基于图正则化和稀疏约束的半监督非负矩阵分解

姜小燕[1] 孙福明[1] 李豪杰[2]
(辽宁工业大学电子与信息工程学院 锦州 121001)[1] (大连理工大学软件学院 大连 116300)[2]

摘 要 非负矩阵分解是在矩阵非负约束下的分解算法。为了提高识别率,提出了一种基于稀疏约束和图正则化的半监督非负矩阵分解方法。该方法对样本数据进行低维非负分解时,既保持数据的几何结构,又利用已知样本的标签信息进行半监督学习,而且对基矩阵施加稀疏性约束,最后将它们整合于单个目标函数中。构造了一个有效的更新算法,并且在理论上证明了该算法的收敛性。在多个人脸数据库上的仿真结果表明,相对于NMF、GNMF、CNMF等算法,GCNMFS具有更好的聚类精度和稀疏性。

关键词 非负矩阵分解,图正则,稀疏约束,半监督
中图法分类号 TP37 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.7.013

Semi-supervised Nonnegative Matrix Factorization Based on Graph Regularization and Sparseness Constraints

JIANG Xiao-yan[1] SUN Fu-ming[1] LI Hao-jie[2]
(School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou 121001, China)[1]
(School of Software Technology, Dalian University of Technology, Dalian 116300, China)[2]

Abstract Nonnegative matrix factorization (NMF) is a kind of matrix factorization algorithm under non-negative constraints. With the aim to enhance the recognition rate, a method called graph regularized and constrained non-negative matrix factorization with sparseness (GCNMFS) was proposed. It not only preserves the intrinsic geometry of data, but also uses the label information for semi-supervised learning and introduces sparseness constraint into base matrix. Finally, they are integrated into a single objective function. An efficient updating approach was produced and the convergence of this algorithm was also proved. Compared with NMF, GNMF and CNMF, experiments on some face databases show that the proposed method can achieve better clustering results and sparseness.

Keywords Nonnegative matrix factorization, Graph regularization, Sparseness constraints, Semi-supervised

1 引言

为高效率地处理存放于矩阵中的数据信息,一般采取将矩阵进行分解的方法。分解后,不但可将用于描述问题的原始矩阵的维数大大消减,同时也可以对原始矩阵中存放的大量数据进行压缩和概括。常用的传统矩阵分解方法有:主成分分析(PCA)、独立成分分析(ICA)、矢量量化(VQ)、奇异值分解(SVD)等,其共同点是允许分解后结果出现负值,从计算的角度看这是正确的,但从应用的角度看负值是没有实际意义的。与上述矩阵分解方法不同的是,非负矩阵分解[1](Non-negative Matrix Factorization, NMF)是目前国际上新的矩阵分解方法,并已初步成功地应用于图像处理、生物医学、文本聚类和语音信号处理等领域。1999年,Daniel D. Lee和H. Sebastian Seung[2,3]在《Nature》上首次提出了非负矩阵分解的概念,其详细介绍了NMF心理和生理学的理论依据,并从直观上展示了NMF"局部构成整体"的特点。非负矩阵分解通过添加"矩阵中所有元素均为非负数"的限制条件,保证了分解结果的可解释性,同时,它还具有实现简便和占用存储空间小的优点,从而更加贴近应用领域。因此,探索矩阵的非负分解方法一直是非常有意义的。

NMF自提出后受到了广泛的研究与应用,为了提高NMF算法的有效性和识别率,人们提出了许多改进算法。Hoyer等[4]提出了非负稀疏编码(Nonnegative Sparse Coding, NNSC)算法。NNSC算法结合了稀疏编码方法和NMF方法。后期Hoyer等进一步提出一种改进算法,在分解原始矩阵的同时,对基矩阵和系数矩阵进行一定程度的稀疏度控制,从而得到一种可以较为精确控制的稀疏性NMF算法(Non-negative Matrix Factorization with Sparseness Constraints, NMFSC)[5],使得分解后的矩阵具有较好的稀疏性,不仅节省了存储空间,同时还提高了运算效率。Wang等人[6]提出了基于Fisher约束的监督的非负矩阵分解方法,其在矩阵分解过程中通过最大化类间散度和类内散度的比值,将监督信息引入到图像表示中。David Guillam等人[7]提出了加权NMF;Ding等人[8]提出了一种对输入数据的正负性

到稿日期:2015-08-11 返修日期:2015-11-01 本文受国家自然科学基金(61572244,61472059),辽宁省高等学校优秀人才支持计划(LR2015030)资助。

姜小燕(1989—),女,硕士生,主要研究领域为图像语义理解,E-mail:1091061380@qq.com;孙福明(1972—),男,博士,教授,CCF会员,主要研究领域为计算机视觉、图像语义理解,E-mail:sunwenfriend@hotmail.com;李豪杰(1973—),男,博士,教授,CCF会员,主要研究领域为计算机视觉、图像语义理解,E-mail:hjli@dlut.edu.cn。

不敏感的 Semi-NMF 的矩阵分解算法。近几年，将流形学习与非负矩阵分解相结合已成为当前热点。Cai 等[9]提出了图正则化非负矩阵分解(GNMF)，在 NMF 的低维表示中，考虑原始样本的近邻几何结构，引进近邻图，使得低维表示很好地保留了原始样本的近邻结构。姜伟等人[10]提出了稀疏约束图正则非负矩阵分解算法(GNMFSC)，其不仅考虑了数据的几何信息，而且对系数矩阵增加稀疏约束，使分解后的人脸图像具有更高的识别率。

上述的标准 NMF 及其改进算法都属于无监督分解方法，没有考虑样本的标签信息。Liu 等[11]提出的 CNMF 算法直接将已知样本标签信息约束到 NMF 算法的目标函数，单一目标函数就实现了半监督分解，不足之处是无法保持样本所在空间的几何结构。本文结合半监督学习和流形学习的思想，并在此基础之上施加稀疏约束，提出一种基于图正则化和稀疏约束的半监督非负矩阵分解(Graph Regularized and Constrainted NMF with Sparseness，GCNMFS)。该方法不仅充分考虑样本的类别信息和数据空间固有的几何流形结构信息，还对进一步稀疏分解结果。实验结果证明了基于图正则化和稀疏约束的半监督非负矩阵分解算法的有效性。下面将重点介绍 GCNMFS 的求解、收敛证明及实验结果。

2 非负矩阵分解算法

对给定的矩阵进行分解，其结果往往是不唯一的。利用 NMF 算法进行分解时往往也是不尽相同的。对于给定的 n 个非负样本 x_i，$i=1,2,\cdots,n$，每一个 $x_i\in\mathbf{R}^m$ 均是列向量，组成非负矩阵 $X=[x_1,x_2,\cdots,x_n]\in\mathbf{R}_+^{m\times n}$。NMF 算法的目的就是寻找两个非负矩阵 $U\in\mathbf{R}^{m\times k}$ 和 $V\in\mathbf{R}^{n\times k}$，$k\leqslant mn/(m+n)$，使得 $X\approx UV$。也就是最小化下列目标函数：

$$O_F=\|X-UV^{\mathrm{T}}\|_F^2 \quad \text{s.t. } U\geqslant 0, V\geqslant 0 \tag{1}$$

其中，$\|\cdot\|_F$ 是 Frobenius 范数。其乘性迭代规则为：

$$u_{ik}\leftarrow u_{ik}\frac{(XV)_{ik}}{(UV^{\mathrm{T}}V)_{ik}} \tag{2}$$

$$v_{jk}\leftarrow v_{jk}\frac{(X^{\mathrm{T}}U)_{jk}}{(VU^{\mathrm{T}}U)_{jk}} \tag{3}$$

其中，$U=[u_{ik}]$，$V=[v_{jk}]$。计算过程中，给定迭代终止条件后，随机选取非负初始矩阵 U_0 和 V_0，按照式(2)和式(3)交替迭代更新直到满足终止条件，可得到最终的 U 和 V。

3 基于图正则化和稀疏约束的半监督非负矩阵分解算法

3.1 图正则化非负矩阵分解

蔡登教授等[9]针对流形数据提出了图正则非负矩阵分解(GNMF)算法。GNMF 非负分解算法要求在矩阵分解过程使降维后的数据(基矩阵)尽可能多地保持原始数据的几何内蕴结构。假设样本分布在高维欧氏空间的低维流形上，通过构建所有样本的近邻图来估计数据，即在矩阵分解过程中明确考虑数据集携带的几何信息：如果数据点 x_i 和 x_j 在原空间是邻近点，那么对应到新的基下 h_{ik} 和 h_{jk} 也是邻近点。

设原始数据点构成的图为 G，其中 S_{ij} 是权矩阵，$N_p(x_i)$ 表示 x_i 的 p 个近邻，则

$$S_{ij}=\begin{cases}\mathrm{e}^{-\frac{\|x_i-x_j\|^2}{\sigma}}, & x_i\in N_p(x_j)\text{或 }x_j\in N_p(x_i)\\ 0, & \text{其他}\end{cases}$$

定义 $L=D-S$，D 是对角矩阵，$D_{ii}=\sum_j S_{ij}$，L 是拉普拉斯矩阵。则 GNMF 算法最小化目标函数：

$$O_F=\|X-UV^{\mathrm{T}}\|_F^2+\lambda\mathrm{Tr}(V^{\mathrm{T}}LV) \quad \text{s.t. } U\geqslant 0, V\geqslant 0 \tag{4}$$

式(4)的迭代更新规则如下：

$$u_{ik}\leftarrow u_{ik}\frac{(XV)_{ik}}{(UV^{\mathrm{T}}V)_{ik}} \tag{5}$$

$$v_{jk}\leftarrow v_{jk}\frac{(X^{\mathrm{T}}U+\lambda SV)_{jk}}{(VU^{\mathrm{T}}U+\lambda DV)_{jk}} \tag{6}$$

3.2 受限的非负矩阵分解

传统的 NMF 是非监督学习算法，没有考虑数据样本的类别信息，因此，Liu 等[11]提出了 CNMF 算法，对部分有标签信息的样本进行监督约束，使得具有同类标签信息的样本在低维空间中投影为一个点，再进行聚类和识别会得到更好的效果。

假设 $X=[x_1,x_2,\cdots,x_n]$ 为 n 个非负样本，前 l 个样本 $x_1,x_2,\cdots,x_l$ 的标签信息已知，剩余的 $n-l$ 个样本的标签信息未知，共包括 c 类样本。定义标签索引矩阵 C，如果样本 x_j 被标记为第 i 类，则令 $c_{ij}=1$，否则令 $c_{ij}=0$。定义标签约束矩阵 A：

$$A=\begin{pmatrix}C_{l\times c} & 0\\ 0 & I_{n-l}\end{pmatrix} \tag{7}$$

其中，矩阵 I_{n-l} 是一个 $(n-l)\times(n-l)$ 维单位矩阵。

在 NMF 目标函数中，引入标签约束矩阵 A，即 $X\approx UV^{\mathrm{T}}=U(AZ)^{\mathrm{T}}$。其中 $Z\in\mathbf{R}^{(c+n-l)\times k}$ 是辅助矩阵，c 为样本的类别数，$k\ll m$ 且 $k\ll n$，同 NMF 算法中的矩阵 V 一样，辅助矩阵 Z 的初始值也是随机产生的。很容易得知，如果 x_i 和 x_j 具有相同的标签，那么其加权系数矢量是相同的。CNMF 算法中的目标函数如式(8)所示。

$$O_F=\|X-UZ^{\mathrm{T}}A^{\mathrm{T}}\|_F^2 \quad \text{s.t. } U\geqslant 0, Z\geqslant 0 \tag{8}$$

式(8)的迭代更新规则为：

$$u_{ik}\leftarrow u_{ik}\frac{(XAZ)_{ik}}{(UZ^{\mathrm{T}}A^{\mathrm{T}}AZ)_{ik}} \tag{9}$$

$$v_{jk}\leftarrow v_{jk}\frac{(A^{\mathrm{T}}X^{\mathrm{T}}U)_{jk}}{(A^{\mathrm{T}}AZU^{\mathrm{T}}U)_{jk}} \tag{10}$$

3.3 基于图正则化和稀疏约束的半监督非负矩阵分解

近年来，稀疏表示在信号处理领域得到了广泛的关注。为了使分解结果尽可能的稀疏以获得主要特征，本文对基矩阵 U 施加正则化稀疏约束。常用的正则化范数有：L_0 范数、L_1 范数和 Frobenius 范数等。L_0 范数很难优化求解，是 NP 难问题。L_1 范数正则化在估计问题中，通常会产生额外的基向量，且在某些特定情况下是无效的。Frobenius 范数不但可以防止过拟合，还可以让优化求解变得稳定和快速。因此，本文选择了 Frobenius 范数。

本文结合流形学习和半监督学习[15]的思想，提出一种基于图正则化[16]和稀疏约束的半监督非负矩阵分解算法(GCNMFS)。在半监督 NMF 算法中加入拉普拉斯图正则化限制，这样不仅能保持样本的类别信息，而且能在低维空间保持样本的几何结构信息。在基于图正则化的半监督非负矩阵分解算法的基础之上对基矩阵进行 Frobenius 范数稀疏约束，那么就可以得到更加有效和稀疏的基矩阵，从而节省了存储空间，提高了分解质量。

综合 GNMF 算法、CNMF 算法和基矩阵稀疏约束条件，获得新的最小化目标函数为：

$$O_F=\|X-UZ^{\mathrm{T}}A^{\mathrm{T}}\|_F^2+\lambda\mathrm{Tr}(Z^{\mathrm{T}}A^{\mathrm{T}}LAZ)+\beta\|U\|_F^2$$
$$\text{s.t. } U\geqslant 0, Z\geqslant 0 \tag{11}$$

其中，矩阵 A 是标签约束矩阵，L 是拉普拉斯矩阵，λ 是正则化参数，β 是稀疏系数，$\beta\in(0,1)$。$\|\cdot\|_F$ 是 Frobenius 范数。

注意，如果基矩阵的稀疏度增加了，学习部件的能力也会提高。在这个意义上，稀疏性对 NMF 是至关重要的。

利用最速下降法和迭代法，可以推导出如下最小化目标函数的乘性迭代规则：

$$\begin{aligned}O_F &= \mathrm{Tr}((X-UZ^{\mathrm{T}}A^{\mathrm{T}})(X-UZ^{\mathrm{T}}A^{\mathrm{T}})^{\mathrm{T}})+\lambda\mathrm{Tr}\\ &\quad (Z^{\mathrm{T}}A^{\mathrm{T}}LAZ)+\beta\|U\|_F^2\\ &= \mathrm{Tr}(XX^{\mathrm{T}})-2\mathrm{Tr}(XAZU^{\mathrm{T}})+\mathrm{Tr}(UZ^{\mathrm{T}}A^{\mathrm{T}}AZU^{\mathrm{T}})+\\ &\quad \lambda\mathrm{Tr}(Z^{\mathrm{T}}A^{\mathrm{T}}LAZ)+\beta\|U\|_F^2\end{aligned}$$

对约束 $u_{ik}\geqslant 0$ 和 $v_{jk}\geqslant 0$，令 ϕ 和 φ 是相对应的拉格朗日乘子，则拉格朗日函数为：

$$L=O_F+\mathrm{Tr}(\phi U^{\mathrm{T}})+\mathrm{Tr}(\varphi Z^{\mathrm{T}})$$

对上式分别求 U 和 Z 的偏导数并令其等于 0，得：

$$\frac{\partial L}{\partial U}=-2XAZ+2UZ^{\mathrm{T}}A^{\mathrm{T}}AZ+\phi+2\beta U=0$$

$$\frac{\partial L}{\partial Z}=-2A^{\mathrm{T}}X^{\mathrm{T}}U+2A^{\mathrm{T}}AZU^{\mathrm{T}}U+2\lambda A^{\mathrm{T}}LAZ+\varphi=0$$

使用 KKT 条件 $\phi_{ij}u_{ij}=0$ 和 $\varphi_{ij}z_{ij}=0$，最终得到下列迭代更新规则：

$$u_{ik}\leftarrow u_{ik}\frac{(XAZ)_{ik}}{(UZ^{\mathrm{T}}A^{\mathrm{T}}AZ)_{ik}+\beta U_{ik}} \tag{12}$$

$$z_{jk}\leftarrow z_{jk}\frac{(A^{\mathrm{T}}X^{\mathrm{T}}U+\lambda A^{\mathrm{T}}SAZ)_{jk}}{(A^{\mathrm{T}}AZU^{\mathrm{T}}U+\lambda A^{\mathrm{T}}DAZ)_{jk}} \tag{13}$$

3.4 GCNMFS 算法的收敛性分析

定义 1 定义 $G(x,x')$ 是 $F(x)$ 的辅助函数，并且满足 $G(x,x')\geqslant F(x)$，$G(x,x)=F(x)$。

引理 1 若 $G(x,x')$ 是 $F(x)$ 的辅助函数，那么 $F(x)$ 对于以下更新规则是单调不增的：。

$$x^{t+1}=\arg\min_x G(x,x^t) \tag{14}$$

证明：

$$F(x^{t+1})\leqslant G(x^{t+1},x^t)\leqslant G(x^t,x^t)=F(x^t)$$

需要注意的是，$F(x^{t+1})=F(x^t)$ 等式成立的一个充分条件是：x^t 是函数 $G(x,x')$ 的一个局部最小值。如果函数 F 存在导数且在 x^t 的一个微小领域内连续，则微分 $\nabla F(x^t)=0$。通过式(14)即可得到收敛到局部极小点 $x_{\min}=\arg\min_x F(x)$ 的序列：

$$F(x_{\min})\leqslant\cdots\leqslant F(x^{t+1})\leqslant F(x^t)\leqslant\cdots\leqslant F(x^1)\leqslant F(x^0)$$

所以，通过定义这样的辅助函数 $G(x,x')$，使得目标函数式(11)的相应的迭代规则满足 $x^{t+1}=\arg\min_x G(x,x^t)$。首先需要证明更新规则式(13)的收敛性。对于 Z 中的任何元素 z_{ij}，用 $F_{z_{ij}}$ 表示目标函数中仅与 z_{ij} 相关的部分。

引理 2 假设 F' 是关于 Z 的一阶微分，函数

$$G(z,z_{ij}^t)=F_{z_{ij}}(z_{ij}^t)+F'_{z_{ij}}(z-z_{ij}^t)+\frac{(A^{\mathrm{T}}AZU^{\mathrm{T}}U+\lambda A^{\mathrm{T}}DAZ)_{ij}}{z_{ij}^t}(z-z_{ij}^t)^2 \tag{15}$$

是 $F_{z_{ij}}$ 的辅助函数。

证明：容易得到 $G(z,z)=F_{z_{ij}}(z)$，根据辅助函数的定义，只需要证明 $G(z,z_{ij}^t)\geqslant F_{z_{ij}}(z)$。$F_{z_{ij}}(z)$ 的泰勒展开式为：

$$F_{z_{ij}}(z)=F_{z_{ij}}(z_{ij}^t)+F'_{z_{ij}}(z_{ij}^t)(z-z_{ij}^t)+[(AA^{\mathrm{T}}U^{\mathrm{T}}U)_{ij}+\lambda(A^{\mathrm{T}}LA)_{ij}](z-z_{ij}^t)^2 \tag{16}$$

与辅助函数相比，$G(z,z_{ij}^t)\geqslant F_{z_{ij}}(z)$ 等价于：

$$\frac{(A^{\mathrm{T}}AZU^{\mathrm{T}}U)_{ij}+\lambda(A^{\mathrm{T}}DAZ)_{ij}}{z_{ij}^t}\geqslant(AA^{\mathrm{T}})_{ii}(U^{\mathrm{T}}U)_{jj}+\lambda(A^{\mathrm{T}}LA)_{ii}$$

将上式等价分解成

$$\frac{(A^{\mathrm{T}}AZU^{\mathrm{T}}U)_{ij}}{z_{ij}^t}\geqslant\frac{1}{2}F''_{ij}=(AA^{\mathrm{T}})_{ii}(U^{\mathrm{T}}U)_{jj} \tag{17}$$

$$\frac{\lambda(A^{\mathrm{T}}DAZ)_{ij}}{z_{ij}^t}\geqslant\lambda(A^{\mathrm{T}}LA)_{ii} \tag{18}$$

由于

$$\begin{aligned}(A^{\mathrm{T}}AZU^{\mathrm{T}}U)&=\sum_l(A^{\mathrm{T}}AZ)(U^{\mathrm{T}}U)\geqslant(A^{\mathrm{T}}AZ)(U^{\mathrm{T}}U)\\ &\geqslant\sum_l(A^{\mathrm{T}}A)z^t(U^{\mathrm{T}}U)\geqslant z^t(A^{\mathrm{T}}A)(U^{\mathrm{T}}U)\end{aligned}$$

故式(17)得证。

由于

$$\begin{aligned}\lambda(A^{\mathrm{T}}DAZ)&=\lambda\sum_l(A^{\mathrm{T}}DA)z^t\geqslant\lambda(A^{\mathrm{T}}DA)z^t\\ &\geqslant\lambda(A^{\mathrm{T}}(D-U)A)z^t=\lambda(A^{\mathrm{T}}LA)z^t\end{aligned}$$

故式(18)得证。由式(17)和式(18)可证 $G(z,z_{ij}^t)\geqslant F_{z_{ij}}(z)$。

引理 3 假设 F' 是关于 U 的一阶微分，函数

$$G(u,u_{ij}^t)=F_{u_{ij}}(u_{ij}^t)+F'_{u_{ij}}(u_{ij}^t)(u-u_{ij}^t)+\frac{(UZ^{\mathrm{T}}A^{\mathrm{T}}AZ+\beta U)_{ij}}{u_{ij}^t}(u-u_{ij}^t)^2 \tag{19}$$

是目标函数 O_F 关于变量 z_{ij} 的辅助函数。其证明过程与引理 2 相似。

定理 1 目标函数式(11)在迭代式(12)和(13)更新条件下是非增的，当且仅当 U 和 Z 是不动点，目标函数式(14)是不变的。

证明：将式(15)中 $G(z,z_{ij}^t)$ 应用到式(14)中，得到：

$$\begin{aligned}z_{ij}^{t+1}&=\arg\min_z G(z,z_{ij}^t)\\ &=z_{ij}^t-z_{ij}^t\frac{F'_{ij}(z_{ij}^t)}{2(A^{\mathrm{T}}AZU^{\mathrm{T}}U)_{ij}+2\lambda(A^{\mathrm{T}}DAZ)_{ij}}\\ &=z_{ij}^t\frac{(A^{\mathrm{T}}X^{\mathrm{T}}U+\lambda A^{\mathrm{T}}SAZ)_{ij}}{(A^{\mathrm{T}}AZU^{\mathrm{T}}U+\lambda A^{\mathrm{T}}DAZ)_{ij}}\end{aligned}$$

根据引理 2 可知式(15)是一个辅助函数，因此 $F_{z_{ij}}(z)$ 在更新过程中是非递增的。

将式(19)中 $G(u,u_{ij}^t)$ 应用到式(14)中，得到：

$$u_{ij}^{t+1}=\arg\min_u G(u,u_{ij}^t)=u_{ij}^t\frac{(XAZ)_{ij}}{(UZ^{\mathrm{T}}A^{\mathrm{T}}AZ+\beta U)_{ij}}$$

同理，由引理 3 可知式(19)是一个辅助函数，因此 $F_{u_{ij}}(u)$ 在更新过程中是非递增的。

由此可知，定理保证目标函数(11)在式(12)和式(13)迭代更新后能收敛到一个局部最优解。

4 实验与结果分析

实验环境：Windows 7，CPU：2.60GHz，内存：4GB。程序环境：Matlab R2012b。

4.1 数据集

本文使用以下数据集验证算法的有效性：ORL 数据集和 PIE_pose27 数据集。

(1)ORL 人脸数据库是由英国剑桥大学 AT&T 验室创建的，包含 40 个人的人脸图像，每人 10 张总共 400 张灰度图像，图像大小为 112×92，其中部分图像包含了光照、姿态、表情和面部饰物的变化。

(2)PIE 数据集(PIE_pose27)是由美国卡耐基梅隆大学创建，包含 68 名自愿者的面部表情图像，每个人的姿态和光照变化都是在严格控制的条件下采集的。PIE_pose27 数据集共有 42 种不同的光照条件共计 2856 幅图片。

两个数据集的对比情况如表 1 所列，部分图片如图 1 所示。

表 1　两个数据集信息

数据集	图片数(n)	维度(m)	类别(class)
ORL	400	1024	40
PIE_pose27	2856	1024	68

(a)ORL 数据库

(b)PIE_pose27 数据库

图 1　数据库中部分图像示例

本文实验方案：选取每类样本的前 20% 作为标记样本，其余作为未标记样本，并从中随机选择 k 类样本进行聚类实验，重复实验 20 次。

下面给出基于 GCNMFS 算法的图像聚类的具体步骤。

算法 1　基于 GCNMFS 算法的图像聚类算法

输入：数据集矩阵 X、选择类别数 k

过程：

(1)初始化参数。设定参数 β(0<β<1)、分解维度 k 和最大迭代次数(nIterMax)，计算索引矩阵 C 并构造标签矩阵 A。随机生成非负矩阵 $U_{m\times k}$ 和 $Z_{n\times k}$，并对原始矩阵 $X_{m\times n}$ 进行归一化。

(2)FOR nI=1:nIterMax(nI 是迭代次数)

(3)运用迭代规则(12)和(13)进行迭代运算；

(4)根据式(11)计算最小化目标函数；

(5)END FOR

(6)返回最终迭代后的矩阵 U 和 Z，利用 k-means 法对系数矩阵 V=AZ 进行聚类可验证该算法的聚类性能。

4.2　评价指标

在实验中，本文利用两个评价指标来验证 GCNMFS 算法的聚类性能，分别是准确率(AC)和归一化互信息(NMI)。

准确率(AC)[12]：通常用来测量具有正确类别信息的样本的比例。假设给定一个包括 n 个数据的样本集，对于每个样本，利用 l_i 来表示通过不同算法获得的样本类别信息，r_i 表示数据集中提供的正确样本类别信息。因此 AC 可以定义为：

$$AC=\frac{\sum_{i=1}^{n}\delta(r_i,map(l_i))}{n} \tag{20}$$

其中，$\delta(x,y)$是一个二值函数，当 $x=y$ 时，$\delta(x,y)=1$；当 $x\neq y$ 时，$\delta(x,y)=0$。$map(l_i)$是映射函数，表示将每个类别信息 l_i 映射到数据集中相同类别的样本。这一映射算法可以用 Kuhn-Munkres 算法来实现。

在聚类应用中，互信息(MI)[13]用来度量两个聚类样本集的相似性。假设给定两种聚类样本集 C 和 C'，则它们的互信息 MI 可以定义为：

$$MI(C,C')=\sum_{c_i\in C,c_j'\in C'}p(c_i,c_j')\cdot\log\frac{p(c_i,c_j')}{p(c_i)\cdot p(c_j')} \tag{21}$$

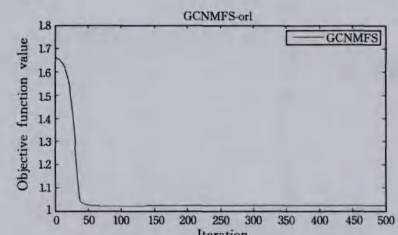
其中，$p(c_i)$和 $p(c_j')$表示从数据集中任意选择一个样本分别属于第 c_i 类和第 c_j'类的概率，$p(c_i,c_j')$表示样本同时属于第 c_i 类和第 c_j'类的联合概率。$H(C)$和 $H(C')$表示聚类 C 和 C'的熵，$MI(C,C')$的值在 0 到 $\max(H(C),H(C'))$之间。当 $MI(C,C')=0$ 时，表示聚类 C 和 C'完全独立。当 $MI(C,C')$达到最大值时，表示聚类 C 和 C'结果一致。$MI(C,C')$的一个重要性质是对于聚类的各种排列其值始终不变。本文实验对 $MI(C,C')$进行了归一化处理，使其值在 0 和 1 之间。

$$NMI(C,C')=\frac{MI(C,C')}{\max(H(C),H(C'))}\in[0,1] \tag{22}$$

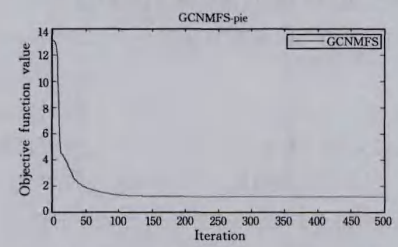
4.3　参数选择

GCNMFS 模型有两个关键的折衷参数：正则项参数 λ 和稀疏系数 β。

为了确定正则项参数 λ，在两个数据集上进行了实验。从实验结果来分析，GCNMFS 在 λ 从 10～1000 变化时，其值对准确率和归一化互信息的影响不大。根据实验经验，取值 $\lambda=100$。同样地，稀疏系数 β 的值也由实验的结果确定。若稀疏系数 β 的值太大，会使得图像过于稀疏以至于不能得到很好的图像表示。β 的取值范围是 0.1 到 0.9，经过多次实验，取值 $\beta=0.3$。



(a) ORL 数据集上目标函数的收敛曲线



(b) PIE_pose27 数据集上目标函数的收敛曲线

图 2　该算法目标函数的收敛曲线

由图 2 可以看出，两个数据集的目标函数在迭代初始时下降非常快；随着迭代次数的增加，目标函数下降越来越缓慢，基本上迭代次数在 50 次以内，就已经收敛得非常好了。考虑到算法的复杂度(程序运行时间)，可设置最大迭代次数为 500。

4.4　实验结果

本文的实验目的是评估 GCNMFS 算法在 ORL 库和PIE_pose27库中的实验效果。因此，需要对表示后的低维数据进行聚类，然后根据聚类的结果来评估数据的表示性能。为了证明所提算法的有效性，在实验结果分析时与几种常用的数据表示算法对比，对比算法主要包括：

(1)NMF：非负矩阵分解算法，能广泛应用于图像和文本

数据的表示；

(2)CNMF：受限的非负矩阵分解算法，充分考虑了样本的类别信息；

(3)GNMF：图正则非负矩阵分解算法，将流形学习与非负矩阵分解相结合；

(4)GCNMFS：本文提出的一种基于图正则和稀疏约束的有监督非负矩阵分解方法。

实验中的对比算法同样都是随机取 $k(k=2,3,\cdots,10)$ 类进行实验，并且对每个 k 值都运行 20 次后取平均值作为最终的结果。

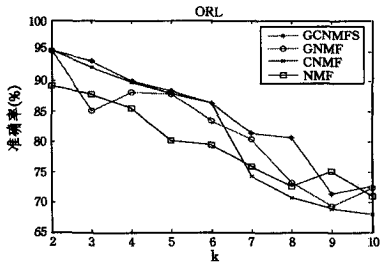
各算法在数据集上聚类的准确率见表 2 和表 3，相对应的聚类准确度的曲线如图 3 所示。

表 2　不同算法在 ORL 数据集上聚类的准确率(AC)

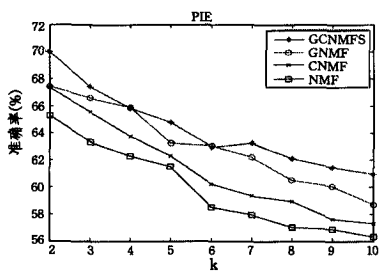
聚类的准确率(AC)(%)				
k	NMF 空间	CNMF 空间	GNMF 空间	GCNMFS 空间
2	89.20	95.00	95.01	95.02
3	87.89	92.28	85.07	93.33
4	85.49	89.79	88.14	90.05
5	80.21	88.00	87.84	88.42
6	79.53	86.36	83.52	86.33
7	75.86	74.28	80.42	81.43
8	72.65	70.75	73.28	80.75
9	75.03	68.89	69.35	71.36
10	71.01	68.02	72.36	72.62
平均值	79.65	81.49	81.67	84.37

表 3　不同算法在 PIE 数据集上聚类的准确率(AC)

聚类的准确率(AC)(%)				
k	NMF 空间	CNMF 空间	GNMF 空间	GCNMFS 空间
2	65.30	67.34	67.45	70.03
3	63.34	65.58	66.60	67.43
4	62.30	63.75	65.88	65.84
5	61.51	62.30	63.24	64.75
6	58.50	60.25	63.02	62.95
7	57.94	59.34	62.20	63.23
8	57.01	58.93	60.52	62.12
9	56.88	57.62	60.03	61.45
10	56.30	57.27	58.68	60.95
平均值	59.90	61.38	63.07	64.30



(a)ORL 数据集



(b)PIE_pose27 数据集

图 3　聚类准确率

由表 2、表 3 可知，GNMF 算法、CNMF 算法和 GCNMFS 算法在两个数据集上的聚类准确率平均值相对于 NMF 算法聚类有较大的改善。在 ORL 数据集中，GCNMFS 算法比 NMF 上的聚类准确率平均高 4.71%，比 GNMF 算法平均高 2.69%，也比 CNMF 算法的聚类效果好。在 PIE_pose27 数据集中，GCNMFS 算法比 NMF 上的聚类准确率平均高 4.4%，比 GNMF 算法平均高 1.23%，比 CNMF 算法平均高 2.92%。

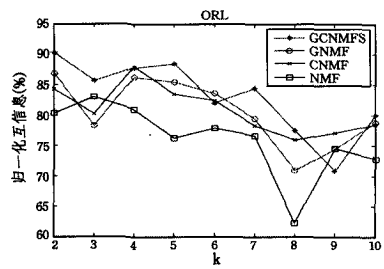
聚类的归一化互信息见表 4 和表 5，相对应的曲线如图 4 所示。

表 4　不同算法在 ORL 数据集上的互信息

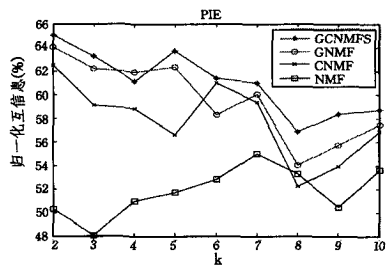
聚类的互信息值(%)				
k	NMF 空间	CNMF 空间	GNMF 空间	GCNMFS 空间
2	80.31	84.30	86.83	90.32
3	83.01	80.20	78.23	85.67
4	80.91	88.00	86.20	87.75
5	76.27	83.48	85.48	88.48
6	77.81	82.33	83.59	81.95
7	76.51	78.28	79.44	84.38
8	62.25	75.88	70.88	77.48
9	74.44	76.99	74.38	70.82
10	72.77	78.43	78.80	80.05
平均值	76.03	80.88	80.43	82.99

表 5　不同算法在 PIE 数据集上的互信息

聚类的互信息值(%)				
k	NMF 空间	CNMF 空间	GNMF 空间	GCNMFS 空间
2	50.31	62.50	64.02	65.01
3	48.02	59.09	62.21	63.26
4	50.99	58.82	61.90	61.11
5	51.70	56.63	62.33	63.72
6	52.80	60.98	58.30	61.41
7	55.00	59.32	60.04	60.95
8	53.30	52.25	54.01	56.85
9	50.43	53.90	55.73	58.33
10	53.60	56.85	57.46	58.70
平均值	51.79	57.82	59.56	61.04



(a)ORL 数据集



(b)PIE_pose27 数据集

图 4　归一化互信息

由表 4 和表 5 可以看出，GNMF 算法、CNMF 算法和 GCNMFS 算法的归一化互信息平均值比 NMF 的要高得多。在 ORL 数据集中，GCNMFS 算法比 NMF 上的聚类互信息

平均高 6.96%，比 GNMF 算法平均高 1.25%，比 CNMF 算法平均高 2.11%。在 PIE_pose27 数据集中，GCNMFS 算法比 NMF 的聚类准确率平均高 9.25%，比 GNMF 算法平均高 1.48%，比 CNMF 算法平均高 3.22%。

总体来看，GCNMFS、CNMF 及 GNMF 在维数约简后数据上的聚类效果远比在 NMF 上的聚类效果好。实验表明，提出的 GCNMFS 取得了最好的聚类效果。

4.5 运行时间对比

在 ORL 人脸数据库上，迭代次数为 500 的情况下，NMF，CNMF，GNMF 以及 GCNMFS 4 种算法分别消耗的运行时间如表 6 所列。在 PIE 人脸数据库上，迭代次数为 500 的情况下，4 种算法的时间消耗情况如表 7 所列。随机取 k（$k=2,3,\cdots,10$）类进行实验，并且对每个 k 值都运行 20 次后取平均值作为最终的结果。

表 6 不同算法在 ORL 数据集上的运行时间

	运行时间(s)			
k	NMF 空间	CNMF 空间	GNMF 空间	GCNMFS 空间
2	0.85	2.76	2.53	2.85
3	1.04	3.14	2.60	3.54
4	1.26	3.81	2.73	4.07
5	1.38	4.52	3.03	4.90
6	1.59	5.52	3.09	5.71
7	1.80	6.13	3.26	6.54
8	1.96	6.54	3.50	6.84
9	2.19	7.17	3.78	8.22
10	2.38	8.16	4.12	10.15
平均值	1.60	5.30	3.18	5.87

表 7 不同算法在 PIE 数据集上的运行时间

	运行时间(s)			
k	NMF 空间	CNMF 空间	GNMF 空间	GCNMFS 空间
2	1.94	6.91	3.11	6.69
3	2.91	9.35	3.63	9.60
4	3.72	12.91	5.15	13.20
5	4.85	16.83	6.47	17.01
6	5.89	20.96	7.79	21.77
7	7.04	27.02	8.34	29.17
8	8.78	32.77	9.63	34.49
9	10.52	41.82	10.37	42.50
10	12.12	48.87	11.36	50.92
平均值	6.42	24.16	7.32	25.04

根据表 6 和表 7 的结果可知，NMF，CNMF，GNMF 和 GCNMFS 的运行时间随着样本类别数的增加呈线性增长。NMF 所用的时间最少，其次是 GNMF，而 GNMF 和 GCNMFS 的运行时间都较长，差距并不大。从表 7 可以看出，NMF 算法在运行时间上具有很明显的优势。

4.6 基图像的稀疏度

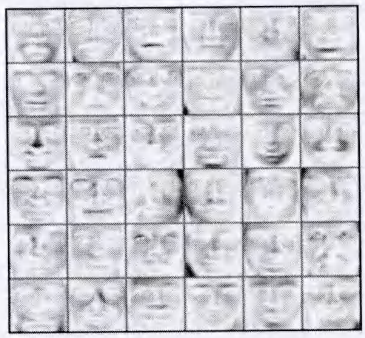
首先定义度量稀疏度的函数为[14]：

$$sparseness(x)=\frac{1}{n-1}\left[n-\left(\frac{\|x\|_1}{\|x\|_2}\right)^2\right] \tag{23}$$

其中，$\|\cdot\|_1$ 是向量的 1 范数，$\|\cdot\|_2$ 是向量的 2 范数，n 是向量 x 的维度，$0\leqslant sparseness(x)\leqslant 1$，当且仅当 x 仅有一个非零元时，$sparseness(x)=1$；当所有元素都相等且不为零时，$sparseness(x)=0$。

实验中 ORL 和 PIE_pose27 的特征维数皆取 36，对于分别用 NMF 算法和 GCNMFS 算法对由训练图像构成的非负矩阵 X 进行矩阵分解得到的基图像，其中用式(23)度量稀疏度，把基矩阵 U 看作向量，计算稀疏度。

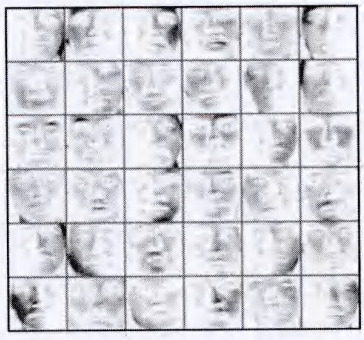
图 5 和图 6 分别是 ORL 数据集的基图像和 PIE_pose27 数据集的基图像在不同稀疏度下的示意图。由图 5 和图 6 可知，在这两个数据库上对比两个算法基图像稀疏度，NMF 算法的稀疏度较差，GCNMFS 算法的稀疏度高于 NMF。由此表明 GCNMFS 算法可以更好地得到图像的局部表示。



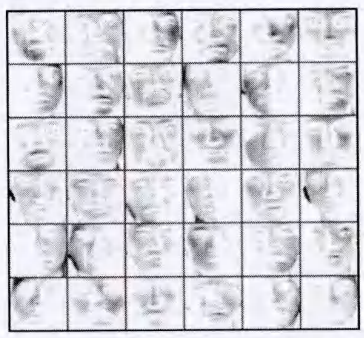
(a) NMF 的稀疏度 0.4133

(b) GCNMFS 的稀疏度 0.4727

图 5 ORL 数据集的基图像



(a)NMF 的稀疏度 0.5093



(b)GCNMFS 的稀疏度 0.6069

图 6 PIE_pose27 数据集的基图像

结束语 本文提出了基于图正则化和稀疏约束的有监督

(下转第 105 页)

Grayhole Attacks in AODV Based MANETs[C]// Proceedings of the 2013 Third International Conference on Advanced Computing & Communication Technologies. IEEE Computer Society,2013:254-260

[11] Zapata M G. Secure ad hoc on-demand distance vector routing. [J]. ACM Mobile Computing & Communication Review Number,2002,6(3):106-107

[12] Sridhar S,Baskaran R,Chandrasekar P. Energy supported AODV (EN-AODV) for QoS routing in MANET[J]. Procedia-Social and Behavioral Sciences,2013,73:294-301

[13] Lian Jian-wu,Ma Xiao-liang,Xu Long-long. Research and optimization of AODV routing protocols in mobile Ad Hoc network [J]. Journal of Chongqing University,2015,38(4):152-158(in Chinese)
梁建武,马晓亮,徐龙龙. 移动 Ad Hoc 网络 AODV 路由协议的研究与优化[J]. 重庆大学学报,2015,38(4):152-158

[14] Rzadca K,Yong J,Datta A. Multi-objective optimization of multicast overlays for collaborative applications[J]. Computer Networks the International Journal of Computer & Telecommunications Networking,2010,54(12):1986-2006

[15] Pati S,Som S K,Chakraborty S. Ant colony optimization algorithm for the Euclidean location-allocation problem with unknown number of facilities[J]. Journal of Intelligent Manufacturing,2013,24(1):45-54

[16] Xiong Y,Kuang Y. Applying an Ant Colony Optimization Algorithm-Based Multi-objective Approach for Time-Cost Trade-Off [J]. Journal of Construction Engineering & Management,2013,134(2):153-156

[17] Yu L,Li M,Yang Y,et al. An Improved Ant Colony Optimization for Vehicle Routing Problem[C]// Logistics@sThe Emerging Frontiers of Transportation and Development in China. ASCE,2015:3360-3366

[18] The Network Simulator-ns-2 [EB/OL]. http://www.isi.edu/nsnam/ns

[19] Nie Zhi,Liu Jing,Gan Xiao-ying,et al. A relay nodel selection technique for opportunistic routing in mobile Ad Hoc networks [J]. Journal of Chongqing University of Posts and Telecommunication(Natural Science Edition),2010,22(4):421-425,449(in Chinese)
聂志,刘静,甘小莺,等. 移动 Ad Hoc 网络中机会路由转发策略的研究[J]. 重庆邮电大学学报(自然科学版),2010,22(4):421-425,449

(上接第 82 页)

非负矩阵分解,并给出了相应的迭代求解公式和收敛性证明。在 ORL 和 PIE_pose27 图像库上,本文对提出的新方法进行了实验,利用两个评价指标即聚类准确率和归一化互信息衡量算法的识别能力。从实验结果可看出 GCNMFS 算法明显优于其它几种 NMF 算法,说明了 GCNMFS 算法的有效性。最后,考察了算法的稀疏度,结果显示提出的算法的稀疏度最高。因此,该算法能够得到图像的最佳局部表示,使得基图像具有更好判别能力。但本文算法中的参数 λ 需要通过搜索得到最优值,因此如何有效选择 λ 是以后研究的重点之一。

参 考 文 献

[1] Paatero P,Tapper U. Positive Matrix Factorization:A Nonnegative Factor Model with Optimal Utilization of Error Estimates of Data Values [J]. Environmetrices,1994,5(2):111-126

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

[3] Lee D D, Seung H S. Algorithms for non-negative matrix factorization [J]. Advances in Neural Information Processing Systems,2001,13:556-562

[4] Hoyer P O. Non-negative sparse coding [C]// Proceedings of IEEE Workshop on Neural Networks for Signal Processing. Martigny,Switzerland,2002:557-565

[5] Hoyer P O. Non-negative matrix factorization with sparseness constrains[J]. Journal of Machine Learning Research, 2004, 5(9):1457-1469

[6] Wang Y, Jia Y. Fisher non-negative matrix factorization for learning local feature [C] // Proceedings of Asian Conf. on Comp. Vision. 2004:27-30

[7] Guillamet D, Vitria J, Schiele B. Introducting a weighted non-negative matrix factorization for image classification [J]. Pattern Recognition,2003,24(14):2447-2454

[8] Ding C, Li T, Jordan M. Convex and semi-nonnegative matrix factorizations[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2010,32(1):45-55

[9] Cai Deng,He Xiao-fei,Han Jia-wei,et al. Graph regularized nonnegative matrix factorization for data representation[J]. IEEE Trans on Pattern Anal Mach Intell,2011,33(8):1548-1560

[10] Jiang Wei,Li Hong,Yu Zhen-guo,et al. Graph regularized Nonnegative Matrix Factorization with Sparseness Constraints. [J]. Computer Science,2013,1(40):218-256(in Chinese)
姜伟,李宏,于震国,等. 稀疏约束图正则非负矩阵分解[J]. 计算机科学,2013,1(40):218-256

[11] Liu Hai-feng,Wu Zhao-hui,Li Xue-long,et al. Constrained nonnegative matrix factorization for image representation [J]. IEEE Trans on Pattern Anal Mach,2012,34(7):1299-1311

[12] Ding C,Li T,Peng W,et al. Orthogonal nonnegative matrix t-factorizations for clustering [C] // Proceedings of SIGKDD. 2006:126-135

[13] Li S Z, Hou X, Zhang H, et al. Learning spatially localized, parts-based representation[C] // Proc of 2001 IEEE Computer Vision and Pattern Recognition. 2001:207-212

[14] Michael W, Shakhina A, Stewart G W. Computing sparse reduced-rank approximations to sparse matrices [J]. ACM Transactions on mathematical software,2004,19(3):231-235

[15] Du Shi-qiang, Shi Yu-qing, Wang Wei-lan, et al. Graph regularized-based semi-supervised non-negative Matrix Factorization. [J]. Computer Engineering and Applications,2012,48(36):194-200(in Chinese)
杜世强,石玉清,王维兰,等. 基于图正则化的半监督非负矩阵分解[J]. 计算机工程与应用,2012,48(36):194-200

[16] Shu Zhen-qiu, Zhao Chun-xia. Graph-regularized Constrained Non-Negative Matrix Factorization algorithm and its application to image representation[J]. Pattern Recognition and Artificial Intelligence,2013,26(3):300-306(in Chinese)
舒振球,赵春霞. 基于图正则化的受限非负矩阵分解算法及在图像表示中的应用[J]. 模式识别与人工智能,2013,26(3):300-306