

# $L_{(r,c)}$ 非负矩阵分解用于图像聚类

陈运川

贵州师范大学, 数学科学学院, 贵州 贵阳

收稿日期: 2023年2月27日; 录用日期: 2023年3月23日; 发布日期: 2023年3月30日

## 摘要

在非负矩阵分解用于聚类的过程中, 将每一张大小为  $r \times c$  灰度图像按列重排成一个样本向量会改变图像数据原有的空间结构, 而目标函数直接度量样本向量与重构向量之间的误差会进一步忽略原始图像与重构后图像列之间的相似度。因此, 为降低图像数据空间结构的改变对聚类效果的不利影响, 本文提出了一种新的损失函数来逐列度量原始图像数据与重构后的图像数据间的误差, 并推导了相应的数值算法。数值实验结果表明本文提出的算法有更好的聚类表现, 图像表示能力更强。

## 关键词

非负矩阵分解, 聚类,  $L_{(r,c)}$  函数, 鲁棒性

# $L_{(r,c)}$ Nonnegative Matrix Factorization for Image Clustering

Yunchuan Chen

School of Mathematical Sciences, Guizhou Normal University, Guiyang Guizhou

Received: Feb. 27<sup>th</sup>, 2023; accepted: Mar. 23<sup>rd</sup>, 2023; published: Mar. 30<sup>th</sup>, 2023

## Abstract

In the process of non-negative matrix factorization for clustering, rearranging each grayscale image with size  $r \times c$  into a sample vector will change the original spatial structure of the image data, and the objective function directly measures the error between the sample vector and the reconstruction vector to further ignore the similarity between the original image and the reconstructed image column. Therefore, in order to reduce the adverse effect of the change of spatial structure of image data on the clustering effect, a new loss function is proposed to measure the error between

文章引用: 陈运川.  $L_{(r,c)}$  非负矩阵分解用于图像聚类[J]. 应用数学进展, 2023, 12(3): 1306-1316.

DOI: 10.12677/aam.2023.123133

the original image data and the reconstructed image data column-by-column, and the corresponding numerical algorithm is derived. Numerical experimental results show that the proposed algorithm has better clustering performance and stronger image representation ability.

## Keywords

Non-Negative Matrix Factorization, Clustering,  $L_{(r,c)}$  Function, Robustness

Copyright © 2023 by author(s) and Hans Publishers Inc.

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

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



Open Access

## 1. 引言

Lee 和 Seung 于 1999 年提出非负矩阵分解(NMF) [1]以来, NMF 已广泛应用于数据挖掘和机器学习等领域。作为一种数据降维方法, 非负矩阵分解将数据矩阵近似表示成两个非负因子矩阵的乘积, 非负性约束使得该方法具有整体由部分表示的可解释性[2]。这使得 NMF 在文本挖掘[3]、模式识别[4]、DNA 基因表达分析[5]等诸多方面得以广泛应用。根据具体问题的不同, 人们提出了具有不同目标函数的 NMF 的衍变算法并将之应用于协同过滤、分类、约束聚类[6]等数据分析任务。C Ding 等理论上分析了 NMF 与 K-means 聚类的潜在联系[7]。

在图像聚类任务中, NMF 以矩阵的 Frobenius 范数作为目标函数度量样本数据矩阵与重构后的矩阵元素间的整体误差, 由于该目标函数对噪声和异常值不稳定[8], 且忽视了原始图像数据与重构后的图像列与列之间的近似关系, 是重构后的数据矩阵对原始图像数据的“粗糙”近似。C Ding [9]等使用矩阵的  $L_{2,1}$  范数作为目标函数, 提出具有鲁棒性(robustness)的  $L_{2,1}$ NMF, 相较于 NMF 而言,  $L_{2,1}$ NMF 的目标函数度量是原始图像数据重排后的样本向量与重构后的矩阵列向量之间的误差, 虽然相较于 NMF 而言,  $L_{2,1}$ NMF 能够获得更为精细的图像表示, 但仍然没有考虑到原始图像数据和重构后图像的列之间的近似关系。

因此, 为使目标函数能够度量原始图像数据和重构后图像列之间的近似关系, 从而减少原始图像数据空间结构的改变产生的影响, 本文定义了向量的  $L_{(r,c)}$  函数并提出相应的  $L_{(r,c)}$  NMF 算法, 该算法具有与  $L_{2,1}$ NMF 相当的鲁棒性的同时, 兼顾了原始图像数据的空间结构, 实验结果上看  $L_{(r,c)}$  NMF 算法也具有更好的图像表示能力和图像聚类表现。

后续内容安排如下: 第二小节介绍了相关工作; 第三小节定义  $L_{(r,c)}$  函数, 并定义  $L_{(r,c)}$  NMF 算法的目标函数, 从不等式推导和数值实验两方面说明  $L_{(r,c)}$  NMF 算法的有效性; 第四节推导出算法的更新公式; 第五小节引入辅助函数分析算法的收敛性; 第六小节展示了  $L_{2,1}$ NMF 算法和三种对比算法在三个数据集上的实验结果。

## 2. 相关工作

NMF 是为了得到数据的部分表示[1]。给定一个非负矩阵  $X \in \mathbb{R}^{m \times n}$ , NMF 要将  $X$  近似表示成两个非负因子矩阵  $U \in \mathbb{R}^{m \times k}$  和  $V \in \mathbb{R}^{k \times n}$  的乘积, 即:

$$X \approx UV \quad (1)$$

为了求解  $U$  和  $V$ , 对如下目标函数最小化:

$$\begin{aligned} Q_{NMF} &= \|X - UV\|_F^2 = \sum_{i,j} (X - UV)_{ij}^2 \\ \text{s.t. } U &\geq 0, V \geq 0 \end{aligned} \quad (2)$$

其中  $\|A\|_F^2 = \sum_{i,j} A_{ij}^2$  表示矩阵  $A$  的 Frobenius 范数。

Lee 和 Seung [2] 给出了相应的更新公式:

$$U_{ij} = U_{ij} \frac{(XV^T)_{ij}}{(UVV^T)_{ij}} \quad (3)$$

$$V_{ij} = V_{ij} \frac{(U^T X)_{ij}}{(U^T UV)_{ij}} \quad (4)$$

然而, NMF 以 Frobenius 范数为目标函数会导致该算法对噪声和异常值敏感[8], 为解决这一问题, D Kong 和 C Ding 等提出了具有鲁棒性的  $L_{2,1}$ NMF, 该算法以矩阵的  $L_{2,1}$  范数为目标函数, 定义如下:

$$\begin{aligned} Q_{L_{2,1}NMF} &= \|X - UV\|_{2,1} = \sum_j \|x_j - Uv_j\| \\ \text{s.t. } U &\geq 0, V \geq 0 \end{aligned} \quad (5)$$

其中  $\|a\| = \sqrt{\sum_i a_i^2}$  表示向量的 2 范数。

文[11]中给出了相应的更新公式:

$$U_{ij} = U_{ij} \frac{(XDV^T)_{ij}}{(UVDV^T)_{ij}} \quad (6)$$

$$V_{ij} = V_{ij} \frac{(U^T XD)_{ij}}{(U^T UVD)_{ij}} \quad (7)$$

其中  $D$  是对角矩阵, 对角元素定义如下:

$$D_{jj} = 1/\|x_j - Uv_j\| \quad (8)$$

在图像聚类任务中, 图像数据矩阵  $X \in \mathbb{R}^{r \times n}$  的每一列  $x_j \in \mathbb{R}^{r \times 1}$  都是由一张分辨率为  $r \times c$  的灰度图像按列重排后的样本向量, 即  $x_j$  的每  $r$  个元素就是灰度图像中的一列。对比式(2)和式(5)可以看出, NMF 度量  $X$  与  $UV$  所有元素间的误差, 而  $L_{2,1}$ NMF 度量的是样本向量  $x_j$  与重构后的数据矩阵列向量  $Uv_j$  间的误差。不难验证下式成立:

$$\sqrt{\sum_{i,j} (X - UV)_{ij}^2} \leq \sum_j \sqrt{\sum_i (X - UV)_{ij}^2} \quad (9)$$

上式表明, 样本数据矩阵与重构后数据矩阵列向量之间的误差达到设定的误差限时, 矩阵元素间的整体误差也能达到误差限, 反之则不然。这使得  $L_{2,1}$ NMF 能够度量样本数据矩阵与重构后数据矩阵列向量间的近似关系, 从实验结果上看,  $L_{2,1}$ NMF 表示出的图像较之 NMF 要更加精细。

然而,  $L_{2,1}$ NMF 度量的仍然忽略了原始图像数据与重构后的图像列之间的近似关系, 这使得样本向量与重构后的样本向量间误差较小时并不能保证原始图像数据的列向量间的误差也足够小。

因此, 本文考虑了新的目标函数来度量原始图像数据与重构后图像数据列之间的近似关系。

### 3. $\|\cdot\|_{(r,c)}$ 目标函数

#### 3.1. $\|\cdot\|_{(r,c)}$ 函数

给定  $x \in \mathbb{R}^{r \times 1}$ ，本文定义  $x$  的一个实值函数如下：

**定义 1**

$$\|x\|_{(r,c)} := \sum_{p=1}^c \sqrt{\sum_{i=r(p-1)+1}^{rp} x_i^2}, \quad r, c \in \mathcal{O}^+ \quad (10)$$

易验证， $\|\cdot\|_{(r,c)}$  满足范数的三条性质：

(1) 正定性 (2) 齐次性 (3) 三角不等式

对  $\forall x, y \in \mathbb{R}^{r \times 1}, a, b \in \mathbb{R}$ ，我们有：

$$\begin{aligned} \|ax\|_{(r,c)} + \|by\|_{(r,c)} &= \sum_{p=1}^c \sqrt{\left( \sqrt{\sum_{i=r(p-1)+1}^{rp} (ax_i)^2} + \sqrt{\sum_{i=r(p-1)+1}^{rp} (by_i)^2} \right)^2} \\ &\geq \sum_{p=1}^c \sqrt{\sum_{i=r(p-1)+1}^{rp} (ax_i)^2 + (by_i)^2 + 2ax_i by_i} \\ &= \sum_{p=1}^c \sqrt{\sum_{i=r(p-1)+1}^{rp} (ax_i + by_i)^2} \\ &= \|ax + by\|_{(r,c)} \end{aligned} \quad (11)$$

当且仅当  $x = y$  时，不等式取等，显然当  $a, b \geq 0$  且  $a + b = 1$ ， $x \neq y$  时式(11)严格成立，因此  $\|x\|_{(r,c)}$  是关于变量  $x$  的严格凸函数。对于向量  $x \in \mathbb{R}^{n \times 1}$ ，若  $r = n, c = 1$ ，该实值函数是  $x$  的  $l_2$  范数；若  $r = 1, c = n$ ，则该实值函数是  $x$  的  $l_1$  范数。限于定义中  $r, c$  是给定的，所以  $\|\cdot\|_{(r,c)}$  并不是严格意义上  $\mathbb{R}^{n \times 1}$  中的向量范数。

#### 3.2. $\|\cdot\|_{(r,c)}$ NMF 目标函数

为了对原始图像数据逐列度量误差，本文提出的  $L_{(r,c)}$  NMF 的目标函数定义如下：

$$\begin{aligned} \mathcal{Q}_{L_{(r,c)}\text{NMF}} &= \sum_j \|x_j - UV_j\|_{(r,c)} \\ \text{s.t. } &U \geq 0, V \geq 0 \end{aligned} \quad (12)$$

对于  $X, UV \in \mathbb{R}^{r \times n}$ ，将  $X, UV$  的每一列对应一张  $r \times c$  的图像，借由式(9)，不难验证下式成立：

$$\sqrt{\sum_i (X - UV)_{ij}^2} \leq \sum_p \sqrt{\sum_{i=r(p-1)+1}^{rp} (X - UV)_{ij}^2} \quad (13)$$

由上式可知，对于  $L_{(r,c)}$  NMF 而言，当原始图像数据与重构后的图像列之间的误差达到设定的误差限时， $L_{2,1}$  NMF 的目标函数值也能达到，反之则不然。这使得本文给出的目标函数相较于 NMF、 $L_{2,1}$  NMF 更能在减少目标函数值的过程中兼顾原始图像数据和重构后图像列之间的近似关系。

### 4. $L_{(r,c)}$ NMF 算法及鲁棒性说明

#### 4.1. $\|\cdot\|_{(r,c)}$ NMF 算法

本节给出  $L_{(r,c)}$  NMF 算法的更新公式及推导过程。重述目标函数如下：

$$\mathcal{Q}_{L(r,c)NMF} = \sum_{j=1}^n \sum_{p=1}^c D_{pj} = \sum_{j,i} \frac{(X_{ij} - UV_{ij})^2}{D_{\lfloor \frac{i}{r+1} \rfloor + 1, j}} \tag{14}$$

$$\text{s.t. } U_{ij} \geq 0, V_{ij} \geq 0$$

其中  $D_{pj} = \sqrt{\sum_{t=r(p-1)+1}^{rp} (X_{tj} - UV_{tj})^2}$ ,  $D \in \mathbb{R}^{c \times n}$ 。式(14)的拉格朗日函数如下:

$$\mathcal{L}_{L(r,c)NMF} = \sum_{j,i} \left[ \frac{(X_{ij} - UV_{ij})^2}{D_{\lfloor \frac{i}{r+1} \rfloor + 1, j}} + U_{ij} \Phi_{ij} + U_{ij} \Psi_{ij} \right] \tag{15}$$

$$\frac{\partial \mathcal{L}_{L(r,c)NMF}}{\partial U_{ab}} = [(\Delta \odot (X - UV))V^T]_{ab} - \Phi_{ab} \tag{16}$$

$$\frac{\partial \mathcal{L}_{L(r,c)NMF}}{\partial V_{ab}} = [U^T ((X - UV) \odot \Delta)]_{ab} - \Psi_{ab} \tag{17}$$

其中  $\Delta \in \mathbb{R}^{rc \times n}$ ,  $\Delta_{ij} = 1/D_{\lfloor \frac{i}{r+1} \rfloor + 1, j}$ ,  $\odot$  是哈达玛积(Hadamard product), 表示矩阵元素间的乘积。

由 KKT 条件  $\frac{\partial \mathcal{L}_{L(r,c)NMF}}{\partial U_{ab}} = 0$ ,  $\Phi_{ab} U_{ab} = 0$ ;  $\frac{\partial \mathcal{L}_{L(r,c)NMF}}{\partial V_{ab}} = 0$ ,  $\Psi_{ab} V_{ab} = 0$  即可得  $U$  和  $V$  的更新公式:

$$U_{ij} = U_{ij} \frac{((\Delta \odot X)V^T)_{ij}}{((\Delta \odot UV)V^T)_{ij}} \tag{18}$$

$$V_{ij} = V_{ij} \frac{(U^T (X \odot \Delta))_{ij}}{(U^T (UV \odot \Delta))_{ij}} \tag{19}$$

$L_{(r,c)}$  NMF 算法见算法 1。

---

**算法 1**  $L_{(r,c)}$  NMF

---

输入非负数据矩阵  $X \in \mathbb{R}^{rc \times n}$ ,  $k$   
 步骤 1 初始矩阵  $U \in \mathbb{R}^{rc \times k}$ ,  $V \in \mathbb{R}^{k \times n}$   
 步骤 2 最大迭代次数 Iter; 计算初始  $\Delta$   
 步骤 3 for  $r = 1$ : Iter  
     3.1 更新  $U$   
     3.2 计算  $\Delta$   
     3.3 更新  $V$   
     3.4 停止条件  
 输出  $U, V$

---

## 4.2. $\|\cdot\|_{(r,c)}$ 函数的鲁棒性说明

我们使用二维数据点来说明  $L_{(r,c)}$  NMF 的鲁棒性。示例的十个数据点中有三个是异常值, 对每一个数据点  $x_i$  我们使用  $UV_i$  进行拟合,  $k$  值取 1, 如此我们把数据点投影到 1 维子空间中, 实验中  $L_{(r,c)}$  NMF 设定参数  $r=1$ ,  $c=2$ 。给定的数据点以及分别使用 NMF、 $L_{2,1}$  NMF 和  $L_{(r,c)}$  NMF 得到的拟合值见图 1, 从图中可以看出,  $L_{(r,c)}$  NMF 和  $L_{2,1}$  NMF 得到的拟合值远离异常值的同时更加贴近原数据点, 而 NMF 的拟

合值则受到三个异常值的影响偏离了原数据点。对应数据点的残差  $\|x_i - Uv_i\|$  见图 2，图中可以清晰地看出  $L_{(r,c)}$  NMF 能够与异常值保持更大的残差从而减小异常值对拟合结果的影响，除开三个异常值处的残差外，相比于 NMF 而言， $L_{(r,c)}$  NMF 和  $L_{2,1}$  NMF 都有更小的残差。

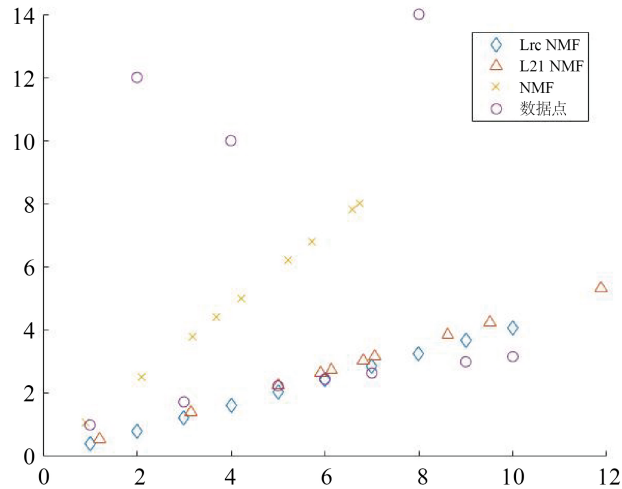


Figure 1. The results of fitting ten data points

图 1. 拟合十个数据点的结果

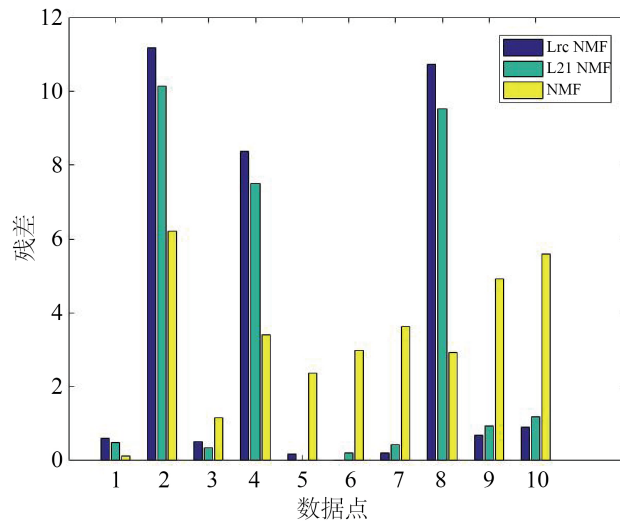


Figure 2. The residual histogram of fitting the data points

图 2. 拟合数据点的残差直方图

为进一步说明  $L_{(r,c)}$  NMF 的有效性。我们在数据集上运行  $L_{(r,c)}$  NMF、 $L_{2,1}$  NMF 和 NMF 算法进行对比。本文展示 Yale 数据集上的对比结果。该数据集中，每一个  $x_i$  都是由一张图像按列向量化所得，因此分解得到的基矩阵  $U = [u_1, u_2, \dots, u_k]$  中的每一个向量  $u_i$  也对应一张图像，将  $u_i$  重排成原始大小的图像，展示见图 3。

图像排成两行，每一行的上、中、下层分别是  $L_{(r,c)}$  NMF、 $L_{2,1}$  NMF 和 NMF 算法收敛后得到的基矩阵  $U$  重排后的图像结果。从图像可见， $L_{(r,c)}$  NMF 表示出的特征脸具有更加清晰的面部特征，例如对比

明显的第一行的第 1、2、5、6 列和第二行的 2、3、6、7、8 列。总的来讲， $L_{(r,c)}$ NMF 和  $L_{2,1}$ NMF 都能降低数据中的噪声和异常值对算法的影响，而  $L_{(r,c)}$ NMF 相较于  $L_{2,1}$ NMF 进一步兼顾了原始图像数据的空间结构，因此能获得比  $L_{2,1}$ NMF 更加清晰的图像表示结果。以上从理论和实验结果两方面说明了  $L_{(r,c)}$ NMF 算法的有效性。



**Figure 3.** The base matrix obtained on the Yale dataset  
**图 3.** Yale 数据集上得到的基矩阵

### 5. 收敛性分析

**定理 1** 对于  $U \geq 0, V \geq 0$ ，目标函数  $Q_{L_{(r,c)}NMF}$  在更新规则式(17)和式(18)下是非增的。为证定理 1，引入辅助函数的定义以及引理 1。

**定义 2** 如果下列条件满足：

$$G(v_{ab}, v'_{ab}) \geq F(v_{ab}) \tag{20}$$

$$G(v'_{ab}, v'_{ab}) = F(v'_{ab}) \tag{21}$$

则称  $G(v_{ab}, v'_{ab})$  是  $F(v_{ab})$  的一个辅助函数。

**引理 1** 若  $G(v_{ab}, v'_{ab})$  是  $F(v_{ab})$  的辅助函数，那么  $F(v_{ab})$  在下列更新规则下是非增的：

$$v_{ab}^{t+1} = \arg \min_{v_{ab}} G(v_{ab}, v_{ab}^t) \tag{22}$$

证明：由定义 2 及引理 1 有

$$F(v_{ab}^{t+1}) = G(v_{ab}^{t+1}, v_{ab}^t) \leq G(v_{ab}^t, v_{ab}^t) = F(v_{ab}^t)$$

接下来给出辅助函数。记  $F(v_{ab})$  是目标函数  $Q_{L_{(r,c)}NMF}$  关于  $v_{ab}$  的部分， $F(v_{ab})$  对  $v_{ab}$  的一阶导  $F'(v_{ab})$  和二阶导  $F''(v_{ab})$  分别如下：

$$F'(v_{ab}) = [U^T ((UV - X) \odot \Delta)]_{ab} \tag{23}$$

$$F''(v_{ab}) = \sum_{p=1}^c \left( \sum_{i=r(p-1)+1}^{rp} \frac{U_{ia}^2}{D_{pb}} - \frac{S}{D_{pb}^3} \right)_{ab} \tag{24}$$

其中  $S = \left( \sum_{i=r(p-1)+1}^p (UV_{ib} - X_{ib})U_{ia} \right)^2$ 。

### 引理 2 函数

$$G(v_{ab}, v'_{ab}) = F(v'_{ab}) + F'(v'_{ab})(v_{ab} - v'_{ab}) + \frac{1}{2} \frac{(U^T(UV \odot \Delta))_{ab}}{v'_{ab}} (v_{ab} - v'_{ab})^2 \quad (25)$$

是  $F(v_{ab})$  在  $v'_{ab}$  处的辅助函数。

证明：取  $F(v_{ab})$  在  $v'_{ab}$  处的泰勒展式

$$F(v_{ab}) = F(v'_{ab}) + F'(v'_{ab})(v_{ab} - v'_{ab}) + \frac{1}{2} F''(v'_{ab})(v_{ab} - v'_{ab})^2 \quad (26)$$

$$\begin{aligned} F''(v_{ab}) &= \sum_{p=1}^c \left( \sum_{i=r(p-1)+1}^p \frac{U_{ia}^2}{D_{pb}} - \frac{S}{D_{pb}^3} \right) \leq \sum_{p=1}^c \left( \sum_{i=r(p-1)+1}^p \frac{U_{ia}^2}{D_{pb}} \right) \\ &\leq \sum_{p=1}^c \left( \sum_{i=r(p-1)+1}^p \frac{UV_{ib}U_{ia}}{D_{pb}} \right) = \frac{(U^T(UV \odot \Delta))_{ab}}{v_{ab}} \end{aligned} \quad (27)$$

故  $G(v_{ab}, v'_{ab}) \geq F(v_{ab})$  成立，显然，当  $v_{ab} = v'_{ab}$ ，有  $G(v'_{ab}, v'_{ab}) = F(v'_{ab})$ ，引理 1 得证。易知  $G(v_{ab}, v'_{ab})$  是关于  $v_{ab}$  的二次函数，可得：

$$v_{ab}^{t+1} = \arg \min_{v_{ab}} G(v_{ab}, v'_{ab}) = v'_{ab} - \frac{F'(v_{ab})}{\frac{(U^T(UV \odot \Delta))_{ab}}{v'_{ab}}} = v'_{ab} \frac{(U^T(X \odot \Delta))_{ab}}{(U^T(UV \odot \Delta))_{ab}} \quad (28)$$

该式恰是更新规则式(19)，故而定理 1 得证。篇幅所限，关于  $U$  的证明不再赘述。

## 6. 数值实验

本节使用本文提出的  $L_{(r,c)}$  NMF 算法进行图像聚类实验，并与 K-means、NMF 算法[2]和  $L_{2,1}$  NMF 算法[9]以及 GNMF 算法[10] (正则项参数根据文献设定为 10)进行图像聚类效果比较，实验使用的数据集是 Yale、ORL。

### 6.1. 数据集介绍

**Yale** 包含 15 个人，每个人不同表情、姿态和光照下的 11 张人脸图像，共 165 张图片，每张图片大小为  $32 \times 32$ 。

**ORL** 包含 40 个人，每个人不同表情、姿态和光照下的 10 张人脸图像，共 400 张图片，每张图片大小为  $32 \times 32$ 。

### 6.2. 评估指标

本文以广泛使用的聚类精度(ACC)、归一化互信息(NMI)作为评估指标[11] [12]。

**聚类精度(ACC)**定义如下：

$$ACC = \frac{\sum_{i=1}^n \delta(s_i, \text{map}(r_i))}{n}$$

其中  $s_i$  是样本点的真实标签， $r_i$  是算法得到的该样本点的聚类标签， $n$  是数据样本总数， $\delta(x, y)$  定义如下：



$$\delta(x, y) = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases}$$

$map(\cdot)$  是置换映射函数(permutation mapping function), Kuhn-Munkres 算法可以得到最佳映射[13]。记  $C$  为真实标签集合,  $C'$  为算法获得的聚类标签, 两者的互信息(mutual information)定义如下:

$$MI(C, C') = \sum_{c_i \in C, c'_j \in C'} p(c_i, c'_j) \log_2 \frac{p(c_i, c'_j)}{p(c_i)p(c'_j)}$$

其中  $p(c_i)$  和  $p(c'_j)$  分别是数据集中任意选取一个样本, 该样本属于  $c_i$  和  $c'_j$  所在类的概率, 而  $p(c_i, c'_j)$  则表示任意选取一个样本, 该样本同属于  $c_i$  和  $c'_j$  所在类的联合概率(joint probability)。

归一化互信息(NMI)定义如下:

$$\overline{MI} = \frac{MI(C, C')}{\max(H(C), H(C'))}$$

$H(C)$ 、 $H(C')$  分别是  $C$ 、 $C'$  的熵(entropies)。 $\overline{MI} \in [0, 1]$ , 若两集合相同, 则  $\overline{MI} = 1$ ; 两集合无关, 则  $\overline{MI} = 0$ 。

### 6.3. 参数设置

本文提出的  $L_{(r,c)}$  NMF 需要根据原始图像数据的大小来确定  $r$  和  $c$  的值。在实验所用的 Yale 和 ORL 数据集中, 每个样本向量表示一张 32 行、32 列的原始图像, 这两个数据集上  $r = 32$ ,  $c = 32$ 。

本文的数值实验环境为: Matlab 2016a, 11th Gen Intel(R) Core(TM) i7-11800H @ 2.30GHz 2.30 GHz, 16.00GB 内存, Windows 1164 位操作系统。

具体步骤如下:

步骤一: 从数据集中随机选取  $k$  类样本;

步骤二: 使用算法 1 对选取的样本数据矩阵进行分解得到  $V$ ;

步骤三: 使用 k-means 算法对  $V$  进行聚类;

步骤四: 计算 ACC 和 NMI 两个聚类指标。

重复上述步骤 15 次, 记录 ACC 和 NMI 的平均值。

**Table 1.** Experimental results of three algorithms on the Yale dataset

**表 1.** 三种算法在 Yale 数据集上的实验结果

$k$	ACC (%)					NMI (%)				
	K-means	NMF	$L_{2,1}$ NMF	GNMF	$L_{(r,c)}$ NMF	K-means	NMF	$L_{2,1}$ NMF	GNMF	$L_{(r,c)}$ NMF
2	71.52	71.52	87.58	72.73	<b>90.91</b>	26.43	18.35	48.91	19.54	<b>57.68</b>
3	52.93	59.19	60.81	58.59	<b>61.21</b>	23.97	35.65	<b>38.29</b>	32.51	33.17
4	45.91	48.48	<b>49.7</b>	49.24	48.48	26.69	27.05	<b>31.71</b>	27.65	30.59
5	42.67	49.82	48.24	53.30	<b>49.21</b>	26.74	33.28	33.47	34.53	<b>37.4</b>
6	41.21	<b>49.8</b>	46.46	47.47	49.7	31.78	36.62	32.34	37.09	<b>39.4</b>
7	44.42	49.18	46.67	47.62	<b>51.69</b>	38.01	40.14	39.54	38.43	<b>43.59</b>
8	49.17	<b>50.45</b>	49.39	48.11	49.47	47.74	<b>48.13</b>	46.08	45.76	46.34
9	39.46	39.12	40.94	41.08	<b>45.66</b>	40.68	38.09	38.73	40.68	<b>41.58</b>

Continued

10	36.12	40.91	41.33	40.61	<b>43.03</b>	36.75	37.73	41.04	38.28	<b>41.59</b>
11	39.61	37.69	42.81	33.88	<b>43.09</b>	42.72	39.35	43.97	35.73	<b>45.47</b>
12	40.25	39.8	41.46	40.15	<b>45.56</b>	43.82	41.57	43.34	41.49	<b>47.66</b>
13	40	40.7	37.76	41.03	<b>42.05</b>	<b>45.68</b>	43.2	41.59	43.39	44.82
14	38.83	39.83	39.96	37.66	<b>40.65</b>	<b>45.78</b>	43.76	44.99	41.83	45.44
15	38.63	39.31	38.55	36.97	<b>42.42</b>	46.64	45.65	44.21	44.86	<b>47.25</b>
Avg.	44.33	46.84	47.97	46.31	<b>50.22</b>	37.38	37.75	40.58	37.26	<b>42.99</b>

Table 2. Experimental results of the three algorithms on the ORL dataset

表 2. 三种算法在 ORL 数据集上的实验结果

$k$	ACC (%)					NMI (%)				
	K-means	NMF	$L_{2,1}$ NMF	GNMF	$L_{(r,c)}$ NMF	K-means	NMF	$L_{2,1}$ NMF	GNMF	$L_{(r,c)}$ NMF
14	55.48	61.9	63.25	57.86	<b>71.05</b>	69.19	73.93	75.17	72.42	<b>80.29</b>
16	53.75	62.87	<b>67.33</b>	62.71	63.25	70.84	76.56	<b>79.23</b>	77.57	76.37
18	57.67	63.78	65.37	64.07	<b>67.3</b>	74.91	78.32	77.78	78.78	<b>80.05</b>
20	54.57	64.7	63.97	64.67	<b>65.67</b>	71.53	78.51	78.31	77.25	<b>81.06</b>
22	56.06	63	63.88	63.64	<b>65.79</b>	73.4	78.51	79.2	78.66	<b>80.18</b>
24	55.11	61.03	<b>62.31</b>	59.03	61.28	73.08	76.99	76.88	75.69	<b>77.09</b>
26	55.33	60.15	61.69	60.38	<b>62.18</b>	73.91	76.76	77.88	75.80	<b>79.1</b>
28	56.76	<b>63.45</b>	62.05	60.71	62.67	76.14	<b>78.82</b>	77.59	77.74	78.58
30	52.02	56.27	57.16	57.56	<b>60.29</b>	72.17	75.2	76.06	76.95	<b>77.91</b>
32	51.73	56.71	57.04	54.90	<b>58.67</b>	72.63	76.28	76	74.99	<b>76.71</b>
34	51.04	56.31	58.53	54.41	<b>62.63</b>	72.65	75.97	77.07	75.28	<b>79.44</b>
36	51.7	57.81	55.65	<b>57.87</b>	56.65	73.44	77.56	76.19	<b>77.70</b>	76.71
38	52.75	59.02	58.84	61.93	<b>61.18</b>	74.62	78.84	78.49	79.80	<b>80.37</b>
40	50.12	55.55	56.15	55.67	<b>58.45</b>	73.6	75.53	75.9	75.76	<b>77.34</b>
Avg.	53.86	60.18	60.94	59.67	<b>62.64</b>	73.00	76.98	77.26	76.74	<b>78.65</b>

## 7. 结论

本文定义了具有鲁棒性且兼顾原始数据图像空间结构的  $L_{(r,c)}$  函数并给出相应的  $L_{(r,c)}$  NMF 算法, 拟合测试数据时目标函数具有较好的抗噪(robustness)表现, 图像数据集上聚类实验结果可看出,  $L_{(r,c)}$  NMF 算法兼顾了原始图像数据结构, 相对于 K-means、NMF 算法和  $L_{2,1}$  NMF 算法以及 GNMF 算法而言在 ACC 和 NMI 指标下具有更好的聚类性能, 实验结果见表 1、表 2。以上均证实了本文算法的有效性。

## 参考文献

- [1] Lee, D.D. and Seung, H.S. (1999) Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature*, **401**, 788-791. <https://doi.org/10.1038/44565>
- [2] Lee, D. and Seung, H.S. (2000) Algorithms for Non-Negative Matrix Factorization. *Proceedings of the 13th Interna-*

- tional Conference on Neural Information Processing Systems*, Barcelona, 4-9 December 2016, 535-541.
- [3] Pauca, V.P., Shahnaz, F., Berry, M.W., *et al.* (2004) Text Mining Using Non-Negative Matrix Factorizations. *Proceedings of the 2004 SIAM International Conference on Data Mining*, Lake Buena Vista, 22-24 April 2004, 452-456. <https://doi.org/10.1137/1.9781611972740.45>
  - [4] Shu, Z., Wu, X.J., You, C., *et al.* (2020) Rank-Constrained Nonnegative Matrix Factorization for Data Representation. *Information Sciences*, **528**, 133-146. <https://doi.org/10.1016/j.ins.2020.04.017>
  - [5] Brunet, J.P., Tamayo, P., Golub, T.R., *et al.* (2004) Metagenes and Molecular Pattern Discovery Using Matrix Factorization. *Proceedings of the National Academy of Sciences*, **101**, 4164-4169. <https://doi.org/10.1073/pnas.0308531101>
  - [6] Li, T., Ding, C. and Jordan, M.I. (2007) Solving Consensus and Semi-Supervised Clustering Problems Using Nonnegative Matrix Factorization. *7th IEEE International Conference on Data Mining (ICDM 2007)*, Omaha, 28-31 October 2007, 577-582. <https://doi.org/10.1109/ICDM.2007.98>
  - [7] Ding, C., He, X. and Simon, H.D. (2005) On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering. *Proceedings of the 2005 SIAM International Conference on Data Mining*, Newport Beach, 21-23 April 2005, 606-610. <https://doi.org/10.1137/1.9781611972757.70>
  - [8] Liu, W., Zheng, N. and You, Q. (2006) Nonnegative Matrix Factorization and Its Applications in Pattern Recognition. *Chinese Science Bulletin*, **51**, 7-18. <https://doi.org/10.1007/s11434-005-1109-6>Kong, D., Ding, C. and Huang, H.
  - [9] (2011) Robust Nonnegative Matrix Factorization Using  $l_{21}$ -Norm. *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, Glasgow Scotland, 24-28 October 2011, 673-682. <https://doi.org/10.1145/2063576.2063676>
  - [10] Cai, D., He, X., Han, J., *et al.* (2010) Graph Regularized Nonnegative Matrix Factorization for Data Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **33**, 1548-1560. <https://doi.org/10.1109/TPAMI.2010.231>
  - [11] Xu, W., Liu, X. and Gong, Y. (2003) Document Clustering Based on Non-Negative Matrix Factorization. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, 28 July-1 August 2003, 267-273. <https://doi.org/10.1145/860435.860485>
  - [12] Cai, D., He, X. and Han, J. (2005) Document Clustering Using Locality Preserving Indexing. *IEEE Transactions on Knowledge and Data Engineering*, **17**, 1624-1637. <https://doi.org/10.1109/TKDE.2005.198>
  - [13] Lovász, L. and Plummer, M.D. (2009) *Matching Theory*. American Mathematical Society, Boston. <https://doi.org/10.1090/chel/367>