



Sparse Graph Regularization Non-Negative Matrix Factorization Based on Huber Loss Model for Cancer Data Analysis

Chuan-Yuan Wang¹, Jin-Xing Liu^{1*}, Na Yu¹ and Chun-Hou Zheng²

¹ School of Information Science and Engineering, Qufu Normal University, Rizhao, China, ² School of Software Engineering, Qufu Normal University, Qufu, China

OPEN ACCESS

Edited by:

Hongmin Cai,
South China University of Technology,
China

Reviewed by:

Wen-Sheng Chen,
Shenzhen University, China
Xiangxiang Zeng,
Xiamen University, China

*Correspondence:

Jin-Xing Liu
sdcavell@126.com

Specialty section:

This article was submitted to
Bioinformatics and
Computational Biology,
a section of the journal
Frontiers in Genetics

Received: 20 July 2019

Accepted: 01 October 2019

Published: 20 November 2019

Citation:

Wang C-Y, Liu J-X, Yu N and
Zheng C-H (2019) Sparse Graph
Regularization Non-Negative Matrix
Factorization Based on Huber Loss
Model for Cancer Data Analysis.
Front. Genet. 10:1054.
doi: 10.3389/fgene.2019.01054

Non-negative matrix factorization (NMF) is a matrix decomposition method based on the square loss function. To exploit cancer information, cancer gene expression data often uses the NMF method to reduce dimensionality. Gene expression data usually have some noise and outliers, while the original NMF loss function is very sensitive to non-Gaussian noise. To improve the robustness and clustering performance of the algorithm, we propose a sparse graph regularization NMF based on Huber loss model for cancer data analysis (Huber-SGNMF). Huber loss is a function between L_1 -norm and L_2 -norm that can effectively handle non-Gaussian noise and outliers. Taking into account the sparsity matrix and data geometry information, sparse penalty and graph regularization terms are introduced into the model to enhance matrix sparsity and capture data manifold structure. Before the experiment, we first analyzed the robustness of Huber-SGNMF and other models. Experiments on The Cancer Genome Atlas (TCGA) data have shown that Huber-SGNMF performs better than other most advanced methods in sample clustering and differentially expressed gene selection.

Keywords: non-negative matrix factorization, Huber loss, sample clustering, graph regularization, robustness

INTRODUCTION

Cancer is considered to be the number one killer of human health. The development of high-throughput sequencing technology has enabled researchers to obtain more comprehensive information about cancer patients (Chen et al., 2019). The gene expression data of cancer patients can be more used for effective data mining through computational methods (Chen et al., 2018). In general, cancer gene expression data are characterized by high dimensionality, which is extremely difficult for data analysis. How to effectively reduce the dimensionality of data is the key to analyzing cancer data. Principal component analysis (PCA) (Feng et al., 2019), locally linear embedding (LLE) (Roweis and Saul, 2000), and non-negative matrix factorization (NMF) (Yu et al., 2017) are common methods for reducing the data dimensionality. Unlike several other methods, NMF can find two non-negative matrices and its product can effectively restore the original matrix. The non-negative constraint guarantees additive combinations between different elements. NMF demonstrates its advantages in facial recognition, speech processing, document clustering, and recommendation systems (Guillamet and Vitrià, 2002; Xu et al., 2003; Schmidt and Olsson, 2006; Luo et al., 2014).

NMF has developed rapidly in recent years, and several variants of NMF have been proposed to improve the effectiveness of the decomposition. Cai et al. proposed graph regularized NMF (GNMF) for data representation (Cai et al., 2011). GNMF considers the association between points to preserve the internal structure of the data. Kim et al. applied the L^1 -norm constraint on the coefficient matrix to introduce sparse NMF for clustering (SNMF) (Kim and Park, 2007). Sparseness is more likely to remove redundant features of data. The most of cancer data have noise and outliers, and the original NMF cannot solve this. Wang et al. introduced Characteristic Gene Selection Based on Robust GNMF (RGNMF) (Wang et al., 2016a) to improve the robustness of the algorithm. RGNMF assumes that the loss follows Laplacian distribution and uses the loss function of the $L^{2,1}$ -norm (Kong et al., 2011) constraint. The $L^{2,1}$ -norm combines the advantages of the L^2 -norm and the L^1 -norm, which impose an L^2 -norm constraint on the entire data space and an L^1 -norm constraint on the sum of the different data points (Ding et al., 2006).

The original NMF model is simple to understand and computationally fast, but the squared loss function is too sensitive to outliers and noise. Mao et al. proposed the correntropy induced metric based GNMF (CGNMF) (Mao et al., 2014) that changed the original loss function. The correntropy uses L^0 -norm approximation for large outliers and noise through kernel function filtering, and the normal data is constrained by the L^2 -norm (Liu et al., 2007), so it is not sensitive to outliers and noise. Du et al. proposed Huber-NMF (Du et al., 2012), which is also a loss function that is insensitive to outliers and noise. It uses approximate L^1 -norm processing for outliers and noise, and L^2 -norm for valuable data. Correntropy uses kernel functions to control weights, and Huber loss uses a different function approximation for different data through threshold adjustment. The robustness analysis of these several non-square loss models is given in the experimental part. To compare the performance of the NMF algorithm, the robust PCA (RPCA) based method for discovering differentially expressed genes proposed by Liu et al. (2013) is added to the experiment.

In this paper, we propose a model called sparse graph regularization NMF based on Huber Loss Model for Cancer Data Analysis (Huber-SGNMF). It effectively combines Huber loss, manifold structure, and sparse constraint. Huber loss is based on the relationship between L^1 -norm and L^2 -norm to approximate different data. In detail, Huber loss adjusts the square loss or linear loss to the data according to the threshold to enhance the robustness of the model to outliers. Geometric information in high-dimensional data should remain locally constant in low-dimensional representations (Cai et al., 2011), so graph regularization is added to preserve the manifold structure of the data. Sparse constraints in the model can remove redundant features contained in the data to reduce the amount of model calculations and improve clustering performance (Kim and Park, 2007).

The contributions of this article are as follows:

1. The squared loss of the original NMF is too sensitive to outliers and noise; so, we use a more robust Huber loss combined with NMF. The Huber loss considers the relationship between the L^1 -norm and the L^2 -norm to effectively handle non-Gaussian noise and large outliers. For the update rules of Huber loss, we use the multiplicative iterative algorithm based on semi-quadratic optimization to find the optimal solution.
2. The NMF model fits the data in Euclidean space but does not consider the intrinsic geometry of the data space. If the data is related in high-dimensional space, then we believe that the data represented by the low-dimensional should also be closely related. Considering the manifolds embedded in the high-dimensional environment space, we add graph Laplacian as a regularization term to the model. Graph regularization takes into account the impact of recent neighbors on data, and retaining graph structure can make NMF more distinguishable.
3. Sparse matrices can remove redundant data, reducing data complexity and model computational difficulty. In data analysis, sparsity can improve clustering performance by reducing the difficulty of feature selection. The $L^{2,1}$ -norm as a sparse constraint is added to the model because the $L^{2,1}$ -norm is robust and can achieve row sparse effect.

The remainder of this paper is organized as follows. The introduction of related work is shown in Section 2. Models and solution optimization are presented in Section 3. The experiment and analysis are arranged in Section 4. Section 5 summarizes the entire paper.

RELATED WORK

Non-Negative Matrix Factorization

NMF is a dimensionality reduction method based on partial representation. For a given dataset $\mathbf{X} = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{m \times n}$, NMF can decompose it into the basic matrix $\mathbf{U} \in \mathbb{R}^{m \times k}$ and the coefficient matrix $\mathbf{V} \in \mathbb{R}^{k \times n}$, with the purpose of approximating the original matrix by two matrix products. In general, the rank of matrix factorization k is selected by the number of larger singular values.

For gene expression data matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, each row represents a gene corresponding to n samples, and each column represents a sample composed of m genes. Moreover, \mathbf{U} contains m rows of metagene and \mathbf{V} contains n rows of metapattern (Liu et al., 2018). Each column of \mathbf{V} is a projection of a corresponding sample vector in \mathbf{X} according to the basic matrix \mathbf{U} (Li et al., 2017). NMF is visualized on gene expression data as shown in **Figure 1**.

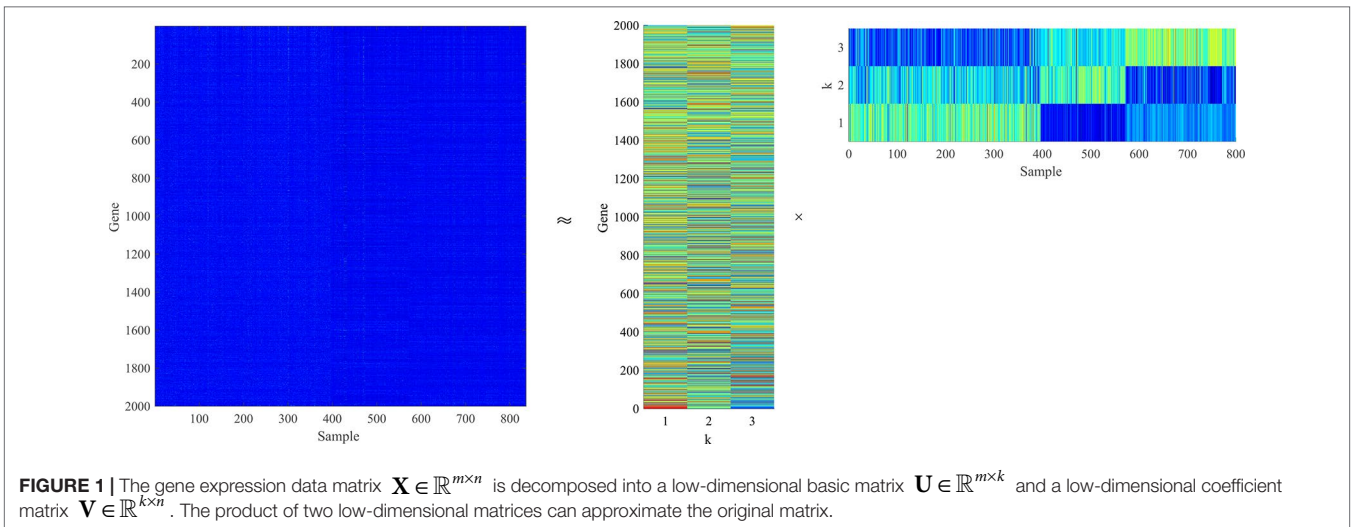
The NMF loss function is minimized as follows:

$$\min \|\mathbf{X} - \mathbf{UV}\|^2, \quad s.t. \mathbf{U} \geq 0, \mathbf{V} \geq 0, \quad (1)$$

where $\|\cdot\|$ represents the application of the Frobenius norm to the matrix.

Lee and Seung proposed the use of multiplicative iterative update rules to solve the optimal solution of NMF (Lee and Seung, 1999). Its update formula is as follows:

$$u_{ik} = u_{ik} \frac{(\mathbf{XV})_{ik}}{(\mathbf{UVV}^T)_{ik}}, \quad (2)$$



$$v_{kj} = v_{kj} \frac{(\mathbf{U}^T \mathbf{X})_{kj}}{(\mathbf{U}^T \mathbf{U} \mathbf{V})_{kj}}, \tag{3}$$

where u_{ik} and v_{kj} are elements belonging to \mathbf{U} and \mathbf{V} , respectively. The non-negative constraints of \mathbf{U} and \mathbf{V} only allow additive combinations between different elements, so NMF can learn part-based representations (Cai et al., 2011).

Huber Loss

Data usually contain a small amount of outliers and noise, which can have a worse effect on model reconstruction. For noise and outliers in the dataset, Huber loss uses weighted L_1 -norm processing because the L_1 -norm is robust and can effectively handle outliers and noise (Guofa et al., 2011; Yu et al., 2016). For other valuable data in the dataset, Huber losses still use L_2 -norm loss to fit the data. Huber loss function $\delta(\cdot)$ is defined as follows:

$$\delta(e) = \begin{cases} e^2 & \text{if } |e| < c, \\ 2c|e| - c^2 & \text{if } |e| \geq c, \end{cases} \tag{4}$$

where c represents the threshold parameter of the data using the L_1 -norm or the L_2 -norm. This function is a bounded and convex function that minimizes the effects of a single anomaly point (Chreiky et al., 2016). Huber losses often apply to the insensitive outliers and noise contained in the data, which are often difficult to find using the squared loss function (Du et al., 2012).

Manifold Regularization

The manifold learning theory (Belkin and Niyogi, 2001) shows that the internal manifold structure of the data can be effectively simulated by the nearest neighbor of the data points. Each data point finds its nearest p neighbors and connects the data points to the neighbors with edges. There are many ways to define the weight of an edge, most commonly 0–1 weighted: $\mathbf{W}_{ij}=1$, if and only if nodes i and j are connected by edges. The advantage of this weighting method is that it is easy to calculate.

Weight matrix \mathbf{W}_{ij} is only used to measure the intimacy between data points. For the low-dimensional representation \mathbf{s}_j of the high dimensional data \mathbf{x}_j , the Euclidean distance $\mathbf{O}(\mathbf{s}_j, \mathbf{s}_i) = \|\mathbf{s}_j - \mathbf{s}_i\|^2$ is typically used to measure the similarity between two low-dimensional data points. According to the intimacy weight \mathbf{W} , the smoothness of the two low-dimensional vectors can be measured as follows:

$$\begin{aligned} R &= \frac{1}{2} \sum_{j,l} \|\mathbf{s}_j - \mathbf{s}_l\|^2 \mathbf{W}_{jl} \\ &= \sum_{j=1}^N \mathbf{s}_j^T \mathbf{s}_j \mathbf{D}_{jj} - \sum_{j,l=1}^N \mathbf{s}_j^T \mathbf{s}_l \mathbf{W}_{jl} \\ &= \text{tr}(\mathbf{V} \mathbf{D} \mathbf{V}^T) - \text{tr}(\mathbf{V} \mathbf{W} \mathbf{V}^T) \\ &= \text{tr}(\mathbf{V} \mathbf{L} \mathbf{V}^T), \end{aligned} \tag{5}$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix. The matrix \mathbf{D} is defined as a diagonal matrix with diagonal elements $\mathbf{D}_{ii} = \sum_{jj} \mathbf{W}_{ji}$. The graph Laplacian (Liu et al., 2014) matrix \mathbf{L} is defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$.

We hope that if the high-dimensional data \mathbf{x}_j and \mathbf{x}_l are very intimate, then \mathbf{s}_j and \mathbf{s}_l should be close enough in low-dimensional representations (Cai et al., 2011). Therefore, minimizing R is added to our model to encode the internal geometry of the data.

METHOD

The Huber-Sgnmf Model

Based on the Huber loss function, we proposed a novel model that preserves the manifold structure and sparsity simultaneously. The Huber loss is combined with NMF to enhance NMF robustness. To further optimize the model, the graph regularization term and the $L_{2,1}$ -norm are added to the loss function as constraints. $L_{2,1}$ -norm mathematical expression is as follows:

$$\|\mathbf{x}\|_{2,1} = \sum_{i=1}^m \sqrt{\sum_{j=1}^n x_{ij}^2} = \sum_{i=1}^m \|\mathbf{x}_{i,*}\|_2. \tag{6}$$

The Huber-SGNMF final model $O_{Huber-SGNMF}$ is as follows:

$$\min_{\mathbf{U} \geq 0, \mathbf{V} \geq 0} \delta(\mathbf{X} - \mathbf{UV}) + \alpha \text{tr}(\mathbf{VLV}^T) + \beta \|\mathbf{V}\|_{2,1}, \quad (7)$$

where $\text{tr}(\cdot)$, α , and β represent the trace of the matrix, the regularization term parameters, and the sparse constraint parameters, respectively. In the experiment, the basic matrix \mathbf{U} and the coefficient matrix \mathbf{V} are used for differential gene selection and cluster analysis, respectively.

Optimization

Obviously, the loss function is a non-quadratic optimization problem, and finding the optimal solution is not simple. Fortunately, the semi-quadratic optimization technique that has been proposed can effectively optimize the loss function. The loss function can be reconstructed to find the optimal solution by introducing auxiliary variables. According to the conjugate function and the semi-quadratic technique (Nikolova and Chan, 2007), the fixed loss function $\sigma(\mathbf{Z})$ can be constructed as follows:

$$\sigma(\mathbf{Z}_{ij}) = \min_{\mathbf{W} \in \mathbb{R}} \tau(\mathbf{Z}_{ij}, \mathbf{W}_{ij}) + \phi(\mathbf{W}_{ij}), \quad (8)$$

where $\mathbf{Z}_{ij} = \mathbf{X}_{ij} - \sum_{k=1}^K \mathbf{U}_{ik} \mathbf{V}_{kj}$ represents the difference between the NMF predicted value and the actual value. $\sigma(\cdot)$ indicates the noise or normal data, which is processed using different loss functions. $\mathbf{W} \in \mathbb{R}^{m \times n}$ is the introduced auxiliary variable. $\phi(\mathbf{W}_{ij})$ is the conjugate function of \mathbf{Z}_{ij} . $\tau(\cdot, \cdot)$ is a quadratic term for \mathbf{Z}_{ij} and \mathbf{W}_{ij} . The NMF model only needs to consider the quadratic term of the multiplication form:

$$\tau(\mathbf{Z}_{ij}, \mathbf{W}_{ij}) = \frac{1}{2} \mathbf{W}_{ij} \mathbf{Z}_{ij}^2. \quad (9)$$

Combine Equation (8) and Equation (9) with the loss function (7):

$$\begin{aligned} & \min_{\mathbf{U} \geq 0, \mathbf{V} \geq 0} \delta(\mathbf{X} - \mathbf{UV}) + \alpha \text{tr}(\mathbf{VLV}^T) + \beta \|\mathbf{V}\|_{2,1} \\ & = \min_{\mathbf{U} \geq 0, \mathbf{V} \geq 0} \mathbf{W} \otimes (\mathbf{X} - \mathbf{UV})^2 + \phi(\mathbf{W}) + \alpha \text{tr}(\mathbf{VLV}^T) + \beta \|\mathbf{V}\|_{2,1}, \end{aligned} \quad (10)$$

where \otimes represents the Hadamard product, which is the product between two matrices' elements. Operator \otimes takes precedence over other matrices operators. Its Lagrangian function expansion is expressed as follows:

$$\mathcal{L}_{Huber-SGNMF}(\mathbf{U}) = \sum_{i=1}^m (\mathbf{X}_{i*} - \mathbf{U}_{i*} \mathbf{V}_{*}) \mathbf{Q}_i (\mathbf{X}_{i*} - \mathbf{U}_{i*} \mathbf{V}_{*})^T + \text{tr}(\boldsymbol{\Psi} \mathbf{U}^T), \quad (11)$$

and

$$\begin{aligned} \mathcal{L}_{Huber-SGNMF}(\mathbf{V}) &= \sum_{j=1}^n (\mathbf{X}_{*j} - \mathbf{U}_{*j} \mathbf{V}_{*j}) \mathbf{R}_j (\mathbf{X}_{*j} - \mathbf{U}_{*j} \mathbf{V}_{*j})^T \\ &+ \alpha \text{tr}(\mathbf{VLV}^T) + \beta \text{tr}(\mathbf{VGV}^T) + \text{tr}(\boldsymbol{\Psi} \mathbf{U}^T) + \text{tr}(\boldsymbol{\Phi} \mathbf{V}^T), \end{aligned} \quad (12)$$

where \mathbf{Q}_i and \mathbf{R}_j are defined as $\mathbf{Q}_i = \text{diag}(\mathbf{W}_{i*})$ and $\mathbf{R}_j = \text{diag}(\mathbf{W}_{*j})$, respectively. $\boldsymbol{\Psi} = [\psi_{ik}]$ and $\boldsymbol{\Phi} = [\phi_{kj}]$ are Lagrangian multipliers of non-negative constraints $\mathbf{U} \geq 0$ and $\mathbf{V} \geq 0$, respectively. \mathbf{G} is a diagonal matrix with diagonal elements, which is given by:

$$G_{jj} = 1 / \sqrt{\sum_{m=1}^k v_{mj}^2 + \omega} \quad (13)$$

where ω is a number that is very close but not equal to zero.

Let $\boldsymbol{\Psi} \mathbf{U} = 0$ and $\boldsymbol{\Phi} \mathbf{V} = 0$ by using Karush–Kuhn–Tucker (KKT) (Qi and Jiang, 1997) conditions. The loss function (10) can be iteratively optimized by the following schemes:

Update \mathbf{W} when \mathbf{U} and \mathbf{V} are fixed. The weight matrix \mathbf{W} according to equation (8) is defined as follows:

$$w_{ij} = \frac{\sigma'(\mathbf{Z}_{ij})}{\mathbf{Z}_{ij}}, \quad (14)$$

where the elements of weight matrix is $w_{ij} \in \mathbf{W}$. Combine the loss function (7) with the equation (14) are as follows:

$$w_{ij} = \begin{cases} 1 & \text{if } |x_{ij} - u_{ik} v_{kj}| \leq c, \\ \frac{c}{|\mathbf{X} - \mathbf{UV}|_{ij}} & \text{otherwise,} \end{cases} \quad (15)$$

Update \mathbf{U} and \mathbf{V} when \mathbf{W} is fixed. The update rules for \mathbf{U} and \mathbf{V} are as follows:

$$\begin{aligned} u_{ik} &= u_{ik} \frac{(\mathbf{X}_{i*} \mathbf{Q}_i \mathbf{V}^T)_{ik}}{(\mathbf{U}_{i*} \mathbf{V} \mathbf{Q}_i \mathbf{V}^T)_{ik}} \\ &= u_{ik} \frac{(\mathbf{W} \otimes \mathbf{XV}^T)_{ik}}{(\mathbf{W} \otimes (\mathbf{UV}) \mathbf{V}^T)_{ik}}, \end{aligned} \quad (16)$$

$$\begin{aligned} v_{kj} &= v_{kj} \frac{(\mathbf{U}^T \mathbf{X}_{*j} \mathbf{R}_j)_{kj}}{(\mathbf{U}^T \mathbf{R}_j \mathbf{UV}_{*j} + \alpha \mathbf{VL} + \beta \mathbf{VG})_{kj}} \\ &= v_{kj} \frac{(\mathbf{U}^T (\mathbf{W} \otimes \mathbf{X}))_{kj}}{(\mathbf{U}^T (\mathbf{W} \otimes \mathbf{UV}) + \alpha \mathbf{VL} + \beta \mathbf{VG})_{kj}}, \end{aligned} \quad (17)$$

The threshold parameter c is set to the median of the reconstruction error,

$$c = \text{median} \left(\left| \mathbf{X} - \mathbf{UV} \right|_{ij} \right). \quad (18)$$

The corresponding algorithm is shown in Algorithm 1.

ALGORITHM 1 | Huber-SGNMF.

Data input: $\mathbf{X} \in \mathbb{R}^{m \times n}$, $\mathbf{L} \in \mathbb{R}^{m \times n}$
 Data output: $\mathbf{U} \in \mathbb{R}^{m \times k}$, $\mathbf{V} \in \mathbb{R}^{k \times n}$ and weight matrix $\mathbf{W} \in \mathbb{R}^{m \times n}$
 Parameters: α, β
 Data initialize: $\mathbf{U} \geq 0, \mathbf{V} \geq 0$
 Repeat
 Update \mathbf{G} by (13);
 Update \mathbf{W} by (15);
 Update \mathbf{u}_k by (16);
 Update \mathbf{v}_k by (17);
 Update c by (18);
 End convergence

Convergence Analysis

According to the update rules of Huber-SGNMF, the loss function $O_{\text{Huber-SGNMF}}$ can converge to the local optimum through theorem 1.

Theorem 1. The loss function (7) is guaranteed to be non-increasing under the update rules (16) and (17). The loss function is constant when the elements u_{ik} and v_{kj} have fixed values.

To prove theorem 1, we introduce the auxiliary function \mathbf{H} in Algorithm.

Lemma 1. Suppose $\mathbf{H}(r, r')$ is an auxiliary function of $F(r)$. If the conditions $\mathbf{H}(r, r') \geq F(r)$ and $\mathbf{H}(r, r) = F(r)$ are satisfied, then it can be concluded that $F(r)$ is non-increasing from iteration t to $t+1$:

$$r^{t+1} = \arg \min_r \mathbf{H}(r, r') \quad (19)$$

Proof:

$$F(r^{t+1}) \leq \mathbf{H}(r^{t+1}, r^t) \leq \mathbf{H}(r^t, r^t) = F(r^t). \quad (20)$$

Suppose loss function $O_{\text{Huber-SGNMF}}$ has a suitable auxiliary function $\mathbf{H}_{\text{Huber}}$. If the minimum updates rule for $\mathbf{H}_{\text{Huber}}$ is equal to (16) and (17), then the convergence of $O_{\text{Huber-SGNMF}}$ can be proved. Furthermore, the parts of the loss function $O_{\text{Huber-SGNMF}}$ associated with the elements $u_{ik} \in \mathbf{U}$ and $v_{kj} \in \mathbf{V}$ are represented by F_{ik} and F_{kj} , respectively. The partial derivative equation of $O_{\text{Huber-SGNMF}}$ can be derived as follows:

$$F'_{ik} = \left(\frac{\partial O_{\text{Huber-SGNMF}}}{\partial \mathbf{U}} \right)_{ik} = \left(-2\mathbf{X}_{i*} \mathbf{Q}_i \mathbf{V}^T + 2\mathbf{U}_{i*} \mathbf{V} \mathbf{Q}_i \mathbf{V}^T \right)_{ik}, \quad (21)$$

$$F''_{ik} = \left(\frac{\partial^2 O_{\text{Huber-SGNMF}}}{\partial \mathbf{U}^2} \right)_{ik} = 2 \left(\mathbf{V} \mathbf{Q}_i \mathbf{V}^T \right)_{kk}, \quad (22)$$

$$F'_{kj} = \left(\frac{\partial O_{\text{Huber-SGNMF}}}{\partial \mathbf{V}} \right)_{kj} = -2 \left(\mathbf{U}^T \mathbf{X}_{*j} \mathbf{R}_j \right)_k + 2 \left(\mathbf{U}^T \mathbf{R}_j \mathbf{U} \mathbf{V}_{*j} \right)_k + 2\alpha \mathbf{V} \mathbf{L} + 2\beta \mathbf{V} \mathbf{G}, \quad (23)$$

$$F''_{kj} = \left(\frac{\partial^2 O_{\text{Huber-SGNMF}}}{\partial \mathbf{V}^2} \right)_{kj} = 2 \left(\mathbf{U}^T \mathbf{R}_j \mathbf{U} \right)_{kk} + (2\alpha \mathbf{L})_{jj} + (2\beta \mathbf{G})_{jj}. \quad (24)$$

Essentially, the algorithm updates each element, which means that if the elements F_{ik} and F_{kj} are non-increasing, then $O_{\text{Huber-SGNMF}}$ is also non-increasing.

Lemma 2. Define $\mathbf{H}_{\text{Huber}}(u, u_{ik}^t)$ and $\mathbf{H}_{\text{Huber}}(v, v_{kj}^t)$ as auxiliary functions for \mathbf{u}_{ik} and \mathbf{v}_{kj} , respectively. The expansion items are as follows:

$$\mathbf{H}_{\text{Huber}}(u, u_{ik}^t) = F_{ik}(u_{ik}^t) + F'_{ik}(u_{ik}^t)(u - u_{ik}^t) + \frac{(\mathbf{U}_{i*} \mathbf{V} \mathbf{Q}_i \mathbf{V}^T)_{ik}}{u_{ik}^t} (u - u_{ik}^t)^2, \quad (25)$$

$$\mathbf{H}_{\text{Huber}}(v, v_{kj}^t) = F_{kj}(v_{kj}^t) + F'_{kj}(v_{kj}^t)(v - v_{kj}^t) + \frac{(\mathbf{U}^T \mathbf{R}_j \mathbf{U} \mathbf{V}_{*j} + \alpha \mathbf{V} \mathbf{L} + \beta \mathbf{V} \mathbf{G})_{kj}}{v_{kj}^t} (v - v_{kj}^t)^2. \quad (26)$$

Proof:

According to the lemma 1, $\mathbf{H}_{\text{Huber}}(u, u) = F_{ik}(u)$ and $\mathbf{H}_{\text{Huber}}(v, v) = F_{kj}(v)$ can be obtained. We have the following formulas through the Taylor series expansion of the auxiliary function.

$$F_{ik}(u) \approx F_{ik}(u_{ik}^t) + F'_{ik}(u_{ik}^t)(u - u_{ik}^t) + \frac{1}{2} F''_{ik}(u_{ik}^t)(u - u_{ik}^t)^2, \quad (27)$$

$$F_{kj}(v) \approx F_{kj}(v_{kj}^t) + F'_{kj}(v_{kj}^t)(v - v_{kj}^t) + \frac{1}{2} F''_{kj}(v_{kj}^t)(v - v_{kj}^t)^2. \quad (28)$$

Next, $\mathbf{H}_{\text{Huber}}(u, u_{ik}^t) \geq F_{ik}(u)$ and $\mathbf{H}_{\text{Huber}}(v, v_{kj}^t) \geq F_{kj}(v)$ need to be guaranteed.

According to (25) and (27), expand $\mathbf{H}_{\text{Huber}}(u, u_{ik}^t) \geq F_{ik}(u)$ is as follows:

$$\frac{(\mathbf{U}_{i*} \mathbf{V} \mathbf{Q}_i \mathbf{V}^T)_{ik}}{u_{ik}^t} \geq \mathbf{V} \mathbf{Q}_i \mathbf{V}^T, \quad (29)$$

since

$$\left(\mathbf{U}_{i^*} \mathbf{V} \mathbf{Q}_i \mathbf{V}^T\right)_{ik} = \sum_{a=1}^K u_{ia} \left(\mathbf{V} \mathbf{Q}_i \mathbf{V}^T\right)_{ka} \geq u_{ik} \left(\mathbf{V} \mathbf{Q}_i \mathbf{V}^T\right)_{kk}. \quad (30)$$

According to (26) and (28), expand $\mathbf{H}_{Huber}(v, v_{kj}^t) \geq F_{kj}(v)$ is as follows:

$$\frac{\left(\mathbf{U}^T \mathbf{R}_j \mathbf{U} \mathbf{V}_{*j} + \alpha \mathbf{V} \mathbf{L} + \beta \mathbf{V} \mathbf{G}\right)_{kj}}{v_{kj}^t} \geq \left(\mathbf{U}^T \mathbf{R}_j \mathbf{U}\right)_{kk} + (\alpha \mathbf{L})_{jj} + (\beta \mathbf{G})_{jj}, \quad (31)$$

since

$$\left(\mathbf{U}^T \mathbf{R}_j \mathbf{U} \mathbf{V}_{*j}\right)_{kj} = \sum_{b=1}^K \left(\mathbf{U}^T \mathbf{R}_j \mathbf{U}\right)_{bk} v_{bj} \geq \left(\mathbf{U}^T \mathbf{R}_j \mathbf{U}\right)_{kk} v_{kj}, \quad (32)$$

$$(\beta \mathbf{V} \mathbf{G})_{kj} = \beta \sum_{b=1}^N v_{kb} \mathbf{G}_{bb} \geq \beta v_{kj} \mathbf{G}_{jj}, \quad (33)$$

and

$$\begin{aligned} (\alpha \mathbf{V} \mathbf{D})_{kj} &= \alpha \sum_{c=1}^N v_{kc} \mathbf{D}_{cc} \geq \alpha v_{kj} \mathbf{D}_{jj} \\ &\geq \alpha v_{kj} (\mathbf{D} - \mathbf{W})_{jj} = \alpha v_{kj} \mathbf{L}_{jj}. \end{aligned} \quad (34)$$

So, $\mathbf{H}_{Huber}(u, u_{ik}^t) \geq F_{ik}(u)$ and $\mathbf{H}_{Huber}(v, v_{kj}^t) \geq F_{kj}(v)$ can be obtained. In other words, the auxiliary functions $F_{ik}(u)$ and $F_{kj}(v)$ of the updated rules (16) and (17) are non-increasing, and the derivation of theorem 1 is completed. Finally, the convergence of the loss function $O_{Huber-SGNMF}$ is proved.

The corresponding convergence analysis curve is shown in Figure 2.

RESULTS AND DISCUSSION

Datasets

Five gene expression datasets downloaded from TCGA are used in the experiment. TCGA is a gene data sharing system that contains information on thousands of cancer patients and has made great contributions to the path of human exploration of cancer genomics. The experiment used five datasets including cholangiocarcinoma (CHOL), colon adenocarcinoma (COAD), head and neck squamous cell carcinoma (HNSC), pancreatic cancer (PAAD), and esophageal cancer (ESCA).

To explore the association between genes and multiple cancers, diseased samples from multiple datasets are integrated into one dataset. In detail, the detesteds PAAD, HNSC, and COAD are integrated into one dataset, which is represented as PHD. The detesteds PAAD, HNSC, and COAD are integrated into one

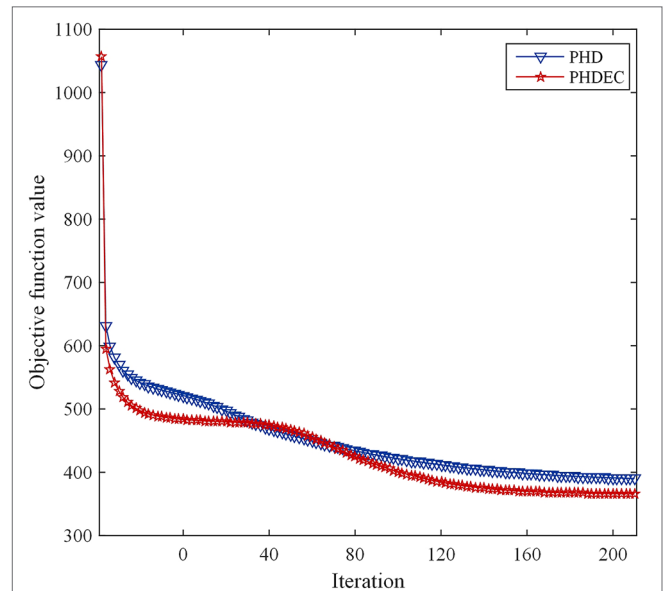


FIGURE 2 | Convergence analysis curve of Huber-SGNMF model. Each curve represents a dataset. PHD and PHDEC are the datasets used in the experiment.

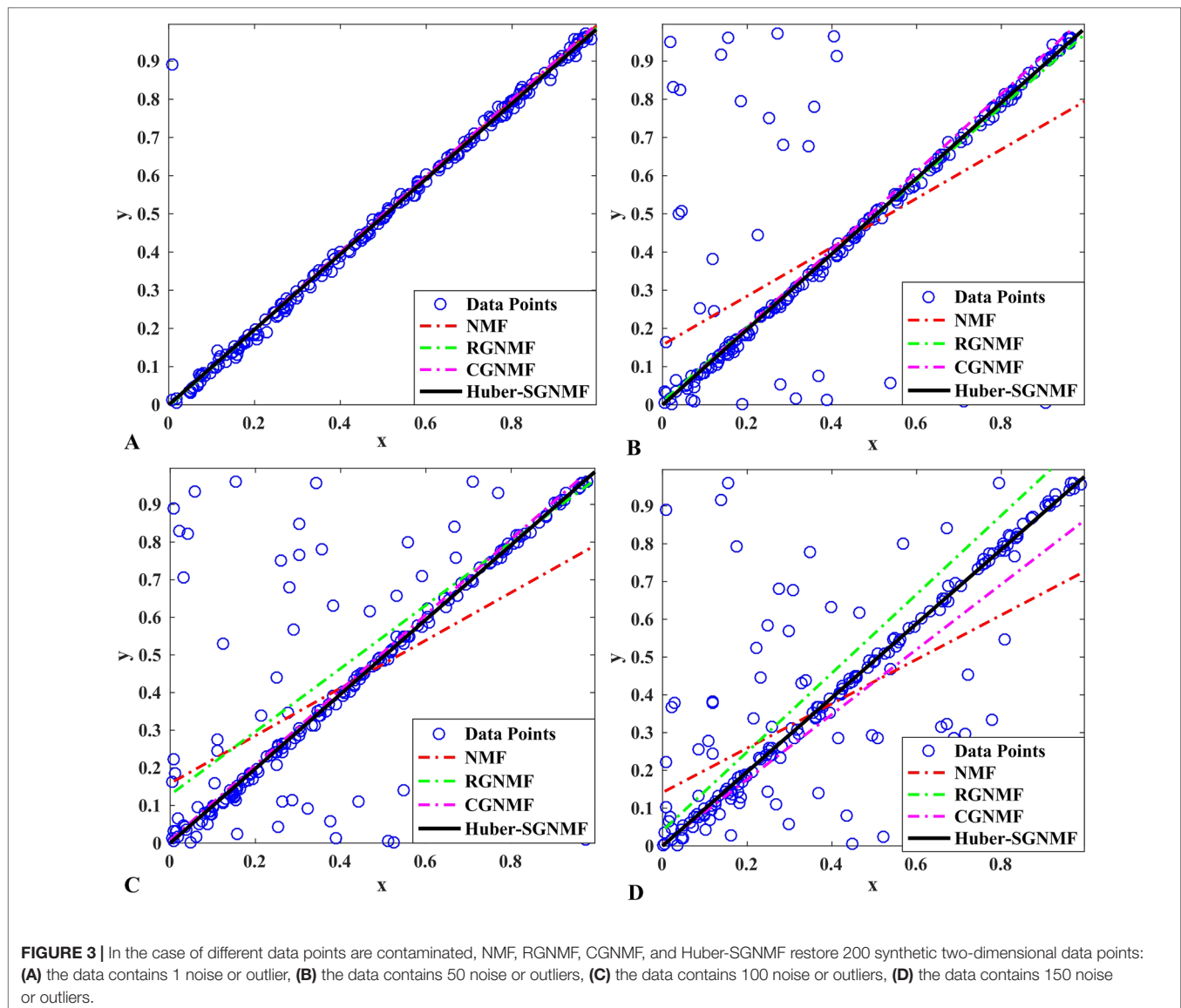
dataset, which is represented as PHD. These two integrated datasets contain only diseased samples of different diseases. Datasets are standardized before using, and the data normalization scales data to specific time intervals. Pre-processing data speeds up searching for the best solution and optimizes convergence speed. Since high-dimensional gene expression data contains a large amount of redundant information, PCA (Wu et al., 2017) is used to reduce the dimensions to 2,000 genes in the pre-processing.

Model Robustness

To analyze the robustness of RGNMF, CGNMF, and Huber-SGNMF, we apply these methods to a composite dataset consisting of 200 two-dimensional data points (Figure 3A). All data points are distributed in one dimensional space. In Figure 3A, there is only one contaminated point, and each model can restore the original data normally. The contaminated points in Figures 3B–D are 50 points, 100 points, and 150 points, respectively. In the case where a part of the data is contaminated, only Huber-SGNMF successfully restores the original data. CGNMF and RGNMF are affected by some noise or outliers when restoring data, while NMF is most affected by noise or outliers.

Parameter Selection

In the experiment, we consider the effect of each parameter on the solution model. A grid search method is used to find the optimal parameters of the model. The grid search range is $[10^{-2} \sim 10^2]$. As shown in Figure 4, the PHD dataset is used as an example to find the optimal parameters of the Huber-SGNMF model. Specifically, the two datasets are set to the same parameters $\alpha = 100$ and $\beta = 0.01$. Other methods in the experiment are set up with prior parameters or grid searches to find the optimal parameters.



Performance Evaluation and Comparisons

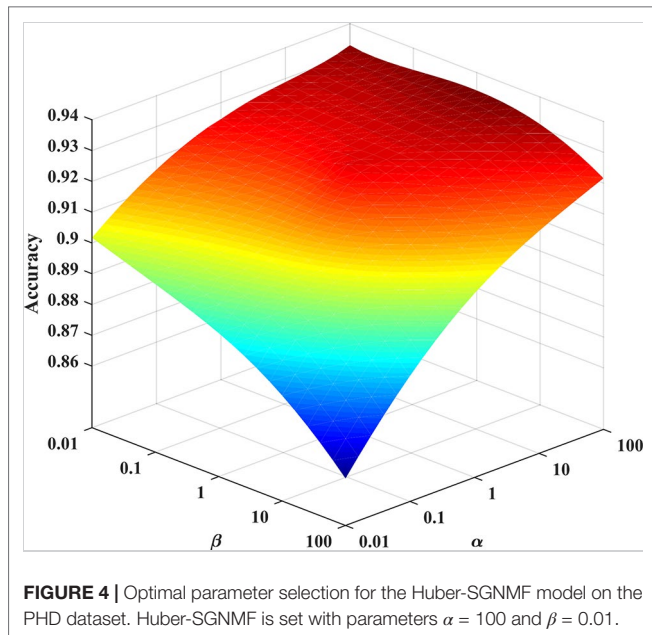
To prove the validity of the performance of the model, six states of the art methods including RPCA (Liu et al., 2013), NMF (Lee and Seung, 1999), SNMF (Kim and Park, 2007), GNMF (Cai et al., 2011), RGNMF (Wang et al., 2016a), CGNMF (Mao et al., 2014), and Huber-NMF (Du et al., 2012) are compared with Huber-SGNMF. In the experiment, the basic matrix and the coefficient matrix are used to differentially gene selection and cluster analysis, respectively.

Feature Selection Results and Analysis

Feature selection is the selection of representative features from multiple feature values (Yu and Liu, 2003). In the analysis of cancer data, the feature selection is to find differentially expressed genes for cancer (that is, pathogenic genes). This is of great significance in exploring the link between cancer and genes (Chen et al., 2017). For each method, the top 500 genes with the greatest differential expression are analyzed.

The GeneCards (<https://www.genecards.org/>) system is used to download all gene libraries associated with the disease. The selected genes are compared with the gene bank to select overlapping genes and obtain a corresponding relevance score. The relevance score is the indicator that GeneCards assesses the association between the gene and the disease. The higher the relevance score is, the greater the intimacy of the gene and the disease. The average relevance score (ARS) and the maximum relevance score (MRS) are used to evaluate the performance of the model.

The specific experimental results of the seven methods are listed in **Table 1**. The results show that the genes selected by Huber-SGNMF model have higher ARS and MRS. This means that the model can effectively find the genes associated with cancer. **Table 2** lists the genes for the top 10 largest relevance scores selected by the Huber-SGNMF model on the PHD dataset. The detailed genetic analysis is as follows.



CTNNB1 is a protein-coding gene from which the protein encoded by the gene forms part of an adhesion-linked protein complex. Mutations in the CTNNB1 proto-oncogene are associated with most human colorectal epithelial tumors, and a significant increase in expression in the same tumor may indirectly or directly lead to intestinal adenocarcinoma (Wang et al., 2011). Moreover, deep sequencing of patients with pancreatic ductal adenocarcinoma also found CTNNB1 mutations (Honda et al., 2013; Javadinia et al., 2019). Multiple studies have shown that CTNNB1 mutation analysis is important for PAAD and COAD (Kubota et al., 2015).

ERBB2, commonly referred to as HER2, may be critical for enhancing the synergistic effect of PI3K inhibitors in HNSC patients (Michmerhuizen et al., 2019). It is generally believed that dysregulated ERBB2 signaling plays a key role in the development of pancreatic cancer (Lin et al., 2019). For the treatment of intestinal adenocarcinoma, ERBB2 mutations and amplification in small intestinal adenocarcinoma patients also make a great contribution (Adam et al., 2019). Recent studies have shown that HER2 targeted therapy has significantly improved outcomes in patients with breast and stomach problems with ERBB2 mutation/amplification (Merik-Bernstam et al., 2019).

The CDH1 gene plays a regulatory role in cell growth (Nagai et al., 2018), and the CDH1 gene located on chromosome 16q22.1 is considered to be a tumor suppressor of diffuse gastric cancer. By measuring the methylation profile of gastric cancer and breast cancer patients, it is found that CDH1 is closely related to low protein expression (Wang et al., 2016b; Wang et al., 2016c). Studies have shown that abnormal expression of CDH1 gene leads to uncontrolled growth of tumor cells (Dial et al., 2007; Chen et al., 2012).

The above experimental results show that Huber-SGNMF model can find pathogenic genes more effectively. Although some genes have not been confirmed, they may be a key part of solving cancer problems in the future.

Clustering Results and Analysis

After the Huber-SGNMF model reduces the dimensions of the data, the coefficient matrix is used for k-means clustering. Sample clustering is a common analytical method for cancer diagnosis and molecular subtype discrimination (Xu et al., 2019). Moreover, multiple evaluation criteria accuracy (ACC), recall (R), precision (P), and F-measure (F) are adopted to judge the model to be feasible and effective. ACC is an evaluation standard

TABLE 1 | Relevance scores for seven methods.

	NMF	SNMF	GNMF	RGNMF	RPCA	CGNMF	Huber-NMF	Huber-SGNMF	
PHD	MRS	116.4	113.99	116.4	164.03	194.01	113.99	164.03	164.03
	ARS	22.64	21.75	22.03	22.16	26.03	21.75	25.56	27.19
PHDEC	MRS	92.51	96.36	153.66	124.37	172.9	164.91	145	172.9
	ARS	29.18	30.05	36.58	27.87	37.83	35.07	35.93	44.97

Bolded texts denoted best experimental results.

TABLE 2 | Detailed analysis of the differentially expressed genes in PHD dataset.

Gene name	Relevance score	Gene official name	Related diseases
CTNNB1	164.03	Catenin beta 1	Colorectal cancer and pilomatrixoma
ERBB2	152.33	Erb-B2 receptor tyrosine kinase 2	Lung cancer and ovary adenocarcinoma
CDH1	149.92	Cadherin 1	Gastric cancer and breast cancer
TGFBR2	102.74	Transforming growth factor beta receptor 2	Colorectal cancer and esophageal cancer
CDK4	93.35	Cyclin dependent kinase 4	Myeloma and melanoma
EPCAM	86.79	Epithelial cell adhesion molecule	Pancreatic cancer and gastrointestinal carcinoma
GNAS	76.17	GNAS complex locus	Osseous heteroplasia
ERBB3	74.35	Erb-B2 receptor tyrosine kinase 3	Transitional cell carcinoma
CEACAM5	59.9	Carcinoembryonic antigen related cell Adhesion molecule 5	Colorectal cancer and lung cancer
MAP2K2	51.51	Mitogen-activated protein kinase kinase 2	Head and neck squamous cell carcinoma

that can visually reflect the clustering of samples. It is defined as follows:

$$ACC = \frac{\sum_{i=1}^n \delta(a_i, \text{map}(b_i))}{n}, \quad (35)$$

Where $\delta(\bullet)$ and $\text{map}(\bullet)$ represent function permutation and delta mapping function, respectively. The actual sample label, the predicted sample label, and the total number of samples are denoted by a , b and n , respectively.

Considering clustering accuracy alone does not fully demonstrate clustering performance, and more evaluation criteria need to be introduced. The clustering results can be divided into true positive (TP), true negative (TN), false positive (FP) cases, and false negative (FN) according to real and predictive labels. These four measures are listed in **Table 3**. The detailed evaluation criteria are as follows.

$$R = \frac{TP}{TP + FN}, \quad (36)$$

$$P = \frac{TP}{TP + FP}, \quad (37)$$

$$F = \frac{2 \times R \times P}{R + P}. \quad (38)$$

Since R , P , and F can only reflect the clustering performance of a certain sample categories, for multi-category problems, the average of each category of indicators is usually used as the evaluation criterions:

$$Macro-R = \frac{1}{n} \sum_{i=1}^n R_i, \quad (39)$$

TABLE 3 | Clustering result confusion matrix.

The true situation	Clustering result	
	Positive	Negative
Positive	TP (true positive)	FN (false negative)
Negative	FP (false positive)	TN (true negative)

TABLE 4 | Clustering effect for seven methods.

Dataset	Evaluation	NMF	SNMF	GGMF	RGNMF	RPCA	CGNMF	Huber-NMF	Huber-SGNMF
PHD	ACC (%)	85.38 ± 1.24	88.93 ± 0.58	86.05 ± 1.97	86.50 ± 1.84	86.37 ± 2.04	87.18 ± 1.43	88.55 ± 0.98	90.36 ± 0.91
	Macro-R (%)	82.99 ± 1.57	86.86 ± 0.82	81.02 ± 1.09	84.28 ± 2.40	84.10 ± 2.79	85.00 ± 1.79	86.41 ± 1.27	88.50 ± 1.19
	Macro-P (%)	84.88 ± 1.74	89.08 ± 0.86	83.55 ± 3.76	85.68 ± 2.74	85.58 ± 2.96	86.77 ± 2.02	88.32 ± 1.36	90.18 ± 1.28
	Macro-F (%)	83.92 ± 1.65	87.95 ± 0.84	82.25 ± 3.51	84.92 ± 2.60	84.83 ± 2.88	85.87 ± 1.90	87.35 ± 1.31	89.33 ± 1.23
PHDEC	ACC (%)	69.84 ± 0.26	71.51 ± 0.31	70.15 ± 0.08	71.86 ± 0.69	75.02 ± 0.32	73.81 ± 0.27	72.53 ± 0.21	75.52 ± 0.20
	Macro-R (%)	63.95 ± 0.18	65.33 ± 0.14	61.98 ± 0.38	64.45 ± 0.87	68.37 ± 0.28	66.74 ± 0.15	67.09 ± 0.07	69.02 ± 0.07
	Macro-P (%)	61.34 ± 0.26	62.45 ± 0.19	58.77 ± 0.10	62.80 ± 0.97	65.81 ± 0.50	64.47 ± 0.27	63.92 ± 0.25	65.56 ± 0.25
	Macro-F (%)	64.17 ± 0.20	63.79 ± 0.21	60.24 ± 0.27	63.49 ± 0.87	66.92 ± 0.29	65.51 ± 0.17	65.34 ± 0.12	67.17 ± 0.10

Bolded texts denoted best experimental results.

$$Macro-P = \frac{1}{n} \sum_{i=1}^n P_i, \quad (40)$$

$$Macro-F = \frac{2 \times Macro-R \times Macro-P}{Macro-R + Macro-P}, \quad (41)$$

where n represents the number of sample categories.

According to the above evaluation criterions, each algorithm is performed 50 times to get an average result. Since the initialization matrix is random, the average value can reduce the chance of the algorithm. **Table 4** lists the comparative experiments of seven methods based on four evaluation criterions. Compared with the other six methods, our proposed model has the excellent clustering performance under the four evaluation criterions. The specific analysis of the clustering results is as follows:

1. Since the squared loss of the original NMF is sensitive to noise and outliers, the squared loss is replaced by Huber loss to improve the robustness of the algorithm. The experimental results show that the clustering performance of RPCA, CGNMF, RGNMF, Huber-NMF, and Huber-SGNMF is higher than standard NMF and GGMF. The reason is that both NMF and GGMF use square loss while other methods use more robust loss function. Moreover, the experimental results show that the robustness of the Huber loss model is higher than the $L_{2,1}$ -norm loss and correntropy loss. The RPCA model has higher performance as a state-of-the-art algorithm and is still lower than Huber-SGNMF. The Huber loss use L_1 -norm or L_2 -norm to different data, which can effectively reduce the influence of noise and outliers and enhance the robustness of the algorithm. Compared with NMF, the clustering accuracy of Huber-SGNMF model on the two datasets increased by 4.90 and 5.68%, respectively.
2. Assuming that data points are related in a high-dimensional state, they should also be relevant in low-dimensional representations. However, the association between data points is difficult to preserve when the data is mapped to low-dimensions. The manifold structure preserves the spatial structure of high-dimensional data in low-dimensional representations, enhancing the correlation between data points. Constructing a sample association graph of gene expression data to preserve the relationship between the samples. The experimental results of several models (NMF

and GNMF, Huber-NMF, and Huber-SGNMF) show that the clustering performance of the model with the addition of graph regularity constraints is improved. Compared with Huber-NMF, Huber-SGNMF has improved clustering accuracy by 1.73 and 2.99% in the two datasets, respectively.

- Matrix sparseness removes redundant data and simplifies model calculations. The sparsity constraint of the coefficient matrix removes redundant features and improves clustering performance. The experimental results of SNMF and Huber-SGNMF prove this. Compared with SNMF, since Huber-SGNMF improves the loss function and manifold structure, the clustering accuracy in the two datasets is increased by 1.35 and 4.02%, respectively.

In summary, the experimental results based on the four evaluation indicators demonstrate the excellent clustering performance of the Huber-SGNMF model. Compared with NMF, the clustering performance of Huber-SGNMF has improved 5.30 and 4.49% on average in PHD dataset and PHDEC dataset, respectively. Huber-SGNMF clustering performance improves 1.93 and 2.07% on average compared to Huber-NMF. The above experimental results strongly prove the effectiveness of Huber-SGNMF in clustering performance.

CONCLUSION

In this paper, we propose a novel model based on Huber loss: Huber-SGNMF, which is dedicated to samples clustering and

differentially expressed gene selection. On the one hand, the squared loss is replaced by Huber loss to enhance algorithm robustness. On the other hand, sparse penalty and graph regularization terms are added to the model to enhance the sparsity of the matrix and preserve data geometry information. Numerous experimental results confirm that the Huber-SGNMF method is more effective. In the future work, we will actively explore more effective constraints based on the traditional NMF method to improve the robustness and sparsity of the method.

DATA AVAILABILITY STATEMENT

The datasets for this study can be downloaded in the The Cancer Genome Atlas [<https://cancergenome.nih.gov/>].

AUTHOR CONTRIBUTIONS

C-YW and NY proposed and designed the algorithm. J-XL demonstrated the robustness of the algorithm and analyzed the experimental data. C-YW and C-HZ drafted the manuscript.

ACKNOWLEDGMENTS

This work was supported in part by the grants provided by the National Science Foundation of China, Nos. 61872220, 61873001, and 61572284.

REFERENCES

- Adam, L., San Lucas, F. A., Fowler, R., Yu, Y., Wu, W., Liu, Y., et al. (2019). DNA sequencing of small bowel adenocarcinomas identifies targetable recurrent mutations in the ERBB2 signaling pathway. *Clin. Cancer Res.* 25 (2), 641–651. doi: 10.1158/1078-0432.CCR-18-1480
- Belkin, M., and Niyogi, P. (2001). “Laplacian eigenmaps and spectral techniques for embedding and clustering,” in *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*. (MIT Press).
- Cai, D., He, X., Han, J., and Huang, T. S. (2011). Graph regularized non-negative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (8), 1548–1560. doi: 10.1109/TPAMI.2010.231
- Chen, J., Han, G., Xu, A., and Cai, H. (2019). Identification of multidimensional regulatory modules through multi-graph matching with network constraints. *IEEE Trans. Biomed. Eng.* doi: 10.1109/TBME.2019.2927157
- Chen, J., Peng, H., Han, G., Cai, H., and Cai, J. (2018). HOGMMNC: a higher order graph matching with multiple network constraints model for gene-drug regulatory modules identification. *Bioinformatics* 35 (4), 602–610. doi: 10.1093/bioinformatics/bty662
- Chen, M., Gutierrez, G. J., and Ronai, Z. A. (2012). The anaphase-promoting complex or cyclosome supports cell survival in response to endoplasmic reticulum stress. *PLoS One* 7 (4), e35520. doi: 10.1371/journal.pone.0035520
- Chen, X., Feng, M., Shen, C., He, B., Du, X., Yu, Y., et al. (2017). A novel approach to select differential pathways associated with hypertrophic cardiomyopathy based on gene coexpression analysis. *Mol. Med. Rep.* 16 (1), 773–777. doi: 10.3892/mmr.2017.6667
- Chreiky, R., Delmaire, G., Puigt, M., Roussel, G., and Abche, A. (2016). “Informed split gradient non-negative matrix factorization using huber cost function for source apportionment,” in: 2016 *IEEE International Symposium on Signal Processing and Information Technology (ISSPIT): IEEE*, 69–74. doi: 10.1109/ISSPIT.2016.7886011
- Dial, J. M., Petrotchenko, E. V., and Borchers, C. H. (2007). Inhibition of APC/Cdh1 activity by Cdh1/Acm1/Bmh1 ternary complex formation. *J. Biol. Chem.* 282 (8), 5237–5248. doi: 10.1074/jbc.M606589200
- Ding, C., Zhou, D., He, X., and Zha, H. (2006). “R 1-PCA: rotational invariant L 1-norm principal component analysis for robust subspace factorization,” in: *Proceedings of the 23rd international conference on Machine learning: ACM*, 281–288.
- Du, L., Li, X., and Shen, Y.-D. (2012). “Robust nonnegative matrix factorization via half-quadratic minimization,” in 2012 *IEEE 12th International Conference on Data Mining (ICDM)*, 201–210. doi: 10.1109/ICDM.2012.39
- Feng, C.-M., Xu, Y., Liu, J.-X., Gao, Y.-L., and Zheng, C.-H. (2019). Supervised discriminative sparse PCA for com-characteristic gene selection and tumor classification on multiview biological data. *IEEE Trans. Neural Netw. Learn. Syst.* doi: 10.1109/TNNLS.2019.2893190
- Guillamet, D., and Vitrià, J. (2002). “Non-negative matrix factorization for face recognition,” in *Catalonian Conference on Artificial Intelligence*. (Springer), 336–344. doi: 10.1007/3-540-36079-4_29
- Guo, S., Ling, X., Feng, L., Mingfeng, J., and Stuart, C. (2011). On epicardial potential reconstruction using regularization schemes with the L1-norm data term. *Phys. Med. Biol.* 56 (1), 57–72. doi: 10.1088/0031-9155/56/1/004
- Honda, S., Okada, T., Miyagi, H., Minato, M., Suzuki, H., and Taketomi, A. (2013). Spontaneous rupture of an advanced pancreaticoblastoma: Aberrant RASSF1A methylation and CTNBN1 mutation as molecular genetic markers. *J. Pediatr. Surg.* 48 (4), e29–e32. doi: 10.1016/j.jpedsurg.2013.02.038
- Javadinia, S. A., Shahidsales, S., Fanipakdel, A., Joudi-Mashhad, M., Mehrnaz, M., Talebian, S., et al. (2019). Therapeutic potential of targeting the Wnt/ β -catenin pathway in the treatment of pancreatic cancer. *J. Cell. Biochem.* 120 (5), 6833–6840. doi: 10.1002/jcb.27835

- Kim, H., and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. *Bioinformatics* 23 (12), 1495–1502. doi: 10.1093/bioinformatics/btm134
- Kong, D., Ding, C. H. Q., and Huang, H. (2011). “Robust nonnegative matrix factorization using L21-norm,” in *Proceedings of the 20th ACM international conference on Information and knowledge management (ACM)* 673–682. doi: 10.1145/2063576.2063676
- Kubota, Y., Kawakami, H., Natsuzaka, M., Kawakubo, K., Marukawa, K., Kudo, T., et al. (2015). CTNNB1 mutational analysis of solid-pseudopapillary neoplasms of the pancreas using endoscopic ultrasound-guided fine-needle aspiration and next-generation deep sequencing. *J. Gastroenterol.* 50 (2), 203–210. doi: 10.1007/s00535-014-0954-y
- Lee, D. D., and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 788. doi: 10.1038/44565
- Li, X., Cui, G., and Dong, Y. (2017). Graph regularized non-negative low-rank matrix factorization for image clustering. *IEEE Trans. Cybern.* 47 (11), 3840–3853. doi: 10.1109/TCYB.2016.2585355
- Lin, T., Ren, Q., Zuo, W., Jia, R., Xie, L., Lin, R., et al. (2019). Valproic acid exhibits anti-tumor activity selectively against EGFR/ErbB2/ErbB3-coexpressing pancreatic cancer via induction of ErbB family members-targeting microRNAs. *J. Exp. Clin. Cancer Res.* 38 (1), 150. doi: 10.1186/s13046-019-1160-9
- Liu, J.-X., Wang, D., Gao, Y.-L., Zheng, C.-H., Xu, Y., and Yu, J. (2018). Regularized non-negative matrix factorization for identifying differentially expressed genes and clustering samples: a survey. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 15 (3), 974–987. doi: 10.1109/TCBB.2017.2665557
- Liu, J.-X., Wang, Y.-T., Zheng, C.-H., Sha, W., Mi, J.-X., and Xu, Y. (2013). “Robust PCA based method for discovering differentially expressed genes,” in *BMC bioinformatics*. (BioMed Central), 14(8) S3. doi: 10.1186/1471-2105-14-S8-S3
- Liu, W., Pokharel, P. P., and Principe, J. C. (2007). Correntropy: properties and applications in non-gaussian signal processing. *IEEE Trans. Signal Process.* 55 (11), 5286–5298. doi: 10.1109/TSP.2007.896065
- Liu, X., Zhai, D., Zhao, D., Zhai, G., and Gao, W. (2014). Progressive image denoising through hybrid graph Laplacian regularization: a unified framework. *IEEE Trans. Image Process.* 23 (4), 1491–1503. doi: 10.1109/TIP.2014.2303638
- Luo, X., Zhou, M., Xia, Y., and Zhu, Q. (2014). An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. *IEEE Trans. Ind. Inform.* 10 (2), 1273–1284. doi: 10.1109/TII.2014.2308433
- Mao, B., Guan, N., Tao, D., Huang, X., and Luo, Z. (2014). “Correntropy induced metric based graph regularized non-negative matrix factorization,” *Proceedings 2014 IEEE international conference on Security, Pattern Analysis, and Cybernetics (SPAC)*. (IEEE), 163–168. doi: 10.1109/SPAC.2014.6982679
- Meric-Bernstam, F., Johnson, A. M., Dumbrava, E. E. I., Raghav, K., Balaji, K., Bhatt, M., et al. (2019). Advances in HER2-targeted therapy: novel agents and opportunities beyond breast and gastric cancer. *Clin. Cancer Res.* 25 (7), 2033–2041. doi: 10.1158/1078-0432.CCR-18-2275
- Michmerhuizen, N. L., Leonard, E., Matovina, C., Harris, M., Herbst, G., Kulkarni, A., et al. (2019). Rationale for using irreversible epidermal growth factor receptor inhibitors in combination with phosphatidylinositol 3-kinase inhibitors for advanced head and neck squamous cell carcinoma. *Mol. Pharmacol.* 95 (5), 528–536. doi: 10.1124/mol.118.115162
- Nagai, M., Shibata, A., and Ushimaru, T. (2018). Cdh1 degradation is mediated by APC/C-Cdh1 and SCF-Cdc4 in budding yeast. *Biochem. Biophys. Res. Commun.* 506 (4), 932–938. doi: 10.1016/j.bbrc.2018.10.179
- Nikolova, M., and Chan, R. H. (2007). The equivalence of half-quadratic minimization and the gradient linearization iteration. *IEEE Trans. Image Process.* 16 (6), 1623–1627. doi: 10.1109/TIP.2007.896622
- Qi, L., and Jiang, H. (1997). Semismooth Karush-Kuhn-Tucker equations and convergence analysis of newton and quasi-newton methods for solving these equations. *Math. Oper. Res.* 22 (2), 301–325. doi: 10.1287/moor.22.2.301
- Roweis, S. T., and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science* 290 (5500), 2323–2326. doi: 10.1126/science.290.5500.2323
- Schmidt, M. N., and Olsson, R. K. (2006). “Single-channel speech separation using sparse non-negative matrix factorization,” in *Ninth International Conference on Spoken Language Processing (IEEE)*.
- Wang, D., Liu, J.-X., Gao, Y.-L., Zheng, C.-H., and Xu, Y. (2016a). Characteristic gene selection based on robust graph regularized non-negative matrix factorization. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 13 (6), 1059–1067. doi: 10.1109/TCBB.2015.2505294
- Wang, H. L., Hart, J., Fan, L., Mustafi, R., and Bissonnette, M. (2011). Upregulation of glycogen synthase kinase β in human colorectal adenocarcinomas correlates with accumulation of CTNNB1. *Clin. Colorectal Cancer* 10 (1), 30–36. doi: 10.3816/CCC.2011.n.004
- Wang, Q., Wang, B., Zhang, Y.-M., and Wang, W. (2016b). The association between CDH1 promoter methylation and patients with ovarian cancer: a systematic meta-analysis. *J. Ovarian Res.* 9, 23. doi: 10.1186/s13048-016-0231-1
- Wang, Y. Q., Yuan, Y., Jiang, S., and Jiang, H. (2016c). Promoter methylation and expression of CDH1 and susceptibility and prognosis of eyelid squamous cell carcinoma. *Tumor Biol.* 37 (7), 9521–9526. doi: 10.1007/s13277-016-4851-2
- Wu, M.-J., Liu, J.-X., Gao, Y.-L., Kong, X.-Z., and Feng, C.-M. (2017). “Feature selection and clustering via robust graph-laplacian PCA based on capped L 1-norm,” in *2017 IEEE international conference on Bioinformatics and Biomedicine (BIBM)* (IEEE), 1741–1745. doi: 10.1109/BIBM.2017.8217923
- Xu, A., Chen, J., Peng, H., Han, G., and Cai, H. (2019). Simultaneous interrogation of cancer omics to identify subtypes with significant clinical differences. *Front. Genet.* 10, 236. doi: 10.3389/fgene.2019.00236
- Xu, W., Liu, X., and Gong, Y. (2003). “Document clustering based on non-negative matrix factorization,” in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval (ACM)*, 267–273.
- Yu, L., and Liu, H. (2003). “Feature selection for high-dimensional data: A fast correlation-based filter solution,” in *Proceedings of the 20th international conference on machine learning (ICML-03)*. (AAAI), 856–863.
- Yu, N., Liu, J.-X., Gao, Y.-L., Zheng, C.-H., Wang, J., and Wu, M.-J. (2017). “Graph regularized robust non-negative matrix factorization for clustering and selecting differentially expressed genes,” in *2017 IEEE international conference on Bioinformatics and Biomedicine (BIBM)* (IEEE), 1752–1756. doi: 10.1109/BIBM.2017.8217925
- Yu, T., Zhao, Z., Yan, Z., and Li, F. (2016). “Robust L1-norm matrixed locality preserving projection for discriminative subspace learning,” in *2016 International Joint Conference on Neural Networks (IJCNN)* (IEEE), 4199–4204.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wang, Liu, Yu and Zheng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.