Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 18, No. 2, 2025

Entropy-Constrained and Sparse Relation-Constrained Subspace Clustering

Yangmeng Wang [1], Haoming Chen [1], Peng Zhang [2, *]

[1] School of Mathematics and Physics, Southwest University of Science and Technology, Mianyang, Sichuan, China
[2] School of Computer Science and Technology, Southwest University of Science and Technology, Mianyang, Sichuan, China
[*] **Corresponding author**: Peng Zhang (Email: 940606758@qq.com)

Abstract: This paper proposes a novel Entropy-Constrained and Sparse Relation-Constrained Subspace Clustering (ECSSC). By integrating information entropy weighting with sparse representation, the proposed method enhances both the accuracy and robustness of high-dimensional data clustering. Key improvements include: the introduction of an information entropy weight matrix to quantify feature discriminability and improve adaptability to noise and redundant features; the use of Frobenius norm constraints on the coefficient matrix to balance computational efficiency and model performance; and the incorporation of block-diagonal constraints to refine the sparsity structure of subspaces. Experiments conducted on three image datasets—MNIST, ORL, and COIL20—demonstrate that EWSSC outperforms traditional methods, classical sparse subspace clustering algorithms, and related state-of-the-art variants in terms of clustering accuracy (ACC), normalized mutual information (NMI), and adjusted Rand index (ARI). Notably, on the ORL dataset, ECSSC achieved an NMI of 0.9031, representing a 6.18% improvement over Sparse Representation-based Clustering (SRR). Ablation studies further confirm the effectiveness of the entropy-weighting module, which contributes to an average performance gain of 10%–18% across different datasets.

Keywords: SSC; Information Entropy; Feature Weighting; Relation-Constraint.

1. Introduction

With the rapid advancement of technology, explosive growth of data has become a common phenomenon in contemporary society. Massive amounts of information are being accumulated across various industries, ranging from real-time data generated by sensors to user behavior trajectories on social media—all flooding into real-world applications at an unprecedented pace and scale. However, this proliferation of information also presents a critical challenge: high-dimensional data. In traditional data analysis, low-dimensional data have often been preferred for modeling and analysis due to the formidable difficulties posed by high-dimensional settings [1].

The challenges of high-dimensional data are mainly manifested in several aspects. Firstly, the sparsity of data makes it difficult for many traditional clustering algorithms to effectively distinguish different data clusters in high-dimensional Spaces. Secondly, high-dimensional data is usually accompanied by a large amount of noise and redundant information, which reduces the performance of traditional methods and makes it difficult to extract truly meaningful patterns [2]. Subspace clustering emerged as a result. By clustering and classifying high-dimensional data, it achieves the purpose of separating subspaces. With its efficient clustering performance, it is widely applied in fields such as computer vision [3], speech signal processing [4], and unmanned driving [5], and has thus become the cornerstone of research hotspots.

Subspace clustering can be roughly divided into sparse subspace clustering (SSC) [6,7], low-rank subspace clustering (LRR) [8,9], and least squares regression subspace clustering (LSR) [10]. All these three methods involve modeling and linear representation of high-dimensional data in subspaces, assuming that data points can be represented by linear combinations of other data points, and inferring the underlying subspace structure from the data to better understand the intrinsic features of the data.

The difference lies in that SSC is dedicated to finding subspaces with shared similar structures in high-dimensional data. Its core idea is based on the sparse representation among data points, assuming that each data point can be linearly represented by others, which can reveal the underlying subspace structure of the data. LRR focuses on handling noise and outliers by representing data in a low-rank subspace, thereby enhancing the robustness of the data. It achieves this by optimizing the low-rank representation while considering the relationships among data points, thus better adapting to the noise and anomalies present in the real world. LSR achieves clustering by minimizing the projection error of data points onto subspaces. By establishing the connection between data points and subspaces through a regression model, it stands out in terms of mathematical simplicity and intuitiveness.

This paper has been organized and improved based on SSC, drawing on the advantages of various sparse subspace clustering methods. Patel et al. applied the kernel function to the data matrix and utilized the alternating direction method of multipliers (ADMM) to solve the model [11], providing a new solution approach for subspace clustering methods. Against this backdrop, we further explored the application of orthogonal matching pursuit (OMP) as an optimization algorithm to find the linear combination that best represents the original data points [12]. Compared with other sparse representation methods used in SSC, OMP can more effectively select important subspace members, thereby enhancing the clustering performance.

To better model the structure of the subspace, Li et al. introduced a regularization term to construct a structural sparse subspace clustering model (SSSC) [13]. In the research of subspace clustering, Haiyang Li et al. introduced the TL_1 norm constraint and theoretically proved that its optimal solution is a coefficient matrix with a block diagonal structure [14]. Lu et al. implemented the LSR algorithm again based on the affinity matrix obtained from LSR to remove noise, making the affinity matrix cleaner and more reliable [15]. At the same time, the augmented Lagrangian multiplier method (ALM) was used to solve the objective function, further improving the robustness of sparse subspace clustering.

To further explore the intrinsic structure of the data, WEI et al. performed second-order self-representation on the original data to obtain the reconstruction coefficient matrix, and introduced a new regularization term to construct the model (SRR) [16], which was used to construct the similarity matrix and perform clustering. Huizhi Hu et al. based on the SRR model to construct a new regularization term to enhance the sparsity and discriminability of the representations of different types of data [17].

Overall, the improvements to the subspace clustering algorithm mainly focus on the weighting or reconstruction of the coefficient matrix, the addition of new regularization terms, the improvement of norm constraints or the solution algorithm. In this paper, a weight matrix is introduced based on the structured sparse subspace [18]. Considering that in practical applications, different variables may have different importance for the clustering task, if they are not distinguished, suboptimal results will be obtained. In addition, there may be noisy variables in the data set, which will interfere with the accuracy of the clustering results. If the weight of these variables can be reduced, it will also have a certain gain on the clustering effect.

In this improvement direction, the GSSC method proposed by Tao Li et al. emphasizes the importance of sparse constraints through the weighting of Gaussian similarity [19], making the data more inclined to be linearly represented by the most similar data, and thus being independent of the dissimilar data. This idea of weighted sparse subspace clustering has brought a significant improvement in clustering performance. The 3DF-SSC method similar to GSSC was proposed by ZHANG et al. [20], by integrating the three-dimensional spectral features weights of hyperspectral images into the sparse subspace model, further expanding the application fields. The SWSSC algorithm proposed by Wenzhou Li et al. uses the Euclidean clustering of pixels to calculate the similarity matrix and determine the weight matrix [21]. However, due to the high dimensionality of the data, the weight effect based on Euclidean clustering may be limited. Yonghong Long et al. introduced an information entropy-based weighted matrix as the penalty term of the sparse coefficient matrix according to the characteristics of pixel data in hyperspectral remote sensing images [22], promoting the linear representation of pixels in the same class, thereby improving the classification accuracy.

This paper begins in Section 1 by introducing the background and development of sparse subspace clustering. Section 2 presents the related work before model establishment. Section 3 builds the model based on the idea of information entropy weighting and the coupling constraint of sparse representation, and solves it using the alternating direction method of multipliers. Section 4 applies MNIST, ORL, and COIL20, and conducts ablation experiments according to relevant evaluation criteria. Section 5 summarizes the entire paper.

2. Related Work

2.1. Subspace clustering via structured sparse relation representation

During the data processing process, the original data is usually affected by noise and other disturbances. Directly performing self-representation on the original data to obtain a similarity matrix is prone to be influenced by noise and interference, resulting in an inability to accurately reflect the true subspace relationship of the data, thereby causing data misclassification. To solve this problem, Subspace clustering via structured sparse relation representation (SRR [16]) adopts a second-order self-representation strategy: Firstly, the original data is self-represented. Since the obtained self-representation matrix reflects a certain correlation between each data and other data, it is called a neighborhood relationship matrix; then, each datas self-representation vector is used as a new feature, and these new features are subjected to second-order self-representation again, also known as a reconstruction coefficient matrix; finally, a similarity matrix is constructed based on this matrix, and the final clustering result is obtained through spectral clustering.

$$\begin{aligned} &\min_{C,Z,E_1,E_2} \| C\|_1 + \| Z\|_* + \lambda_1 E_{11} + \lambda_2 E_{2F}^{\,2} \\ &\text{s.t.} \quad diag(C) = 0, X = XC + E_1, C = CZ + E_2 \end{aligned} \tag{1}$$

In the formula, X represents the original data matrix, with each column corresponding to one data sample. C is the neighborhood relationship matrix; Z is the reconstruction coefficient matrix; $\|E_1\|_1$ is used to describe outliers in the data; $\|E_2\|_F^2$ is used to handle Gaussian noise in the neighborhood relation matrix; λ_1 and λ_2 are weight parameters. In the SSR model: Firstly, each column of the neighborhood relationship matrix C represents a certain correlation between the corresponding column in the original data matrix X and other data, and it is constrained by the l_1-norm. Then, C is used as the new feature of the original data X for self-representation or self-reconstruction, and the F-norm is used to constrain the reconstruction coefficient matrix Z. Finally, calculate the similarity matrix using Z and obtain the clustering results through spectral clustering.

2.2. Weighted Block Sparse Subspace Clustering Algorithm Based on Information Entropy (EBSSC)

Under the influence of the Gaussian similarity weighted sparse subspace algorithm, the information entropy based on the correlation coefficient is introduced to determine the weight matrix of the sparse constraint. This method can impose a greater penalty on the coefficients corresponding to two uncorrelated pixels, thereby positively influencing the sparse model. Meanwhile, the block diagonal constraint of the expression matrix is introduced to enhance the stability of the model during the self-expression process. The weighted block sparse subspace clustering algorithm (EBSSC[22]) model is as follows.

$$\begin{aligned} &\min_{C} \| W \odot C\|_1 + \frac{\lambda}{2}\| X - XC\|_F^2 + \beta \| C\|_k \\ &\text{s.t.} \quad diag(C) = 0, C = C^T, C \neq 0 \end{aligned} \tag{2}$$

In the formula, $\|W \odot C\|_1$ represents the l_1-norm of the element-wise multiplication of the two matrices. W represents the weight matrix, which is determined by information entropy, that is, $W_{ij} = \sum_k p_{ij,k} \log_2 p_{ij,k}$. If $k =$

1, $p_{ij,k}$ represents the probability that two pixel points belong to the same category, and $p_{ij,k}$ is also the correlation coefficient between the two pixels. The first term in equation (1) is the weighted sparse constraint, whose entropy weight represents the penalty term of the l1 norm constraint of the corresponding coefficients of the two pixels. The second term represents the self-expression of the data, using the F-norm constraint, aiming to minimize the error between the data set and its sparse representation. The third term is the block diagonal constraint, aiming to make the coefficient matrix of the representation become a block diagonal sparse form.

3. Entropy-Constrained and Sparse Relation-Constrained Subspace Clustering

3.1. Model Building

Several common subspace clustering models are shown in Table 1. Sparse Relationship Representation (SRR) is a subspace clustering algorithm with good performance. It utilizes the neighborhood relationship between a data sample and all samples as a new feature to learn the self-representation coefficients, and then constructs a similarity matrix. The existing SRR model improves clustering accuracy by using second-order self-representation, but its nuclear norm constraint leads to high computational complexity [16], and it does not distinguish the importance of different variables.

In practical data, noise variables or redundant features may interfere with the construction of the similarity matrix [7, 22]. Therefore, this paper proposes two improvements: (1) Introduce the information entropy weighting matrix W, quantifying the discriminability between variables through the correlation coefficient $p_{ij,k}$. The theoretical basis is that information entropy can measure the uncertainty of probability distribution, and the lower the entropy value, the stronger the correlation, which is suitable as the weight of sparse penalty; (2) Replace the nuclear norm constraint with the F-norm to reconstruct the coefficient matrix Z, balancing computational efficiency and robustness. The theoretical basis is that the F-norm is more sensitive to Gaussian noise and has a closed-form solution [22]. Finally, combine the block diagonal constraint ($\|C\|_*$) to enhance the sparsity of the subspace structure, forming a unified optimization framework. The subspace clustering model based on information entropy and sparse relationship coupling constraints is as follows.

Table 1. SSC Classic Model and Extended Model

Model	Objective Function	Constraint Condition
SSC/SSC-OMP	$\|C\|_1+\lambda\|X-XC\|_F^2$	$X=XC,\operatorname{diag}(C)=0,C^T1=1$
LSR	$\|E_X\|_F^2+\lambda\|Z\|_F^2$	$X=XZ+E_X$
GSSC	$\sum_{i\neq j}\frac{1}{\omega_{ij}}\|C_{ij}\|+\lambda\|X-XC\|_F^2$	$\operatorname{diag}(C)=0$
KSSC	$\|C\|_1+\lambda\|\Phi(Y)-\Phi(Y)C\|_F^2$	$\operatorname{diag}(C)=0,C^T1=1$
3DF-SSC	$\sum_{i\neq j}\frac{1}{\omega_{ij}}\|C_{ij}\|+\lambda\|x_ic_i-x_i\|_2^2$	$\operatorname{diag}(C)=0$
TL_1	$\|C\|_{TL_1}$	$X=XC,\operatorname{diag}(C)=0$
SWSSC	$\|WC\|_1+\lambda\|E\|_F^2$	$X=XC+E,\operatorname{diag}(C)=0$
EBSSC	$\|W\odot C\|_1+\lambda\|X-XC\|_F^2+\beta\|C\|_k$	$\operatorname{diag}(C)=0,C=C^T,C\leq 0$
DSLSR	$\|E_X\|_F^2+\lambda_1\|E_Z\|_F^2+\lambda_2\|C\|_F^2$	$X=XZ+E_X,Z=ZC+E_Z$
MSCD-DSC	$\sum_{l=1}^{L}\|X_{\theta_e}^l-\widehat{X_{\theta_e}^l}\|_F^2+\lambda_1\sum_{l=1}^{L}\|Z_{\theta_e}^l-Z_{\theta_e}^l(C+D^l)\|_F^2 +\lambda_2\sum_{l=1}^{L}\|C+D^l\|_F^2+\lambda_3\sum_{l\neq\omega}^{L}\|D^l\odot D^\omega\|_1$	$\operatorname{diag}(C+D^l)=0$

$$\min_{C,Z,E_1,E_2}\|W\odot C\|_1+\|Z\|_F+\lambda_1\|E_1\|_F+\lambda_2\|E_2\|_F^2$$
$$\text{s.t.}\quad \operatorname{diag}(C)=0,X=XC+E_1,C=CZ+E_2 \tag{3}$$

3.2. Model Convergence Analysis and Solution

The given problem is solved by using the alternating direction method of multipliers (ADMM) iterative algorithm. An auxiliary transformation variable J is introduced, and the original problem is transformed as follows.

$$\min_{C,Z,J,E_1,E_2}\|W\odot J\|_1+\|Z\|_F+\lambda_1\|E_1\|_F+\lambda_2\|E_2\|_F^2$$
$$\text{s.t.}\quad X=XC+E_1,C=J,\operatorname{diag}(J)=0,C=CZ+E_2 \tag{4}$$

Then the augmented Lagrangian function of the model is:

$$\begin{aligned}\mathcal{L}(C,Z,J,E_1,E_2,Y_1,Y_2,Y_3) = \ & \|W \odot J\|_1 + \|Z\|_F^2 \\ & + \lambda_1 \|E_1\|_F + \lambda_2 \|E_2\|_F^2 \\ & + \frac{\mu_1}{2}\|X - XC - E_1 + \frac{Y_1}{\mu_1}\|_F^2 \\ & + \frac{\mu_2}{2}\|C - CZ - E_2 + \frac{Y_2}{\mu_2}\|_F^2 \\ & + \frac{\mu_3}{2}\|J - C + \frac{Y_3}{\mu_3}\|_F^2 \end{aligned} \quad (5)$$

Among them, Y_1, Y_2, Y_3, are Lagrange multipliers, and μ_1, μ_2, μ_3 are penalty terms. According to Boyd et al.s ADMM convergence theorem, two conditions need to be verified: the convexity of the objective function and the block convexity of the augmented Lagrangian function. Since in the main objective function L, $\||W \odot C\|_1$ is an l_1-norm, which is a convex function, an $\|Z\|_F$, $\|E_1\|_F$ and $\|E_2\|_F^2$ are F-norm squares, which are strongly convex functions. Therefore, L is closed and convex. In the augmented Lagrangian function, the variables groups (J, E_1, E_2) and (C, Z) are strictly convex when block-wise. Although $C = CZ$ introduces nonlinear coupling, the convergence is forced through the penalty term $\frac{\mu_2}{2}\|C - CZ - E_2\|_F^2$. In conclusion, the model converges and there exists a feasible solution.

Because the structure of the augmented Lagrangian function L is independent, C, Z, J, E_1 and E_2 can be solved by fixing other variables. Fixing other variables, the sub-problem regarding J is:

$$J^{t+1} = \arg\min_J (\|W \odot J\|_1 + \frac{\mu_3^t}{2}\|C^t - J + \frac{Y_3^t}{\mu_3^t}\|_F^2) \quad \text{s.t.} \quad \text{diag}(J) = 0 \quad (6)$$

The solution of J obtained by using the soft threshold operator is,

$$\begin{aligned} J^{t+1} &= W \odot (\tilde{J} - \text{diag}(\tilde{J})) \\ \tilde{J} &= \text{SoftThreshold}(C^t + \frac{Y_3^t}{\mu_3^t}, \frac{1}{\mu_3^t}) \end{aligned} \quad (7)$$

Fixing other variables, the sub-problem concerning C is,

$$C^{t+1} = \arg\min_C \begin{pmatrix} \frac{\mu_1^t}{2}\|X - XC - E_1^t + \frac{Y_1^t}{\mu_1^t}\|_F^2 \\ + \frac{\mu_2^t}{2}\|C - CZ^t - E_2^t + \frac{Y_2^t}{\mu_2^t}\|_F^2 \\ + \frac{\mu_3^t}{2}\|J^t - C + \frac{Y_3^t}{\mu_3^t}\|_F^2 \end{pmatrix} \quad (8)$$

This update can be obtained by using the gradient descent method to take the partial derivative of C_{t+1} with respect to C and setting it to zero.

$$C^{(t+1)} = C^t - \alpha \begin{pmatrix} -\mu_1^t X^T (X - XC - E_1^t + \frac{Y_1^t}{\mu_1^t}) \\ +\mu_2^t (1 - (Z^t)^T)(C - CZ^t - E_2^t + \frac{Y_2^t}{\mu_2^t}) \\ -\mu_3^t (J^t - C + \frac{Y_3^t}{\mu_3^t}) \end{pmatrix} \quad (9)$$

Here, α represents the learning rate of gradient descent. Fixing other variables, the sub-problem regarding Z is:

$$Z^{t+1} = \arg\min_Z \left(\|Z\|_F^2 + \frac{\mu_2^t}{2}\|C^t - C^t Z - E_2^t + \frac{Y_2^t}{\mu_2^t}\|_F^2 \right) \quad (10)$$

Taking the derivative of variable Z and setting it to 0, the closed-form solution of Z is,

$$Z^{t+1} = (1 + \mu_2 C^T C)^{-1} \mu_2 C^T (C - E_2 + \frac{Y_2}{\mu_2}) \quad (11)$$

Fixing other variables, the sub-problem regarding E_1 is,

$$E_1^{t+1} = \arg\min_{E_1} \begin{pmatrix} \lambda_1 \|E_1\|_F \\ + \frac{\mu_1^t}{2}\|X - XC^t - E_1 + \frac{Y_1^t}{\mu_1^t}\|_F^2 \end{pmatrix} \quad (12)$$

This update can be obtained by solving with a soft threshold operator.

$$E_1^{t+1} = sign(G) \max\left(|G| - \frac{\lambda_1 I}{\mu_1}, 0 \right) \quad (13)$$

$$G = X - XC + \frac{Y_1}{\mu_1} \quad (14)$$

Fixing other variables, the sub-problem regarding E_2 is:

$$E_2^{t+1} = \arg\min_{E_2} \begin{pmatrix} \lambda_2 \|E_2\|_F^2 \\ + \frac{\mu_2^t}{2}\|C^t - C^t Z^t - E_2 + \frac{Y_2^t}{\mu_2^t}\|_F^2 \end{pmatrix} \quad (15)$$

This update can be obtained through a simple closed-form solution.

$$E_2^{t+1} = \frac{\mu_2^t}{2\lambda_2 + \mu_2^t} \left(C^t - C^t Z^t + \frac{Y_2^t}{\mu_2^t} \right) \quad (16)$$

Update Lagrange multiplier,

$$Y_1^{t+1} = Y_1^t + \mu_1^t (X - XC^{t+1} - E_1^{t+1}) \quad (17)$$

$$Y_2^{t+1} = Y_2^t + \mu_2^t (C^{t+1} - C^{t+1} Z^{t+1} - E_2^{t+1}) \quad (18)$$

$$Y_3^{t+1} = Y_3^t + \mu_3^t (C^{t+1} - C^{t+1} Z^{t+1} - E_2^{t+1}) \quad (19)$$

These steps will be iterated until convergence. In each iteration step, some optimization algorithms are used to update the variables in the problem separately. This is a basic ADMM solution step. The specific optimization algorithms and solution steps may vary depending on the specific form and characteristics of the problem.

3.3. Computation Analysis

The time complexity of the subspace clustering algorithm proposed in this paper, which combines information entropy and sparse relationship coupling constraints, mainly consists of two parts: ADMM iterations and spectral clustering. Let the data matrix $X \in R^{d \times n}$ contain n d-dimensional samples. The time complexity analysis of each step of the algorithm is as follows.

During the ADMM iteration stage, in step 2 of variable update, (1) the soft-threshold operation of J takes $O(n^2)$ time; (2) the closed-form solution of Z involves the inversion of an $n \times n$ matrix, which is implemented using Cholesky decomposition and has a complexity of $O(n^3)$; (3) the gradient update of C includes matrix operations such as $X^T X$ and $C^T C$, which takes $O(max(d n^2, n^3))$ time; (4) the updates of E_1 and E_2 are linear operations and only require $O(n^2)$.

The parameter updates in Step 3 and the convergence judgment in Step 4 are both of $O(n^2)$ complexity. Therefore, the complexity of a single ADMM iteration is $O(max(d\,n^2, n^3))$. When the algorithm converges after T iterations, the total complexity of the ADMM stage is $O(T \cdot max(d\,n^2, n^3))$. In practical applications, the following strategies are adopted to improve efficiency: (1) Use truncated SVD to accelerate eigenvalue decomposition; (2) Optimize storage by using sparse matrix operations; (3) Set a reasonable maximum iteration number Tmax to control the ADMM loop. Table 2 details the execution flow of the algorithm, where steps 2-4 constitute the main ADMM loop, and the convergence and computational efficiency have been theoretically guaranteed through the aforementioned complexity analysis. The actual performance of the algorithm will be verified through experiments in Section 4.

Table 2. Entropy-Constrained and Sparse Relation-Constrained Subspace Clustering

Input: Data matrix X, number of categories k, trade-off parameter $\alpha, \lambda_1, \lambda_2$.
Output: Prediction label.
Step 1 Initialize $C^0, Z^0, J^0, E_1^0, E_2^0, Y_1^0, Y_2^0, Y_3^0, \mu_{\max}, \rho, \mu_1^0, \mu_2^0, \mu_3^0, \mu^0, T_{\max}, \varepsilon$
Step 2 Update $J, C, Z, E_1, E_2, Y_1, Y_2, Y_3$
Step 3 Update $\mu_1, \mu_2, \mu_3 : \mu_1^{t+1} = \mu_2^{t+1} = \mu_3^{t+1} = \mu^{t+1} = \min(\mu_{\max}, \rho\mu^t)$
Step 4 If $\max(\Vert X - XC - E_1 \Vert_\infty, \Vert C - J \Vert_\infty, \Vert C - CZ - E_2 \Vert_\infty) < \varepsilon$ or $t > T_{\max}$: step 5 Else: repeat steps 2 to 4.
Step 5 Calculate the similarity matrix $G = \vert J \vert + \vert J^T \vert$, and obtain the clustering results using spectral clustering.

4. Experiment

This paper compares the performance of the proposed EWSSC with K-means, spectral clustering, SSC, LRR, and SRR on different datasets through experiments. K-means performs well in partitioning spherical clusters, while spectral clustering is good at handling graph-structured data. Sparse subspace clustering stands out for its adaptability to the sparsity of high-dimensional data. The SRR algorithm enhances robustness to noise through a second-order self-representation strategy, thereby more accurately uncovering the subspace structure of the data. Through meticulous experimental design and evaluation, the aim is to deeply understand the performance of these algorithms under different data characteristics, providing practical guidance for the selection of clustering algorithms, and to indirectly observe the effect of the sparse subspace clustering algorithm.

4.1. Experiment Setup

To verify the effectiveness of the proposed algorithm, three widely used standard datasets were selected for this experiment. These datasets cover different application scenarios in image processing and have been widely applied in unsupervised learning and clustering tasks, making them important benchmarks for evaluating new algorithms. Since our method is based on an unsupervised learning clustering model, the true labels of the samples were not used during the training process. The three datasets used in this experiment correspond to three types of images: the MNIST handwritten digit image dataset, the ORL face image dataset, and the COIL20 object dataset.

The MNIST dataset consists of 10 handwritten digit images ranging from 0 to 9. Each digit has approximately 6000 different handwritten samples. The training set contains 60,000 images, and the test set contains 10,000 images. Each image has a size of 28×28 pixels and is a widely used benchmark dataset for handwritten digit recognition. Its classification performance may be affected in cases where the handwriting styles are diverse and the shapes of the digits vary significantly, as shown in Figure 1a.

The ORL dataset consists of 40 facial images of different individuals, each with 10 different expressions, resulting in a total of 400 images. Each image has a size of 32×32 pixels and is a relatively challenging facial image dataset. Its clustering performance may be affected when there are significant posture variations and different lighting conditions, as shown in Figure 1b.

The COIL20 dataset is an image dataset containing 20 types of objects. Each object has 72 images taken from different angles, with the image size being 32×32 pixels. The image content includes items such as toys and daily necessities, as shown in Figure 1c. This dataset is a small object image dataset and is suitable for testing clustering algorithms for object recognition and classification.

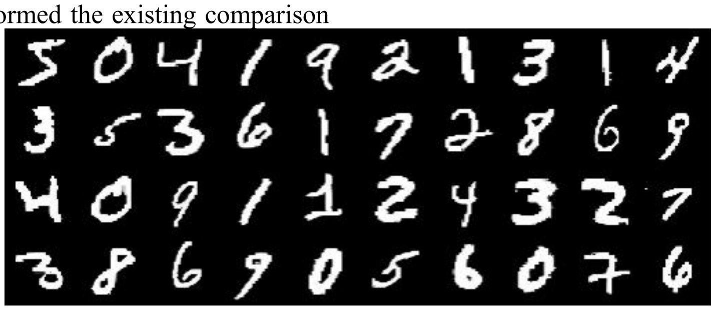
Overall, choosing these classic and diverse datasets not only helps to comprehensively evaluate the improved sparse subspace clustering algorithm, but also makes the experimental results more interpretable, providing strong theoretical support for the practical application of the algorithm. Detailed information of the datasets is shown in Table 3. If the algorithm is not tested on the datasets, the experimental results will be obtained using the hyperparameters mentioned in the text.

In the final evaluation section of the clustering results, the following evaluation indicators were selected: Accuracy (ACC), which measures the classification accuracy of the model for all samples; Normalized Mutual Information (NMI), which is used to evaluate the consistency between the clustering results and the true labels; Adjusted Rand Index (ARI), an adjusted version of the Rand index, which is used to measure the similarity between two data distributions.

These indicators cover the evaluation of clustering effects in various aspects. Among them, ACC is an intuitive measure, while NMI and ARI take into account the relative nature of the clustering results. The reason for choosing these evaluation indicators is that they can comprehensively consider the accuracy and consistency of clustering, providing a comprehensive assessment of the algorithms performance.

4.2. Algorithm Evaluation

After conducting systematic tests on three datasets, EWSSC significantly outperformed the existing comparison methods in all evaluation metrics, further verifying its effectiveness and robustness in high-dimensional data clustering tasks. To comprehensively evaluate the algorithms performance, we conducted multiple experiments on each dataset and took the average values of each metric as the final results. The optimal parameter configuration is shown in Table 4. On the MNIST dataset, this algorithm achieved an ACC score of 0.7628, which was 9.48% higher than the SRR method; on the ORL dataset, the NMI metric reached 0.9031, which was 6.18% higher than SRR; on the COIL20 dataset, the ARI metric was 0.7972, which was 1.91% higher than SRR.



(a)

(b)

(c)

Figure 1. Visualization examples of some samples from three datasets

The experimental results show that the EWSSC algorithm not only leads comprehensively in all benchmarks, but also demonstrates stronger discrimination ability in high-dimensional face datasets like ORL, indicating that the algorithm can effectively capture data structure information and improve clustering consistency. The detailed comparison results are shown in Table 4.

Furthermore, by conducting a grid search on the hyperparameters λ_1 and λ_2, we plotted the classification accuracy surface of the ACC metric on the MNIST dataset (see Figure 2). It can be observed in the figure that when $\lambda_1 = 0.05$ and $\lambda_2 = 0.01$, the ACC reaches a significant peak, which is in the same high response area as MNIST when the optimal parameters $\alpha = 0.1$ and $\rho = 1.053$ are used in Table 5. This indicates the robustness of the parameter selection. The smooth variation of the surface also indicates that the algorithm is insensitive to parameter changes and has good generalization ability. In conclusion, the EWSSC algorithm consistently outperforms existing mainstream clustering methods on multiple typical datasets. Its superiority stems from the joint optimization of data structure and representation learning. In the future, the parameter adaptive mechanism can be further studied to reduce the parameter tuning cost and extend to larger-scale datasets.

Table 3. Statistical information of the dataset

Dataset	Classes	Number of images/class	Samples	Size
MNIST	10	7000	70000	28×28
ORL	40	10	400	32×32
COIL20	20	72	1440	32×32

Table 4. Experimental results of evaluation indicators corresponding to each dataset

Dataset	Clustering Algorithm	ACC	NMI	ARI
MNIST	Kmeans	0.5902	0.5811	0.4450
	Spectral clustering	0.6301	0.7170	0.4902
	SSC	0.6359	0.5713	0.5858
	LSR	0.5755	0.4812	0.4999
	SRR	0.6680	0.6635	0.6111
	EWSSC	**0.7628**	**0.7218**	**0.7009**
ORL	Kmeans	0.6175	0.7843	0.4792
	Spectral clustering	0.7275	0.8586	0.5295
	SSC	0.7750	0.7878	0.6523
	LSR	0.7380	0.7215	0.7123
	SRR	0.7955	0.8413	0.7001
	EWSSC	**0.8175**	**0.9031**	**0.7272**
COIL20	Kmeans	0.5708	0.5179	0.5246
	Spectral clustering	0.6203	0.5989	0.6002
	SSC	0.6918	0.5961	0.6223
	LSR	0.6896	0.5588	0.6132
	SRR	0.7956	0.7601	0.7781
	EWSSC	**0.8122**	**0.7831**	**0.7972**

Table 5. The optimal parameters corresponding to each dataset

Dataset	α	λ_1	λ_1	ρ	μ_1	μ_2	μ_3
MNIST	0.005	0.05	0.01	1.053	0.1	0.1	1.0
ORL	0.001	0.016	0.011	1.04	0.1	0.1	1.0
COIL20	0.001	0.02	0.05	1.08	0.1	0.1	1.0

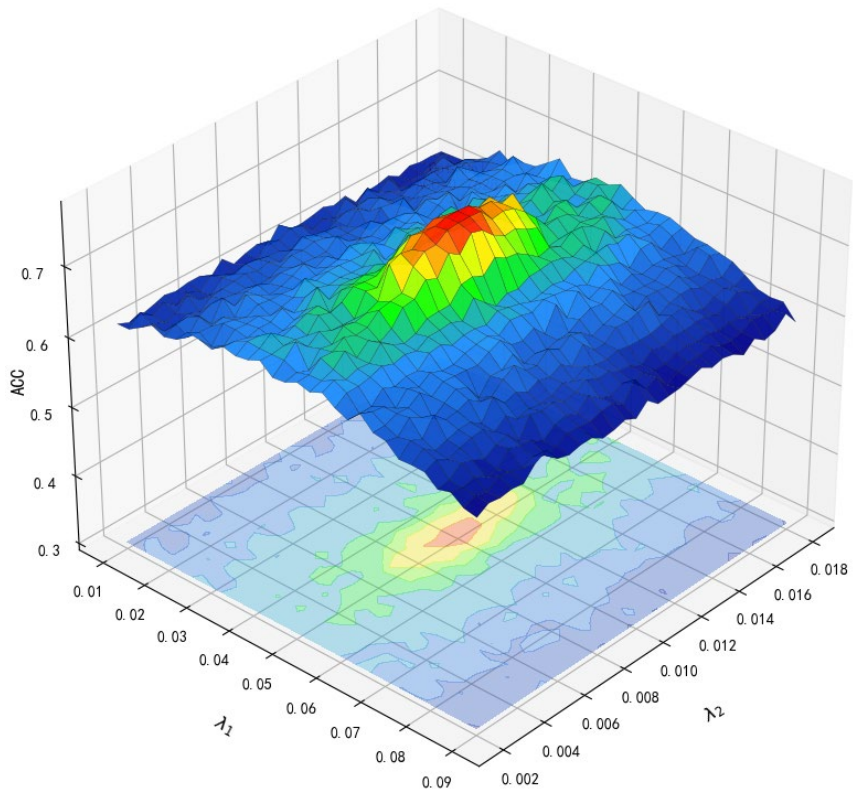


Figure 2. The classification accuracy (ACC) surface under fixed parameters (α = 0.005, ρ = 1.053) on the MNIST dataset.

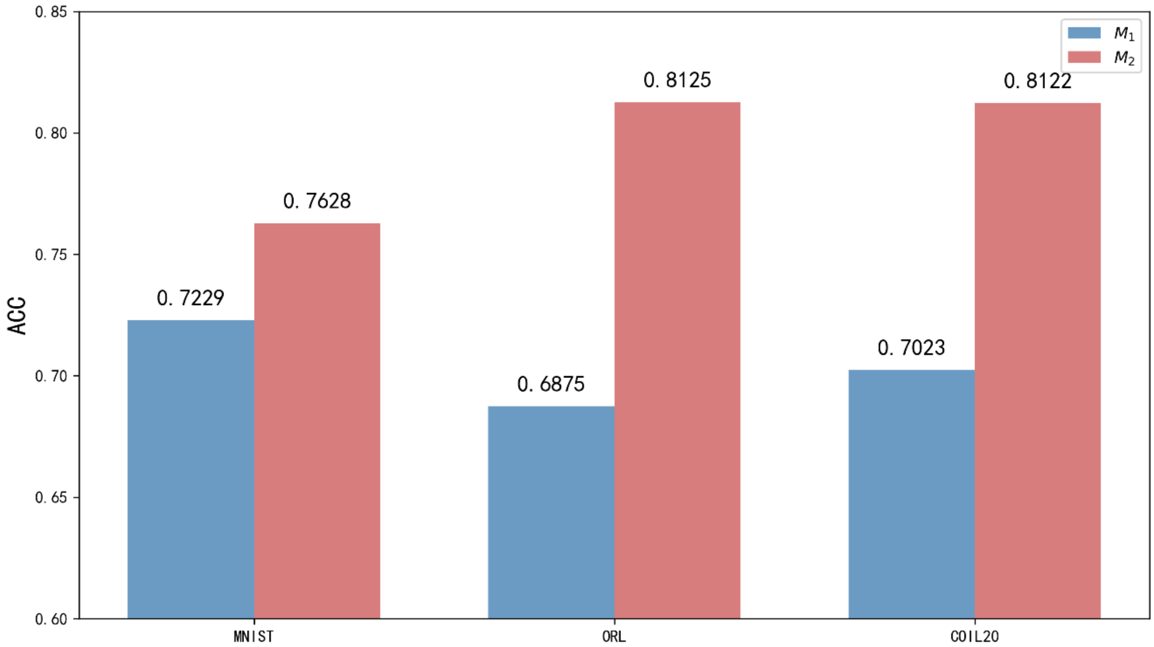


Figure 3. The comparison results of M_1 and M_2 on the MNIST, ORL and COIL20 datasets

4.3. Ablation Experiment

To verify the effectiveness of the information entropy weight in the model proposed in this paper, an ablation experiment was conducted for comparison. Two control groups were set up: M_1: the model without weight W; M_2: the model with weight. The comparison results on three datasets, namely MNIST, ORL, and COIL20, are shown in Figure 3. Figure 3 illustrates the impact of the information entropy weight module on the model performance. The experimental results show that the model with the weight (M_2) outperforms the baseline model without weight (M_1) significantly on all three datasets. Specifically: on the MNIST dataset, M_2s ACC reaches 0.7628, which is 5.52% higher than M_1, verifying the effective distinction of the weight module on the importance of handwritten digit features; on the ORL dataset, the performance improvement of M_2 is the most significant, with ACC increasing from 0.6875 to 0.8125 (a relative increase of 18.18%), indicating that the information entropy weight has stronger adaptability to illumination and posture changes in face images; on the COIL20 dataset: the 7 ACC of M_2 increases by 15.65%, indicating that the weight mechanism can better capture key features in multi-view object recognition tasks.

These results consistently demonstrate that the information entropy weight module significantly enhances the models ability to represent complex data structures by quantifying the importance of features. It is particularly effective in scenarios with noise or significant intra-class differences (such as ORL and COIL20).

5. Conclusion

The EWSSC proposed in this paper significantly improves the clustering performance of high-dimensional data through the joint optimization of information entropy weights and sparse representation. Theoretical analysis shows that the information entropy weights can effectively distinguish the importance of features, and the introduction of the F-norm constraint reduces the computational complexity. The experimental part verified the superiority of the algorithm on multiple standard datasets: 1) On MNIST and COIL20, the algorithm shows stronger robustness for data with large intra-class differences; 2) The significant improvement on the ORL dataset indicates that the algorithm has good adaptability to changes in illumination and posture; 3) The ablation experiments confirmed that the weight module is a key source of performance improvement. Future work will explore the influence of different weights and expand the experimental datasets.

References

[1] Wang W W, Li X P, Feng X C, et al. A survey on sparse subspace clustering[J]. IEEE/CAA Journal of Automatica Sinica, 2015, 41(8): 1373-1384.

[2] Ouyang P P. Research on Improved Sparse Subspace Clustering Algorithm[D]. Qingdao University, 2015.

[3] Xue X Q. Research on Low-Rank Subspace Clustering Algorithm and Its Application[D]. Southwest University of Science and Technology, 2021.

[4] Dong W H. Research on Sparse Subspace Clustering and Its Application[D]. Jiangnan University, 2019. (in Chinese)

[5] Chen H Z. Research on Subspace Clustering Analysis and Application of High-Dimensional Data[D]. Xidian University, 2019.

[6] Vidal R. Sparse subspace clustering[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009: 2790-2797.

[7] Elhamifar E, Vidal R. Sparse subspace clustering: Algorithm, theory, and applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2765-2781.

[8] Liu G, Lin Z, Yu Y. Robust subspace segmentation by low-rank representation[C]//Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel: OmniPress, 2010: 663-670.

[9] Liu G, Lin Z, Yan S, et al. Robust recovery of subspace structures by low-rank representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 171-184.

[10] Lu C, Min H, Zhao Z, et al. Robust and efficient subspace segmentation via least squares regression[C]//Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer, 2012: 347-360.

[11] Patel V M, Vidal R. Kernel sparse subspace clustering[C]//2014 IEEE International Conference on Image Processing. Paris, France: IEEE, 2014: 2849-2853.

[12] [12] You C, Robinson D, Vidal R. Scalable sparse subspace clustering by orthogonal matching pursuit[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 3918-3927.

[13] Li C G, Vidal R. Structured sparse subspace clustering: a unified optimization framework[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 277-286.

[14] Li H Y, Wang H Y. Subspace clustering method based on TL_1 norm constraint[J]. Journal of Electronics & Information Technology, 2017, 39(10): 2428-2436.

[15] Lu G F, Tang R, Yao L. Dual structure least squares regression for subspace clustering[J]. Journal of Nanjing University (Natural Science), 2022, 58(6): 1050-1058.

[16] Wei L, Ji F, Liu H, et al. Subspace clustering via structured sparse relation representation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(9): 4610-4623.

[17] Hu H Q, Zhang W Q, Xu C. Discriminatively enhanced sparse subspace clustering[J]. Computer Engineering and Applications, 2023, 49(2): 98-104.

[18] Liu Z Y, Wang H W, Zhao Q. Self-weighted scaled simplex representation for subspace clustering[J]. Journal of Beijing University of Aeronautics and Astronautics. https://doi.org/10.13700/j.bh.1001-5965.2023.0617

[19] Li T, Wang W W, Zhai D, et al. Weighted sparse subspace clustering for image segmentation[J]. Journal of Systems Engineering and Electronics, 2014, 36(3): 580-585.

[20] Zhang H, Zhai H, Zhang L, et al. Spectral-spatial sparse subspace clustering for hyperspectral remote sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(6): 3672-3684.

[21] Li W Z, Deng X Q, Liu F C. Subspace clustering algorithm fused three-dimensional spatial spectral features of hyperspectral images[J]. Application Research of Computers, 2019, 36(11): 3496-3498.

[22] Long Y H, Deng X Q, Wang Z W, et al. Weighted block sparse subspace clustering algorithm based on information entropy[J]. Journal of Data Acquisition and Processing, 2021, 36(3): 544-555.