The Application of Improved Local Preserving Projection Dimension Reduction Method in Spare Parts Classification

Qiang Wang\textsuperscript{1,2,a}, Zhonghua Cheng\textsuperscript{1,*}, Qian Wang\textsuperscript{1,b}, Min Du\textsuperscript{1,c}, Yadong Wang\textsuperscript{1,d}

\textsuperscript{1}Department of Equipment Command and Administration, Army Engineering University, Shijiazhuang 050003, China
\textsuperscript{2}Army Military Transportation University, Tianjin, 300161, China

*Corresponding author e-mail: 511091065@qq.com, qw_up@foxmail.com, 18003131595@163.com, 734234341@qq.com, 1161741609@qq.com

Abstract. In order to classify spare parts by machine learning method, it is necessary to reduce the dimension of spare parts because of the large number of characteristic indexes. An improved dimensionality reduction method for local preserving projection is presented in this paper. The model optimizes the parameters in local preserving projection dimensionality reduction by using kernel function parameter estimation method. The accuracy and efficiency of dimension reduction classification model are improved. Finally, through an example analysis, the simulation results show that the proposed dimension reduction method can solve the problem of spare parts classification better. It also improves the accuracy and efficiency of traditional classification methods.

1. Introduction

With the development of modern industry, machinery and equipment are becoming larger and more complex. As an important material basis for maintenance and support of machinery and equipment, how to classify and select among the numerous spare parts is the key to reasonably plan and reserve spare parts. Classical spare parts classification problem is mainly through the construction of multi-attribute decision-making evaluation system, and then according to the comprehensive evaluation of spare parts to complete the classification of reserves \cite{1-4}. However, the classification method of multi-attribute evaluation will inevitably affect the accuracy of classification results because of the subjective decision of weight assignment \cite{5}. Compared with the traditional classification methods, which are subjective in weight calculation, machine learning classification can effectively avoid the influence of subjective decision on classification accuracy. Therefore, the use of machine learning methods for spare parts classification increasingly highlights its advantages.

However, with the continuous application and research of machine learning in classification, how to further improve the classification accuracy and efficiency has become the focus of machine learning classification methods. Among them, the classification speed is closely related to the number of data feature indicators. It is necessary to reduce the dimensionality of data to remove redundant information and improve the classification speed. Therefore, many scholars focus on the dimensionality reduction of feature indicators. Document \cite{6} considers the preference of spare parts attributes, uses dominance
rough set method to make logical judgment on feature indexes, and finally generates "if-then" decision rules for spare parts classification. In Document [7], principal component analysis method (PCA) was used to reduce dimensions of nuclear plant accident characteristics, which improved the efficiency of accident classification decision. However, the dimension reduction methods proposed in the above literature are all feature selection, that is, to select a subset of the original data feature indicators as the basis for decision. It will result in the loss of data features and the deviation of classification accuracy in forecasting classification. The dimension reduction method of Local Preservation Projection (LPP) can retain the original index characteristics to the greatest extent.

In view of this, a new dimension reduction method based on Kernel Density Estimation (KDE-LPP) is proposed in this paper. Firstly, the characteristic indexes of spare parts are obtained according to the influencing factors of spare parts reserve. Then the kernel density estimation method is used to adjust the parameters of the dimensionality reduction method adaptively to improve the classification accuracy. Finally, the data are input into the spare parts classifier to realize the rapid classification of spare parts. The reasonableness and practicability of the classification model are verified by simulating the spare parts of large-scale equipment in enterprises.

2. Problem Description

![Figure 1. The factors influencing the classified reserve of the spare parts.](image-url)
There are many characteristic indexes that affect spare parts reserve. If the spare parts are classified directly by machine learning method, not only will the calculation amount of classification be too large and the speed be slowed down, but also there may be information redundancy between the feature indexes, which will cause data over-fitting because of excessive dimension. Therefore, characteristic indexes of data should be screened before classification, and dimension reduction of spare parts data of characteristic indexes should be conducted. In addition, it can reduce the feature dimension of the data and effectively improve the speed of classification. It can be seen from the analysis in Figure 1 that there are many characteristic indexes affecting the classification of spare parts reserve, and the feature dimension reduction method can be used to improve the classification efficiency of spare parts.

3. Feature Index Dimensionality Reduction Method Based on Local Projection

3.1. The principle of local projection (LPP) dimensionality reduction method

Local Preserving Projection (LPP) is a linear approximation method of Laplace Mapping (LE) [8-9]. Let data sample \( X = (x_1, x_2, ..., x_n), x_i \in \mathbb{R}^M \) be the initial high-dimensional data set. The dimension of data is \( M \). The dimension reduction method of LPP is to use a mapping matrix \( A \) to transform high-dimensional data set \( X \) into low-dimensional data set \( Y = (y_1, y_2, ..., y_n), y_i \in \mathbb{R}^d \) through \( y_i = A^T \cdot x_i \). Where the dimension is \( d(d << M) \). Mapping matrix \( A \) can be obtained by satisfying the following objective equation.

\[
Z = \min_{i,j} \sum_{i,j} w_{ij} \left\| y_i - y_j \right\|^2
\]  

In the formula, \( w_{ij} \) is the element of the weight matrix \( W \), and \( w_{ij} \) represents the weight of the relationship between the original data point \( x_i \) and \( x_j \).

In the dimension reduction method of LPP, the key is to determine the weight matrix \( W \), and the element \( w_{ij} \) of the weight matrix is related to the location of the original data points. The specific solving steps are as follows:

1. Constructing adjacency graphs
   The distance formula is used to calculate the distance between the original data, and then the adjacent data points are connected to form a sparse adjacency graph. This paper chooses \( \epsilon \)-neighbourhood method as the nearest neighbor rule to judge the relationship between data.

2. Calculating the weights between data points
   The edge weights between two original data points are calculated to measure the similarity between two data points.
   \[
   w_{ij} = \begin{cases} 
   \exp\left(-\frac{\left\| x_i - x_j \right\|^2}{t}\right) & \left\| x_i - x_j \right\|^2 \leq \epsilon \\
   0 & \text{other}
   \end{cases}
   \]  
   In the formula, parameter \( t > 0 \) is the width of the core.

3. Computing eigenvalues
   Constraint \( YDY^T = I \) is introduced. Then, \( XAX^T = \lambda DAX^T \).
In the formula, \( \lambda \) is the generalized eigenvalue, and the column vector \( a_1, a_2, \ldots, a_d \) is the eigenvector corresponding to the first \( d \) minimum non-zero eigenvalues in \( \lambda \). Then the dimension of the mapping matrix \( A \) is \( D \times d \), and its expression is: \( A = \{a_1, a_2, \ldots, a_d\} \).

3.2. LPP dimension reduction method based on kernel density estimation

Formula (2) shows that when selecting neighborhood data points in high-dimensional data space, the neighborhood is determined by the value of parameter \( \varepsilon \). And directly affects the performance of dimension reduction methods. In view of this, this paper uses the Kernel Density Estimation (KDE) method [10] to estimate the sample probability density of the original data, and then adaptively adjusts the neighborhood parameter \( \varepsilon \) [11].

Let \( X = (x_1, x_2, \ldots, x_n), x_i \in \mathbb{R}^M \) be the original high-dimensional data set, then its kernel probability density estimation formula is:

\[
p(x_i) = \frac{1}{nV} \sum_{j=1}^{n} \varphi\left( \frac{x_i - x_j}{a} \right)
\]  

(3)

In the formula, \( \varphi() \) is a kernel function and \( \varphi(x) \geq 0, \int \varphi(x)dx = 1 \) satisfies it. In this paper, Gauss window function is selected. \( \varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{||x-x||^2}{2}\right) \), Then the probability density of data point \( x_i \) is estimated as:

\[
\hat{p}(x_i) = \frac{1}{n(2\pi)^{\frac{M}{2}}a^M} \sum_{j=1}^{n} \exp\left(-\frac{||x_i - x_j||^2}{2a^2}\right)
\]  

(4)

Let the initial neighbourhood parameter be \( \varepsilon_0 \), then the neighbourhood of the original sample point \( x_i \) with respect to \( \varepsilon_0 \) is \( N_{\varepsilon_0}(x_i) \). The number of nearest neighbour data of data point \( x_i \) is \( n(x_i) \). The window width \( a \) of the Gauss window is a variable to adjust the probability density. Then the original data points can adjust the neighbourhood parameters adaptively according to the global density distribution of high-dimensional data sets. Its parameter adjustment formula is as follows:

\[
\varepsilon(x_i) = \varepsilon_0 \frac{\hat{p}(x_i)}{\overline{p}}, \overline{p} = \frac{1}{n(x_i)} \sum_{i=1}^{n} \hat{p}(x_i)
\]  

(5)

3.3. Spare parts classification based on KDE-LPP dimension reduction method

The process of spare parts classification based on KDE-LPP dimensionality reduction method is shown in Figure 2. The steps are as follows:
Figure 2. Spare parts classification model based on KDE-LPP.

(1) According to the characteristic indexes of spare parts classification, data collection is conducted for spare parts information and the classification results of spare parts are recorded. Finally, all data are divided into training sample A and test sample B.

(2) For each spare parts sample, KDE-LPP dimension reduction method is used to extract feature, and the dimension of the index is reduced. When calculating the value of each reduced dimension $d(d=1,2,...,12)$, the corresponding mapping matrix $A$ and low-dimensional eigenvector $Y$, $y_i \in R^d$ are obtained.

(3) In step (2), when dimension $d$ takes different values, the low-dimensional eigenvector $Y$ is taken as a known quantity. Input to the classifier, repeat steps (2), and compare the maximum accuracy of spare parts classification with different dimension reduction $d$. Then the optimal accuracy $\eta^*$ of spare parts classification model is determined, and the corresponding optimal dimension $d^*$ is obtained. At this time, the data dimension is low and the speed of spare parts classification is fast.

(4) The optimal dimension $d^*$ obtained in step (3) is taken as input. Using KDE-LPP to reduce the dimension of test sample data, the dimension-reduced eigenvector $Y^*$ and the optimal dimension-reduced dimension $d^*$ are obtained. Finally, input the classifier to get the final result of spare parts classification.
3.4. Case Study and Analysis

In order to manufacture a product, an enterprise needs many kinds of equipment to cooperate with each other, and there are many kinds of spare parts needed by the equipment. This requires the maintenance personnel to select the appropriate classification method of spare parts. At the same time, the classification accuracy of spare parts is the best. Finally, according to the classification results of spare parts, the spare parts are classified and stored. According to the characteristics of spare parts, spare parts are divided into three categories: key reserve spare parts (A), conditional reserve spare parts (B), and non-reserve spare parts (C). Fifty sets of spare parts in each category were selected for verification and analysis. Thirty samples of each category were randomly selected as training samples, and another 20 samples were selected as test samples. The spare parts information is collected according to the characteristic indexes in Fig.1. The specific data are shown in table 1.

| number | index 1 | index 2 | index 3 | index 4 | index 5 | index 6 | index 7 | index 8 | index 9 | index 10 | index 11 | index 12 | classification |
|--------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------------|
| 1      | 1       | 0.95    | 0       | 1       | 38      | 2       | 2.3     | 1       | 1       | 5000     | 1       | 1       | A               |
| 2      | 1       | 0.92    | 0       | 1       | 30      | 5       | 1.5     | 1       | 1       | 18000    | 1       | 1       | A               |
| 3      | 2       | 0.68    | 0       | 7       | 26      | 136     | 0.9     | 4       | 1       | 600      | 0       | 3       | B               |
| ...    | ...     | ...     | ...     | ...     | ...     | ...     | ...     | ...     | ...     | ...      | ...     | ...     | ...             |

LPP algorithm is used to reduce the dimension of the feature data set of spare parts. In order to verify the validity of the KDE-LPP dimension reduction method proposed in this paper, the feature data of spare parts are analysed by LPP dimension reduction method and principal component analysis (PCA), respectively, as shown in Figure 3. The initial neighbourhood parameter $\epsilon_0 = 2.5$ and the thermonuclear width $a = 6$ are set in the LPP dimension reduction method. When using PCA dimension reduction method to find the optimal dimension, the dimension of PCA dimension reduction is determined according to the principle that the cumulative contribution rate of PCA principal component is not less than 95%.

In Figure 3, the curves show the advantages and disadvantages of three dimension reduction methods. At the initial stage, with the continuous increase of dimension reduction, the accuracy of spare parts classification improves gradually. When the accuracy of spare parts classification reaches the maximum, the optimal dimension reduction is reached. When the dimension continues to increase, the classification accuracy of spare parts decreases, but the curve fluctuation tends to be stable.

Compared with the three dimension reduction methods, the KDE-LPP dimension reduction method has the highest accuracy in the classification of spare parts. When dimension reduction is 6, the accuracy of spare parts classification reaches 0.92. LPP dimension reduction method is second, when dimension reduction is 8, the accuracy of spare parts classification is 0.88. This is because KDE-LPP adaptively adjusts the neighbourhood range according to the density of high-dimensional data points, which can reflect the distance relationship of data points more appropriately. In contrast, the neighbourhood parameters in the LPP dimension reduction method are fixed values, and the effect is not as good as the former. The accuracy of spare parts classification obtained by PCA dimension reduction method is the smallest. However, compared with the other two methods, the optimal dimension reduction method is the smallest. When the dimension reduction is 4, the classification accuracy of spare parts is 0.86. Because PCA dimension reduction method changes high-dimensional data into low-dimensional data through linear transformation, which makes the non-linear relationship between data missing when high-dimensional data is mapped to low-dimensional space. Combining the accuracy of spare parts classification, the dimension reduction method of KDE-LPP is the best for spare parts classification.
Figure 3. The relationship between classification accuracy of spare parts and dimension reduction.

4. Conclusion
When classifying spare parts by machine learning, the traditional method of dimensionality reduction does not reflect the position between data. In this paper, KDE-LPP dimension reduction method is used to deal with the feature indexes of spare parts. The method based on kernel density estimation can adjust the neighborhood parameters adaptively and ensure the popular structure between the original data to the greatest extent. According to the relationship between the classification accuracy of spare parts and the dimension reduction, the optimal dimension reduction is obtained and the classification of spare parts is finally realized. The reasonableness of dimension reduction method is verified by an example, and the cross comparison shows that the proposed method can improve the speed and accuracy of classification model operation compared with other methods.

Acknowledgments
This work was financially supported by National natural science foundation of China (No.71871219) and National social science foundation of China (No.16gj003-069).

References
[1] Roda I, Macchi M, Fumagalli L, et al. A review of multi-criteria classification of spare parts [J]. Journal of Manufacturing Technology Management, 2014, 25 (4): 1118 - 1124.
[2] Sarmah S P, Moharana U C. Multi-criteria classification of spare parts inventories – a web based approach [J]. Journal of Quality in Maintenance Engineering, 2015, 21 (4): 456 - 477.
[3] BU Xin-wang, ZHANG Yang, YANG Yi-fan. Research on the Variety and Quantity Optimization of Repair Maintenance Spare Parts Based on Comprehensive Fuzzy Evaluation Method [J]. Value Engineering, 2016, 35 (30): 125 - 128 (in Chinese).
[4] ZHOU Yi-ming, WANG Qian, RONG Peng-hui, et al. Determination Method of Spare Part Varieties for Air Ammunition Support Equipment Based on Grey Situation Decision-making [J]. Shipboard Electronic Countermeasure, 2017, 40 (05): 56 - 60 (in Chinese).
[5] Roda I, Macchi M, Fumagalli L, et al. On the classification of spare parts with a multi-criteria perspective [J]. Ifac Proceedings Volumes, 2012, 45 (31): 19 - 24.
[6] Hu Q, Chakhar S, Siraj S, et al. Spare parts classification in industrial manufacturing using the dominance-based rough set approach[J]. European Journal of Operational Research, 2017, 3 (262): 1136 - 1163.
[7] Prusty M R, Jayanthi T, Chakraborty J, et al. Feasibility of ANFIS towards multiclass event classification in PFBR considering dimensionality reduction using PCA [J]. Annals of Nuclear Energy, 2017, 99:311 - 320.

[8] Zhai Y, Zhang L, Wang N, et al. A Modified Locality-Preserving Projection Approach for Hyperspectral Image Classification [J]. IEEE Geoscience & Remote Sensing Letters, 2017, 13 (8): 1059 - 1063.

[9] Xu Y, Zhong A, Yang J, et al. LPP solution schemes for use with face recognition [J]. Pattern Recognition, 2010, 43 (12): 4165 - 4176.

[10] Tang X, Xu A. Multi-class classification using kernel density estimation on K-nearest neighbours [J]. Electronics Letters, 2016, 52 (8): 600 - 602.

[11] Abuela H, Ozdemir M K. Automatic Modulation Classification Based on Kernel Density Estimation [J]. Canadian Journal of Electrical & Computer Engineering, 2016, 39 (3): 203 - 209.