

DOI: 10.11992/tis.201811021

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20190902.1105.004.html>

## 图正则化稀疏判别非负矩阵分解

徐慧敏, 陈秀宏

(江南大学数字媒体学院, 江苏 无锡 214000)

**摘 要:** 非负矩阵分解是一种流行的数据表示方法, 利用图正则化约束能有效地揭示数据之间的局部流形结构。为了更好地提取图像特征, 给出了一种基于图正则化的稀疏判别非负矩阵分解算法 (graph regularization sparse discriminant non-negative matrix factorization, GSDNMF- $L_{2,1}$ )。利用同类样本之间的稀疏线性表示来构建对应的图及权矩阵; 以  $L_{2,1}$  范数进行稀疏性约束; 以最大间距准则为优化目标函数, 利用数据集的标签信息来保持数据样本之间的流形结构和特征的判别性, 并给出了算法的迭代更新规则。在若干图像数据集上的实验表明, GSDNMF- $L_{2,1}$  在特征提取方面的分类精度优于各对比算法。

**关键词:** 非负矩阵分解; 特征提取; 降维; 流形学习; 最大间距准则; 判别信息; 稀疏约束; 线性表示

**中图分类号:** TP391.4    **文献标志码:** A    **文章编号:** 1673-4785(2019)06-1217-08

中文引用格式: 徐慧敏, 陈秀宏. 图正则化稀疏判别非负矩阵分解 [J]. 智能系统学报, 2019, 14(6): 1217-1224.

英文引用格式: XU Huimin, CHEN Xiuhong. Graph-regularized, sparse discriminant, non-negative matrix factorization[J]. CAAI transactions on intelligent systems, 2019, 14(6): 1217-1224.

## Graph-regularized, sparse discriminant, non-negative matrix factorization

XU Huimin, CHEN Xiuhong

(School of Digital Media, Jiangnan University, Wuxi 214000, China)

**Abstract:** Non-negative matrix factorization is a popular data representation method. Using graph regularization constraints can effectively reveal the local manifold structure between data. In order to better extract image features, a graph-regularized, sparse-discriminant, non-negative matrix factorization algorithm is proposed in this paper. The sparse linear representation between similar samples was used to construct the corresponding graph and weight matrix. The objective function using the maximum margin criterion with  $L_{2,1}$ -norm constraint was optimized, using the tag information of the dataset to maintain the manifold structure of samples and discrimination of characteristics, and the iterative update rules of the algorithm are given. Experiments were carried out on the ORL, AR, and COIL20 datasets. Compared with other algorithms, GSDNMF- $L_{2,1}$  showed higher classification accuracy in feature extraction.

**Keywords:** non-negative matrix factorization; feature extraction; dimensionality reduction; manifold learning; maximum margin criterion; discriminant information; sparse constraints; linear representation

有效的数据低维表示不仅能够发现数据集中的潜在信息, 还能去除原始数据中的冗余特征, 快速处理高维数据, 其中最具代表性的方法包括主成分分析<sup>[1]</sup>、线性判别分析<sup>[2]</sup>及非负矩阵分解<sup>[3]</sup>。然而 PCA 和 LDA 得到的数据低维表示中可能会

包含负值元素, 这在实际的应用中缺少合理的解释; 而 NMF 是一种非负约束下的矩阵分解方法, 仅允许原始数据的加法而非减法组合, 因此它有利于稀疏的且基于部分的表示, 比非稀疏的全局表示更稳健。

经典的 NMF 是一种无监督学习算法, 不能充分利用原始数据的类别信息。为了提高 NMF 的

收稿日期: 2018-11-26. 网络出版日期: 2019-09-02.

基金项目: 2018 年江苏省研究生科研创新计划项目 (KYCX18\_1871).

通信作者: 徐慧敏. E-mail: [1215625771@qq.com](mailto:1215625771@qq.com).

判别能力, Wang 等<sup>[4]</sup>提出了 Fisher-NMF(FNMF), 而 Zafeiriou 等<sup>[5]</sup>和 Kotsia 等<sup>[6]</sup>在 NMF 的目标函数中增加了散度差项, 建立起判别子空间法。由于 LDA<sup>[2]</sup>中 Fisher 准则的比值形式很难处理, 集成在 NMF 模型中会出现“小样本问题”。为此, 可考虑最大间距准则 (max margin criterion, MMC)<sup>[7-8]</sup>, 该准则是以类中心及类散度为度量依据, 使得降维后不同类样本之间距离越远而同类样本之间距离越近, 在充分使用原始数据类别信息的基础上使得改进后的 NMF(如 NDMF<sup>[9]</sup>)具有良好的判别性质。

另外, 为了揭示和利用数据的内在几何结构, 出现了许多基于流形特征的 NMF, 例如 GNMF<sup>[10]</sup>、GDNMF<sup>[11]</sup>、LCGNMF<sup>[12]</sup>和 NLMF<sup>[13]</sup>等, 这些方法通过添加图拉普拉斯正则化项达到了提高图像聚类或分类性能的目的。Shang 等<sup>[14]</sup>同时考虑数据流形和特征流形, 提出了图对偶正则化 NMF(DNMF)。而 Meng 等<sup>[15]</sup>则在图对偶模型中加入了稀疏性 (DSNMF) 和正交性的约束<sup>[16]</sup>, 有效地简化了高维计算, 揭示数据和特征之间的双流形结构。

由于 NDMF<sup>[9]</sup>具有良好的判别性质, 而 NLMF<sup>[13]</sup>又利用了数据的图结构来提升其分类性能, 故本文将两种方法结合起来提出一种基于图正则化的稀疏判别非负矩阵分解算法 (GSDNMF-L<sub>2,1</sub>), 在充分利用数据的类别信息的同时, 不仅保持了数据样本之间的局部流形结构, 而且能提取稀疏的特征, 进而提升了其分类性能。若干图像数据集上的实验结果表明, 该方法在特征提取与分类性能等方面优于其他各对比算法。

## 1 相关工作

### 1.1 非负矩阵分解 (NMF)

假设由  $n$  个非负样本数据  $\mathbf{x}_i \in \mathbb{R}^m (i = 1, 2, \dots, n)$  组成的非负数据矩阵  $\mathbf{X} = [\mathbf{x}_1 \mathbf{x}_2 \dots \mathbf{x}_n] \in \mathbb{R}_+^{m \times n}$ , 非负矩阵分解<sup>[3]</sup>的目标是将非负数据矩阵  $\mathbf{X}$  分解为基矩阵  $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_r] \in \mathbb{R}_+^{m \times r}$  和系数矩阵  $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_n] \in \mathbb{R}_+^{r \times n}$  的乘积, 即  $\mathbf{X} \approx \mathbf{WH}$ , 当  $r \leq \min(m, n)$  时, 矩阵  $\mathbf{X}$  就达到了有效压缩。NMF 可表示为以下最小化问题:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}} \quad & \|\mathbf{X} - \mathbf{WH}\|_F^2 \\ \text{s.t.} \quad & \mathbf{W} \geq 0, \mathbf{H} \geq 0 \end{aligned} \quad (1)$$

其中  $\|\cdot\|_F$  表示 Frobenius 范数。该问题的求解采用以下乘性迭代规则:

$$\begin{aligned} \mathbf{W}_{ik} \leftarrow \mathbf{W}_{ik} \frac{(\mathbf{XH}^T)_{ik}}{(\mathbf{WHH}^T)_{ik}}, \mathbf{H}_{kj} \leftarrow \mathbf{H}_{kj} \frac{(\mathbf{W}^T \mathbf{X})_{kj}}{(\mathbf{W}^T \mathbf{WH})_{kj}} \\ (i = 1, 2, \dots, m; k = 1, 2, \dots, r; j = 1, 2, \dots, n) \end{aligned} \quad (2)$$

### 1.2 图正则化非负矩阵分解 (GNMF)

图正则化非负矩阵分解 (graph regularized

non-negative matrix factorization, GNMF)<sup>[9]</sup>考虑了数据样本之间的流形结构, 希望数据样本在低维空间中尽可能多地保持其近邻结构, 从而寻找数据在低维空间中的紧致嵌入。该问题可用以下优化模型表示:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}} \quad & \|\mathbf{X} - \mathbf{WH}\|_F^2 + \lambda \text{tr}(\mathbf{HLH}^T) \\ \text{s.t.} \quad & \mathbf{W} \geq 0, \mathbf{H} \geq 0 \end{aligned} \quad (3)$$

其中  $\mathbf{L} = \mathbf{D} - \mathbf{S}$  为拉普拉斯矩阵,  $\mathbf{S}$  是权矩阵,  $\mathbf{D}$  为对角矩阵,  $D_{ii} = \sum_j S_{ij}$ 。

求解该问题的乘性迭代规则为

$$\begin{aligned} \mathbf{W}_{ik} \leftarrow \mathbf{W}_{ik} \frac{(\mathbf{XH}^T)_{ik}}{(\mathbf{WHH}^T)_{ik}}, \mathbf{H}_{kj} \leftarrow \mathbf{H}_{kj} \frac{(\mathbf{W}^T \mathbf{X} + \lambda \mathbf{HS}_H)_{kj}}{(\mathbf{W}^T \mathbf{WH} + \lambda \mathbf{HD}_H)_{kj}} \\ (i = 1, 2, \dots, m; k = 1, 2, \dots, r; j = 1, 2, \dots, n) \end{aligned} \quad (4)$$

### 1.3 最大间距准则 (MMC)

设有  $C$  类样本, 最大间距准则的主要思想是寻找一个投影矩阵  $\mathbf{P}$  使得投影后同类样本之间的散度最小而不同类样本之间的散度最大, 达到保持原始数据判别信息的目的, 它可表示为以下优化问题:

$$\max \text{tr}(\mathbf{P}^T (\mathbf{S}_b - \mathbf{S}_w) \mathbf{P}) \quad (5)$$

式中:  $\mathbf{S}_b = \sum_{c=1}^C (\bar{\mathbf{x}} - \bar{\mathbf{x}}_c)(\bar{\mathbf{x}} - \bar{\mathbf{x}}_c)^T$  为类间散度矩阵;  $\mathbf{S}_w = \sum_{c=1}^C \sum_{j=1}^{n_c} (\mathbf{x}_j^c - \bar{\mathbf{x}}_c)(\mathbf{x}_j^c - \bar{\mathbf{x}}_c)^T$  为类内散度矩阵;  $\bar{\mathbf{x}}$  表示所有样本的均值, 而  $\bar{\mathbf{x}}_c$  为  $c$  类样本的均值,  $n_c$  为  $c$  类样本数。

## 2 本文算法

根据流形学习的图嵌入思想<sup>[17-19]</sup>, 数据在高维空间中的局部邻域关系在低维样本空间中应得到保持, 同时也希望原始数据集的散度关系得到保持, 因此在 NMF 的目标函数中通过添加图正则化约束和判别约束可达到提高算法分类性能的目的。

### 2.1 构建权重矩阵

基于图嵌入的流形学习法其前提是构建相应的邻接图及权矩阵。传统的流形学习方法通常是利用  $k$ -近邻法或  $\varepsilon$ -邻域法来构建图和权矩阵, 但此类方法的性能依赖于参数  $k$  或  $\varepsilon$ 。本文利用基于同类样本的稀疏表示方法选择邻域样本, 该方法是自适应且与参数无关的, 对噪声数据具有鲁棒性。

图 1 比较了两种方法选择的邻域样本。其中,  $k$ -近邻法选择的邻域样本包含了数据集中不同类别的人脸, 人脸角度变化与标注图像一致。相比而言, 本文方法所选的样本不仅和标注人脸具有相似角度, 而且最接近其正面的人脸图像, 且

相应的稀疏表示系数越大, 越接近标注样本。所以, 基于同类样本的稀疏表示方法构建的权重矩阵不仅包含了样本的类别信息, 而且还包含了样本间的局部邻域结构, 从而具有良好的判别性质。



图 1  $k$ -近邻法与稀疏表示法确定的近邻

Fig. 1 Neighbor samples determined by  $k$ -nearest neighbor method and sparse representation

对于某一类数据样本  $\mathbf{x}_i^c$ , 将它表示为所有同类的其他样本的稀疏线性组合, 这可以用以下优化模型来表示:

$$\begin{aligned} \min \quad & \|\mathbf{s}_i^c\|_1 \\ \text{s.t.} \quad & \mathbf{x}_i = \mathbf{X}_c \mathbf{s}_i^c, c(\mathbf{x}_i) = c \end{aligned} \quad (6)$$

式中:  $\mathbf{X}_c$  为第  $c$  类所有样本组成的矩阵, 表示系数向量  $\mathbf{s}_i^c$  中的每一个分量描述了与  $\mathbf{x}_i^c$  同类的其他相应样本在表示  $\mathbf{x}_i^c$  时的贡献大小, 从而可以用来表示样本之间的相似性。  $c(\mathbf{x}_i)$  表示样本  $\mathbf{x}_i^c$  的类标签,  $n_c$  表示第  $c$  类样本的数量, 且  $n = \sum_{c=1}^C n_c$ 。从而, 有以下基于稀疏表示的样本权重矩阵:

$$\mathbf{S}_H = (\mathbf{S}_{ij})_{n \times n} = \text{diag}\{\mathbf{S}^1, \mathbf{S}^2, \dots, \mathbf{S}^C\} \quad (7)$$

其中: 子矩阵  $\mathbf{S}^c \in \mathbb{R}^{n_c \times n_c}$  是由第  $c$  类中每个训练样本  $\mathbf{x}_i^c$  的稀疏表示系数  $\mathbf{s}_i^c$  所组成的分块矩阵。

## 2.2 正则化约束

图正则化是基于邻域保持嵌入的, 它将训练样本之间的局部邻域关系通过权重矩阵  $\mathbf{S}$  在由 NMF 的系数矩阵  $\mathbf{H}$  的列向量  $\mathbf{h}_j$  所张成的子空间中得到保持, 因此得到以下图正则化约束项:

$$\begin{aligned} \frac{1}{2} \sum_{i,j=1}^n \|\mathbf{h}_i - \mathbf{h}_j\|_2^2 \mathbf{S}_{ij} = & \text{tr}(\mathbf{H} \mathbf{D}_H \mathbf{H}^T) - \text{tr}(\mathbf{H} \mathbf{S}_H \mathbf{H}^T) = \\ & \text{tr}(\mathbf{H} \mathbf{L}_H \mathbf{H}^T) \end{aligned} \quad (8)$$

式中:  $\mathbf{S}_H$  为权重矩阵;  $\mathbf{L}_H = \mathbf{D}_H - \mathbf{S}_H$  为拉普拉斯矩

阵;  $\mathbf{D}_H$  为对角矩阵且对角元素为  $\mathbf{D}_{Hii} = \sum_j \mathbf{S}_{Hij}$ 。

以 NMF 中的基矩阵  $\mathbf{W}$  为投影矩阵, 使得投影后同类样本之间的散度最小而不同类样本之间的散度最大<sup>[19]</sup>, 以保留训练样本的判别特征, 从而有以下基于 MMC 准则的判别约束项:

$$\text{tr}(\mathbf{W}^T (\mathbf{S}_B - \mathbf{S}_W) \mathbf{W}) \quad (9)$$

为满足 NMF 的非负约束的要求, 式 (9) 可改写为

$$\text{tr}(\mathbf{W}^T ((\mathbf{S}_B^+ - \mathbf{S}_{|B|}^-) - (\mathbf{S}_W^+ - \mathbf{S}_{|W|}^-)) \mathbf{W}) \quad (10)$$

式中:  $\mathbf{S}_B$  表示由所有训练样本的类间散度矩阵;  $\mathbf{S}_B^+$  和  $\mathbf{S}_{|B|}^-$  分别表示  $\mathbf{S}_B$  的正元素和非正元素的绝对值构成的矩阵;  $\mathbf{S}_W$  表示类内散度矩阵;  $\mathbf{S}_W^+$  和  $\mathbf{S}_{|W|}^-$  分别表示  $\mathbf{S}_W$  的正元素和非正元素的绝对值构成的矩阵。

对于图像数据, 其特征维度和冗余程度都很高, 基于稀疏编码的思想<sup>[20]</sup>, 将样本  $\mathbf{x}_i$  表示为基图像的稀疏线性表示, 可以减少高维样本的计算量, 得到更具鲁棒性的特征:

$$\mathbf{x}_i \approx \mathbf{w}_1 \mathbf{h}_{1,i} + \mathbf{w}_2 \mathbf{h}_{2,i} + \dots + \mathbf{w}_r \mathbf{h}_{r,i} \quad (11)$$

通常使用  $L_0$  范数对系数矩阵进行稀疏性约束, 但  $L_0$  范数的求解属于 NP 难问题, 常用  $L_1$  范数作为其凸近似进行近似求解:

$$\|\mathbf{H}\|_1 = \sum_{i=1}^r \sum_{j=1}^n |\mathbf{h}_{ij}| \quad (12)$$

系数矩阵  $\mathbf{H}$  的每一列可看作原始数据  $\mathbf{X}$  在子空间中的低维表示,  $L_1$  范数使得系数矩阵  $\mathbf{H}$  整体保持稀疏, 本文使用  $L_{2,1}$  范数对系数矩阵  $\mathbf{H}$  进行列稀疏约束, 使得每一个低维表示  $\mathbf{h}_j$  保持稀疏, 能够在特征空间中用尽可能少的特征重构原始数据:

$$\|\mathbf{H}^T\|_{2,1} = \sum_{j=1}^n \|\mathbf{h}_j\|_2 = \sum_{j=1}^n \sqrt{\sum_{i=1}^r \mathbf{h}_{ij}^2} \quad (13)$$

其中  $\mathbf{h}_j$  为矩阵  $\mathbf{H}$  的第  $j$  列向量。

## 2.3 优化模型与求解

由以上分析, 利用图正则化约束来保持样本之间的邻域结构; 以最大间距准则保持数据集的判别性; 以  $L_{2,1}$  范数进行稀疏性约束<sup>[21]</sup>, 建立优化模型:

$$\begin{aligned} \min \quad & \|\mathbf{X} - \mathbf{W} \mathbf{H}\|_F^2 + \lambda \text{tr}(\mathbf{H} \mathbf{L}_H \mathbf{H}^T) + \\ & \beta \text{tr}(\mathbf{W}^T ((\mathbf{S}_W^+ - \mathbf{S}_{|W|}^-) - (\mathbf{S}_B^+ - \mathbf{S}_{|B|}^-)) \mathbf{W}) + \mu \|\mathbf{H}^T\|_{2,1} \\ \text{s.t.} \quad & \mathbf{W} \geq 0, \mathbf{H} \geq 0 \end{aligned} \quad (14)$$

令  $\Phi_{ik}$  与  $\Psi_{kj}$  为拉格朗日乘子, 构造拉格朗日函数:



$$L = \|X - WH\|_F^2 + \lambda \text{tr}(HLH^T) + \beta \text{tr}(W^T((S_W^+ - S_{|W|}^-) - (S_B^+ - S_{|B|}^-))W) + \mu \text{tr}(H^T QH) - \text{tr}(\Phi W^T) - \text{tr}(\Psi H) \quad (15)$$

式中:  $Q$  为对角矩阵, 且  $Q_{ii} = \frac{1}{2\|H_i^T\|_2}$ ,  $H_i^T$  为系数矩阵  $H$  的转置的第  $i$  列。

使用交替迭代方法求解。先固定  $H$ , 更新  $W$ ; 然后固定  $W$ , 更新  $H$ 。式 (15) 对  $W$  求偏导并令导数等于 0, 得:

$$\frac{\partial L}{\partial W} = -2XH^T + 2WHH^T + 2\beta((S_W^+ - S_{|W|}^-) - (S_B^+ - S_{|B|}^-))W - \Phi = 0 \quad (16)$$

由互补松弛条件  $\Phi_{ik}W_{ik} = 0$ , 得:

$$(-XH^T + WHH^T + \beta((S_W^+ - S_{|W|}^-) - (S_B^+ - S_{|B|}^-))W)_{ik}W_{ik} = 0 \quad (17)$$

于是, 有以下关于  $W$  的更新规则:

$$W_{ik}^{t+1} \leftarrow W_{ik}^t \frac{(XH^T + \beta(S_{|W|}^- + S_B^+))_{ik}}{(WHH^T + \beta(S_W^+ + S_{|B|}^-))_{ik}} \quad (18)$$

同理, 得到关于  $H$  的更新规则:

$$H_{kj}^{t+1} \leftarrow \frac{(W^T X + \lambda HS_H + \beta \nabla_H \text{tr}(W(S_{|W|}^- + S_B^+)W^T))_{kj}}{H_{kj}^T (W^T WH + \lambda HD_H + \beta \nabla_H \text{tr}(W(S_W^+ + S_{|B|}^-)W^T) + \mu HQ)_{kj}} \quad (19)$$

在每次迭代得到基图像矩阵  $W$  后, 对其每一列进行归一化处理, 使非负矩阵分解得到的结果保持缩放不变性。设置最大迭代次数  $T_0$ , 使算法在最大迭代次数下收敛。

综合以上讨论, 得到图正则化稀疏判别非负矩阵分解法:

**算法 1** 图正则化稀疏判别非负矩阵分解算法 (GSDNMF- $L_{2,1}$ )

**输入** 训练样本  $X \in \mathbb{R}_+^{m \times n}$ , 样本类别  $Y$ , 特征数  $r$ , 正则化参数  $\lambda, \beta, \mu$ , 最大迭代次数  $T_0$ 。

1) 利用式 (6) 和式 (7) 计算权重矩阵  $S_H$ ;

2) 利用式 (9) 计算类内散度  $S_W$  和类间散度  $S_B$ ;

3) 随机初始化  $W_0 \in \mathbb{R}_+^{m \times k}$ ,  $H_0 \in \mathbb{R}_+^{k \times n}$ ;

4)

For  $t = 1: T_0$

利用式 (18) 得到更新后的  $W$ , 并对  $W$  的每一列进行归一化处理, 利用公式 (19) 得到更新后的  $H$ ;

End

**输出** 基矩阵  $W$  和系数矩阵  $H$ 。

## 2.4 算法复杂度

设  $m$  为样本维度,  $n$  为样本个数,  $k$  为特征

数。算法 GSDNMF- $L_{2,1}$  主要由 1)、2) 和 4) 组成, 其中步骤 1) 是构建数据图的权重矩阵, 其复杂度为  $O(n^2m)$ ; 步骤 2) 则是计算类内散度  $S_W$  和类间散度  $S_B$ , 其复杂度为  $O(m^2)$ ; 对步骤 4), 一次迭代计算  $W$  和  $H$  的复杂度  $O(mnk)$ , 假设算法在  $T_0$  次迭代后收敛, 则整个步骤 4) 的复杂度为  $O(T_0mnk)$ 。综上可知, GSDNMF- $L_{2,1}$  算法的总体时间复杂度为  $O(T_0mnk + n^2m + m^2)$ 。

## 3 实验分析

为了说明 GSDNMF- $L_{2,1}$  算法在图像特征提取中的有效性, 分别在 ORL 和 AR 人脸数据集和 COIL20 实物数据集上进行实验, 并将之与 NMF<sup>[3]</sup>、GNMF<sup>[10]</sup>、NDMF<sup>[9]</sup> 等算法进行比较, 同时比较利用  $L_1$  范数进行稀疏约束的模型 (GSDNMF- $L_1$ )。将数据集按 0.5 的比例随机划分为训练集和测试集, 利用训练集进行特征提取, 获得基矩阵  $W$ 。利用测试集计算分类精度: 测试样本  $x \in \mathbb{R}^m$  在低维子空间中的表示为  $y \approx W^+x \in \mathbb{R}^r$ , 其中  $W^+ = (W^T W)^{-1} W^T$  为矩阵  $W$  的 Pseudo 逆, 使用传统的  $k$ -近邻分类器 ( $k$ -NN) 预测类标签。每次实验独立随机, 重复 20 次取平均识别率和方差作为最后实验结果。

### 3.1 数据集

ORL 人脸数据库包含了 40 个人的人脸图像, 每个人有 10 张, 图像具有不同面部表情 (睁/闭眼, 笑/不笑)、面部细节 (眼镜/无眼镜) 及不同光照条件。

AR 数据集包含了 120 个人的人脸图像, 每个人有 26 张图像, 图像具有不同的面部表情、照明条件和遮挡 (太阳眼镜/围巾)。

COIL20 数据集包含了 20 个不同的物体 (玩具小鸭、招财猫、杯子等), 其中, 每个物体在水平面上每隔  $5^\circ$  拍摄一张图片, 因此每个物体一共有 72 幅图片, 整个数据集共计 1440 幅图片。

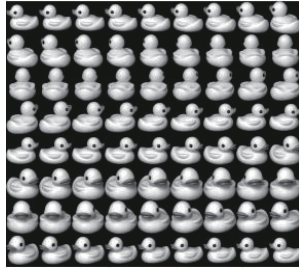
实验中, 所有图像均压缩成  $32 \times 32$  大小的灰度图, 将其每列相连构成大小为 1024 维的向量, 并做归一化处理。如图 2 所示。



(a) ORL数据集



(b) AR数据集



(c) COIL20数据集

图 2 数据集示例

Fig. 2 The instances of datasets

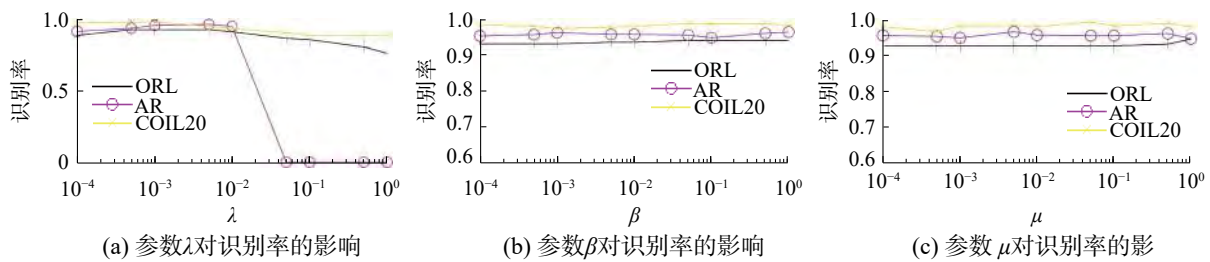


图 3 参数对识别率的影响

Fig. 3 The influence of parameters

### 3.3 实验结果及分析

由图 3 可知, 图正则化参数  $\lambda$  对识别率的影响较大, 在  $[0, 0.01]$  的范围内, 识别率随参数的增大而增大, 在  $[0.01, 1]$  的范围内, 识别率随参数的增大而减少。判别约束参数和稀疏参数对识别率的影响较小, 在  $[0, 1]$  范围内一直保持较高识别率。经过多次实验, 最终确定算法的最优参数组合为  $\lambda=0.005$ ,  $\beta=1$ ,  $\mu=0.5$ 。

取基图数分别为 3、4、5、6、7、8 的平方, 在 3 种数据集上独立随机地进行 20 次实验, 计算平均识别率及方差。表 1~3 给出了 5 种算法在 3 种数据集上的平均识别率随基图像个数的增加而变

### 3.2 参数选择

GSDNMF- $L_{2,1}$  算法包含了图正则化约束参数  $\lambda$ 、判别约束参数  $\beta$  和稀疏参数  $\mu$  等 3 个关键参数。为说明 3 个参数对识别率的影响, 分别在两个数据集上采取固定两个参数来确定另一个参数的方法进行讨论, 设参数的取值范围为  $[0, 1]$ , 取数据集的类别数作为基图数 (ORL 取 40, AR 取 120, COIL20 取 20), 最大迭代次数为 300 次, 实验所得的平均识别率与正则化参数的关系如图 3 所示。

化的情况。在大多数情况下, GSDNMF- $L_{2,1}$  的识别率要优于其他几种算法, 且方差更小, 说明在 NMF 的目标函数中结合图正则化约束和判别约束能够有效提高图像特征提取的准确率和稳定性。图 4 比较了在 ORL 数据集上 5 种算法的训练时间 (取类别数 40 作为基图数目), 可以发现 NMF 和 GNMF 的训练时间较短, NDMF 训练时间最长, GSDNMF- $L_{2,1}$  和 GSDNMF- $L_1$  比 NDMF 时间短。可知, 利用稀疏编码的思想对系数矩阵  $H$  添加稀疏性约束, 在每一迭代更新的过程中能够尽可能少地选择基图像重构训练样本, 提高了运算速度。

表 1 ORL 数据集上的平均识别率 (方差)

Table 1 Average recognition rate (variance) on ORL dataset

特征维数 $r$	NMF	GNMF	NDMF	GSDNMF- $L_1$	GSDNMF- $L_{2,1}$
9	79.30(2.36)	79.48(1.73)	79.70(2.83)	82.05(2.34)	<b>82.25(1.51)</b>
16	85.15(1.69)	85.96(1.35)	86.18(1.58)	90.83(1.58)	<b>90.83(1.38)</b>
25	86.33(1.52)	85.00(2.14)	85.90(1.67)	89.83(1.36)	<b>92.13(1.25)</b>
36	87.18(1.50)	87.30(1.30)	86.85(1.69)	<b>92.50(1.06)</b>	93.13(1.03)
49	90.14(1.04)	90.85(1.17)	90.63(0.69)	95.43(0.69)	<b>95.43(0.56)</b>

表2 AR数据集上的平均识别率(方差)

Table 2 Average recognition rate (variance) on AR dataset

特征维数 $r$	NMF	GNMF	NDMF	GSDNMF-L <sub>1</sub>	GSDNMF-L <sub>2,1</sub>
16	62.68(1.86)	67.80(2.66)	61.12(2.17)	76.01(1.80)	<b>76.10(1.83)</b>
25	74.66(1.57)	77.92(1.94)	74.01(1.98)	85.31(1.17)	<b>85.83(1.02)</b>
36	77.57(0.84)	80.48(1.25)	76.74(1.10)	85.96(0.81)	<b>86.49(1.02)</b>
49	78.29(1.49)	80.92(0.96)	77.84(1.15)	88.42(0.75)	<b>88.79(0.81)</b>
64	90.62(1.03)	93.05(0.70)	90.47(0.91)	94.92(0.51)	94.91(0.28)

表3 COIL20数据集上的平均识别率(方差)

Table 3 Average recognition rate (variance) on COIL20 dataset

特征维数 $r$	NMF	GNMF	NDMF	GSDNMF-L <sub>1</sub>	GSDNMF-L <sub>2,1</sub>
9	90.45(1.07)	91.97(1.32)	91.03(1.43)	90.56(1.67)	<b>92.05(1.35)</b>
16	95.05(0.81)	96.62(0.83)	95.26(0.90)	95.35(1.43)	<b>97.11(0.74)</b>
25	98.19(0.49)	98.76(0.39)	97.93(0.64)	98.97(0.47)	<b>99.23(0.38)</b>
36	98.44(0.87)	98.65(0.59)	98.37(0.63)	98.91(0.83)	<b>99.34(0.35)</b>

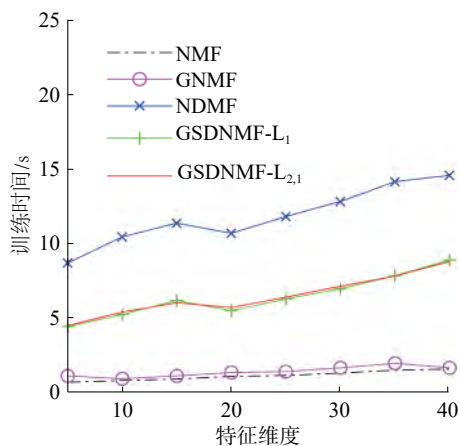


图4 训练时间随基图数变化的曲线

Fig. 4 Variation curve of training time

图5给出了GSDNMF-L<sub>2,1</sub>算法在3种数据集上的收敛情况,当迭代次数小于10次时损失函数值迅速下降;而当迭代次数大于100时,随着迭代次数的增加其损失函数值的下降趋于平缓,并收敛于一个固定值。本文所有实验均设置最大迭代次数为300次,以保证算法收敛。

为了刻画和评价基图像矩阵的特点, Hoyer<sup>[20]</sup>提出采用向量的L<sub>1</sub>范数和L<sub>2</sub>范数的差异来度量向量的稀疏度,其计算方式为

$$s(x) = \frac{\sqrt{n} - \left( \sum_i x_i / \sqrt{\sum_i x_i^2} \right)}{\sqrt{n} - 1} \quad (20)$$

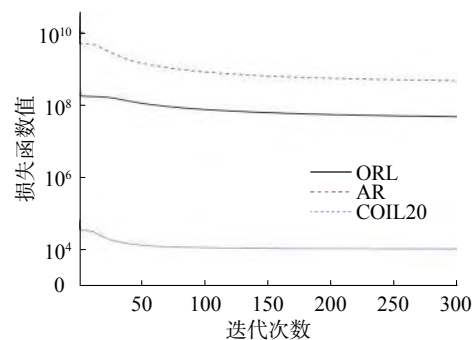


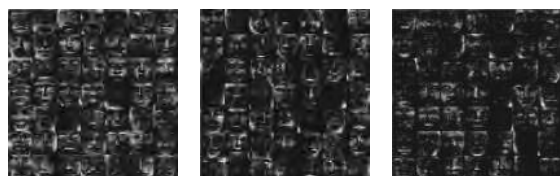
图5 损失函数的变化曲线

Fig. 5 Variation curve of loss function

其中向量 $x_i$ 是其第 $i$ 个分量。稀疏度值域是 $[0,1]$ ,值越大,说明向量越稀疏。表4比较了5种算法在3种数据集上迭代300次后生成的基图的平均稀疏度(基矩阵的每一列表示一个基图)。图6比较了3种具有判别性质的NMF算法:NDMF<sup>[9]</sup>、GSDNMF-L<sub>1</sub>和GSDNMF-L<sub>2,1</sub>得到的基图。对比发现,3种算法都能得到的图像的局部化表示,且都包含了每类物体的判别信息,但GSDNMF-L<sub>2,1</sub>和GSDNMF-L<sub>1</sub>得到的基图比NDMF<sup>[9]</sup>得到的基图更稀疏。由实验结果可知,本文提出的算法,能够在保证稀疏性的条件下,保持良好分类水平,说明算法分解得到了更有效的基图特征。

表 4 比较基图像  $W$  的稀疏度  
Table 4 Comparison of sparsity of the base matrix  $W$

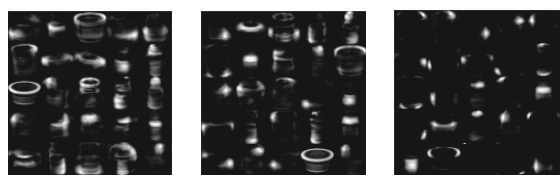
数据集	NMF	GNMF	NDMF	GSDNMF- $L_1$	GSDNMF- $L_{2,1}$
ORL	0.25	0.32	0.26	0.30	<b>0.33</b>
AR	0.52	0.55	0.52	0.55	<b>0.56</b>
COIL	0.49	0.56	0.49	0.54	<b>0.65</b>



(a) 3种算法在ORL数据集上的基图



(b) 3种算法在AR数据集上的基图



(c) 3种算法在COIL数据集上的基图

图 6 比较 NDMF, GSDNMF- $L_1$  和 GSDNMF- $L_{2,1}$  算法在 ORL、AR 和 COIL20 数据集上得到的基图

Fig. 6 Comparison of basic images computed by NDMF, GSDNMF- $L_1$  and GSDNMF- $L_{2,1}$  algorithm on ORL, AR and COIL20 dataset

## 4 结束语

本文给出了一种新的图正则化稀疏判别 NMF 算法 (GSDNMF- $L_{2,1}$ )。该方法在对图像进行特征提取时,充分利用了数据集的标签信息来构造权重矩阵和判别约束项。与已有的图正则化方法不同的是, GSDNMF- $L_{2,1}$  算法用同类样本的稀疏线性组合来构建权重矩阵,很好地保持了样本内的相似性和样本间的差异性,在最大间距准则下使得类内离散性最小而类间分离性最大,从而获得很强的判别能力。另外, GSDNMF- $L_{2,1}$  在稀疏性约束条件下得到了稀疏系数矩阵,从而提高了矩阵分解的速度。下一步研究的方向是如何处理新样本,及自动选择最佳特征数量,使算法更加完善。

## 参考文献:

[1] WOLD S, ESBENSEN K, GELADI P. Principal compon-

ent analysis[J]. Chemometrics and intelligent laboratory systems, 1987, 2(1/2/3): 37–52.

- [2] BELHUMEUR P N, HESPAHHA J P, KRIEGMAN D J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection[C]//Proceedings of the 4th European Conference on Computer Vision. Cambridge, UK, 1996: 45–58.
- [3] LEE D D, SEUNG H S. Learning the parts of objects by non-negative matrix factorization[J]. *Nature*, 1999, 401(6755): 788–791.
- [4] WANG Yuan, JIA Yunde, HU Changbo, et al. Fisher non-negative matrix factorization for learning local features[C]//Proceedings of the 6th Asian Conference on Computer Vision. Jeju, Korea, 2004.
- [5] ZAFEIRIOU S, TEFAS A, BUCIU I, et al. Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification[J]. *IEEE transactions on neural networks*, 2006, 17(3): 683–695.
- [6] KOTSIA I, ZAFEIRIOU S, PITAS I. A novel discriminant non-negative matrix factorization algorithm with applications to facial image characterization problems[J]. *IEEE transactions on information forensics and security*, 2007, 2(3): 588–595.
- [7] GU Quanquan, ZHOU Jie. Two dimensional maximum margin criterion[C]// IEEE International Conference on Acoustics, Speech and Signal Processing. Taipei, Taiwan, 2009: 1621–1624.
- [8] LI Haifeng, JIANG Tao, ZHANG Keshu. Efficient and robust feature extraction by maximum margin criterion[J]. *IEEE transactions on neural networks*, 2006, 17(1): 157–165.
- [9] LU Yuwu, LAI Zhihui, XU Yong, et al. Nonnegative discriminant matrix factorization[J]. *IEEE transactions on circuits and systems for video technology*, 2017, 27(7): 1392–1405.
- [10] CAI Deng, HE Xiaofei, HAN Jiawei, et al. Graph regularized nonnegative matrix factorization for data representation[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2011, 33(8): 1548–1560.
- [11] LONG Xianzhong, LU Hongtao, PENG Yong, et al. Graph regularized discriminative non-negative matrix factorization for face recognition[J]. *Multimedia tools and applications*, 2014, 72(3): 2679–2699.
- [12] LIAO Qing, ZHANG Qian. Local coordinate based graph-regularized NMF for image representation[J]. *Sig-*



- nal processing, 2016, 124: 103–114.
- [13] LI Xuelong, CUI Guosheng, DONG Yongsheng. Graph regularized non-negative low-rank matrix factorization for image clustering[J]. *IEEE transactions on cybernetics*, 2017, 47(11): 3840–3853.
- [14] SHANG Fanhua, JIAO L C, WANG Fei. Graph dual regularization non-negative matrix factorization for co-clustering[J]. *Pattern recognition*, 2012, 45(6): 2237–2250.
- [15] MENG Yang, SHANG Ronghua, JIAO Licheng, et al. Feature selection based dual-graph sparse non-negative matrix factorization for local discriminative clustering[J]. *Neurocomputing*, 2018, 290: 87–99.
- [16] EGGERT J, KORNER E. Sparse coding and NMF[C]// *Proceedings of 2004 IEEE International Joint Conference on Neural Networks*. Budapest, Hungary, 2004: 2529–2533.
- [17] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples[J]. *Journal of machine learning research*, 2006, 7(1): 2399–2434.
- [18] HOU C, JING W, YI W, et al. Local linear transformation embedding[J]. *Neurocomputing*, 2009, 72(10-12): 2368–2378.
- [19] LI H, LIU D, WANG D. Manifold regularized reinforcement learning[J]. *IEEE transactions on neural networks & learning systems*, 2017, 29(4): 932–943.
- [20] HOYER P O. Non-negative matrix factorization with sparseness constraints[J]. *Journal of machine learning research*, 2004, 5: 1457–1469.
- [21] NIE Feiping, HUANG Heng, CAI Xiao, et al. Efficient and robust feature selection via joint  $\ell_2$ ,  $\ell_1$ -norms minimization[C]// *Proceedings of the 23rd International Conference on Neural Information Processing Systems*. Vancouver, British Columbia, Canada, 2010: 1813–1821.

#### 作者简介:



徐慧敏, 女, 1994 年生, 硕士研究生, 主要研究方向为数字图像处理、模式识别。



陈秀宏, 男, 1964 年生, 教授, 博士后, 主要研究方向为数字图像处理 and 模式识别、优化理论与方法等。发表学术论文 120 余篇。