



# Unsupervised Feature Selection Based on Low-Rank Regularized Self-Representation

Wenyuan Li, Lai Wei\*

College of Information Engineering, Shanghai Maritime University, Shanghai, China

Email: lwy770783093@163.com, \*weilai@shmtu.edu.cn

**How to cite this paper:** Li, W.Y. and Wei, L. (2020) Unsupervised Feature Selection Based on Low-Rank Regularized Self-Representation. *Open Access Library Journal*, 7: e6274. <https://doi.org/10.4236/oalib.1106274>

**Received:** March 27, 2020

**Accepted:** April 14, 2020

**Published:** April 17, 2020

Copyright © 2020 by author(s) and Open Access Library Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Feature selection aims to find a set of features that are concise and have good generalization capabilities by removing redundant, uncorrelated, and noisy features. Recently, the regularized self-representation (RSR) method was proposed for unsupervised feature selection by minimizing the  $L_{2,1}$  norm of residual matrix and self-representation coefficient matrix. In this paper, we find that minimizing the  $L_{2,1}$  norm of the self-representation coefficient matrix cannot effectively extract the features with strong correlation. Therefore, by adding the minimum constraint on the kernel norm of the self-representation coefficient matrix, a new unsupervised feature selection method named low-rank regularized self-representation (LRRSR) is proposed, which can effectively discover the overall structure of the data. Experiments show that the proposed algorithm has better performance on clustering tasks than RSR and other related algorithms.

## Subject Areas

Computer Engineering

## Keywords

Feature Selection, Unsupervised Learning, Low-Rank Representation, Feature Correlation

## 1. Introduction

In recent years, the field of scientific research has been facing more and more high-dimensional data. The high-dimensionality of data has greatly increased the time and space complexity of data processing, and has made the performance of some algorithms for analyzing problems in low-dimensional spaces greatly declining, this is the so-called “dimensional disaster” [1] problem. The

method of extracting typical features from high-dimensional data was proposed to alleviate this problem [2] [3] [4], that is, to remove some irrelevant or redundant information from the original high-dimensional data, and find out the mapping relationship from complex data sources to low-dimensional space, and then use this relationship to extract the typical features in high-dimensional data, instead of the original data for subsequent processing [5] [6] [7] [8]. In recent years, many novel feature selection methods have been proposed successively, such as FSFS (feature selection using feature similarity) [9], LS (Laplacian score) [10], MRFS (feature selection via minimum redundancy) [11], RSR (regularized self-representation) [12] and so on.

Feature selection methods are mainly divided into three categories: Filter, Wrapper [13] and Embedding [14] [15] [16] [17] [18]. Among them, the initial features are filtered by the filter method firstly, and then the filtered features are used to train the model [19]. The Relief algorithm [20], which was first proposed by Kira, is a feature weighting algorithm, and it is one of the more well-known filter methods. In this algorithm, the weight of each feature is obtained through the correlation between the feature and the category, and the features with a weight higher than a certain threshold are selected as the feature selection result. For the wrapper method, it uses the clustering or classification performance of the learner as an evaluation criterion to select feature [14], so as to select a set of feature subsets that are most beneficial to the learner. Since the selected features need to be evaluated by a classifier, the time complexity of the wrapper method is much higher than that of the filter method. In addition, the performance of filter and wrapper methods are affected by the search strategy. The embedding method is currently the most widely used, which integrates the feature selection process and the learner training process. Both steps are completed in the same optimization process, that is, the feature selection is automatically performed during the learner training process [19]. The recently proposed RSR method is one of the better performances of the embedding methods.

The RSR method believes that any feature of the sample can be linearly represented by other relevant features, and the weight of the linear representation reflects the importance of each feature in describing the feature. That is to say, in the RSR method, if one feature is important, it will participate in the reconstruction representation of most other features, thereby generating a row of larger self-representation coefficients in the process of self-representation reconstruction [12], vice versa. For the entire attribute set, all self-representation coefficient vectors form a reconstructed coefficient matrix. The RSR method makes the matrix rows sparse by minimizing the  $L_{2,1}$  norm of the self-representation coefficient matrix, thereby facilitating the selection of features of higher importance. Experiments show that RSR has proved its effectiveness in comparison with other unsupervised feature selection methods.

Nevertheless, we find that only by minimizing the  $L_{2,1}$  norm of the self-representation coefficient matrix cannot effectively extract the features with

strong correlation. For example, if there are two important features with strong correlation, then the RSR method cannot effectively extract them simultaneously. In order to solve this problem, this paper proposes a new unsupervised feature selection method, that is, low-rank regularized self-representation (LRRSR), by adding a constraint on minimizing the kernel norm of the self-representation coefficient matrix, and gives the optimization of the LRRSR algorithm. Experiments on some UCI datasets and standard face datasets confirm the efficiency of the LRRSR in clustering tasks.

## 2. Regularized Self-Representation

As one of unsupervised feature selection methods, the regularized self-representation (RSR) method aims to select a set of representative features, which can effectively reconstruct the remaining features. The more important features can participate in the reconstruction of most other features and correspondingly generate self-representation coefficients with larger values. Therefore, the RSR method selects the more important features by finding the row vectors with larger self-representation coefficients.

Let  $X \in \mathbb{R}^{n \times m}$  be a data matrix containing  $n$  samples, each of which has  $m$  features. Let  $X = XW + E$ , where  $W$  is a self-representation matrix and  $E$  is a residual matrix. RSR selects features by minimizing the  $\|E\|_{2,1}$  and the  $\|W\|_{2,1}$ , which is transformed into the following formula:

$$\hat{W} = \arg \min_W \|E\|_{2,1} + \lambda \|W\|_{2,1} \quad \text{s.t., } X = XW + E. \quad (1)$$

It can be seen from the formula that RSR learns the optimal  $W$  by solving this minimization problem.

Among them, for the residual matrix  $E$ , the common way to minimize the residual is to make the  $F$  norm of the residual term as small as possible, that is, to minimize  $\|E\|_F$ . However, considering that there may be outliers in the sample, and  $F$  norm is sensitive to outliers. Therefore, RSR minimizes the  $L_{2,1}$  norm of  $E$  to make it line sparsity, that is, minimizes  $\|E\|_{2,1}$ , so as to both minimize the reconstruction error in the feature selection process and enhance the robustness to abnormal attributes. For the self-representation coefficient matrix  $W$ ,  $W = [w_1; w_2; \dots; w_m]$ , the size of  $w_i$  can reflect the importance of the  $i^{\text{th}}$  feature of the data. The RSR uses  $\|w_i\|_{2,1}$  as a weight to measure the importance of the  $i^{\text{th}}$  feature, and selects the features with higher importance by minimizing  $\|w\|_{2,1}$ .

## 3. Low-Rank Regularized Self-Representation

As mentioned in the previous chapter, the RSR algorithm can learn the more important features in the data. However, the data in nature often contains a lot of redundant information and the correlation between the data is relatively strong. The RSR ignores the correlation between the features of the data when minimizing the residual. Therefore, on the basis of the RSR model, a low-rank constraint on the self-representation coefficient matrix is added in this paper to

find the correlation between the sample data, so as to more effectively find the overall structure information of the sample.

Given a data matrix  $X$ ,  $X = [x_1; x_2; \dots; x_n] \in \mathbb{R}^{n \times m}$ , that is, the data set has  $n$  samples in total, and each sample has  $m$  features. The low-rank regularized self-representation (LRRSR) model can be described as the following minimization problem:

$$\hat{Z} = \arg \min_Z \|E\|_{2,1} + \lambda \|Z\|_{2,1} + \beta \|Z\|_* \quad \text{s.t., } X = XZ + E. \quad (2)$$

Among them,  $Z \in \mathbb{R}^{m \times m}$  is the self-representing coefficient matrix of the data,  $Z = [z_1; z_2; \dots; z_m]$ ,  $z_i$  represents the  $i^{\text{th}}$  self-representation vector of  $Z$ ;  $E$  represents the residual matrix,  $E = X - XZ = [e_1; e_2; \dots; e_n]$ ,  $e_i \in \mathbb{R}^{1 \times m}$  is the residual vector of the  $i^{\text{th}}$  sample;  $\|\cdot\|_{2,1}$  represents the  $L_{2,1}$  norm,  $\|\cdot\|_*$  represents the kernel norm. Here we find the whole structure of the feature space by minimizing the kernel norm.  $\lambda$  and  $\beta$  are empirical parameters.

As shown in the above formula (2), solving the optimal self-representation coefficient matrix problem is transformed into solving the problem of minimizing the sum of residual and coefficient regularization constraints. Among them, minimizing  $\|E\|_{2,1}$  makes the residuals as small as possible when training  $Z$ ; the value of  $\|z_i\|_2$  can directly reflect the importance of the  $i^{\text{th}}$  feature in the self-representation process. Therefore, minimizing  $\|Z\|_{2,1}$  can effectively extract important features; the rank of a matrix is a measure of the global structure of the matrix and contains the overall structure information of the matrix. Therefore, by adding a low rank constraint to the self-expression coefficient matrix  $Z$ , that is,  $\|Z\|_*$ , we can better find the overall structure information of the data and the correlation between its features, so as to effectively extract the features with strong correlation.

#### 4. Algorithm Optimization and Solution

In this section, we optimize and solve the LRRSR. First, we first convert formula (2) into the following equivalence problem:

$$\min_{Z, E, W, J} \|E\|_{2,1} + \lambda \|W\|_{2,1} + \beta \|J\|_* \quad \text{s.t., } X = XZ + E, Z = J, Z = W. \quad (3)$$

Then, through the augmented Lagrange method (ALM), formula (3) can be transformed into the following problem:

$$\begin{aligned} \min_{Z, E, W, J, Y_1, Y_2, Y_3} & \|E\|_{2,1} + \lambda \|W\|_{2,1} + \beta \|J\|_* + \text{tr}[Y_1^T (X - XZ - E)] \\ & + \text{tr}[Y_2^T (Z - J)] + \text{tr}[Y_3^T (Z - W)] \\ & + \frac{\mu}{2} (\|X - XZ - E\|_F^2 + \|Z - J\|_F^2 + \|Z - W\|_F^2). \end{aligned} \quad (4)$$

Among them,  $\mu > 0$  is the penalty parameter, and  $Y_1, Y_2, Y_3$  are Lagrange multipliers. The above unknown terms can be solved by iterative updating. During the iteration process,  $J, W, Z, E$  can in turn update a single unknown term by fixing other variables. The specific update process can be summarized as follows:

1) Fix other unknowns and update  $J$ . Ignoring the terms unrelated to  $J$  in formula (4), then at the  $(k + 1)^{\text{th}}$  iteration,  $J$  can be expressed as

$$\begin{aligned} J_{k+1} &= \arg \min_{J_k} \beta \|J_k\|_* + \langle Y_2^k, Z_k - J_k \rangle + \frac{\mu_k}{2} \left( \|Z_k - J_k\|_F^2 \right) \\ &= \arg \min_{J_k} \frac{\beta}{\mu_k} \|J_k\|_* + \frac{1}{2} \left\| J_k - \left( Z_k + \frac{Y_2^k}{\mu_k} \right) \right\|_F^2. \end{aligned} \tag{5}$$

Then,  $J_{k+1} = U_k \Theta_{1/\mu_k} (S_k) V_k^T$ , where the  $U_k (S_k) V_k^T$  is a singular value decomposition matrix of the  $(Z_k + Y_2^k / \mu_k)$ , and the  $\Theta$  is a singularity value threshold [21].

2) Fix other unknowns and update  $W$ . Ignoring the terms unrelated to  $W$  in formula (4), then at the  $(k + 1)^{\text{th}}$  iteration,  $W$  can be expressed as

$$\begin{aligned} W_{k+1} &= \arg \min_{W_k} \lambda \|W_k\|_{2,1} + \langle Y_3^k, Z_k - W_k \rangle + \frac{\mu_k}{2} \left( \|Z_k - W_k\|_F^2 \right) \\ &= \arg \min_{W_k} \frac{\lambda}{\mu_k} \|W_k\|_{2,1} + \frac{1}{2} \left\| W_k - \left( Z_k + \frac{Y_3^k}{\mu_k} \right) \right\|_F^2. \end{aligned} \tag{6}$$

Here we can solve the above problem by the following Theorem 1 [22].

**Theorem 1** Suppose for a given matrix  $Q$ ,  $Q = [q_1, q_2, \dots, q_i, \dots]$ , there are the following problems:

$$\hat{W} = \arg \min_W \lambda \|W\|_{2,1} + \frac{1}{2} \|W - Q\|_F^2. \tag{7}$$

Then the  $i^{\text{th}}$  column of the optimal solution  $W$  can be expressed as:

$$W(:, i) = \begin{cases} \frac{\|q_i\|_2 - \lambda}{\|q_i\|_2} q_i, & \text{if } \lambda < \|q_i\|_2, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

3) Fix other unknowns and update  $Z$ . Ignoring the terms unrelated to  $Z$  in formula (4), then at the  $(k + 1)^{\text{th}}$  iteration,  $Z$  can be expressed as

$$\begin{aligned} Z_{k+1} &= \arg \min_{Z_k} \langle Y_1^k, X - XZ_k - E_k \rangle + \langle Y_2^k, Z_k - J_{k+1} \rangle + \langle Y_3^k, Z_k - W_{k+1} \rangle \\ &\quad + \frac{\mu_k}{2} \left( \|X - XZ_k - E_k\|_F^2 + \|Z_k - J_{k+1}\|_F^2 + \|Z_k - W_{k+1}\|_F^2 \right) \\ &= \arg \min_{Z_k} \left\| X - XZ_k - E_k + \frac{Y_1^k}{\mu_k} \right\|_F^2 + \left\| J_{k+1} - \left( Z_k + \frac{Y_2^k}{\mu_k} \right) \right\|_F^2 \\ &\quad + \left\| W_{k+1} - \left( Z_k + \frac{Y_3^k}{\mu_k} \right) \right\|_F^2. \end{aligned} \tag{9}$$

Then,

$$Z_{k+1} = \left( 2I + X^T X \right)^{-1} \left( X^T X - X^T E_k + X^T \frac{Y_1^k}{\mu_k} + J_{k+1} - \frac{Y_2^k}{\mu_k} + W_{k+1} - \frac{Y_3^k}{\mu_k} \right). \tag{10}$$

4) Fix other unknowns and update  $E$ . Ignoring the terms unrelated to  $E$  in formula (4), then at the  $(k + 1)^{\text{th}}$  iteration,  $E$  can be expressed as

$$\begin{aligned}
E_{k+1} &= \arg \min_{E_k} \|E_k\|_{2,1} + \langle Y_1^k, X - XZ_{k+1} - E_k \rangle + \frac{\mu_k}{2} \left( \|X - XZ_{k+1} - E_k\|_F^2 \right) \\
&= \arg \min_{E_k} \frac{1}{\mu_k} \|E_k\|_{2,1} + \frac{1}{2} \left\| E_k - \left( X - XZ_{k+1} + \frac{Y_1^k}{\mu_k} \right) \right\|_F^2.
\end{aligned} \tag{11}$$

Consistent with the solution formula (6), the above problem is solved by Theorem 1.

5) Fixed other unknowns and updated parameters. The parameter update in the  $(k+1)^{\text{th}}$  iteration is summarized as follows:

$$\begin{cases}
Y_1^{k+1} = Y_1^k + \mu_k (X - XZ_{k+1} - E_{k+1}) \\
Y_2^{k+1} = Y_2^k + \mu_k (Z_{k+1} - J_{k+1}) \\
Y_3^{k+1} = Y_3^k + \mu_k (Z_{k+1} - W_{k+1}) \\
\mu_{k+1} = \min(\mu_{\max}, \rho\mu_k)
\end{cases} \tag{12}$$

Here  $\mu_{\max}$  and  $\rho$  are two positive parameters, and  $k$  represents the number of iterations.

We summarize the algorithm process of LRRSR. Algorithm 5( )@ describes the complete LRRSR algorithm process.

---

**Algorithm 1** low-rank regularized self-representation (LRRSR)

---

**Input:** data matrix  $X \in \mathbb{R}^{n \times m}$ , parameters  $\lambda$  and  $\beta$ , maximum iteration  $MaxIter$ .

**Initial:**  $J_0 = Z_0 = W_0 = E_0 = 0$ ,  $\mu_0 = 10^{-4}$ ,  $\mu_{\max} = 10^{30}$ ,  $\rho = 1.1$ ,  $Y_1^0 = Y_2^0 = Y_3^0 = 0$ ,  $\varepsilon = 10^{-6}$ ,  $k = 0$ .

1: **while**  $\|X - XZ - E\|_{\infty} > \varepsilon$ ,  $\|Z - J\|_{\infty} > \varepsilon$ ,  $\|Z - W\|_{\infty} > \varepsilon$ ,  $k < MaxIter$  **do**

2: Update  $J_{k+1}$  by solving formula (5);

3: Update  $W_{k+1}$  by solving formula (6);

4: Update  $Z_{k+1}$  by solving formula (10);

5: Update  $E_{k+1}$  by solving formula (11);

6: Update  $Y_1^{k+1}$ ,  $Y_2^{k+1}$ ,  $Y_3^{k+1}$ ,  $\mu_{k+1}$  by solving formula (12);

7:  $k = k + 1$ ;

8: **return**  $Z_* = Z_k$ ,  $E_* = E_k$ .

**Output:** self-representing coefficient matrix  $Z_*$ , residual matrix  $E_*$ .

---

## 5. Experiments

We perform experiments on two UCI datasets and face datasets to compare the performance of the proposed LRRSR model with other unsupervised feature selection methods in clustering tasks. In our comparison experiment, since the algorithms LS, UDFS, and MCFS need to construct the nearest neighbor graph, the number of nearest neighbors  $k$  is set as 5 according to experience.

### 5.1. Experiments on the UCI Datasets

We compare the performance of the proposed LRRSR and RSR in the clustering task by using the zoo data set and the dermatology data set. The data statistics of

the two data sets are shown in **Table 1**.

In the comparison experiment of the UCI datasets, we compared the clustering accuracy of the two methods with different features. In order to better show the performance difference between the two methods, we select (5,10) and (2,15) from all the experimental results as the feature quantity interval of Zoo dataset and Dermatology dataset, respectively. The experimental results are shown in **Figure 1**. It can be seen from the figure that the clustering accuracy of LRRSR is higher than RSR in most cases. Therefore, the LRRSR proposed in this paper performs better than the RSR model in clustering tasks.

## 5.2. Experiments on the Face Datasets

### 5.2.1. Introduction to Experimental Datasets

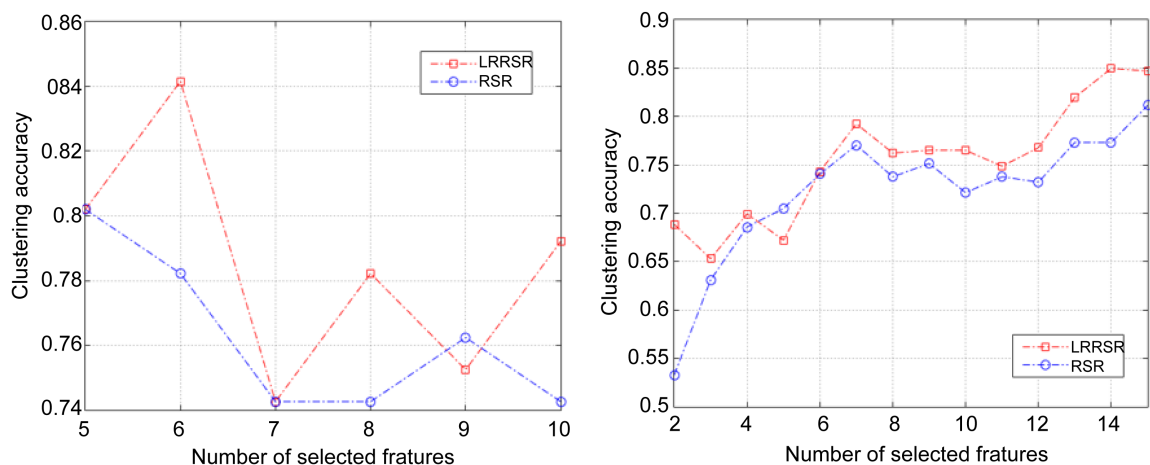
In this section, we use three commonly used standard face databases for experiments, namely ORL face database, YALEB face database and AR face database.

The ORL face database contains 400 face images taken under different expressions, lighting and shooting angles, including 40 people and 10 images for each person. We used the first 100 images, including 10 people, 10 images for each person, and normalized the images to  $32 \times 32$ . Some samples are shown in **Figure 2**.

The YALEB face database contains 38 people, 64 face images for each person. These images are also produced under different lighting and different fixed shooting angles. We selected 10 people's images for experiments. Each of them

**Table 1.** Some attributes of the zoo dataset and the dermatology dataset.

Dataset	Number of samples	Number of features	Number of categories
<i>Zoo</i>	101	16	7
<i>Dermatology</i>	366	33	6



**Figure 1.** Experimental results of two different methods on the zoo dataset (Left) and the dermatology dataset (Right).



**Figure 2.** Some examples of the images of a class in the ORL face database.

selected 6 images at the same shooting angle and normalized the images to  $32 \times 32$ . Some samples are shown in **Figure 3**.

The AR face database contains 14 people and 100 images for each person. We used a total of 140 images of the first 10 people for experiments and normalized the images to  $60 \times 43$ . Some samples are shown in **Figure 4**.

### 5.2.2. Experimental Analysis and Results

We compare the feature weight maps learned by the proposed LRRSR and RSR, as shown in **Figure 5**. As can be seen from the figure, the feature weight map learned by LRRSR is smoother and more similar to real face features than RSR. The reason is that LRRSR analyzes the correlation between sample data based on RSR, which can better extract the overall structure information. This result also verifies the effectiveness of our proposed algorithm.

We compare the proposed LRRSR method with the feature weight map learned by the RSR method, as shown in **Figure 5**. It can be seen that the feature weight maps learned by the LRRSR method are smoother than the RSR method is more similar to real face features, because the LRRSR method analyzes the correlation between sample data based on the RSR method, which can better extract the overall structure information, which also verifies the effectiveness of our proposed algorithm.

In the comparison experiment of the face databases, we use RSR, LS, UDFS, MCFS and other algorithms for comparison. The experiments compare the clustering accuracy rates of different algorithms in different feature dimensions. We set the number of selected features to  $\{60, 70, 80, 90, 100, 110\}$ . Among them, LS and MCFS need to build a neighbor graph, and we set the number of neighbors to 5; for the algorithm RSR, we set the parameter  $\lambda = 10$ ; For the proposed algorithm LRRSR, in the experiments of the ORL face database and the YALEB face database, the parameters are set to  $\lambda = 5$  and  $\beta = 1$ , and the comparison experimental results are plotted in **Figure 6** and **Figure 7**, respectively. In the experiment of the AR face database,  $\lambda = 11$  and  $\beta = 1$  are used as experimental parameters, and the comparison experimental results are shown in **Figure 8**. It can be seen from the experimental comparison experiments of three different face databases that the method LRRSR has a higher clustering accuracy rate in most cases than the other comparison algorithms. Among them, LS and MCFS are designed to preserve the similarity of the original feature space as much as possible; UDFS introduces discriminative information; MCFS, RSR and LRRSR use regression models for feature selection. Compared with other algorithms, the method LRRSR retains similarity information in the original feature space, and also explores the overall structure of the data and the correlation between data features as much as possible. Therefore, features that can better describe the

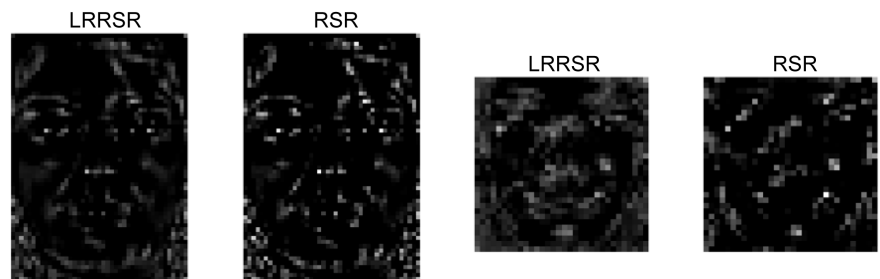




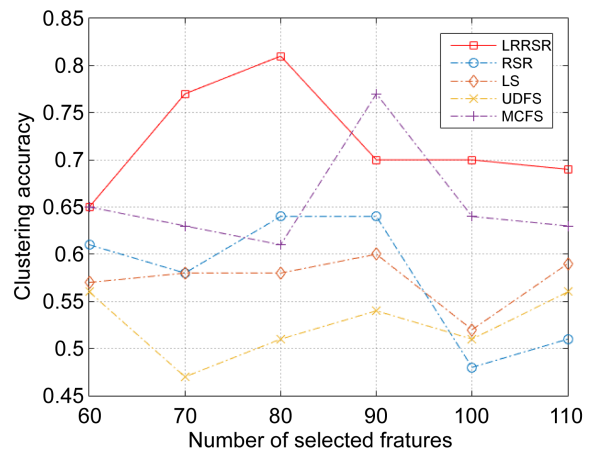
**Figure 3.** Some examples of the images of a class in the YALEB face database.



**Figure 4.** Some examples of the images of a class in the AR face database.

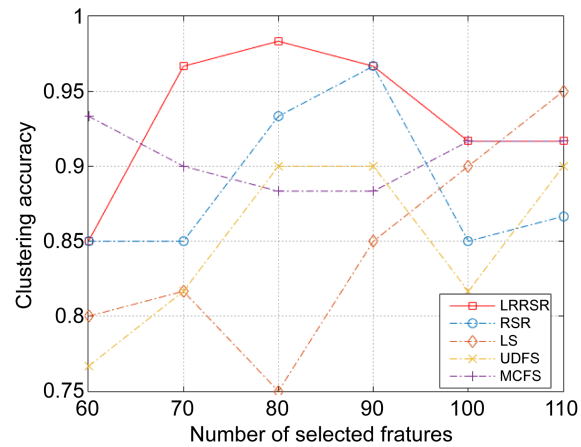


**Figure 5.** Feature weight maps learned under the AR face database (two pictures on the left) and the ORL face database (two pictures on the right).

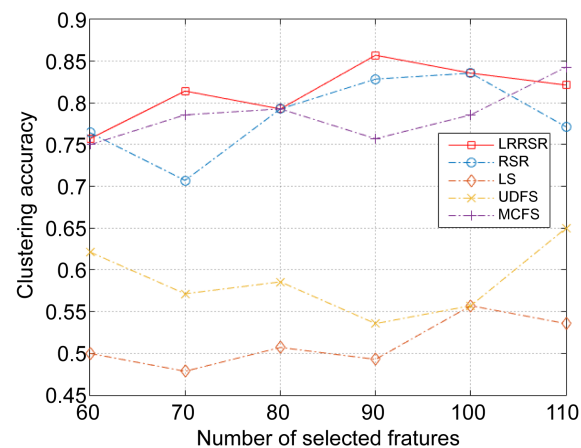


**Figure 6.** Experimental results of various methods on the ORL face database.

original data can be extracted. Experimental results can prove that the features learned by the proposed LRRSR are more effective in clustering tasks.



**Figure 7.** Experimental results of various methods on the YALEB face database.



**Figure 8.** Experimental results of various methods on the AR face database.

## 6. Conclusion

In this paper, we propose a new feature selection method, that is, the unsupervised feature selection method guided by low-rank regularized self-representation (LRRSR). The regularized self-representation (RSR) selects features by minimizing the  $L_{2,1}$  norm of the residual matrix and the self-representation coefficient matrix. Since the minimization of  $L_{2,1}$  norm of self-representation coefficient matrix cannot effectively extract features with strong correlation, LRRSR adds a low-rank constraint to self-representation coefficient matrix to better extract the correlation information of the data. In LRRSR, by minimizing the  $L_{2,1}$  norm of the residual matrix, the residual matrix is minimized and robust to outliers; minimizing the  $L_{2,1}$  norm of the self-representation coefficient matrix makes the rows of the self-representation coefficient matrix sparse, so as to better extract important features; at the same time, minimizing the kernel norm of the self-representation coefficient matrix can discover the overall structure of the sample data and the correlation information among the data features, which can

effectively extract the features with strong correlation. Experiments show that the LRRSR has better performance in clustering tasks than RSR and other related algorithms.

### Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

### References

- [1] Bishop, C.M. (2006) Pattern Recognition and Machine Learning. Springer, Berlin, 738.
- [2] Zhu, X., Zhang, S., Jin, Z., Zhang, Z. and Xu, Z. (2011) Missing Value Estimation for Mixed-Attribute Data Sets. *IEEE Transactions on Knowledge and Data Engineering*, **23**, 110-121. <https://doi.org/10.1109/TKDE.2010.99>
- [3] Zheng, M., Bu, J., Chen, C., Wang, C., Zhang, L., Qiu, G. and Cai, D. (2011) Graph Regularized Sparse Coding for Image Representation. *IEEE Transactions on Image Processing*, **20**, 1327-1336. <https://doi.org/10.1109/TIP.2010.2090535>
- [4] Ma, Z., Yang, Y., Sebe, N. and Hauptmann, A.G. (2014) Knowledge Adaptation with Partially Shared Features for Event Detection Using Few Exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36**, 1789-1802. <https://doi.org/10.1109/TPAMI.2014.2306419>
- [5] Cai, D., Zhang, C. and He, X. (2010) Unsupervised Feature Selection for Multi-Cluster Data. In: *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD Press, Washington DC, 333-342. <https://doi.org/10.1145/1835804.1835848>
- [6] Gao, S., Tsang, I.W.-H., Chia, L.-T. and Zhao, P. (2010) Local Features Are Not Lonely—Laplacian Sparse Coding for Image Classification. In: *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition*, IEEE Press, San Francisco, 3555-3561. <https://doi.org/10.1109/CVPR.2010.5539943>
- [7] Yang, Y., Zhuang, Y.-T., Wu, F. and Pan, Y.-H. (2008) Harmonizing Hierarchical Manifolds for Multimedia Document Semantics Understanding and Cross-Media Retrieval. *IEEE Transactions on Multimedia*, **10**, 437-446. <https://doi.org/10.1109/TMM.2008.917359>
- [8] Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R.P., Tang, J. and Liu, H. (2018) Feature Selection: A Data Perspective. *ACM Computing Surveys*, **50**, 94. <https://doi.org/10.1145/3136625>
- [9] Mitra, P., Murthy, C.A. and Pal, S.K. (2002) Unsupervised Feature Selection Using Feature Similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**, 301-312. <https://doi.org/10.1109/34.990133>
- [10] He, X., Cai, D. and Niyogi, P. (2005) Laplacian Score for Feature Selection. In: *International Conference on Neural Information Processing Systems*, MIT Press, Cambridge, 507-514.
- [11] Zhao, Z., Wang, L. and Liu, H. (2010) Efficient Spectral Feature Selection with Minimum Redundancy. *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*, AAAI Press, Atlanta, 673-678.
- [12] Zhu, P., Zuo, W., Zhang, L., Hu, Q. and Shiu, S.C.K. (2015) Unsupervised Feature Selection by Regularized Self-Representation. *Pattern Recognition*, **48**, 438-446.

<https://doi.org/10.1016/j.patcog.2014.08.006>

- [13] Kohavi, R. and John, G.H. (1997) Wrappers for Feature Subset Selection. *Artificial Intelligence*, **97**, 273-324. [https://doi.org/10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)
- [14] Kotsiantis, S.B. (2011) Feature Selection for Machine Learning Classification Problems: A Recent Overview. *Artificial Intelligence Review*, **42**, 1-20. <https://doi.org/10.1007/s10462-011-9230-1>
- [15] Zhang, S., Zhou, H., Jiang, F. and Li, X. (2015) Robust Visual Tracking Using Structurally Random Projection and Weighted Least Squares. *IEEE Transactions on Circuits and Systems for Video Technology*, **25**, 1749-1760. <https://doi.org/10.1109/TCSVT.2015.2406194>
- [16] Wang, W., Cai, D., Wang, L., Huang, Q., Xu, X. and Li, X. (2016) Synthesized Computational Aesthetic Evaluation of Photos. *Neurocomputing*, **172**, 244-252. <https://doi.org/10.1016/j.neucom.2014.12.106>
- [17] Abualigah, L.M. and Khader, A.T. (2017) Unsupervised Text Feature Selection Technique Based on Hybrid Particle Swarm Optimization Algorithm with Genetic Operators for the Text Clustering. *The Journal of Supercomputing*, **73**, 4773-4795. <https://doi.org/10.1007/s11227-017-2046-2>
- [18] Han, K., Wang, Y., Zhang, C., Li, C. and Xu, C. (2018) Autoencoder Inspired Unsupervised Feature Selection. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, 15-20 April 2018, 2941-2945. <https://doi.org/10.1109/ICASSP.2018.8462261>
- [19] Zhou, Z. (2016) *Machine Learning*. Tsinghua University Press, Beijing, 247-263.
- [20] Kira, K. and Rendell, L.A. (1992) The Feature Selection Problem: Traditional Methods and a New Algorithm. In: *Proceedings of the 10th National Conference on Artificial Intelligence*, AAAI Press, Atlanta, 129-134.
- [21] Cai, J.-F., Candes, E.J. and Shen, Z. (2010) A Singular Value Thresholding Algorithm for Matrix Completion. *SIAM Journal on Optimization*, **20**, 1956-1982. <https://doi.org/10.1137/080738970>
- [22] Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y. and Ma, Y. (2012) Robust Recovery of Subspace Structures by Low-Rank Representation. *IEEE Transactions on Software Engineering*, **35**, 171-184. <https://doi.org/10.1109/TPAMI.2012.88>