappings is as large as possible, that is, the distance between the central mappings of the two types of data $\left\|w^T \mu_0 - w^T \mu_1\right\|_2^2$ is as large as possible, where μ_i is the mean. In order to balance the two conditions, the criterion for maximizing is

$$J = \frac{\left\|w^T \mu_0 - w^T \mu_1\right\|_2^2}{w^T \sum_0 w + w^T \sum_1 w} = \frac{w^T (\mu_0 - \mu_1)(\mu_0 - \mu_1)^T w}{w^T \left(\sum_0 + \sum_1\right) w} \tag{5}$$

the formula for ''intra-class divergence matrix'' is

$$S_w = \sum_0 + \sum_1 = \sum_{x \in X_0} (x - \mu_0)(x - \mu_0)^T + \sum_{x \in X_1} (x - \mu_1)(x - \mu_1)^T \tag{6}$$

''Interclass divergence matrix'' is

$$S_b = (\mu_0 - \mu_1)(\mu_0 - \mu_1)^T \tag{7}$$

then equation (5) can be rewritten as

$$J = \frac{w^T S_b w}{w^T S_w w} \tag{8}$$

when $w^T S_w w = 1$, then equation (8) is equivalent to

$$\begin{aligned} &\min_{w} -w^T S_b w \\ &s.t.\ w^T S_w w = 1 \end{aligned} \tag{9}$$

using the Lagrangian multiplier method, the formula (9) can be written as follow

$$S_b w = \lambda S_w w \tag{10}$$

where λ is the Lagrange multiplier. Taking all the feature vectors of the first d' eigenvalues of $S_w^{-1} S_b$ to form the projection matrix.

C. AUTO ENCODER

AE is based on neural network. Input data is encoded in an unsupervised learning manner to obtain the characteristics of the hidden layer, and output data is obtained by decoding. The goal of AE is to make the difference between input and output data as small as possible. Figure 1 shows the network model of the AE. In the AE, features number of input layer is equal to features number of output layer, and features number of the hidden layer has three relationships with the number of features of the input and output layers. The first type is that the number of hidden layer features is greater than the number of input layer features, that is, the dimension is increased. The second is that the number of hidden layer features is equal to the number of input layer features, that is, the same dimension. The third is that the number of hidden layer features is less than the number of input layer features, that is, dimensionality reduction. In this paper, the dimension reduction is mainly used.

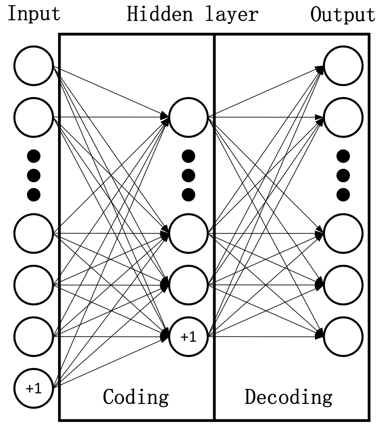

FIGURE 1. Auto encoder network model.

The input data is $\left\{x^{(n)} \in \mathbb{R}^u\right\}_{n=1}^{N}$, and input data is the desired output. The model is

$$\left\{ \begin{array}{l} X = \sigma_a (W_a \cdot x + b_a) \in \mathbb{R}^v \\ \hat{x} = \sigma_s (W_s \cdot X + b_s) \in \mathbb{R}^u \end{array} \right\} \tag{11}$$

where the parameter of the analysis phase is $W_a \in \mathbb{R}^{v \times u}$, $b_a \in R^v$, the activation function is $\sigma_a(\cdot)$, where the letter ''a'' is the initial letter of the analysis; the parameter of the synthesis phase is $W_s \in \mathbb{R}^{u \times v}$, $b_s \in R^u$, the activation function is $\sigma_s(\cdot)$, and the same letter ''s'' is the initial letter of the synthesis; The predicted value of output is $\hat{x}$. The optimization objective equation based on energy loss can be expressed as follows:

$$\min_{\theta} J(\theta) = \frac{1}{N} \sum_{n=1}^{N} \left\|\hat{x}^{(n)} - x^{(n)}\right\|_2^2 + \lambda \cdot R(\theta) \tag{12}$$

where $\hat{x}^{(n)}$ is the predicted value of the ideal target value $x^{(n)}$, and the parameters and regular terms are:

$$\left\{ \begin{array}{c} \theta = [W_a, b_a; W_s, b_s] \\ R(\theta) = \|W_a\|_F^2 + \|W_s\|_F^2 \end{array} \right\} \tag{13}$$

D. EVALUATION OF FEATURE DIMENSION REDUCTION METHOD

The distance-based evaluation criteria are used in the evaluation of the dimensionality reduction method, and the distance evaluation function is also called the separability criterion.

The distance between different types of samples increases with the divergence between different types of samples, and the spatial distribution of the samples is farther and farther. At the same time, the intra-class distance is as small as possible, and the intra-class aggregation is high. For a classification problem, the distance evaluation function can evaluate the degree of differentiation of features from different samples.

The distance evaluation function includes a class-to-class distance criterion, an inter-class distance criterion, and a distance ratio criterion.

1) Class-to-Class Distance Criterion: The class-to-class distance criterion evaluates the degree to which features are distinguished from different classes of samples based on the distance between the classes and the classes. The greater the distance between classes, the stronger the classification ability of the feature set. The centroid is used in the calculation of the distance. The centroid refers to the physical mass center. The average value of the similar sample is used as the centroid of the sample, and the centroid is used to represent the sample. For the data set, the centroid of each class is to add all the sample feature values of the sample to find the average value, that is, the centroid of the class. The size of the distance between classes represents the degree of separation between classes. Therefore, different features can be used to calculate the distance between centroids of the class. The larger the distance, the more obvious the classification of different types of samples, indicating that the feature has better separability for different types of samples. A typical method of calculating the distance is the Euler distance formula. The distance criterion between classes is: first calculate the centroid of each type of sample; secondly calculate the distance between the two types of samples, calculate the average of the distance again, and finally evaluate the degree of separation.
2) Inter-Class Distance Criterion: The inter-class distance criterion is used to evaluate the degree of convergence of features on similar samples. The smaller the sum of the distances between similar samples, the smaller the maximum distance and the closer the aggregation, indicating that the characteristics of the same sample are more powerful. The calculation method of the inter-class distance criterion is as follows: firstly calculate the distance between two pairs of the same kind of sample, then calculate the distance average, and finally evaluate the degree of convergence.
3) Distance Ratio Criterion: The two distance criteria described above have different evaluation angles for feature separability. The class-to-class distance criterion evaluates the separability of features to different types of samples. The inter-class distance criterion evaluates the degree of aggregation of features to similar samples, so that the feature space distances of different types of samples are as large as possible, and similar samples are gathered as much as possible. Therefore, the distance ratio criterion can be designed to comprehensively evaluate the separability of the feature. The distance ratio criterion is calculated as follows:

$$R_d = \frac{D_b}{D_w} \tag{14}$$

where D_b is the evaluation value of the distance criterion between classes, D_w is the evaluation value of the distance criterion within classes, and R_d is the ratio of the former to the latter. The larger the ratio, the stronger the separability.

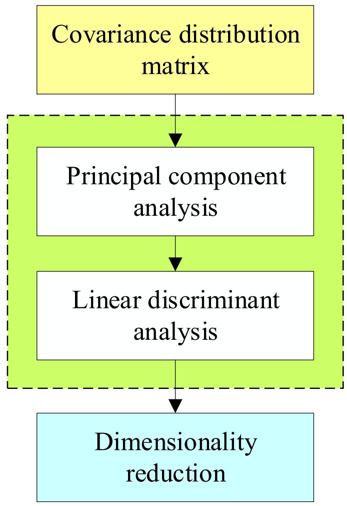

FIGURE 2. Schematic diagram of PCA-LDA method.

In summary, the distance-based evaluation criterion can evaluate the separability of features, and has the advantages of simple and intuitive calculation principle. However, if the number of samples is larger, the amount of calculation will be larger.

III. PRINCIPAL COMPONENT ANALYSIS-LINEAR DISCRIMINANT ANALYSIS

The traditional PCA approach is an unsupervised learning approach that does not consider the correlation between data and categories. When a part of the component is discarded, some principal components with a small contribution rate may contain important information that affects the sample difference. The biggest disadvantage of the traditional LDA method is that it is limited by the type of sample. The dimension of the projection space is smaller than the number of categories, and there are singular problems. Aiming at the problems existing in the traditional PCA method and LDA method, the PCA-LDA dimension reduction method is studied. The schematic diagram of this method is shown in Figure 2.

The PCA-LDA method can overcome the singular problem. In practice, this problem is avoided by projecting data set into the lower dimensional space, so that intra-class scatter matrix is non-singular. PCA-LDA is implemented by using PCA to reduce the size of the feature space to d_{pca} dimensions and then applying a standard FLD to reduce the size to the target dimension. W_{opt}^{T} means the following:

$$W_{opt}^{T} = W_{fld}^{T} W_{pca}^{T} \tag{15}$$

$$W_{pca} = \arg\max_{W} \left| W^{T} S_{T} W \right| \tag{16}$$

$$W_{fld} = \arg\max_{W} \frac{\left| W^{T} W_{pca}^{T} S_{B} W_{pca} W \right|}{\left| W^{T} W_{pca}^{T} S_{W} W_{pca} W \right|} \tag{17}$$

IV. SIMULATION EXPERIMENT AND ANALYSIS

A. EXPERIMENTAL DATA AND METHODS

The experimental data set is the signal of 8 amplifiers collected in the real environment. Each amplifier has 120 samples and a total of 960 samples, each of which has a signal length of 20000 sample points. In the experiment, the data set is divided into training set, verification set and test set, the ratio is set to 4:1:1, that is, the training set contains 640 samples, the verification set contains 160 samples, and the test set contains 160 samples. Gaussian white noise is added to the signal to simply simulate the influence of the environment on the signal, and the signal-to-noise ratio (SNR) ranges from 0 dB to 20 dB. The feature extraction method used is a covariance distribution feature with a feature dimension of 14850. Therefore, the dataset has a total of 20160 samples, of which the training set contains 13,440 samples, the validation set contains 3360 samples, and the test set contains 3360 samples.

Four feature dimension reduction methods studied in this chapter were applied in the experiment, including PCA method, LDA method, AE method and PCA-LDA method. The parameters of the automatic encoder are as follows: the number of input layer nodes is 14850, the number of hidden layer 1 nodes is 128, the number of hidden layer 2 nodes is set to the target dimensionality reduction dimension, that is, from 1 to 30, the number of hidden layer 3 nodes is 128, and the optimizer we select is Adadelta. The activation function of the middle layer is ReLU, ReLU, and the output layer activation function is Linear. In the PCA-LDA method, the PCA dimension reduction dimension is initially set to 37. If the experimental results are not good, the dataset is adapted to the adjustment. The classifier uses the K-nearest neighbor method (KNN), and the number of neighbors is 6.

B. EXPERIMENTAL RESULTS AND ANALYSIS

This part of the simulation experiment mainly includes three parts: 1) classification experiment of different dimensionality reduction methods with characteristic dimension under the same condition of SNR; 2) different feature selection methods with SNR under the same conditions of dimensionality reduction Changed classification experiments; 3) visualization experiments; 4) separability evaluation.

1) CLASSIFICATION EXPERIMENT OF DIFFERENT DIMENSIONALITY REDUCTION METHODS WITH CHARACTERISTIC DIMENSION CHANGES UNDER THE SAME CONDITIONS OF SNR

The classification accuracy rate is the most important indicator to measure the effectiveness of the dimension reduction method. Figure 3 shows the classification accuracy curves of PCA method, LDA method, AE method and PCA-LDA method with the change of feature dimension when the SNR is 5dB, 10dB, 15dB and 20dB respectively.

In Figure 3, classification accuracy of PCA generally increases first and then gradually decreases and finally stabilizes. With the increase of the dimensionality of the dimension reduction, the classification accuracy of the LDA method is gradually increased. After the decline, it gradually rises and then falls, and finally stabilizes. The classification accuracy rate of the AE method generally increases gradually. Because of the randomness of the training, the curve fluctuation is more obvious. The classification accuracy of the PCA-LDA method increases gradually and then gradually decreases. Finally, the trend tends to be stable, but the overall fluctuation range is small compared with PCA, LDA and AE, and the classification accuracy rate is 98.75% when SNR is 20 dB and the dimensionality reduction dimension is 2. It shows that the PCA-LDA method can effectively reduce the dimensions and reduce redundancy, and improve classification performance of KNN.

2) CLASSIFICATION EXPERIMENT OF DIFFERENT FEATURE SELECTION METHODS WITH SNR UNDER THE SAME CONDITIONS OF DIMENSIONALITY REDUCTION

Figure 4 shows the classification accuracy curves of PCA, LDA, AE and PCA-LDA methods with SNR changes when the feature dimensions are 2, 10, 20 and 30, respectively.

It can be seen from Figure 4 that with the increase of SNR, the classification accuracy of PCA, LDA, AE and PCA-LDA methods generally increases, and the AE method is affected by the dimensionality of dimensionality reduction. The more feature dimensions of dimensionality reduction, the better the classification effect of the AE method. The classification effect of PCA-LDA method is better than that of PCA method, LDA method and AE method under most SNR, and the classification accuracy is higher than 90% at high SNR.

3) VISUAL EXPERIMENT

Figure 5 shows the sample distribution of different dimensionality reduction methods down to 2D when the SNR is 20dB. The horizontal and vertical axes represent the projected axes.

It can be seen from Figure 5 that the sample distribution of the AE method is the densest and the most difficult to distinguish. The sample distribution of the PCA method is relatively scattered and has a certain ability to distinguish, but different types of samples are easy to overlap. The sample distribution of LDA method and PCA-LDA method is the

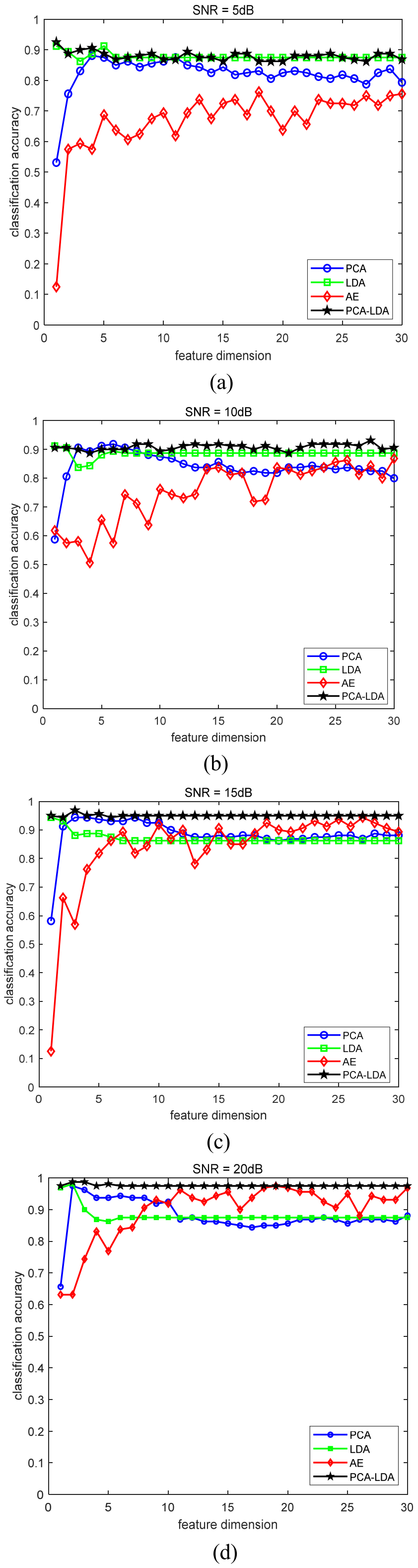

FIGURE 3. Classification accuracy curve of different dimensionality reduction methods with feature dimension at different SNR, and with the sequences: (a) SNR = 5 dB, (b) SNR = 10 dB, (c) SNR = 15 dB, (d) SNR = 20dB.

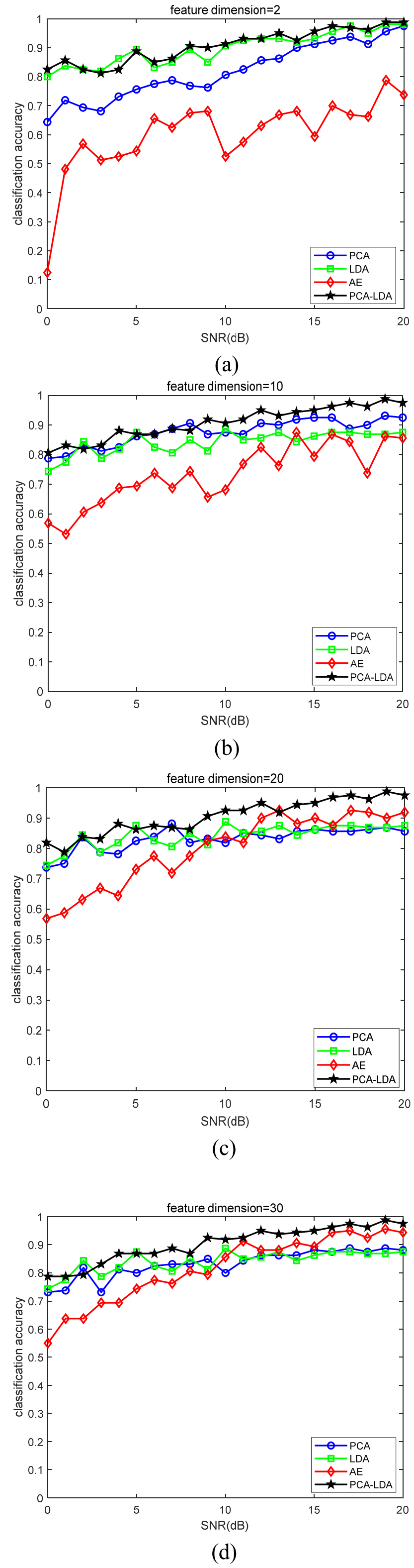

FIGURE 4. Classification accuracy curve of different dimensionality reduction methods with SNR, and with the sequences: (a) feature dimension = 2, (b) feature dimension = 10, (c) feature dimension = 20, (d) feature dimension = 30.

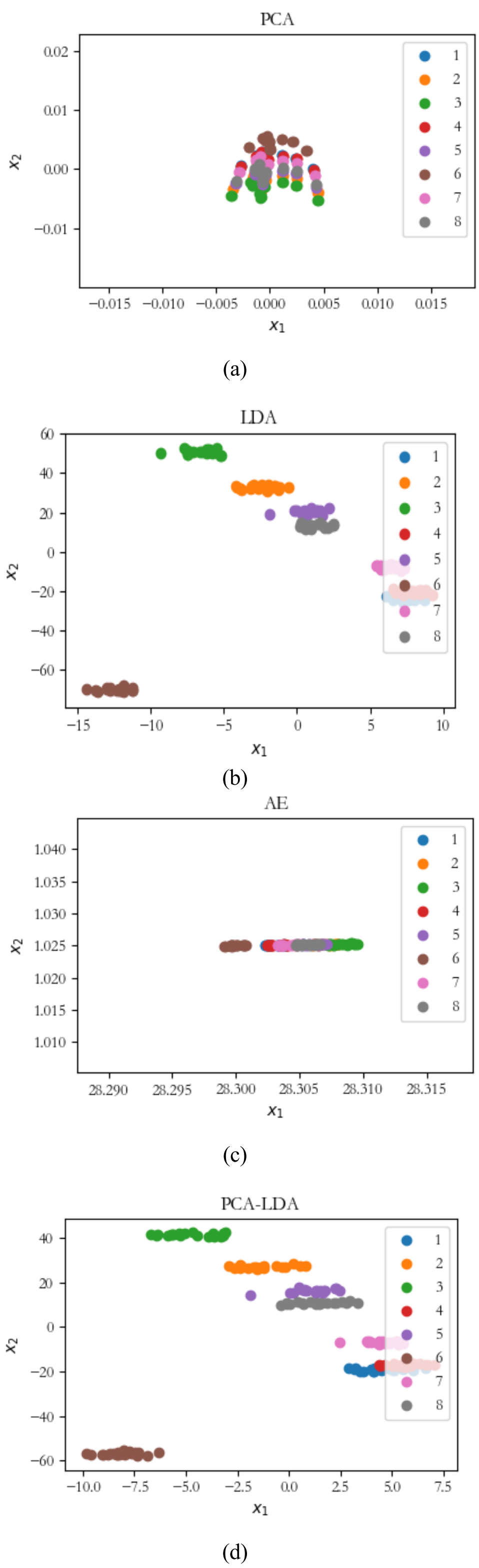

FIGURE 5. Sample distribution of different dimensionality reduction methods down to 2 dimensions (SNR = 20 dB), and with the sequences: (a) PCA, (b) LDA, (c) AE, (d) PCA-LDA.

most loose, and there is a clear distinction between each type of sample, and the sample distinguishing ability is the strongest.

TABLE 1. Data summary of granite rock burst experimental data.

Dimensionality reduction method	Distance ratio
PCA	0.8505
LDA	23.6548
AE	5.0661
PCA-LDA	21.1303

4) SEPARABILITY EVALUATION

When the SNR is 20 dB and the feature dimension is reduced to 2 dimensions, the distance ratios of the different feature dimensionality reduction methods are shown in Table 1.

It can be seen from Table 1 that when the dimensionality reduction dimension is 2, the distance ratio of the PCA method is the smallest, and the distance ratio of the AE method is larger than the PCA method, and the separability between the two is small. The distance ratio of the LDA and PCA-LDA methods are relatively large and the score ability is strong.

V. CONCLUSION

This paper focuses on the method of RF fingerprint dimension reduction. The three traditional dimensionality reduction methods, including PCA method, LDA method and AE method, are studied. The PCA-LDA method is studied for the shortcomings of PCA and LDA. The feature selected in simulation experiment is the covariance distribution feature, and the feature number is 14850. When the SNR is 5dB, 10dB, 15dB, 20dB respectively, the classification effect of PCA-LDA is better than three traditional methods. When feature dimension is 2 and the SNR is 20dB, the classification accuracy rate can reach 98.75%. When the feature dimension is 2, 10, 20, 30, the classification accuracy of PCA-LDA method is higher than PCA method, LDA method and AE method under most SNR. And the classification accuracy rate is higher than 90% under high SNR. Therefore, the PCA-LDA method can effectively reduce dimensions and reduce redundancy, and improve the classification performance of KNN. The distance ratio criterion is used to prove that the PCA-LDA method has strong feature separability.

REFERENCES

[1] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.

[2] Y. Lin, H. Tao, Y. Tu, and T. Liu, "A node self-localization algorithm with a mobile anchor node in underwater acoustic sensor networks," *IEEE Access*, vol. 7, pp. 43773–43780, 2019.

[3] G. Ding, Q. Wu, L. Zhang, T. A. Tsiftsis, and Y. Yao, "An amateur drone surveillance system based on cognitive Internet of Things," *IEEE Commun. Mag.*, vol. 56, no. 1, pp. 29–35, Jan. 2018.

[4] Z. Zhang, X. Guo, and Y. Lin, "Trust management method of D2D communication based on RF fingerprint identification," *IEEE Access*, vol. 6, pp. 66082–66087, 2018.

[5] J. Wang, C. Ju, Y. Gao, A. K. Sangaiah, and G.-J. Kim, "A PSO based energy efficient coverage control algorithm for wireless sensor networks," *Comput. Mater. Continua*, vol.56, no. 3, pp. 433–446, 2018.

[6] R. Sun, L. Shi, C. Yin, and J. Wang, "An improved method in deep packet inspection based on regular expression," *J. Supercomput.*, vol.75, no. 6, pp. 3317–3333, 2019.
[7] Y. Lin, X. Zhu, Z. Zheng, Z. Dou, and R. Zhou, "The individual identification method of wireless device based on dimensionality reduction and machine learning," *J. Supercomput.*, vol. 5, pp. 1–18, Dec. 2017.
[8] J. Wang, Y. Gao, W. Liu, W. Wu, and S.-J. Lim, "An asynchronous clustering and mobile data gathering schema based on timer mechanism in wireless sensor networks," *Comput. Mater. Continua*, vol. 58, no. 3, pp. 711–725, 2019.
[9] B. C. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Autom. Control*, vol. AC-26, no. 1, pp. 17–32, Feb. 1981.
[10] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
[11] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
[12] S. Mika, G. Ratsch, and J. Weston, "Fisher discriminant analysis with kernels," in *Proc. 9th IEEE Signal Process. Soc. Workshop Neural Netw. Signal Process.*, Aug. 1999, pp. 41–48.
[13] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
[14] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[15] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.
[16] B. Tang, Y. Tu, Z. Zhang, and Y. Lin, "Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks," *IEEE Access*, vol. 6, pp. 15713–15722, 2018.
[17] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[18] M. Liu, T. Song, and G. Gui, "Deep cognitive perspective: Resource allocation for NOMA-based heterogeneous IoT with imperfect SIC," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2885–2894, Apr. 2019.
[19] J. Li, D. Bi, Y. Ying, K. Wei, and B. Zhang, "An improved algorithm for extracting subtle features of radiation source individual signals," *Electronics*, vol. 8, no. 2, p. 246, 2019.
[20] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DoA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.
[21] Y. Li, X. Cheng, and G. Gui, "Co-robust-ADMM-Net: Joint ADMM framework and DNN for robust sparse composite regularization," *IEEE Access*, vol. 6, pp. 47943–47952, 2018.
[22] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Statist.*, vol. 15, no. 2, pp. 265–286, Jun. 2006.
[23] L. Clemmensen, T. Hastie, D. Witten, and B. Ersboll, "Sparse discriminant analysis," *Technometrics*, vol. 53, no. 4, pp. 406–413, 2011.
[24] M. Jia, Z. Gao, G. Qing, X. Gu, and Y. Lin, "Sparse feature learning for correlation filter tracking toward 5G-enabled tactile internet," *IEEE Trans. Ind. Inform.*, to be published.
[25] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.
[26] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H.-J. Zhang, "Multilinear discriminant analysis for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 212–220, Jan. 2007.
[27] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.
[28] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Sparse transfer manifold embedding for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1030–1043, Feb. 2014.
[29] Y. Yuan, Z. Huang, X. Wang, and F. Wang, "Radio Specific Emitter Identification based on nonlinear characteristics of signal," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw.*, May 2015, pp. 77–81.
[30] D. R. Reising, M. A. Temple, and J. A. Jackson, "Authorized and rogue device discrimination using dimensionally reduced RF-DNA fingerprints," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 6, pp. 1180–1192, Jun. 2015.

XIANG CHEN received the B.S. degree in electrical and information engineering from the First Aeronautical College, Air Force, Xinyang, China, in 2007, the M.S. degree in electromagnetic field and microwave technology, and the Ph.D. degree in weapon systems and operational engineering from the Mechanical Engineering College, Shijiazhuang, China, in 2010 and 2013, respectively. He is currently an Assistant Professor with the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), Luoyang, Henan. His current research interests include wireless network security, machine learning, and EMC.

XIAOJUN HAO received the B.S., M.S., and Ph.D. degrees in electrical and information engineering from the National University of Defence Technology (NUDT), Changsha, China, in 2000, 2002, and 2008, respectively. He is currently an Associate Professor with the State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), Luoyang, Henan. His current research interests include machine learning, strong field environment, and EMC.