**Frontiers in Computing and Intelligent Systems**
ISSN: 2832-6024 | Vol. 14, No. 2, 2025

# Application of Sparse Kernel Graph-regularized Discriminant Non-negative Matrix Factorization in Image Clustering

**Ran Zhou, Jianlei Li ***

Electronic Information Major, School of Mathematics and Statistics, North China University of Water Resources and Electric Power, Zhengzhou, Henan Province, 450046, China

*** Corresponding author:** Jianlei Li (Email: lijianlei@ncwu.edu.cn)

**Abstract:** Non-negative Matrix Factorization (NMF) is widely used in image clustering; however, it has inherent limitations, including its unsupervised nature, lack of sparsity constraints, inability to leverage label information, and difficulty in capturing the geometric structure and nonlinear characteristics of data. To address these limitations, this paper proposes a Sparse Kernel Graph-regularized Discriminant Non-negative Matrix Factorization (SKGDNMF) algorithm. The algorithm innovatively adopts a dual normalization strategy, which involves column normalization for the basis matrix and row normalization for the coefficient matrix, and integrates spectral clustering to construct an end-to-end deep clustering framework. By applying sparse regularization to the coefficient matrix, the model's robustness is significantly improved; this regularization forms complementary optimization with graph regularization, thereby effectively alleviating overfitting. A three-dimensional golden-section parameter optimization method is employed to determine key parameters, which enhances the algorithm's practicality. Comparative experiments conducted on multiple datasets show that SKGDNMF significantly outperforms mainstream algorithms in terms of sparsity, robustness, and clustering performance, indicating its superior effectiveness for image clustering tasks.

**Keywords:** Graph Regularization; Image Clustering; Non-negative Matrix Factorization; Sparsity; Spectral Clustering.

## 1. Introduction

In image clustering algorithms, data dimensionality reduction is a key focus of researchers. Data usually changes before and after dimensionality reduction, and based on these changes, dimensionality reduction methods are categorized into nonlinear dimensionality reduction and linear dimensionality reduction. In the real world, however, there exists a large amount of data, and such data is often non-negative; otherwise, it would be meaningless. Therefore, the Non-negative Matrix Factorization (NMF) algorithm is widely applied in data feature extraction and has also been proven equivalent to many clustering methods.

Inspired by the Non-negative Matrix Factorization algorithm, this paper proposes a Kernel-based Sparse Discriminative Graph-regularized Non-negative Matrix Factorization (SKGDNMF) algorithm. This algorithm is a semi-supervised method that mainly uses label information to construct a label matrix. Meanwhile, it incorporates a graph regularization term to capture the geometric structure of data, adopts a kernel method to handle the nonlinear nature of data, and adds an $L_1$ regularization term to reduce the impact of noise on clustering results and enhance the discriminative power of clustering.

## 2. Sparse Kernel Graph Discriminative Non-negative Matrix Factorization Algorithm

### 2.1. SKGDNMF Algorithm

This paper combines three algorithms, namely GNMF [1], DNMF [2], and KNMF [3], and incorporates a sparsity regularization term into the model to propose a new algorithm. Since the NMF algorithm is solved through an iterative update approach, and the matrix for initial iteration is randomly generated, this paper, inspired by Reference [4], adopts a "warm-start" strategy [4] at the initial stage of the algorithm. The warm-start strategy first performs K-means clustering on the data, and uses the obtained cluster centers as the initial basis matrix of the model. Based on the above, Sparse Kernel Graph Discriminative Non-negative Matrix Factorization (SKGDNMF) is proposed. The model of the algorithm is presented as follows:

$$\min_{U\geq 0,V\geq 0}\left\|\phi(X)-\phi(U)V^T\right\|_F^2+\alpha\left\|Q-AV_1^T\right\|_F^2+ \beta\mathrm{Tr}(V^TLV)+\gamma\left\|U-U_0\right\|_F^2+\lambda\left\|V\right\|_1 \tag{1}$$

Similar to the standard NMF model, the optimization problem of solving U and V simultaneously is non-convex. Therefore, this paper still adopts the alternating iterative update approach to solve the iterative formulas. The objective formula can be simplified to:

$$\begin{aligned} o=&\mathrm{Tr}(K_{XX})-2\mathrm{Tr}(K_{XU}V^T)+\mathrm{Tr}(VK_{UU}V^T)\\ &+\alpha\mathrm{Tr}(QQ^T)-2\alpha\mathrm{Tr}(QV_lA^T)\\ &+\alpha\mathrm{Tr}(AV_l^TV_lA^T)+\beta\mathrm{Tr}(V^TLV)\\ &+\gamma\mathrm{Tr}(U^TU)-2\gamma\mathrm{Tr}(U^TU_0)+\gamma\mathrm{Tr}(U_0^TU_0) \end{aligned} \tag{2}$$

By constructing the Lagrangian function, we can obtain:

$$\begin{aligned} L=&\mathrm{Tr}(K_{XX})-2\mathrm{Tr}(K_{XU}V^T)+\mathrm{Tr}(VK_{UU}V^T)\\ &+\alpha\mathrm{Tr}(QQ^T)-2\alpha\mathrm{Tr}(QV_lA^T)+\alpha\mathrm{Tr}(AV_l^TV_lA^T)\\ &+\beta\mathrm{Tr}(V^TLV)+\gamma\mathrm{Tr}(U^TU)-2\gamma\mathrm{Tr}(U^TU_0)\\ &+\gamma\mathrm{Tr}(U_0^TU_0)+\mathrm{Tr}(\Phi U^T)+\mathrm{Tr}(\Psi V^T) \end{aligned} \tag{3}$$

Herein, $\phi_{ik}$ and $\psi_{ik}$ are the Lagrange multipliers corresponding to constraints $u_{ik} \geq 0$ and $v_{ik} \geq 0$. Therefore, by differentiating the Lagrangian function and based on the KKT conditions (Karush-Kuhn-Tucker conditions): $\phi_{ik}u_{ik} = 0$, $\psi_{ik}v_{ik} = 0$, $\frac{\partial L}{\partial U}=0$, $\frac{\partial L}{\partial V} = 0$, and $\frac{\partial L}{\partial A} = 0$, the update rules can be derived as follows:

$$u_{ik} \leftarrow u_{ik} \frac{(K'_{XU}V + \gamma U_0)_{ik}}{(UBK'_{UU} + \gamma U)_{ik}} \tag{4}$$

$$v_{ik} \leftarrow v_{ik} \frac{(K_{XU} + \alpha(Q^T A)^+ + \alpha(V_l A^T A)^- + \beta WV)_{ik}}{(VK_{UU} + \alpha(Q^T A)^- + \alpha(V_l A^T A)^+ + \beta DV + \lambda)_{ik}} \tag{5}$$

$$A \leftarrow QV_l(V_l^T V_l + regAI)^{-1} \tag{6}$$

Here, a regularization parameter $regA$ is incorporated into the iterative formula of $A$ to avoid matrix singularity.

Regarding the update rules, Theorem 1 holds as follows: The optimization problem (1) is non-increasing under the update rules (4), (5), and (6). The value of the objective function remains unchanged if and only if the matrices $U$, $V$, and $A$ reach a stable point.

## 2.2. Convergence Analysis

The following proof refers to Reference [8].Let Model (1.1) be defined as the following objective function:

$$\begin{aligned} M = &\left\|\phi(X) - \phi(U)V^T\right\|_F^2 + \alpha\left\|Q - AV_1^T\right\|_F^2 \\ &+\beta \mathrm{Tr}(V^T LV) + \gamma\left\|U - U_0\right\|_F^2 + \lambda\left\|V\right\|_1 \end{aligned} \tag{7}$$

Definition 1: Function $G(h,h^t)$, which is the auxiliary function of Function $M(h)$, satisfies the following two conditions:

$$G(h,h^t) \geq M(h), G(h,h) = M(h) \tag{8}$$

Lemma 1: If function $G(h,h^t)$ is the auxiliary function of function $M(h)$, then $M(h)$ is non-increasing under the iterative rule:

$$m^{t+1} = \arg\min_h G(h,h^t) \tag{9}$$

It can be proven as follows:

$$M(h^{t+1}) \leq G(h^{t+1},h^t) \leq G(h^t,h^t) = M(h^t) \tag{10}$$

Moreover, $M(h^{t+1}) = M(h^t)$ holds if and only if $m^t$ is a local minimum point of $G(h,h^t)$.

If the derivative of $M$ exists and is continuous within a small neighborhood of $m'$, then the differential $\nabla M(h') = 0$ exists. Through Equation (9), a series of sequences that converge to the local minimum point $m_{\min} = \arg\min_h M(h)$ can be obtained:

$$\begin{aligned} &M(h_{\min}) \leq \cdots \leq M(h^{t+1}) \leq M(h^t) \\ &\leq \cdots \leq M(h^2) \leq M(h^1) \leq M(h^0) \end{aligned} \tag{11}$$

For an element $v_{ij}$ in matrix $V$ (from the objective function), the formulas for the first-order partial derivative $M'_{v_{ij}}$ and the second-order partial derivative $M''_{v_{ij}}$ of the objective function with respect to $v_{ij}$ can be obtained as follows:

$$\begin{aligned} M'_{v_{ij}} = &(-2K_{XU} + 2VK_{UU} - 2\alpha Q^T A \\ &+ 2\alpha V_l A^T A + 2\beta LV + \lambda)_{ij} \end{aligned} \tag{12}$$

$$M''_{v_{ij}} = \begin{cases} 2(K_{UU})_{jj} + 2\alpha(A^T A)_{jj} + 2\beta L_{ii}, if\ x_i\ has\ labels \\ 2(K_{UU})_{jj} + 2\beta L_{ii} \quad ,others \end{cases} \tag{13}$$

Lemma 2: Function

$$\begin{aligned} G(v,v^t_{ij}) = &M_{v_{ij}}(v^t_{ij}) + M'_{v_{ij}}(v^t_{ij})(v - v^t_{ij}) + \\ &\frac{(VK_{UU} + \alpha(Q^T A)^- + \alpha(V_l A^T A)^+ + \beta DV + \lambda)_{ij}}{v^t_{ij}}(v - v^t_{ij})^2 \end{aligned} \tag{14}$$

is the auxiliary function of $M_{v_{ij}}$.

Proof: As can be seen from Definition 1, to prove a function is the auxiliary function (of another function), it must satisfy both $G(v,v^t_{ij}) \geq M_{ij}(v)$ and $G(v,v) = M_{ij}(v)$ simultaneously. Obviously, $G(v,v) = M_{ij}(v)$ can be obtained directly. Next, we conduct a detailed proof for $G(v,v^t_{ij}) \geq M_{ij}(v)$. First, perform a Taylor expansion on $M_{ij}(v)$; by comparing it with Equation (14), to prove $G(v,v^t_{ij}) \geq M_{ij}(v)$, it is equivalent to proving that:

$$\frac{(VK_{UU} + \alpha(Q^T A)^- + \alpha(V_l A^T A)^+ + \beta DV + \lambda)_{ij}}{v^t_{ij}} \geq \frac{1}{2}M''_{v_{ij}}(v^t_{ij}) \tag{15}$$

Substitute Equation (13) into Equation (15):

$$\begin{cases} (VK_{UU} + \alpha(Q^T A)^- + \alpha(V_l A^T A)^+ + \beta DV + \lambda)_{ij} \\ \quad \geq v^t_{ij}[(K_{UU})_{jj} + \alpha(A^T A)_{jj} + \beta L_{ii}] \quad if\ x_i\ has\ labels \\ (VK_{UU} + \beta DV + \lambda)_{ij} \geq v^t_{ij}[(K_{UU})_{jj} + \beta L_{ii}] \quad others \end{cases} \tag{16}$$

Given $(VK_{UU})_{ij} = \sum_k v^t_{ik}(K_{UU})_{kj}$, thus it is obvious that:

$$(VK_{UU})_{ij} = \sum_k v^t_{ik}(K_{UU})_{kj} \geq v^t_{ij}(K_{UU})_{jj} \tag{17}$$

Similarly, it can be obtained that:

$$\begin{aligned} \beta(DV)_{ij} &= \beta\sum_k D_{ik}v^t_{kj} \geq \beta D_{ii}v^t_{ij} \\ &\geq \beta(D_{ii} - W_{ii})v^t_{ij} = \beta L_{ii}v^t_{ij} \end{aligned} \tag{18}$$

Moreover, if $x_a$ has labels, then both $(V_l)_{ab} = V_{ab}$ and $(Q^T A)^- \geq 0$ hold; thus it follows that:

$$\begin{aligned} &[\alpha(Q^T A)^- + \alpha(V_l A^T A)^+]_{ab} \geq [\alpha(V_l A^T A)^+]_{ab} \\ &\geq [\alpha(V_l A^T A)]_{ab} = \alpha\sum_k V^t_{ik}(A^T A)_{kj} \geq \alpha V^t_{ij}(A^T A)_{jj} \end{aligned} \tag{19}$$

From Equations (17), (18), and (19), it can be proven that Equation (17) holds. Thus, Equation (14) is the auxiliary function of $M_{v_{ij}}$, and Theorem 2 is proven.

The above is the convergence proof for the update rule of Matrix $V$, and the update rule of Matrix $U$ can be proven similarly. Since objective function $A$ (with respect to the matrix) in Algorithm Model (1) is a convex function, the update rule can be obtained directly by setting $\frac{\partial L}{\partial A} = 0$ and solving it; thus, the update formula of Matrix $A$ is also convergent.

In conclusion, Theorem 1 is proven.

# 3. Experiments and Results Analysis

This paper tests the clustering performance of the

SKGDNMF algorithm on the face dataset PIE, the object dataset COIL20, and the handwritten digit dataset MNIST, and compares it with other algorithms including K-means [5], KPCA [6], NMF [7], GNMF, GDNMF [8], and KGDNMF [9] in terms of clustering performance.

## 3.1. Experimental Datasets

In the COIL20 dataset, all images are uniformly resized to a size of 32×32 pixels. In the PIE dataset, 68 subjects are randomly selected, and 42 facial images of each subject under different lighting conditions are used; these images are uniformly cropped to 32×32 grayscale images. For the MNIST dataset, 200 images are randomly selected for each digit (0-9) from the training set, resulting in a total of 2000 images.

## 3.2. Experimental Process

At the start of the experiment, we first load and standardize the three datasets, then divide them into a preset number of categories. For semi-supervised clustering algorithms, we randomly select k categories of data subsets each time, randomly extract 20% of the samples from each category to construct labeled data, and then run the algorithm model for semi-supervised decomposition. For the other algorithms, we set the dimensionality reduction dimension to the corresponding number of clusters k. In the algorithm post-processing stage, clustering labels are obtained through Spectral Clustering [10]. We calculate the Accuracy and Mutual Information (MI) metrics, repeat the experiment 10 times for each k to obtain the mean and standard deviation, and finally output the result table.

The algorithm proposed in this paper has a total of four parameters, and different parameters exert different impacts on the experimental results. Therefore, this paper adopts the control variable method to investigate the selection of experimental parameters on the three datasets. With the number of neighbors fixed, we conduct parameter sensitivity analysis for $\alpha$ , $\beta$ , and $\gamma$ respectively, repeat the experiment ten times to calculate the classification accuracy, and take the average value. The relevant results are shown in the figure below:

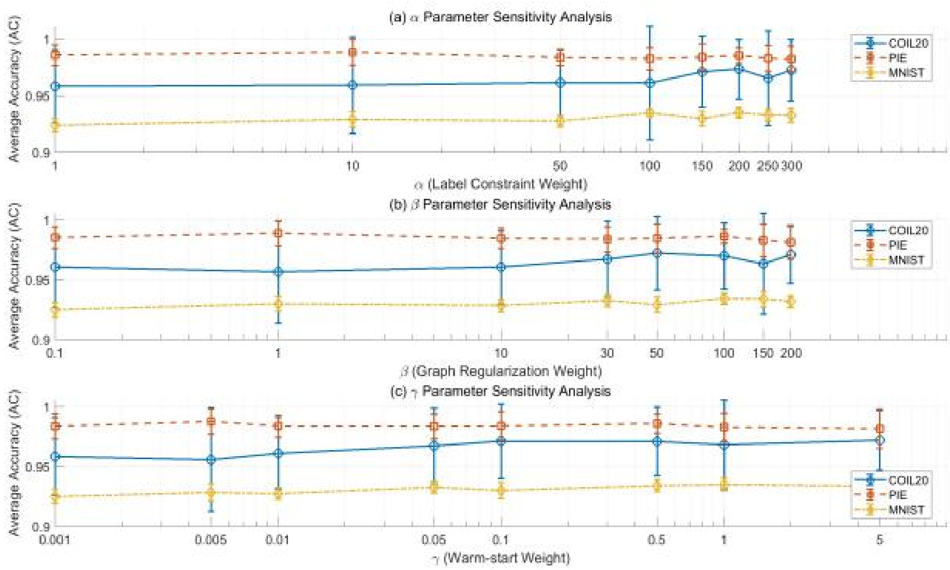


**Figure 1.** Parameter Sensitivity Analysis

From the three groups of experiments, it can be observed that the proposed algorithm exhibits extremely high robustness to the label constraint weight $\alpha$ and the warm-start weight $\gamma$ ; parameter variations hardly affect the clustering performance. The graph regularization weight $\beta$ only has slight optimization potential for specific datasets, such as the COIL20 dataset with high graph structure dependence. This characteristic also indicates that most parameters of the SKGDNMF algorithm do not require strict tuning to ensure model stability, further enhancing the algorithm's practicality.

## 3.3. Analysis of Experimental Results

**Table 1.** Three Scheme comparing

| | COL20 | | PIE | | MINST | |
|---|---|---|---|---|---|---|
| | AC | MIH | AC | MIH | AC | MIH |
| K-means | 0.6920 | 0.7455 | 0.3140 | 0.3765 | 0.6400 | 0.5449 |
| KPCA | 0.6979 | 0.7506 | 0.3162 | 0.3801 | 0.6461 | 0.5548 |
| NMF | 0.6840 | 0.7129 | 0.4904 | 0.5712 | 0.5992 | 0.4698 |
| GNMF | 0.8356 | 0.8910 | 0.8332 | 0.8840 | 0.6830 | 0.6588 |
| GDNMF | 0.8813 | 0.9267 | 0.9389 | 0.9498 | 0.9527 | 0.8736 |
| KGDNMF | 0.8756 | 0.9250 | 0.9344 | 0.9470 | 0.8914 | 0.8133 |
| SKGDNMF | **0.9584** | **0.9534** | **0.9833** | **0.9724** | **0.9543** | **0.8769** |

Table 1 presents the experimental results of the six algorithms on the three datasets. Among them, the results for the COIL20 and PIE datasets are the average values when the number of clusters ranges from 2 to 20 (with a step size of 2); the results for the MNIST dataset are the average values when the number of clusters ranges from 2 to 10. It can be seen that the proposed algorithm exhibits significant performance advantages in the clustering tasks of the three datasets, achieving higher Accuracy and Mutual Information (MI). This indicates that the algorithm can stabilize the discriminative structure of manifold data across different image datasets, proving its effectiveness in aspects such as preserving manifold structures and suppressing noise interference. In conclusion, the proposed algorithm provides a new solution for semi-supervised clustering tasks, is more suitable for feature learning and structure mining of high-dimensional nonlinear data, and has important theoretical reference value and practical application prospects.

# 4. Conclusion

This paper proposes the Sparse Kernel Graph Discriminative Non-negative Matrix Factorization (SKGDNMF) algorithm, which integrates semi-supervised learning, sparse constraints, label information, capture of the inherent geometric structure of data, and nonlinear data processing capabilities into a comprehensive model. Experimental results show that compared with existing algorithms, the proposed algorithm exhibits significant comprehensive advantages in image clustering tasks, with improvements in clustering accuracy, robustness, and sparsity. Through its innovative design, the algorithm effectively overcomes the limitations of traditional methods, provides a more effective solution for the clustering analysis of complex image data, and holds important theoretical significance and promising application prospects.

# References


[1] Cai D, He X, Wu X, et al. Non-negative matrix factorization on manifold[C]//2008 Eighth IEEE International Conference on Data Mining. IEEE, 2008: 63-72.

[2] Babaee M, Tsoukalas S, Babaee M, et al. Discriminative nonnegative matrix factorization for dimensionality reduction [J]. Neurocomputing, 2016, 173: 212-223.

[3] Zhang D, Zhou Z H, Chen S. Non-negative matrix factorization on kernels[C]//PRICAI 2006: Trends in Artificial Intelligence: 9th Pacific Rim International Conference on Artificial Intelligence Guilin, China, August 7-11, 2006 Proceedings 9. Springer Berlin Heidelberg, 2006: 404-412.

[4] Sun L, Zhao K, Han C Y, et al. Enhancing hyperspectral unmixing with two-stage multiplicative update nonnegative matrix factorization[J]. IEEE Access, 2019,7:171023-171031.

[5] Ashabi A, Sahibuddin S B, Salkhordeh Haghighi M. The systematic review of K-means clustering algorithm[C]// Proceedings of the 2020 9th International Conference on Networks, Communication and Computing. 2020: 13-18.

[6] Schölkopf B, Smola A ,Müller, K.Nonlinear Component Analysis as a Kernel Eigenvalue Problem[J].Neural Computation, 1998, 10(5):1299-1319.

[7] Lee D D , Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401(6755):788-791.

[8] Li H, Zhang J, Shi G, et al. Graph-based discriminative nonnegative matrix factorization with label information[J]. Neurocomputing, 2017, 266: 91-100.

[9] Li X L, Zhang Y. Kernel-based discriminative graph regularized non-negative matrix factorization [in Chinese][J]. Journal of Computer Science and Exploration, 2020,14 (11): 1899-1907.

[10] Ng A Y, Jordan M I , Weiss Y .On Spectral Clustering: Analysis and an algorithm[J].proc nips, 2002.