计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200200106

基于预处理的超图非负矩阵分解算法

李向利[1,2,3]　贾梦雪[1,4]
1 桂林电子科技大学数学与计算科学学院　广西 桂林 541004
2 广西密码学与信息安全重点实验室　广西 桂林 541004
3 广西自动检测技术与仪器重点实验室　广西 桂林 541004
4 广西高校数据分析与计算重点实验室　广西 桂林 541004

摘　要　随着多媒体技术的发展,信息越来越多的以图片的形式出现。如何对海量的无标签图片进行聚类,是机器学习领域的热点问题。而图像聚类在人脸识别、手写数字识别等领域也有着重要的作用。由于图片数据通常以非负矩阵的形式存储,因此非负矩阵分解算法(NMF)在图像聚类领域得到了广泛的应用。但是NMF算法直接在数据的原始空间进行处理,这就导致NMF算法所得的图片标签易受到数据采集过程中含有的噪声等不利因素的影响。为了解决这些问题,提出了一种基于预处理的超图非负矩阵分解算法(Nonnegative Matrix Factorization with Hypergraph Based on Per-treatments,PHGNMF)。PHGNMF算法将预处理操作和超图的思想引入到NMF算法。在预处理的过程中,使用灰度处理来去除图片中不同光线条件所带来的影响,采用小波分析来提取图片的低时频子图,同时降低了算法所处理的矩阵维度。采取构建超图的方法来进一步保留对聚类结果有重要影响的数据局部结构。最后在5个主流数据集上的实验验证了PHGNMF算法相对于传统算法的有效性,结果显示聚类精度提升了2%～7%,标准互信息在部分数据集上提升了2%～5%。

关键词:图像聚类;非负矩阵分解;灰度处理;小波分析;超图
中图法分类号　TP391.4

Nonnegative Matrix Factorization Algorithm with Hypergraph Based on Per-treatments

LI Xiang-li[1,2,3] and JIA Meng-xue[1,4]
1 School of Mathematics & Computing Science,Guilin University of Electronic Technology,Guilin,Guangxi 541004,China
2 Guangxi Key Laboratory of Cryptography and Information Security,Guilin,Guangxi 541004,China
3 Guangxi Key Laboratory of Automatic Testing Technology and Instrument,Guilin,Guangxi 541004,China
4 Guangxi University Key Laboratory of Data Analysis and Calculation,Guilin,Guangxi 541004,China

Abstract　With the development of the media technology,more information is stored as the pictures. It is a topic problem in the machine learning field that how to distribute the right label to lots of unsigned pictures. And the image clustering has wide application on the face recognition and the handwriting number recognition field. Because the pictures are always stored as nonnegative matrices,the nonnegative matrix factorization algorithm (NMF) plays an important role in the image clustering. But the disadvantage in NMF algorithm is that the algorithm processes the data in the original data space which may produce a terrible result when the data have errors. To address this problem,the proposed algorithm is the nonnegative matrix factorization algorithm with a hypergraph based on per-treatments (PHGNMF). The PHGNMF algorithm introduces the per-treatments and the hypergraph into the NMF algorithm. In the per-treatments,the algorithm uses the grayscale normalization to eliminate the influence of the different illuminations firstly and then the algorithm can extract the low-frequency information of the pictures by the wavelet analysis. The wavelet procession could also reduce the dimensions of the data. The algorithm constructs a hypergraph for the data to save the neighboring information which has an important influence in the clustering procession. At last the results in five fundamental data sets confirm the effectiveness of the algorithm compared with fundamental algorithms. The results show the in-

到稿日期:2020-01-24　返修日期:2020-04-28　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(11961010,61967004),广西自然科学基金(2018GXNSFAA138169),广西密码学与信息安全重点实验室研究课题(GCIS201708),广西自动检测技术与仪器重点实验室基金(YQ19111),桂林电子科技大学研究生教育创新计划资助项目(2020YCXS087).
This work was supported by the National Natural Science Foundation of China (11961010,61967004),Guangxi Natural Science Foundation (2018GXNSFAA138169),Guangxi Key Laboratory of Cryptography and Information Security (GCIS201708),Guangxi Key Laboratory of Automatic Testing Technology and Instruments (YQ19111) and Innovation Project of GUET Graduate Education (2020YCXS087).
通信作者:李向利(lixiangli@guet.edu.cn)

crease of accuracy is 2%～7% and the increase of normalized mutual information on some data sets is 2%～5%.

Keywords Image clustering, Nonnegative matrix factorization, Grayscale normalization, Wavelet analysis, Hypergraph

1 引言

随着多媒体技术和网络的发展，越来越多的数据以图像的形式出现。图像聚类问题是机器学习领域的一个热点问题。而图片数据通常是以高维的非负的矩阵形式进行储存，因此直接对原始图片数据进行聚类并不是最好的方法。通常的做法是将图片映射到一个低维的空间中，对图片在低维空间中的代表进行操作以得到最终的聚类结果。因此，矩阵降维的方法在图像聚类领域有着广泛的应用，例如主成分分析(PCA)[1]、奇异值分解(SVD)[2]、线性判别分析(LDA)[3]等。但是以上方法通常会将原本非负的图片数据矩阵近似分解成含有负元素的子矩阵的乘积形式，从而降低了算法的可解释性。Lee 等[4]提出了非负矩阵分解算法(Nonnegative Matrix Factorization，NMF)。NMF 算法是将非负矩阵 $\boldsymbol{X}$ 由两个非负子矩阵 $\boldsymbol{U}$，$\boldsymbol{V}$ 乘积的形式逼近，基本形式为 $\boldsymbol{X}\approx\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}$，其中 $\boldsymbol{U}$ 一般称为基矩阵，$\boldsymbol{V}$ 是数据的低维表示，或称系数矩阵。NMF 算法更加符合人类对部分构成整体的认知。目前，NMF 算法已经在诸多领域得到了广泛的应用，例如人脸识别[5]、目标识别[6]、文本挖掘[7]、语音信号处理[8]等。

虽然 NMF 算法具有很强的解释性，但是其依旧存在以下问题：1)NMF 算法是在原始数据空间内进行处理，但是图像数据的采集存在噪音等不利因素，这会对聚类结果产生不利的影响；2)NMF 算法未考虑数据间的局部几何结构，NMF 算法是在全体数据空间内进行处理，而 Cai 等[9]指出数据的局部几何结构对聚类结果会产生重大的影响；3)NMF 算法未能利用已知的一些先验信息来获取更好的聚类结果。

针对问题 1)，Cui 等[10]提出将凸非负矩阵分解算法与子空间聚类算法相结合的算法；Xu 等[11]提出概念分解算法(Concept Factorization，CF)，将核技巧应用于 NMF 算法。针对问题 2)，Cai 等[9]提出了图非负矩阵分解算法(Graph Regularized Nonnegative Matrix Factorization，GNMF)，将图正则项引入 NMF 算法中，取得了较好的效果，此后图正则项在非负矩阵分解算法中得到了广泛应用，例如 Hu 等[12]提出带有图正则项的凸非负矩阵分解算法(Graph Regularized And Convex Nonnegative Matrix Factorization，GCNMF)。针对问题 3)，目前已有众多基于 NMF 算法的半监督算法。数据的先验信息通常分为两类，一类是标签信息，另一类是成对信息。对于标签信息，Babaee 等[13]提出了判别非负矩阵分解算法(Discriminative Nonnegative Matrix Factorization，DNMF)，Peng 等[14]提出了基于支持向量机的半监督最小二乘非负矩阵分解算法(Semi-Supervised Least Squares NMF，SLSNMF)及其带有图正则项的拓展形式(Graph-Based Least Squares NMF，GLSNMF)。对于成对信息，Wang 等[15]提出了惩罚矩阵分解算法(Penalized Matrix Factorization，PMF)，Li 等[16]提出了成对约束情形下的带有图正则项的凸非负矩阵分解算法(Pairwise Constrained Graph Regularized Convex Matrix Factorization，PGCNMF)。

本文研究图像在无先验信息情形下的聚类。为了获得更好的聚类结果，我们对非负矩阵分解算法进行改进以解决问题 1)和问题 2)造成的聚类结果下降。本文的主要贡献如下：

1)通过构建超图来更好地保存数据的局部结构，消除简单图不能处理复杂数据联系的弊端。

2)在 Zhou[17]提出预处理的过程后，我们将预处理过程应用于带有超图正则项的非负矩阵分解算法，并分析了均值和标准差的变动对算法聚类结果的影响。

本文第 2 节介绍基于预处理的超图非负矩阵分解算法(PHGNMF)所需的预备知识；第 3 节提出 PHGNMF 算法的模型并求解；第 4 节是 PHGNMF 算法的数值实验部分；最后对全文进行总结。

2 预备知识

2.1 NMF 算法与 GNMF 算法

NMF 算法由 Lee 等[4]提出，该算法采用欧氏距离作为误差衡量标准时所求解的问题模型如下：

$$\begin{aligned}&\min\ \|\boldsymbol{X}-\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\|_F^2\\&\text{s.t. }\boldsymbol{U}\geqslant 0,\boldsymbol{V}\geqslant 0\end{aligned}\tag{1}$$

其中，$\boldsymbol{X}\geqslant 0$，$\boldsymbol{X}\in R^{m\times n}$ 表示非负矩阵，$\boldsymbol{U}\geqslant 0$，$\boldsymbol{U}\in R^{m\times k}$ 是算法分解所得的子矩阵，一般称为基矩阵。$\boldsymbol{V}\geqslant 0$，$\boldsymbol{V}\in R^{n\times k}$ 也是算法分解所得的子矩阵，一般称为数据的低维代表或系数矩阵。k 表示数据所含有的类别数，是一个预先设定的值。

在该模型下算法的迭代公式为：

$$\begin{aligned}\boldsymbol{U}&=\boldsymbol{U}\otimes\frac{\boldsymbol{X}\boldsymbol{V}}{\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}}\\\boldsymbol{V}&=\boldsymbol{V}\otimes\frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}}{\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}}\end{aligned}\tag{2}$$

其中，$\otimes$ 表示矩阵点乘。

GNMF 算法由 Cai 等[9]提出，在欧氏距离的误差衡量标准下，其目标函数为：

$$\begin{aligned}&\min\ \|\boldsymbol{X}-\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\|_F^2+\lambda\mathrm{tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{V})\\&\text{s.t. }\boldsymbol{U}\geqslant 0,\boldsymbol{V}\geqslant 0\end{aligned}\tag{3}$$

其中，λ 表示第一项与第二项的平衡系数，$\boldsymbol{L}$ 表示拉普拉斯矩阵，$\boldsymbol{L}=\boldsymbol{D}-\boldsymbol{W}$。$\boldsymbol{W}$ 表示数据间的相似性矩阵，$\boldsymbol{D}$ 是对角矩阵，其主对角线元素 $\boldsymbol{D}_{ii}$ 是 $\boldsymbol{W}$ 的第 i 行所有元素的和。

GNMF 算法的更新公式为：

$$\begin{aligned}\boldsymbol{U}&=\boldsymbol{U}\otimes\frac{\boldsymbol{X}\boldsymbol{V}}{\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}}\\\boldsymbol{V}&=\boldsymbol{V}\otimes\frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{U}+\lambda\boldsymbol{W}\boldsymbol{V}}{\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}+\lambda\boldsymbol{D}\boldsymbol{V}}\end{aligned}\tag{4}$$

2.2 预处理过程

2.2.1 灰度处理

由于图片在采集的过程中通常处于不同的光照条件，因此如何处理光照条件不同所带来的影响是值得考虑的。Zhou[17]提出添加灰度的预处理过程，即将所有图片数据调整为相同的均值与方差。具体调整的方法如下：一个 $m_1\times n_1$ 维的矩阵 $\boldsymbol{I}$ 代表一个图片数据，矩阵 $\boldsymbol{I}$ 所对应的均值 μ

和方差 σ^2 的计算方法为：

$$\mu=\frac{1}{m_1\times n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1}\boldsymbol{I}_{ij} \tag{5}$$

$$\sigma^2=\frac{1}{m_1\times n_1}\sum_{i=1}^{m_1}\sum_{j=1}^{n_1}(\boldsymbol{I}_{ij}-\mu)^2 \tag{6}$$

若要将所有图片的均值与标准差调整到所指定的均值 μ_0 与标准差 σ_0，则需要对每个图片矩阵 $\boldsymbol{I}$ 的每个元素进行如下操作：

$$\boldsymbol{I}_{ij}=\frac{\sigma_0}{\sigma}(\boldsymbol{I}_{ij}-\mu)+\mu_0 \tag{7}$$

灰度处理后所有图片处于相同的均值与标准差。

2.2.2 小波分析

小波分析是图像压缩的一种手段，Zhou[17] 提出采用小波分析对图片数据处理后再进行分解。确定小波分析的层数是其中的关键。在对原始图像进行一层小波分析后可以得到一阶低频子图，对一阶低频子图再进行一次小波分析可以得到二阶低频子图，示例如图 1 所示。

(a)原始图像 (b)一阶低频子图 (c)二阶低频子图

图 1 JAFFE 数据集中的原始图像及其一阶、二阶低频子图

Fig. 1 Original picture and its low frequence pictures on JAFFE set

由图 1 可见，随着小波分析层数的增加，图片的维度降低，而二阶低频子图较原始图像丢失了大部分的细节，因此 Zhou[17] 采用一阶低频子图作为算法使用的数据。

2.3 超图理论

超图的提出消除了简单图的一些弊端，在许多领域得到了广泛应用。下面对超图理论进行简单介绍[18]。

假设现有超图 $G=(V,e)$，其中 $V=\{v_1,v_2,\cdots,v_k\}$ 是有限个数的数据点的集合，$e=\{e_1,e_2,\cdots,e_t\}$ 是超边集合。因为 G 是定义在 V 上的超图，所以超边集合 e 满足如下条件：

1) $e_j\notin\emptyset, j=1,2,\cdots,t$；

2) $e_1\cup e_2\cup\cdots\cup e_t=V$。

集合 V 中的元素称为顶点，集合 e 中的元素称为超边。每一条超边 e_j 都有相应的权重 w_j。顶点与超边会形成一个关联矩阵 $\boldsymbol{H}$，$\boldsymbol{H}\in R^{|V|\times|e|}$。$\boldsymbol{H}$ 的元素的计算方式如下：

$$\boldsymbol{H}_{ij}=\begin{cases}1, & v_i\in e_j\\ 0, & \text{其他}\end{cases} \tag{8}$$

为了更进一步地说明超图理论，我们给出了如下例子。假设有超图 $G=\{V,e\}$，其中顶点集合 $V=\{v_1,v_2,\cdots,v_8\}$，超边集合 $e=\{e_1,e_2,e_3\}$。图 2 给出了超图 G 的示例。

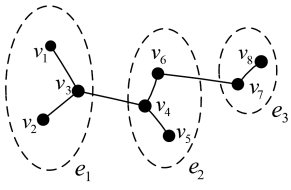


图 2 超图 G 的示例

Fig. 2 Example of hypergraph G

图 2 中超边通过虚线标识，可见超图 G 含有 3 条超边，其中 $e_1=\{v_1,v_2,v_3\}$，$e_2=\{v_4,v_5,v_6\}$，$e_3=\{v_7,v_8\}$。G 的邻接矩阵 $\boldsymbol{H}$ 如表 1 所列。

表 1 超图 G 的关联矩阵 $\boldsymbol{H}$

Table 1 Incidence matrix $\boldsymbol{H}$ of hypergraph G

V	e_1	e_2	e_3
v_1	1	0	0
v_2	1	0	0
v_3	1	0	0
v_4	0	1	0
v_5	0	1	0
v_6	0	1	0
v_7	0	0	1
v_8	0	0	1

超图中每个顶点的度 d_i 定义为其所属超边的权重之和，而超边的度 δ_j 定义为超边所属的节点个数。其计算方式如下：

$$d_i=\sum_{j=1}^{t}w_j\boldsymbol{H}_{ij} \tag{9}$$

$$\delta_j=\sum_{i=1}^{k}\boldsymbol{H}_{ij} \tag{10}$$

记 $\boldsymbol{D}_v$ 表示一个对角矩阵，其主对角元素 $\boldsymbol{D}_{v_{ii}}=d_i$，其中 $i=1,2,\cdots,k$。相似地，记 $\boldsymbol{D}_e$ 与 $\boldsymbol{W}$ 分别为由 δ_j 和 w_j 生成的对角矩阵，其中 $j=1,2,\cdots,t$。那么非正则化的超图拉普拉斯矩阵的计算方法如下：

$$\boldsymbol{L}^H=\boldsymbol{D}_v-\boldsymbol{H}\boldsymbol{W}\boldsymbol{D}_e^{-1}\boldsymbol{H} \tag{11}$$

$\boldsymbol{L}^H$ 的计算方法类似于简单图的拉普拉斯矩阵的计算方法[19]，显然 $\boldsymbol{L}^H$ 是一个对称矩阵。

3 基于预处理的超图非负矩阵分解算法

本节介绍了关于基于预处理的超图非负矩阵分解算法(PHGNMF)。表 2 列出了一些符号所代表的含义。

表 2 一些符号及其所代表的含义

Table 2 Some notions and means

符号	维数	含义
$\boldsymbol{X}$	$M_{\boldsymbol{X}}\times N$	非负矩阵，代表算法的输入矩阵，每个图片为一列
$\boldsymbol{X}^G$	$M_{\boldsymbol{X}}\times N$	非负矩阵，代表 $\boldsymbol{X}$ 经过灰度处理后的矩阵
$\boldsymbol{X}^W$	$M\times N$	非负矩阵，代表 $\boldsymbol{X}^G$ 经过小波分析后提取的低频图像所成矩阵
K	1	图片所含类别数
$\boldsymbol{U}$	$M\times K$	非负矩阵，由算法输出
$\boldsymbol{V}$	$N\times K$	非负矩阵，由算法输出
μ_0	1	灰度处理时指定的均值
σ_0	1	灰度处理时制定的标准差
t	1	$\boldsymbol{X}^w$ 形成的超图所含的超边数目
s	1	每条超边所含的顶点数目
$\boldsymbol{H}$	$N\times t$	超图的关联矩阵
$\boldsymbol{W}$	$t\times t$	t 个超边权重所形成的对角矩阵
$\boldsymbol{L}^H$	$N\times N$	超图的非正则超图拉普拉斯矩阵
$\boldsymbol{D}_v$	$N\times N$	超图中每个顶点的度所形成的对角矩阵
$\boldsymbol{D}_e$	$t\times t$	每条超边所含顶点数目所形成的对角矩阵

3.1 PHGNMF 算法的描述

PHGNMF 算法的输入为 $\boldsymbol{X}$，其中的每一列代表一个样本。首先根据给定的 μ_0，σ_0，利用式(7)对 $\boldsymbol{X}$ 的每一列进行灰度处理，而此时式(7)中的 μ 与 σ 是指该列数据的均值与标准差。为了避免灰度处理后的数据出现负值，采用投影的方法

将所有的负值改为 0 得到$\boldsymbol{X}^G$。所采用的投影方法是将处理后的数据中的负值改为 0，而非负的值保持不变。

在灰度处理以后进行小波分析以提取图片的低频子图。将$\boldsymbol{X}^G$的每一列还原至原图像大小进行小波分析以提取低频子图。为了避免出现负值，采用与灰度处理相同的投影操作将负值改为 0 得到$\boldsymbol{X}^W$。

之后构建$\boldsymbol{X}^W$的非正则化的超图拉普拉斯矩阵$\boldsymbol{L}^H$。我们定义$\boldsymbol{X}^W$所形成的超图中超边的数目 t 与数据的数目 N 相等，每条超边所含的顶点数目为 s。每条超边所含顶点是$\boldsymbol{X}_n^W$自身及其最近的 $s-1$ 个邻居所产生的，其中$\boldsymbol{X}_n^W$指$\boldsymbol{X}^W$的第 n 列。根据式(11)计算得到$\boldsymbol{L}^H$，其中w_j的计算方法采用 Wang 等[18]的方法，计算公式如下：

$$w_j=\sum_{\boldsymbol{x}_{n_1}^W,\boldsymbol{x}_{n_2}^W\in e_j}\exp\left(-\frac{\|\boldsymbol{X}_{n_1}^W-\boldsymbol{X}_{n_2}^W\|}{\delta^2}\right) \tag{12}$$

其中，$\delta=\frac{1}{s\times t}\sum_{j=1}^{t}\sum_{\boldsymbol{x}_{n_1}^W,\boldsymbol{x}_{n_2}^W\in e_j}\|\boldsymbol{X}_{n_1}^W-\boldsymbol{X}_{n_2}^W\|$。

在得到X^W和L^H后，PHGNMF 算法所求解的问题如下：

$$\begin{aligned}&\min\ \|\boldsymbol{X}^W-\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\|_F^2+\lambda\mathrm{tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}^H\boldsymbol{V})\\&\text{s.t. }\boldsymbol{U}\geqslant 0,\boldsymbol{V}\geqslant 0\end{aligned} \tag{13}$$

式(13)中，第一项代表对$\boldsymbol{X}^W$进行非负分解所产生的误差；第二项是在$\boldsymbol{X}^W$构成超图的情形下所形成的超图正则项，是为了保存数据的局部几何结构来提升算法最后的聚类结果。

3.2 问题(13)的求解

由于直接求解式(13)是一个 NP 难问题，因此采用迭代求解的方法。式(13)的拉格朗日函数为：

$$\boldsymbol{L}=\|\boldsymbol{X}^W-\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\|_F^2+\lambda\mathrm{tr}(\boldsymbol{V}^T\boldsymbol{L}^H\boldsymbol{V})+\mathrm{tr}(\boldsymbol{\Psi}\boldsymbol{U}^{\mathrm{T}})+\mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{V}^{\mathrm{T}}) \tag{14}$$

其中，$\boldsymbol{\Psi}$是$\boldsymbol{\Psi}_{mk}$对于$\boldsymbol{U}_{mk}\geqslant 0$的拉格朗日乘子所形成的矩阵，$\boldsymbol{\Phi}$是$\boldsymbol{\Phi}_{nk}$对于$\boldsymbol{V}_{nk}\geqslant 0$的拉格朗日乘子所形成的矩阵。将 L 改写为矩阵的形式为：

$$\begin{aligned}\boldsymbol{L}&=\mathrm{tr}(\boldsymbol{X}^{W^T}\boldsymbol{X}^W)-\mathrm{tr}(\boldsymbol{X}^{W^T}\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}})-\mathrm{tr}(\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^W)+\\&\quad\mathrm{tr}(\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}})+\lambda\mathrm{tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}^H\boldsymbol{V})+\mathrm{tr}(\boldsymbol{\Psi}\boldsymbol{U}^{\mathrm{T}})+\mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{V}^{\mathrm{T}})\\&=\mathrm{tr}(\boldsymbol{X}^{W^T}\boldsymbol{X}^W)-2\mathrm{tr}(\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{X}^W)+\mathrm{tr}(\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}})+\\&\quad\lambda\mathrm{tr}(\boldsymbol{V}^{\mathrm{T}}\boldsymbol{L}^H\boldsymbol{V})+\mathrm{tr}(\boldsymbol{\Psi}\boldsymbol{U}^{\mathrm{T}})+\mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{V}^{\mathrm{T}})\end{aligned} \tag{15}$$

其中，$\|\boldsymbol{A}\|_F^2=\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$，$\mathrm{tr}(\boldsymbol{A})=\mathrm{tr}(\boldsymbol{A}^{\mathrm{T}})$。

由拉格朗日函数 L 对 $\boldsymbol{U}$，$\boldsymbol{V}$ 求偏导，可得：

$$\frac{\partial L}{\partial\boldsymbol{U}}=-2\boldsymbol{X}^W\boldsymbol{V}+2\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V}+\boldsymbol{\Psi} \tag{16}$$

$$\frac{\partial L}{\partial\boldsymbol{V}}=-2\boldsymbol{X}^{W^{\mathrm{T}}}\boldsymbol{U}+2\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}+2\lambda\boldsymbol{L}^H\boldsymbol{V}+\boldsymbol{\Phi} \tag{17}$$

根据 KKT 条件$\boldsymbol{\Psi}_{mk}\boldsymbol{U}_{mk}=0$，$\boldsymbol{\Phi}_{nk}\boldsymbol{V}_{nk}=0$，可以得到：

$$-(\boldsymbol{X}^W\boldsymbol{V})_{mk}\boldsymbol{U}_{mk}+(\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})_{mk}\boldsymbol{U}_{mk}=0 \tag{18}$$

$$-(\boldsymbol{X}^{W^T}U)_{nk}\boldsymbol{V}_{nk}+(\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}U)_{nk}\boldsymbol{V}_{nk}+\lambda(\boldsymbol{L}^H\boldsymbol{V})_{nk}\boldsymbol{V}_{nk}=0 \tag{19}$$

由此，可以得到如下的更新公式：

$$\boldsymbol{U}_{mk}\leftarrow\boldsymbol{U}_{mk}\otimes\frac{(\boldsymbol{X}^W\boldsymbol{V})_{mk}}{(\boldsymbol{U}\boldsymbol{V}^{\mathrm{T}}\boldsymbol{V})_{mk}} \tag{20}$$

$$\boldsymbol{V}_{nk}\leftarrow\boldsymbol{V}_{nk}\otimes\frac{(\boldsymbol{X}^{W^T}\boldsymbol{U})_{nk}+(\lambda\boldsymbol{H}\boldsymbol{W}\boldsymbol{D}_e^{-1}\boldsymbol{H}\boldsymbol{V})_{nk}}{(\boldsymbol{V}\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U})_{nk}+(\lambda\boldsymbol{D}_v\boldsymbol{V})_{nk}} \tag{21}$$

由于非正则化的超图拉普拉斯矩阵L^H与图拉普拉斯矩阵具有相似性，因此问题(13)在式(20)和式(21)的更新公式下是下降的，其证明请参考文献[9]。

PHGNMF 算法的迭代流程如算法 1 所示。

算法 1 PHGNMF 算法

输入：$\mathbf{X}$，K，μ_0，σ_0，λ，s

输出：$\mathbf{U}$，$\mathbf{V}$

1. 在 $\mathbf{X}$ 上利用式(7)进行灰度处理并投影得到$\mathbf{X}^G$；
2. 在$\mathbf{X}^G$上利用小波分析处理并投影得到$\mathbf{X}^W$；
3. 根据$\mathbf{X}^W$利用式(11)计算$\mathbf{L}^H$；
4. 设定最大循环次数 trynumber；
5. 对 $\mathbf{U}$，$\mathbf{V}$ 赋初值；
6. for k←1 to trynumber do
   通过式(20)更新 $\mathbf{U}$；
   通过式(21)更新 $\mathbf{V}$；
   k=k+1；
end

4 数值实验

本节通过数值实验说明 PHGNMF 算法的有效性。

数值实验在 5 个数据集上进行，分别是：Yale 数据集、YaleB 数据集、JAFFE 数据集、ORL 数据集和 USPS 数据集。

Yale 数据集，包含 165 幅人脸图像，共 15 类，每幅图片的大小为 32×32。

YaleB 数据集，从中选出 38 个类别，每个类包含 59 幅图片，图片大小为 32×32。

JAFFE 数据集，采集自 10 位日本女性，每个类别选取 20 张图片，图片大小为 256×256。

ORL 数据集，共 400 幅图片，有 10 个类别，图片大小为 112×92。

USPS 数据集，选取 10 个数字的手写体，每个类别 200 张图片，图片大小为 16×16。

实验中，对照算法选取 Kmeans 聚类算法、谱聚类算法[20]、主成分分析算法(PCA)[1]和图非负矩阵分解算法(GNMF)[9]。

4.1 评价准则

在衡量聚类结果时，我们采用聚类精度(ACC)[21]和标准互信息(NMI)[21]作为标准。

聚类精度(ACC)[21]是每个样本聚类所得标签与真实标签的比值，其计算方法为：

$$ACC=\frac{\sum_{i=1}^{n}\delta(map(r_i),l_i)}{n} \tag{22}$$

其中，r_i代表第 i 个数据聚类后所得到的样本标签，映射$map(r_i)$代表将聚类标签r_i映射到真实标签集上的结果，l_i代表第 i 个数据的真实标签。$\delta(a,b)$在 $a=b$ 时为 1，其余情况为 0。n 代表样本个数。

归一化互信息(NMI)[21]也是衡量聚类结果的标准，其计算方法为：

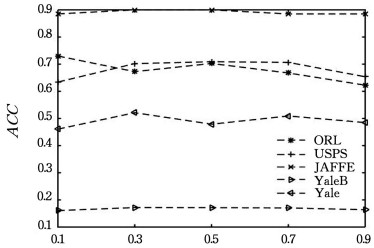
$$NMI=\frac{\sum_{i=1}^{c}\sum_{j=1}^{c}n_{i,j}\log\frac{n_{i,j}}{n_i\hat{n}_j}}{\sqrt{(\sum_{i=1}^{c}n_i\log\frac{n_i}{n})(\sum_{j=1}^{c}n_j\log\frac{\hat{n}_j}{n})}} \tag{23}$$

其中，c 表示类别个数，n_i 表示属于第 C_i 个聚类类别的样本数，$n_j^{\wedge}$ 表示属于真实标签 L_j 的数据个数，$n_{i,j}$ 表示同时属于 C_i 与 L_j 的样本数目。

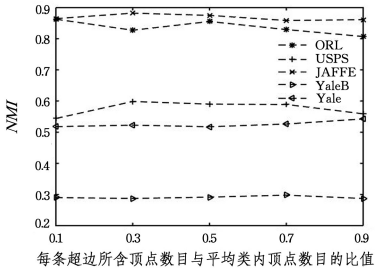
4.2 PHGNMF 算法的参数选择

由于我们设定数据的超边个数 t 与样本数目 N 相等，因此可以改变超边所含顶点个数 s、预先给定的均值 μ_0 和标准差 σ_0，以及平衡系数 λ 这 4 个变量。

我们固定 μ_0，σ_0 不变，其值分别为数据 $\boldsymbol{X}$ 的均值与标准差。根据本文所提的 PHFGNMF 模型与 GNMF 算法[9]的相近性，将 λ 设定为 10。改变 s，采用 s 与 $\frac{N}{K}$ 的比值发生变动的方法进行实验，其中称 $\frac{N}{K}$ 为平均类内顶点数目。实验在各数据集的结果如图 3 所示。



(a)*ACC* 变化



(b)*NMI* 变化

图 3 超边所含顶点数目变动时聚类的变化

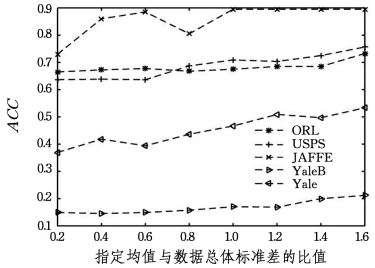
Fig. 3 Result changes when s changes

由图 3 可以看到，超边内顶点数目的不同对于最终聚类结果会产生一定的影响，但并不是超边内所含顶点数目越多，聚类效果就越优，因为超边内顶点数目太多就有可能包含欧氏距离很近但不属于同一类的样本，而保存这些不利的局部结构就会降低算法的聚类效果。为了更好地平衡算法在各个数据集上的表现，我们设定超边内顶点数目为 $\frac{N}{K}$ 的 50%，若此时计算的数目不为整数，则向下取整。

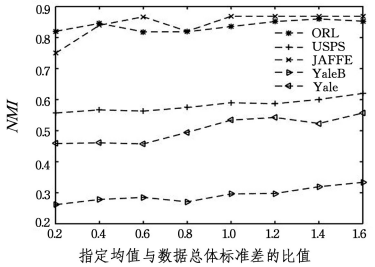
在确定 $s=\frac{N}{2K}$ 后，我们对预处理中需要预先指定的均值 μ_0 进行调整。设定 μ_0 的值在总体数据均值的 0.2 倍到 1.6 倍之间浮动。在各数据集上 ACC 与 NMI 的变动如图 4 所示。

从图 4 可以明显看出，随着比值的增大，在各个数据集上的实验结果都呈上升态势。这说明将 μ_0 设定为比原始数据总体均值适度大一些的值有提升聚类效果的作用。在后续实验中，设定 μ_0 是原始数据均值的 1.6 倍。

在确定 $s=\frac{N}{2K}$，μ_0 是数据总体均值的 1.6 倍后，我们对 σ_0 进行调整，并指定 σ_0 的值在数据总体标准差的 0.2 倍到 1.6 倍之间浮动。在各数据集上 ACC 与 NMI 的变动如图 5 所示。



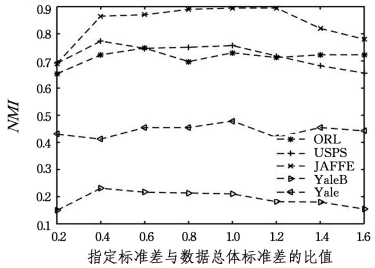
(a)*ACC* 变化



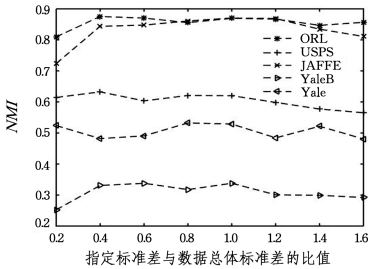
(b)*NMI* 变化

图 4 指定均值 μ_0 变动时聚类的变化

Fig. 4 Result changes when μ_0 changes



(a)*ACC* 变化



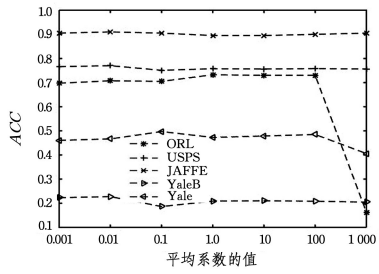
(b)*NMI* 变化

图 5 指定标准差 σ_0 变动时聚类的变化

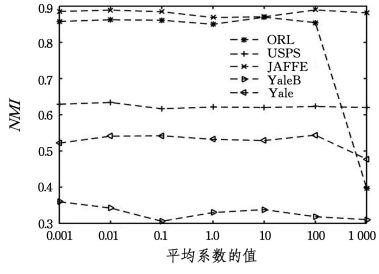
Fig. 5 Result changes when σ_0 changes

由图 5 可见，指定标准差 σ_0 与原始数据标准差之间比值的不断增大并不能使聚类结果稳定地上升，在某些数据集上还会起到负面作用，例如 Yale 数据集与 JAFFE 数据集。因此我们固定 σ_0 等于数据总体标准差。

最后我们调整式(13)中的平衡系数 λ，λ 在[0.001，0.01，0.1，1，10，100，1000]范围内浮动，其在各数据集上的聚类表现如图 6 所示。



(a)*ACC* 变化



(b)*NMI* 变化

图 6 平衡系数 λ 变动时聚类的变化

Fig. 6 Result changes when λ changes

由图 6 可知，在 λ 不断变化时聚类表现总体平稳，在 Yale 数据集上呈现波动的态势，在其余 4 个数据集上均较稳定。但是 λ 过大时，PHGNMF 算法在 ORL 数据集、Yale 数据集上的表现下滑较为明显。总体来看，PHGNMF 算法对于 λ 的取值并不十分敏感，在接下来的实验中我们设定 $\lambda=10$。

4.3 对比实验

设定 PHGNMF 算法的各参数为 $s=\frac{N}{2K}$，μ_0 为数据总体均值的 1.6 倍，σ_0 为原始数据的标准差，在 $\lambda=10$ 的情况下在 5 个数据集上进行实验。谱聚类算法在 Yale 数据集、YaleB 数据集、ORL 数据集、USPS 数据集上的 $\sigma=1$，在 JAFFE 数据集上的 $\sigma=4$。GNMF 算法的参数 λ 根据文献[9]设定为 100。Kmeans 聚类算法可以直接得到标签结果，而谱聚类算法、PCA 算法、GNMF 算法以及 PHGNMF 算法均得到数据的低维表示，我们在这些数据低维表示上运行 Kmeans 算法得到最终的聚类结果。为避免随机性的影响，每个实验重复 20 次取平均。各算法聚类结果的 ACC 均值及相应标准差如表 3 所列，NMI 均值及相应标准差如表 4 所列。表 3 和表 4 中加粗的数据为最优结果。

表 3 20 次实验的 *ACC*

Table 3 *ACC* of 20 experiments

数据集	Kmeans	谱聚类	PCA	GNMF	PHGNMF
Yale	0.4070(0.0225)	0.2661(0.0154)	0.4027(0.0398)	0.3127(0.0170)	**0.4758(0.0317)**
YaleB	0.1048(0.0051)	0.1634(0.0052)	0.1016(0.0059)	0.0874(0.0021)	**0.1973(0.0081)**
JAFEE	0.8332(0.0494)	0.3297(0.0198)	—	0.8297(0.0579)	**0.8698(0.0357)**
ORL	0.6440(0.0358)	0.2931(0.0094)	0.6756(0.0422)	0.6313(0.0231)	**0.7156(0.0282)**
USPS	0.6522(0.0130)	0.2414(0.0038)	0.6519(0.0232)	0.5933(0.0155)	**0.7187(0.0376)**

表 4 20 次实验的 *NMI*

Table 4 *NMI* of 20 experiments

数据集	Kmeans	谱聚类	PCA	GNMF	PHGNMF
Yale	0.4593(0.0216)	0.3299(0.0131)	0.4777(0.0345)	0.3567(0.0151)	**0.5203(0.0262)**
YaleB	0.1435(0.0049)	0.2756(0.0047)	0.1428(0.0055)	0.1572(0.0054)	**0.3144(0.0076)**
JAFEE	0.8616(0.0321)	0.3562(0.0244)	—	**0.9001(0.0198)**	0.8522(0.0294)
ORL	0.8198(0.0133)	0.5640(0.0072)	0.8490(0.0170)	0.7907(0.0116)	**0.8560(0.0125)**
USPS	0.5915(0.0063)	0.1741(0.0062)	0.5813(0.0065)	**0.6267(0.0204)**	0.6088(0.0174)

由于 JAFFE 数据集原图片的尺寸较大，不适用于 PCA 算法的实验，因此未在 PCA 算法上进行实验。观察表 3 和表 4 的数据可知，PHGNMF 方法在 ACC 方面占据优势，但是在 NMI 方面在两个数据集上的表现并不是最优的。但总体来看，PHGNMF 方法较对比算法实现了聚类质量的提升。

结束语 针对图片聚类问题，本文提出了基于预处理的超图非负矩阵分解算法(PHGNMF)。预处理的过程主要是剔除光照、噪音等不利因素的影响，而采用超图的思想是为了进一步保存数据的局部结构，消除简单图的一些弊端。在 5 个数据集上的实验结果说明了算法的有效性。由于并没有理论上的保证，我们并不能确保文中所提的参数是最优值，因此如何选择参数仍值得进一步研究。

参 考 文 献

[1] BELHUMEUR P N, HESPANHA J P, KRIEGMAN D J. Eigenfaces vs, Recongnition using class specific linear projection [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1997, 7: 711-720.

[2] KALMAN D. A singularly valuable decomposition: the SVD of a matrix [J]. The College Mathematics Journal, 1996, 27(1): 2-23.

[3] FISHER R A. The use of multiple measurements in taxonomic problems [J]. Annals of Eugenics, 1936, 7(2): 179-188.

[4] LEE D D, SEUNG H S. Learning the parts of objects by non-negative matrix factorization [J]. Nature, 1999, 401(6755): 788.

[5] GUILLAMET D, JORDI V. Non-negative matrix factorization for face recognition[C]// Castellón, Spain, lecture notes in artifi-

cial intelligence. Berlin:Springer-Verlag,2002:336-344.

[6] ZHANG X,GAO H,LI G,et al. Multi-view clustering based on graph-regularized nonnegative matrix factorization for object recognition [J]. Information Sciences,2018,432:463-478.

[7] ZURADA J M,ENSARI T,ASL E H,et al. Nonnegative matrix factorization and its application to pattern analysis and text mining[C]// Krakow,Poland,Federated Conference on Computer Science and Information Systems. New York,IEEE,2013:11-16.

[8] GAO B,WOO W L,DLAY S S. Variational regularized 2-D nonnegative matrix factorization [J]. IEEE Transactions on Neural Networks and Learning Systems,2012,23(5):703-716.

[9] CAI D,HE X,HAN J,et al. Graph regularized nonnegative matrix factorization for data representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(8): 1548-1560.

[10] CUI G,LI X,DONG Y. Subspace clustering guided convex nonnegative matrix factorization [J]. Neurocomputing, 2018, 292: 38-48.

[11] XU W,GONG Y. Document clustering by concept factorization [C]// Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,Association for Computing Machinery. New York,2004: 202-209.

[12] HU W,CHOI K S,WANG P,et al. Convex nonnegative matrix factorization with manifold regularization [J]. Neural Networks,2015,63:94-103.

[13] BABAEE M,TSOUKALAS S. Discriminative nonnegative matrix factorization for dimensionality reduction [J]. Neurocomputing,2015,173:212-223.

[14] PENG X,CHEN D,XU D. Semi-supervised least squares nonnegative matrix factorization and graph-based extension [J]. Neurocomputing,2018,320:98-111.

[15] WANG F,LI T,ZHANG C S. Semi-supervised clustering via matrix factorization [C]// Atlanta,Georgia,Proceedings of the 2008 SIAM International Conference on Data Mining. SIAM, America,2008:1-12.

[16] LI G,ZHANG X,ZHANG S,et al. Semi-supervised convex nonnegative matrix factorizations with graph regularized for image representation [J]. Neurocomputing,2017,237:1-11.

[17] ZHOU J. Research of SWNMF with new iteration rules for facial feature extraction and recognition [J]. Symmetry, 2019, 11(3):354.

[18] WANG W H,QIAN Y T,TANG Y Y. Hypergraph-regularized sparse NMF for hyperspectral unmixing [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2016,9(2):681-694.

[19] ZHOU D Y,HUANG J Y,BERNHARD S. Learning with hypergraphs: clustering, classification, and embedding [C] // British,Advances in Neural Information Processing Systems 19: Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems,MIT Press,2007:1601-1608.

[20] SHI J B,MALIK J. Normalized cuts and image segmentation [J]. IEEE Transactions on pattern analysis and machine intelligence,2000,22(8):888-905.

[21] HE X,WANG Q,LI X L. Robust adaptive graph regularized non-negative matrix factorization [J]. IEEE Access, 2019, 7: 83101-83110.

LI Xiang-li, born in 1977, Ph.D, professor. Her main research interests include image clustering, nonnegative matrix factorization, and optimization.