Research on Communication Individual Identification Method Based on PCA-NCA and CV-SVM

Xinghao Guo[1](✉) and Shuai Liu[2]

[1] Harbin Engineering University, Harbin, China
s317080023@hrbeu.edu.cn
[2] Inner Mongolia University, Hohhot, China

Abstract. In recent years, high-dimensional data has often appeared in the fields of science and industry, such as computer vision, pattern recognition, biological information, and aerospace. Feature dimension reduction and selection are the process of reducing data from high dimensionality to low dimensionality to reveal the nature of the data. In the field of wireless communication, in view of the feature redundancy caused by the high-dimensional features of wireless device startup transient signals, this paper converts the high-dimensional features of signals into low-dimensional features that are conducive to classification through the feature dimensionality reduction and selection method based on PCA-NCA. In addition, this paper also carried out parameter optimization for SVM classifier, and the established CV-SVM classifier improved the classification performance. This paper also carries out simulation devices on the measured start-up signals of ten identical walkie-talkies. When the SNR is greater than 0 dB, the recognition accuracy of the PCA-NCA algorithm is 10% higher than recognition accuracy of the PCA algorithm alone; when the SNR is greater than 10 dB.

Keywords: PCA · Individual identification · Feature selection

1 Introduction

Feature selection refers to selecting a subset of features from the original feature set that optimizes certain evaluation criteria. With feature selection, some task-independent or redundant features are removed, and simplified data sets are often more accurately modeled and easier to understand [1]. Feature selection is a hot issue in related fields such as machine learning and pattern recognition. Many scholars at home and abroad have studied this and proposed a series of algorithms [1–3].

The feature selection methods are mainly divided into 2 types according to whether the classification accuracy is used as an evaluation function: a filtering method and a wrapper method [4]. Since the filtering method does not use the classification accuracy as evaluation index. However, since the method evaluates the classification effect during each cycle, it leads to a long time. Jain [5] experiments show that genetic algorithms are prone to premature convergence, and when the structure of genetic algorithms is applied to large-scale data, the operating efficiency will decline [6].

G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 301, pp. 12–19, 2019.
https://doi.org/10.1007/978-3-030-36402-1_2

Neighborhood components analysis (NCA) is a supervised learning method proposed by Goldberger et al. [7]. It measures the sample data according to a given distance measurement algorithm, and then classifies the multivariate data. In terms of function, it has the same purpose as k-nearest neighbor algorithm, and directly USES the concept of random nearest neighbor to determine the labeled training samples adjacent to the test samples. It has been widely used in speech recognition [8], image recognition [9, 10] and text recognition [11, 12]. Based on the research of PCA and NCA methods, this paper combines the two algorithms and improves the traditional SVM algorithm. The remaining papers are arranged as follows: in the second chapter, this paper will introduce the basic theory of PCA, NCA and CV-SVM algorithm; In chapter 3, this paper takes the measured startup transient signals of ten interphones of the same model as the data set to verify the effect of the method proposed in this paper. In the fourth chapter, the thesis will put forward the conclusion.

2 Algorithm Theory

2.1 PCA Feature Dimensionality Reduction

PCA is widely used in various fields because it can concentrate information, simplify the structure of indexes and make the process of analyzing problems simple, intuitive and effective. Suppose, a training set is $X = \{x_1, x_2, \cdots, x_m\}$, x_i is a n-dimensional vector. Meantime, a lower dimensional data set is $\mathbf{Y} = \{y_1 y_2, \cdots, y_m\}$, y_i is a one-dimensional vector.

(1) Data preprocessing

Make $X = \{x_1, x_2, \cdots, x_m\}$ have a mean of 0.

$$\mu_x = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{1}$$

$$x_i' = x_i - \mu_x \tag{2}$$

The processed matrix is $X' = \left\{x_1', x_2', \ldots x_m'\right\}$

(2) Obtain the covariance matrix

$$C_x = X' X'^T \tag{3}$$

(3) Calculate feature values and feature vectors

$$C_x a = \lambda a \tag{4}$$

The eigenvalues are arranged in descending order according to the value, the eigenvector related to the first eigenvalue is selected, and then used to form the transformation matrix A.

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \cdot & 0 & \\ & 0 & \cdot & \\ & & & \lambda_n \end{bmatrix}, \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \tag{5}$$

$$\frac{\sum_{i=1}^{l} \lambda_i}{\sum_{i=1}^{n} \lambda_i} \geq k \tag{6}$$

$$A = \{a_1, a_2 \cdots, a_l\} \tag{7}$$

k represents the proportion of energy in the reduced data to the original data. We usually set k to 85%$\sim$90%.

(4) Calculate the dimensionality reduction matrix Y

The sample $\mathbf{Y} = \{y_1 y_2, \cdots, y_m\}$ is multiplied by the transpose matrix of the feature vector. The sample $\mathbf{Y} = \{y_1 y_2, \cdots, y_m\}$ is subtracted from the mean of the training samples to ensure that the training sample and the test sample are converted to the same sample space.

$$Y = A^T \times X' \tag{8}$$

2.2 NCA Feature Selection

Neighborhood component analysis is a distance measurement learning method. Its purpose is to obtain a linear space transfer matrix by learning on the training set, and to maximize the classification effect of LOO in the new transformation space. The key of this algorithm is to find A positive definite matrix A related to the spatial transformation matrix, which can be obtained by defining A differentiable objective function of A and using iterative methods (such as conjugate gradient method, conjugate gradient descent method, etc.). One of the benefits of this algorithm is that the number of categories K can be defined by a function f (determining scalar constants). Therefore, the algorithm can be used to solve the problem of model selection.

To define the transformation matrix A, we first define an objective function that represents the classification accuracy in the new transformation matrix, and try to determine that A* maximizes this objective function.

$$A^* = \text{argmax}_A\, f(A) \tag{9}$$

When classifying a single data point, we need to consider the k nearest neighbors determined by a given distance metric, and get the class of the sample according to the category label of the k neighbors.

In the new conversion space, we do not use the left-sort method to find k nearest neighbors for each sample point, but consider the entire data set as a random nearest neighbor in the new space. We use a squared Euclidean distance function to define the distance between a data point and other data in the new conversion space. The function is defined as follows:

$$p_{ij} = \begin{cases} \frac{e^{-\left\|Ax_i - Ax_j\right\|^2}}{\sum_k e^{-\left\|Ax_i - Ax_j\right\|^2}}, & \text{if } j \neq i \\ 0, & \text{if } j = i \end{cases} \tag{10}$$

The classification accuracy of the input point i is the classification accuracy of the nearest neighbor set C_i adjacent to it: $p_i = \sum_j^n p_{ij}$, Where p_{ij} denotes the probability that j is the nearest neighbor of i. The objective function defined by the global data set as the nearest neighbor classification method of random nearest neighbors is defined as follows:

$$f(A) = \sum_i \sum_{j \in C_i} p_{ij} = \sum_i p_i \tag{11}$$

The objective function can be better chosen as:

$$\frac{\partial f}{\partial A} = -2A \sum_i \sum_{C_i} p_{ij} \left(x_{ij} x_{ij}^T - \sum_k p_{ik} x_{ik} x_{ik}^T \right) \tag{12}$$

A continuous gradient descent algorithm is used here.

2.3 CV-SVM Classifier

The SVM classifier uses the nonlinear mapping function to optimally classify the sample signals in high-dimensional space to separate the training sample points and maximize the distance from the optimal separation surface [13, 14]. The SVM classifier is used to construct the optimal classification hyperplane in high dimensional space.

$$f(x) = a\rho(x) + b = \sum_{k=1}^{m} a_k \rho(x_k) + b = 0 \tag{13}$$

Where a is a weight vector, b is a threshold, and a and b determine the location of the classification plane. Introduce a relaxation factor to measure the distance between the SVM output value and the practice indicator value, thereby transforming the problem of optimizing the sample signal separation surface into

$$\begin{cases} \min \frac{1}{2}\|a\|^2 + C\sum_{i=1}^{m}\mu_i; i = 1, \cdots, m \\ s.t. \begin{cases} y_i(ax_i + b) \geq 1 - \mu_i \\ \mu_i \geq 0 \end{cases} \end{cases} \tag{14}$$

Among them, the penalty parameter C is used to control the degree of penalty of the wrong sample. Secondly, the Lagrangian multiplier is introduced to make the above problem a secondary plan. Then introduce the RBF kernel function to get RBF-SVM.

$$f(x) = sign\left(\sum_{i,j=1}^{m}\alpha_i y_i \exp\left\{-g\|x - x_i\|^2\right\} + b\right) \tag{15}$$

C and g that take values within a certain range are used as the original data set. Then, step by step, gradually increase the values of C and g, and calculate the classification accuracy under this group C and g. The final parameters are obtained by C and g with the highest classification accuracy. Since the search process for this optimal parameter is a cross-validation, grid search process, the final SVM is called CV-SVM.

3 Simulation Analysis

In order to verify the effect of Hilbert transform, NCA-PCA feature selection method and CV-SVM classifier, this paper collects the startup transient signals for the same model of the same manufacturer, and uses MATLAB to simulate noise interference.

This simulation verifies the method proposed in this paper from two aspects: the classification results between different methods and the classification results of different individuals. In this paper, the antenna of the walkie-talkie is directly connected to the input of the oscilloscope cable to reduce the fading of the signal, ignoring the multipath phenomenon, the delay phenomenon, and the temperature during the transmission. A high-performance Agilent oscilloscope is used to connect the communication station with a cable, and the startup transient signal is directly collected at a sampling rate of 40 MHz, and the number of sampling points is 159901. For 10 devices, each device collects 50 noise-free signals, and divides the 500 signals into training samples and test samples in a ratio of 2:3. It is necessary to artificially add Gaussian white noise after acquiring the signal. Because the time domain length of the captured transient signal is long, the computer memory will be required in the subsequent processing. So, we did a simple sampling at every 50 points, and the length of the signal Hilbert envelope dropped to 3187.

Firstly, this paper uses five methods of 'PCA+SVM', 'PCA+NCA+SVM', 'PCA +NCA+CVSVM', 'PCA+KNN' and 'PCA+NCA+KNN' to identify the startup transient signals of 10 walkie-talkies. The results are shown in Fig. 1. After the Hilbert transformation and sampling of the experimental data used in this paper, PCA dimension reduction is carried out. In this paper, the dimension when the contribution rate is 90% is taken as the dimension after dimension reduction. In the three methods of 'PCA+NCA+SVM', 'PCA+NCA+CVSVM' and 'PCA+NCA+KNN', we also use the NCA feature selection method on the basis of PCA to select features with feature weights greater than 1 for classification. We can see that: (1) when 'PCA+SVM' and 'PCA+NCA+SVM' are less than 10 dB, the recognition rate is always greater than 10%, which proves that the feature selection method proposed in this paper can effectively improve the performance of the classifier. Not only that, from the 'PCA+ KNN' and 'PCA+NCA+KNN' these two methods of the results of comparison, can also explain the superiority of the method proposed in this paper; (2) from a general view of the recognition rate curve of 'PCA+NCA+CVSVM' method, it can be seen that when the traditional gaussian kernel function SVM is modified to CV-SVM, the recognition accuracy is improved a little. Thus, the PCA-NCA and CV-SVM methods proposed in this paper can both effectively improve the accuracy of individual communication recognition.

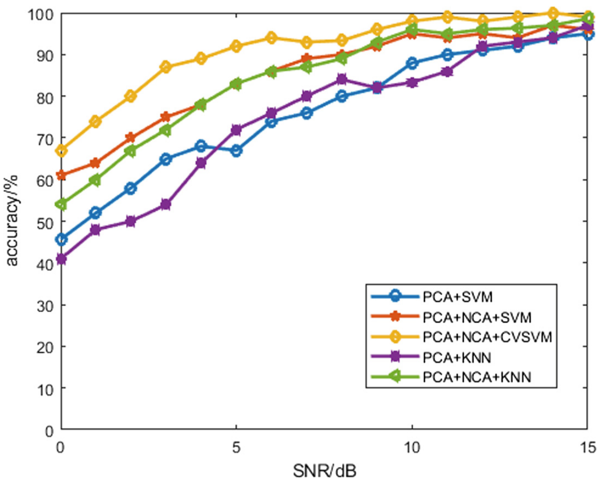
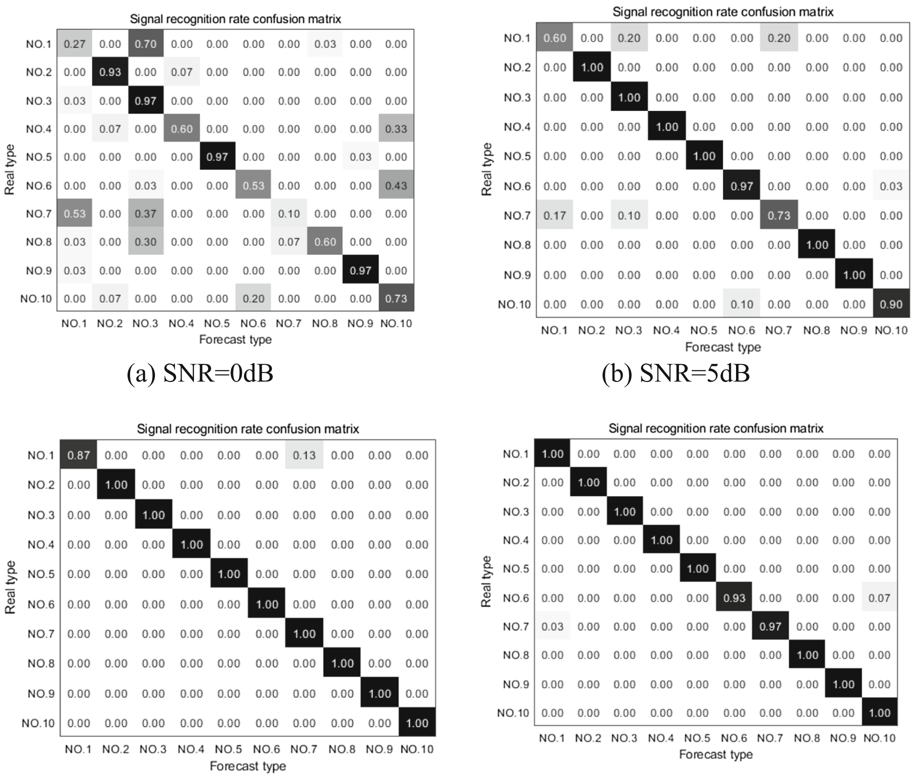


Fig. 1. Accuracy of five methods

Next is the confusion matrix of the recognition results of 'PCA+NCA+CVSVM' method under the SNR of 0, 5, 10 and 15 dB, respectively. It can be seen that when SNR = 0 dB, the recognition rate of devices NO.1, NO.4, NO.6, NO.7 and NO.8 does not exceed 70%, which is caused by the overlapping feature distribution between them, which makes it difficult to correctly classify the classifier. When SNR is greater than 5 dB, the feature distribution is gradually clear, and the location of different types of feature distribution is also gradually far away. Therefore, when the SNR increases, the classification accuracy of each device also increases gradually (Fig. 2).



(a) SNR=0dB (b) SNR=5dB

(c) SNR=10dB (d) SNR=15dB

Fig. 2. Classification result confusion matrix under four SNRs

4 Conclusion

In the field of radio frequency fingerprint identification, this paper proposes a feature selection method of PCA and NCA, and improves the SVM parameter selection method to select the optimal parameters of SVM by cross validation method [15]. In this paper, PCA is used to remove the correlation between features, and then NCA feature selection method is used to select the features that are most conducive to classification, so as to solve the problem of high feature dimension of power-on transient signal. Experimental results show that this algorithm can effectively reduce feature redundancy and improve recognition accuracy while reducing data dimensions.

Acknowledgment. The authors would like to thank State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System Director Fund (CEMEE2019K0104B).

References

1. Yu, L., Liu, H.: Efficient feature selection via analysis of relevance and redundancy. J. Mach. Learn. Res. **5**(12), 1205–1224 (2004)
2. Mao, K.Z.: Fast orthogonal forward selection algorithm for feature subset selection. IEEE Trans. Neural Netw. **13**(5), 1218–1224 (2002)
3. Hua-Liang, W., Billings, S.A.: Feature subset selection and ranking for data dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 162–166 (2007)
4. Liu, Y., Zheng, Y.F.: FS_SFS: a novel feature selection method for support vector machines. Pattern Recogn. **39**, 1333–1345 (2006)
5. Jain, A.K., Duin, R., Mao, J.C.: Statistical pattern recognition a review. IEEE Trans. Pattern Anal. Mach. Intell. **22**(1), 4–37 (2000)
6. Martin-Bautista, M.J., Vila, M.A.: A survey of genetic feature selection in mining issues. In: Proceeding of the 1999 Congress on Evolutionary Computation, pp. 13–23. IEEE Press (1999)
7. Goldberger, J., Roweis, S., Hinton, G., et al.: Neighbourhood components analysis. In: Proceedings of 2005 Conference on Neural Information Processing Systems. MIT Press (2005)
8. Singh-Miller, N., Collins, M., Hazen, T.J.: Dimensionality reduction for speech recognition using neighborhood components analysis. In: Proceedings of Interspeech 2007, Antwerp, Belgium (2007)
9. Butman, M., Goldberger, J.: Face recognition using classification based linear projections. EURASIP J. Adv. Signal Process. **2008**, 1–7 (2008)
10. Xing, E.P., Ng, A.Y., Jordan, M.I., et al.: Distance metric learning, with application to clustering with side-information. In: Proceedings of 2002 Conference on Neural Information Processing Systems. MIT Press, Cambridge (2002)
11. Globerson, A., Roweis, S.T.: Metric learning by collapsing classes. In: Proceedings of 2006 Conference on Neural Information Processing Systems. MIT Press, Cambridge (2006)
12. Liu, S., Bai, W., Liu, G., et al.: Parallel fractal compression method for big video data. Complexity **2018**, 2016976 (2018)
13. Zhang, X., Lu, X., Shi, Q., Xu, X.Q., Leung, H.C.E., Harris, L.N., et al.: Recursive SVM feature selection and sample classification for mass-spectrometry and microarray data. BMC Bioinform. **7**(1), 1–13 (2006)
14. Miao, L., Shuai, L., Weina, F., et al.: Distributional escape time algorithm based on generalized fractal sets in cloud environment. Chin. J. Electron. **24**(1), 124–127 (2015)
15. Lu, M., Liu, S., Sangaiah, A.K., et al.: Nucleosome positioning with fractal entropy increment of diversity in telemedicine. IEEE Access **6**, 33451–33459 (2018)