计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200800190

基于高斯场和自适应图正则的半监督聚类

赵 敏 刘惊雷
烟台大学计算机与控制工程学院 山东 烟台 264005
(ytdxzhaomin@163.com)

摘 要 聚类是将给定的样本分成几个不同的簇,它在机器学习、数据挖掘等领域得到了广泛应用,并受到研究人员的广泛关注。但是,传统的聚类方法仍然存在3个方面的不足。首先,由于一些数据中存在噪声和异常值,传统的聚类方法容易产生误差较大的目标函数。其次,传统的聚类方法没有使用监督信息来指导构建相似矩阵。最后,加入图正则的聚类方法在计算相似度矩阵时,邻居关系都是确定的,一旦计算错误就会导致构造图的质量低,进而影响聚类性能。因此,提出了一种基于高斯场和自适应图正则化的半监督聚类(SCGFAG)模型。该模型通过高斯场及谐波函数法引入监督信息,来指导构建相似度矩阵,实现半监督学习,还引入稀疏误差矩阵来表示稀疏噪声,如脉冲噪声、死线和条纹,并且使用 l_1 范数来缓解稀疏噪声。此外,所提模型还引入 $l_{2,1}$ 范数来处理异常值的影响。因此,SCGFAG对数据噪声和异常值不敏感。更重要的是,SCGFAG通过引入自适应图的正则化提高了聚类性能。为了实现优化聚类的目标,提出了一种迭代更新算法—增广拉格朗日法(Augmented Lagrangian Method,ALM),分别对优化变量进行更新。在4个数据集上进行的实验表明,所提方法优于相比较的8种经典聚类方法获得了更好的聚类性能。

关键词:自适应图正则;半监督聚类;$l_{2,1}$ 的旋转不变性;噪声和异常值;增广拉格朗日法
中图法分类号 TP311

Semi-supervised Clustering Based on Gaussian Fields and Adaptive Graph Regularization

ZHAO Min and LIU Jing-lei
School of Computer and Control Engineering,Yantai University,Yantai,Shandong 264005,China

Abstract Clustering is to divide a given sample into several different clusters,which is a widely used tool,has been applied in machine learning,data mining and so on,and has received extensive concern by researchers. However,there are still three main limitations. Firstly,usually there are noises and outliers in the data,which will bring about significant errors in the clustering results. Secondly,traditional clustering methods do not use supervision information to guide the construction of similarity matrices. Finally,in the graph-based clustering method,when constructing graphs,the neighbor relationship is determined. Once the calculation is wrong, it will result in poor quality of the constructed graph,which will affect the clustering performance. Therefore,a semi-supervised clustering model based on Gaussian field and adaptive graph regularization (SCGFAG) is proposed in this paper. In this model,supervised information is introduced by gaussian field and harmonic function to guide the construction of similarity matrix to realize semi-supervised learning. Sparse error matrix is introduced to represent sparse noise,such as impulse noise,dead line,stripes,and l_1 norm is introduced to alleviate the sparse noise. In addition,the $l_{2,1}$ norm is also introduced by the proposed model to mitigate the effects of outliers. Therefore,our SCGFAG is insensitive to data noise and outliers. More importantly,the regularization of adaptive graph is introduced into SCGFAG to improve the clustering performance. In order to realize the goal of optimization clustering,an iterative updating algorithm-Augmented Lagrangian Method (ALM) is proposed to update the optimization variables respectively. Experimental results on four datasets show that the proposed method is superior to the eight classical clustering methods,and has better clustering performance.

Keywords Adaptive graph regularization,Semi-supervised clustering,Rotation invariance property of $l_{2,1}$,Noise and outliers,Augmented Lagrangian method

到稿日期:2020-08-28 返修日期:2020-10-19 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(61572419,61773331,61801414,62072391)
This work was supported by the National Natural Science Foundation of China(61572419,61773331,61801414,62072391).
通信作者:刘惊雷(jinglei_liu@sina.com)

1 引言

聚类是将对象集划分为由相似对象组成的多个类的过程。聚类生成的簇是一组数据对象，它们与同一簇中的对象相关，但与其他簇中的对象不同。在机器学习和数据挖掘中，数据聚类是一种很有价值的数据分析工具。聚类被应用于多个场景，如多媒体分割[1-2]、人脸识别[3]等。

近年来，多种聚类方法被提出，如 K-means[4]、谱聚类[5-7]、基于非负矩阵分解(Non-negative Matrix Factorization，NMF)的聚类[8-9]、半监督聚类[10-11]。K-means 表示旨在学习最小化簇内数据距离的 c 簇中心。谱聚类是一种基于图论的聚类方法，它通过聚类样本数据的拉普拉斯矩阵的特征向量来达到聚类样本数据的目的，在样本之间进行了低维的关联矩阵嵌入。基于非负矩阵分解的聚类就是将 NMF 作为一种聚类方法，NMF 的一般形式是 $\boldsymbol{X}\approx\boldsymbol{W}\boldsymbol{H}^{\mathrm{T}}$ s. t. $\boldsymbol{W}\geqslant 0$，$\boldsymbol{H}\geqslant 0$[12]，它使分解后的所有分量非负，同时实现了非线性降维。为了提高 NMF 的性能，各种正则化的 NMF 相继被提出。GNMF(Graph Regularized Non-negative Matrix Factorization)[13]利用数据簇分配矩阵的图正则化约束得到数据的几何结构。文献[14]提出了一种流形非负矩阵分解方法 RMNMF(Robust Manifold Non-negative Matrix Factorization)，使用拉普拉斯正则约束。文献[15]提出了一种超图正则化非负矩阵分解算法 HNMF(Image Clustering by Hyper-Graph Regularized Non-negative Matrix Factorization)，它通过构造超图而不是简单图来捕获数据空间的固有几何结构，这是结合图拉普拉斯基于图的 NMF 的几种典型方法。

半监督聚类就是通过先验知识对样本间的相似关系进行约束。传统的聚类算法没有有效地利用标记信息。但是，现在许多聚类问题中除了未标记的数据之外，通常还包含一些有标记的信息。根据问题的不同，有标记信息出现的形式也不同。例如，可以知道聚类的数量，或一些必连约束(must-link)，即样本必属于同一簇，和勿连约束(cannot-link)，即样本不属于同一个簇。半监督学习在许多不同领域引起了广泛关注，因为无标记数据比有标记数据更容易获得，使用半监督学习方法可减少所需的精力、专业知识和时间。因此，有必要使用这些已标记的信息。比如，用于半监督学习的非负低秩稀疏图(Non-negative Low-Rank and Sparse，NNLRS)通过寻找一个 NNLRS 矩阵来学习图中边的权值，该 NNLRS 矩阵将每个数据点表示为其他数据点的线性组合[16]。

本文需要解决以下 3 个问题。

(1)图构造时需要计算所有数据样本对之间的距离，如果计算得到的数据样本之间的距离不够精确，则会影响生成图的质量。

(2)以往的方法没有使用标签信息来指导构建相似矩阵，并且由于标记成本很高，标记后的训练样本往往不足。因此，有必要利用有限的标签信息来指导相似度矩阵的构造。而传统聚类没有使用标签信息来指导相似度矩阵的构建。

(3)传统的 NMF 对噪声和异常值非常敏感，因为它使用平方误差函数来测量损失。这就导致只有少数误差较大的异常值容易占据目标函数的主导地位。

本文的特点和贡献如下。

(1)相较于传统的聚类方法，本文提出了一种基于高斯场和自适应图正则的半监督聚类的方法。本文将非负矩阵分解、自适应正则化、高斯场和谐波函数(Gaussian Fields and Harmonic Functions，GFHF)集成到一个统一的框架中，如式(5)所示，通过引入标签信息来指导相似矩阵的构建。

(2)为了解决数据被噪声和异常值破坏的问题，本文分别利用 l_1 范数和 $l_{2,1}$ 范数来缓解噪声和异常值的影响。同时，本文还利用自适应图正则化，自适应地从数据矩阵中学习拉普拉斯矩阵，以提高聚类性能。

(3)为了解决优化问题，提出了一种基于增广拉格朗日法的有效算法，并证明了该算法的时间复杂度和收敛性。在基准数据集上与几种经典的聚类方法进行实验比较，SCGFAG 的聚类性能有显著提高，证明了该方法的优越性。

本文第 2 节介绍相关工作、半监督学习方法和相关的自适应图正则化；第 3 节介绍一种基于高斯场和自适应图正则的半监督聚类算法框架(SCGFAG)；第 4 节对所提 SCGFAG 进行理论分析；第 5 节通过实验验证 SCGFAG 的聚类效果，并给出实验结果分析；最后总结全文并对未来相关工作进行展望。

2 相关定义及方法

本节首先介绍本文中使用的一些基本符号，随后对本文提出的算法所涉及的相关定义和概念进行简单的介绍。

2.1 相关符号

具体的符号描述如表 1 所列。矩阵用大写粗体字母(例如 $\boldsymbol{X}$)表示，向量用粗体小写字母表示。$\boldsymbol{X}_{:j}$ 为第 j 列，$\boldsymbol{X}_{i:}$ 为第 i 行，$\boldsymbol{X}_{ij}$ 为 $\boldsymbol{X}$ 第 i 行第 j 列上的元素。

表 1 基本符号

Table 1 Basic symbols

符号	含义
$\boldsymbol{X}$	数据矩阵
$\boldsymbol{Z}$	相似矩阵
$\boldsymbol{L}$	拉普拉斯矩阵
$\boldsymbol{L}_{\boldsymbol{Z}}$	学习 $\boldsymbol{Z}$ 计算的拉普拉斯矩阵
$\boldsymbol{D}$	度矩阵
$\mathrm{tr}(\cdot)$	矩阵的迹
$\boldsymbol{I}$	单位矩阵
$\|\boldsymbol{X}\|_{2,1}$	矩阵 $\boldsymbol{X}$ 的 $l_{2,1}$ 范数
$\|\boldsymbol{X}\|_{\mathrm{F}}$	矩阵 $\boldsymbol{X}$ 的 Frobenius 范数
$\|\boldsymbol{X}\|_1$	矩阵 $\boldsymbol{X}$ 的 l_1 范数

2.2 高斯场和谐波函数

高斯场和谐波函数是一种半监督的学习方法[17-18]。它使用预测标签 $\boldsymbol{F}\in\boldsymbol{R}^{n\times c}$ 在图上对标签适应度和流形平滑度进行估计。GFHF 的目标函数为：

$$\min_{\boldsymbol{F}}\frac{1}{2}\|\boldsymbol{F}_i-\boldsymbol{F}_j\|^2\boldsymbol{Z}_{ij}+\lambda_\infty\sum_{i=1}^{u}\|\boldsymbol{F}_i-\boldsymbol{Y}_i\|^2 \tag{1}$$

其中，λ_∞ 是一个非常大的数字，$\boldsymbol{F}$ 是所有样本的预测标签，$\boldsymbol{Z}$ 为相似度矩阵。式(1)可以改写为：

$$\min_{\boldsymbol{F}}\frac{1}{2}\mathrm{tr}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{F})+\mathrm{tr}((\boldsymbol{F}-\boldsymbol{Y})^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{F}-\boldsymbol{Y})) \tag{2}$$

其中，$\boldsymbol{L}$ 是图拉普拉斯矩阵，$\boldsymbol{U}$ 是一个对角矩阵，它的前 u 行和后 $n-u$ 行对应的元素分别是 λ_∞ 和 0。

GFHF 采用了一种简洁的方法将标签信息纳入半监督学习，标记和未标记的数据被表示为加权图中的顶点，图中边权

值是编码实例之间的相似度。学习问题在此图上以高斯随机场表示，其中场的均值以谐波函数表示，并使用矩阵方法获得。

2.3 自适应图调节

图的构建在 NMF 算法中起着至关重要的作用，但是图参数（如邻居数）需要预先人工定义，因此选择最优的图参数比较困难。现有的许多图正则化的 NMFs 不能有效地捕捉数据的结构。首先，大多数图构造时需要计算数据样本之间的距离。但是，如果计算出的距离不够精确，那么得到的图质量就会很差。然后，一旦基于错误的计算建立了图，它在后续步骤中就会保持不变，因此得到的图不是最优的，从而影响 NMF 的聚类性能。因此，构建高质量的图是非常必要的[19]。自适应图调节[20]假设每个样本 x_i 连接到它的邻居 x_j 的概率是 z_{ij}。z_{ij} 是相似矩阵 $\boldsymbol{Z}$ 的元素。$\|x_i-x_j\|^2$ 的值小，表示 x_i 和 x_j 相似，此时 z_{ij} 的值就大。为了得到图 $\boldsymbol{Z}$，可以解决下列问题：

$$\begin{aligned}&\min_{z}\sum_{j=1}^{n}\left(\frac{1}{2}\|\boldsymbol{x}_i-\boldsymbol{x}_j\|^2\boldsymbol{z}_{ij}+\gamma\boldsymbol{z}_{ij}^2\right)\\&\text{s.t. }\boldsymbol{z}_i^{\mathrm{T}}1=1,0\leqslant\boldsymbol{z}_{ij}\leqslant1\end{aligned}\tag{3}$$

其中，γ 是权衡参数。因为拉普拉斯矩阵 $\boldsymbol{L}=\boldsymbol{D}-(\boldsymbol{Z}+\boldsymbol{Z}^{\mathrm{T}}/2)$，其中 $\boldsymbol{D}$ 是对角矩阵，$d_{ii}=\sum_j[(z_{ij}+z_{ji})/2]$，所以式(3)可以改写为：

$$\begin{aligned}&\min_{\boldsymbol{Z}}\operatorname{tr}(\boldsymbol{XLX}^{\mathrm{T}})+\gamma\|\boldsymbol{Z}\|_{\mathrm{F}}^2\\&\text{s.t. }\boldsymbol{Z}1=1,0\leqslant\boldsymbol{Z}\leqslant1\end{aligned}\tag{4}$$

通过优化上述问题，可以自适应地从数据中学习 $\boldsymbol{Z}$。$\|\boldsymbol{Z}\|_{\mathrm{F}}^2$ 是为了防止产生平凡解。

自适应图调节具有参数不敏感、尺度不变、操作简单等优点。因为自适应正则化只包含最近邻数的参数，所以其对参数不敏感。当每个点被缩放时，z_{ij} 保持不变，这使得它具有尺度不变性。自适应正则化只涉及加、减、乘、除的基本运算，操作简单[21]。

本文将高斯场及谐波函数和自适应图正则结合起来，通过带有标记的信息来指导图的构建，并且对噪声和异常值也做了处理。

3 基于高斯场和自适应图正则化的半监督聚类

本节首先详细描述 SCGFAG 方法，然后给出 SCGFAG 的优化算法和算法代码。

3.1 SCGFAG 算法

本文主要研究基于自适应图正则的半监督聚类问题。如何合理地利用有限的标签信息，消除噪声和异常值的影响，提高聚类性能，是一个重要研究问题。GFHF 将标签信息纳入半监督学习中，指导构建相似矩阵；l_1 范数、$l_{2,1}$ 范数减小噪声和异常值的影响；自适应图正则提高聚类准确度。

本文将自适应图正则、l_1 范数、$l_{2,1}$ 范数与半监督聚类集合到一个统一的框架中，得到目标函数：

$$\begin{aligned}&\arg\min_{\boldsymbol{H},\boldsymbol{W},\boldsymbol{S},\boldsymbol{Z}}\|\boldsymbol{X}-\boldsymbol{WH}^{\mathrm{T}}-\boldsymbol{S}\|_{2,1}+\lambda\|\boldsymbol{S}\|_1+\gamma\|\boldsymbol{Z}\|_{\mathrm{F}}^2+\\&2\beta\operatorname{tr}(\boldsymbol{HL}_{\boldsymbol{Z}}\boldsymbol{H}^{\mathrm{T}})+\operatorname{tr}((\boldsymbol{H}-\boldsymbol{Y})^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{H}-\boldsymbol{Y}))\\&\text{s.t. }\boldsymbol{Z}1=1,0\leqslant\boldsymbol{Z}\leqslant1,\boldsymbol{H}\geqslant0,\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}=\boldsymbol{I}\end{aligned}\tag{5}$$

其中，$\|\boldsymbol{X}\|_{2,1}=\sum_{i=1}^{n}\sqrt{\sum_{j=1}^{m}\boldsymbol{X}_{ij}^2}$[22]，$\|\boldsymbol{S}\|_1=\sum_{i=1}^{m}\sum_{j=1}^{n}|\boldsymbol{S}_{ij}|$，$\|\boldsymbol{Z}\|_{\mathrm{F}}=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}\boldsymbol{Z}_{ij}^2}$；其他变量和符号的定义见式(2)和式(4)。

3.2 SCGFAG 算法求解

近年来，许多方法被提出以解决这类优化问题，如 ALM[23]，APG[24]和 LADM[25]等。本文使用的方法是 ALM 方法。首先，引入辅助变量 $\boldsymbol{E}=\boldsymbol{X}-\boldsymbol{WH}^{\mathrm{T}}-\boldsymbol{S}$ 和 $\boldsymbol{G}=\boldsymbol{S}$，式(5)可以改写为：

$$\begin{aligned}&\arg\min_{\boldsymbol{E},\boldsymbol{G},\boldsymbol{H},\boldsymbol{W},\boldsymbol{S},\boldsymbol{Z}}\|\boldsymbol{E}\|_{2,1}+\lambda\|\boldsymbol{S}\|_1+\gamma\|\boldsymbol{Z}\|_{\mathrm{F}}^2+\\&\beta\operatorname{tr}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{Z}}\boldsymbol{H})+\operatorname{tr}((\boldsymbol{H}-\boldsymbol{Y})^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{H}-\boldsymbol{Y}))\\&\text{s.t. }\boldsymbol{E}=\boldsymbol{X}-\boldsymbol{WH}^{\mathrm{T}}-\boldsymbol{S},\boldsymbol{G}=\boldsymbol{S}\\&\quad0\leqslant\boldsymbol{Z}\leqslant1,\boldsymbol{H}\geqslant0,\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}=\boldsymbol{I}\end{aligned}\tag{6}$$

式(6)可转化为增广拉格朗日函数：

$$\begin{aligned}&\Gamma(\boldsymbol{E},\boldsymbol{W},\boldsymbol{H},\boldsymbol{Z},\boldsymbol{S},\boldsymbol{G},\boldsymbol{C}_1,\boldsymbol{C}_2)\\&=\|\boldsymbol{E}\|_{2,1}+\lambda\|\boldsymbol{S}\|_1+\gamma\|\boldsymbol{Z}\|_{\mathrm{F}}^2+\beta\operatorname{tr}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{Z}}\boldsymbol{F})+\\&\quad\operatorname{tr}((\boldsymbol{H}-\boldsymbol{Y})^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{H}-\boldsymbol{Y}))+\langle\boldsymbol{C}_1,\boldsymbol{X}-\boldsymbol{WH}^{\mathrm{T}}-\boldsymbol{S}\rangle+\langle\boldsymbol{C}_2,\\&\quad\boldsymbol{S}-\boldsymbol{G}\rangle\\&=\|\boldsymbol{E}\|_{2,1}+\lambda\|\boldsymbol{S}\|_1+\gamma\|\boldsymbol{Z}\|_{\mathrm{F}}^2+\beta\operatorname{tr}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{Z}}\boldsymbol{F})+\\&\quad\frac{\mu}{2}\left(\left\|\boldsymbol{X}-\boldsymbol{WH}^{\mathrm{T}}-\boldsymbol{S}-\boldsymbol{E}+\frac{\boldsymbol{C}_1}{\mu}\right\|_{\mathrm{F}}+\left\|\boldsymbol{S}-\boldsymbol{G}+\frac{\boldsymbol{C}_2}{\mu}\right\|_{\mathrm{F}}\right)\end{aligned}\tag{7}$$

其中，$\mu>0$，$\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$ 是拉格朗日乘数。交替求解这些未知变量，每一步求解一个，其他变量保持不变。

更新矩阵 $\boldsymbol{E}$：固定其他变量，更新 $\boldsymbol{E}$。

$$\Gamma=\arg\min_{\boldsymbol{E}}\|\boldsymbol{E}\|_{2,1}+\frac{\mu}{2}\left\|\boldsymbol{X}-\boldsymbol{WH}^{\mathrm{T}}-\boldsymbol{E}+\frac{\boldsymbol{C}_1}{\mu}\right\|_{\mathrm{F}}^2\tag{8}$$

根据文献[14]，将式(8)改写为：

$$\arg\min_{\boldsymbol{E}}\frac{1}{2}\|\boldsymbol{E}-\boldsymbol{Y}\|_{\mathrm{F}}^2+\frac{1}{\mu}\|\boldsymbol{E}\|_{2,1}\tag{9}$$

其中，$\boldsymbol{Y}=\boldsymbol{X}-\boldsymbol{WH}^{\mathrm{T}}-\boldsymbol{E}+\frac{\boldsymbol{C}_1}{\mu}$。

可以得到如下解：

$$\boldsymbol{E}(:,\boldsymbol{i})=\begin{cases}\dfrac{\|\boldsymbol{y}_i\|-\lambda}{\|\boldsymbol{y}_i\|}\boldsymbol{y}_i, & \text{如果}\dfrac{1}{\mu}<\|\boldsymbol{y}_i\|\\0, & \text{其他}\end{cases}\tag{10}$$

其中，y_i 是 $\boldsymbol{Y}$ 的第 i 列。

更新矩阵 $\boldsymbol{G}$：固定其他变量，更新 $\boldsymbol{G}$。引出下述问题：

$$\Gamma=\arg\min_{\boldsymbol{G}}\frac{\mu}{2}\left\|\boldsymbol{S}-\boldsymbol{G}+\frac{\boldsymbol{C}_2}{\mu}\right\|_{\mathrm{F}}^2\tag{11}$$

令

$$\frac{\partial(\Gamma_G)}{\partial\boldsymbol{G}}=0\tag{12}$$

$\boldsymbol{G}$ 的解为：

$$\boldsymbol{G}=\boldsymbol{S}+\frac{\boldsymbol{C}_2}{\mu}\tag{13}$$

更新矩阵 $\boldsymbol{W}$：固定其他变量，更新 $\boldsymbol{W}$。

$$\Gamma=\arg\min_{\boldsymbol{W}}\left\|\boldsymbol{X}-\boldsymbol{WH}^{\mathrm{T}}-\boldsymbol{E}-\boldsymbol{S}+\frac{\boldsymbol{C}_1}{\mu}\right\|_{\mathrm{F}}^2\tag{14}$$

令

$$\boldsymbol{M}=\boldsymbol{X}-\boldsymbol{E}-\boldsymbol{S}+\frac{\boldsymbol{C}_1}{\mu}\tag{15}$$

得到 $\boldsymbol{W}$ 的解：

$$\boldsymbol{W}=\boldsymbol{MH}\tag{16}$$

更新矩阵 $\boldsymbol{H}$：固定其他变量，更新 $\boldsymbol{H}$。

$$\Gamma=\arg\min_{\boldsymbol{H}}\beta\,\mathrm{tr}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{L}_{\boldsymbol{Z}}\boldsymbol{H})+\frac{\mu}{2}\left\|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}^{\mathrm{T}}-\boldsymbol{E}+\frac{\boldsymbol{C}_1}{\mu}\right\|_{\mathrm{F}}^2+\mathrm{tr}((\boldsymbol{H}-\boldsymbol{Y})^{\mathrm{T}}\boldsymbol{U}(\boldsymbol{H}-\boldsymbol{Y})) \tag{17}$$

令其一阶导数为 0，得到：

$$\boldsymbol{H}=(\boldsymbol{L}_{\boldsymbol{Z}}+\mu\boldsymbol{W}^2+\boldsymbol{U})^{-1}\left(\boldsymbol{U}\boldsymbol{Y}+\mu\left(\boldsymbol{X}-\boldsymbol{E}-\boldsymbol{S}+\frac{\boldsymbol{C}_1}{\mu}\right)\boldsymbol{W}\right) \tag{18}$$

更新矩阵 $\boldsymbol{S}$：固定其他变量，更新 $\boldsymbol{S}$。

$$\Gamma=\arg\min_{\boldsymbol{S}}\lambda\|\boldsymbol{S}\|_1+\frac{\mu}{2}\left(\left\|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}^{\mathrm{T}}-\boldsymbol{S}-\boldsymbol{E}+\frac{\boldsymbol{C}_1}{\mu}\right\|_{\mathrm{F}}+\left\|\boldsymbol{S}-\boldsymbol{G}+\frac{\boldsymbol{C}_2}{\mu}\right\|_{\mathrm{F}}\right) \tag{19}$$

令 $\boldsymbol{M}_1=\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}^{\mathrm{T}}-\boldsymbol{E}+\frac{\boldsymbol{C}_1}{\mu}$，$\boldsymbol{M}_2=\boldsymbol{G}+\frac{\boldsymbol{C}_2}{\mu}$，那么通过软阈值化可以得到 $\boldsymbol{S}$：

$$\boldsymbol{S}=\max\left(\max\left(\boldsymbol{M}_1+\boldsymbol{M}_2-\frac{\lambda}{\mu},0\right)+\min\left(\boldsymbol{M}_1+\boldsymbol{M}_2-\frac{\lambda}{\mu},0\right),0\right) \tag{20}$$

更新矩阵 $\boldsymbol{Z}$：固定其他变量，更新 $\boldsymbol{Z}$。令 $f_{ij}=\|h_i-h_j\|^2$，得到：

$$\arg\min_{z}\left\|z_i+\frac{\beta}{2\gamma}\boldsymbol{f}_i\right\| \tag{21}$$

s. t. $\boldsymbol{Z}1=1,0\leqslant\boldsymbol{Z}\leqslant1$

更新参数 $\boldsymbol{C}_1,\boldsymbol{C}_2,\mu$：

$$\boldsymbol{C}_1=\boldsymbol{C}_1+\mu(\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}^{\mathrm{T}}-\boldsymbol{S}-\boldsymbol{E}) \tag{22}$$

$$\boldsymbol{C}_2=\boldsymbol{C}_2+\mu(\boldsymbol{S}-\boldsymbol{G}) \tag{23}$$

$$\mu=\rho\mu \tag{24}$$

算法 1 描述了 SCGFAG 算法的过程。

算法 1 SCGFAG 算法

输入：数据矩阵 **X**，标签指示矩阵 **Y**，簇的个数 c，μ，ρ，最大迭代次数 T

输出：**H**

1. 步骤 1 计算 **H**
2. 初始化：**E**=0，**S**=**G**=0，**W**=0，**Z**=0，**H** 用 K-means 初始化。
3. WHILE 没有达到最大迭代次数 DO
4. 用式(10)更新 **E**；
5. 用式(13)更新 **G**；
6. 用式(16)更新 **W**；
7. 用式(18)更新 **H**；
8. 用式(20)更新 **S**；
9. 用式(21)更新 **Z**；
10. 用式(22)－式(24)更新 $\mathbf{C}_1,\mathbf{C}_2,\mu$；
11. END WHILE
12. 步骤 2 对 **H** 进行聚类
13. 记录 **H** 每一行的最大值的列号。

4 算法的性质分析

本节主要分析 SCGFAG 算法的时间复杂度和收敛性。

4.1 算法的复杂度分析

定理 1 算法时间复杂度为：$O(m^3+n^3+mc^2+mnc+mn+mc\cdot\max(m,n))$，$m$ 为数据集的行数，n 为数据集的列数，c 为聚类簇个数。

证明：$\boldsymbol{E}$ 的计算复杂度包括 $\boldsymbol{Y}$ 的计算和更新，分别是 $O(mn+c^3)$ 和 $O(mn)$；矩阵 $\boldsymbol{Z}$ 需要 $O(mn^2)$ 来计算；$\boldsymbol{W}$ 的计算复杂度为 $O(c^2)$；$\boldsymbol{G}$ 的计算复杂度为 $O(c^2)$；$\boldsymbol{S}$ 的计算复杂度为 $O(m^3+mc^2+mnc+mc\cdot\max(m,n))$；$\boldsymbol{H}$ 的计算复杂度为 $O(n^3)$。每次迭代的总成本为 $O(m^3+n^3+mc^2+mnc+mn+mc\cdot\max(m,n))$。SCGFAG 的计算复杂度为多项式时间。

4.2 算法的收敛性分析

ALM 的收敛性已得到证明[25]。然而，本文中有 6 个变量：$\boldsymbol{W},\boldsymbol{H},\boldsymbol{E},\boldsymbol{Z},\boldsymbol{G},\boldsymbol{S}$。而且，目标函数式(6)并不是绝对平滑的。这些因素不能保证本文方法是收敛的。定理 2 证明了收敛的 3 个充分条件[3]。

定理 2 $L(x)=f(x)+h(x)$的形式可以用 ALM 方法求解。ALM 方法收敛需要满足的 3 个条件为：

(1)ALM 问题的参数 λ 需要有上界；

(2)原始数据矩阵为全列秩；

(3)每个迭代步骤产生的误差是单调递减的，即误差 υ_k 是单调递减的。

定理 3 算法 1 是收敛的。

证明：根据定理 2，本文提出的目标函数求解方法满足下列条件：

(1)式(7)的参数 μ 需要有上界；

(2)数据矩阵 $\boldsymbol{X}$ 为全列秩；

(3)每个迭代步骤产生的误差，即 $\upsilon_k=\|(\boldsymbol{E}_k,\boldsymbol{W}_k,\boldsymbol{S}_k,\boldsymbol{H}_k,\boldsymbol{Z}_k,\boldsymbol{G}_k)-\arg\min_{\boldsymbol{E},\boldsymbol{W},\boldsymbol{S},\boldsymbol{H},\boldsymbol{Z},\boldsymbol{G}}\|^2$ 单调减少。其中，$\boldsymbol{E}_k,\boldsymbol{W}_k,\boldsymbol{S}_k,\boldsymbol{H}_k,\boldsymbol{Z}_k,\boldsymbol{G}_k$ 分别代表 $\boldsymbol{E},\boldsymbol{W},\boldsymbol{S},\boldsymbol{H},\boldsymbol{Z},\boldsymbol{G}$ 在第 k 步的值。前两个条件已经满足[3]，但第三个条件在理论上很难证明。可以通过实验证明第三个条件。在 k 次迭代目标函数的值为 $\upsilon_k=\|\boldsymbol{E}_k-\boldsymbol{E}_{k-1}\|_{\mathrm{F}}^2+\|\boldsymbol{W}_k-\boldsymbol{W}_{k-1}\|_{\mathrm{F}}^2+\|\boldsymbol{S}_k-\boldsymbol{S}_{k-1}\|_{\mathrm{F}}^2+\|\boldsymbol{H}_k-\boldsymbol{H}_{k-1}\|_{\mathrm{F}}^2+\|\boldsymbol{Z}_k-\boldsymbol{Z}_{k-1}\|_{\mathrm{F}}^2+\|\boldsymbol{G}_k-\boldsymbol{G}_{k-1}\|_{\mathrm{F}}^2$。由图 1 可以看出，目标函数的值先急剧下降，然后迅速收敛到一个稳定的值，说明其在一定程度上满足了第三个条件。综上所述，算法的收敛性得到了保证。

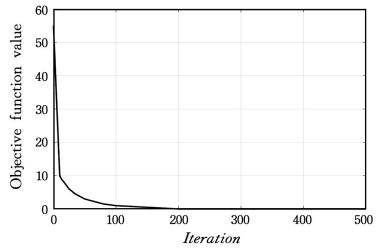


图 1 SCGFAG 在 COIL20 上的收敛曲线

Fig. 1 Convergence curve of SCGFAG on COIL20

5 实验

为了验证 SCGFAG 聚类的有效性，在大量基准数据集上将其与几种代表性的方法进行了比较。

5.1 实验数据集

表 2 列出了实验中所用数据集的特征。

(1) COIL20 数据集包含 20 个物体，每个物体水平旋转 360°，每 5°拍摄一次，每个物体总共 72 张图片。

(2)YaleB 数据集由 38 个人的 2 432 张人脸图像组成。每人有大约 64 张在不同光照条件下拍摄的照片，其中一半的图像被阴影或反射损坏。

(3)AR 人脸数据集包含 4 000 多张图像，对应于 126 人。

这些照片是在不同的面部表情、光照和遮挡（太阳镜和围巾）条件下拍摄的。

（4）Yale 数据集包含 15 个人的 165 张图像，每个人提供 11 张不同的图像，这些图像具有不同的面部表情，并且是在不同光照条件下拍摄的。

表 2　实验数据集

Table 2　Experimental datasets

数据集	# instances	# features	# classes
COIL20	1 440	1 024	20
YaleB	165	1 024	15
AR	840	768	126
Yale	165	4 096	15

各数据集的部分示例如图 2－图 5 所示。

图 2　COIL20 数据集示例

Fig. 2　COIL20 dataset samples

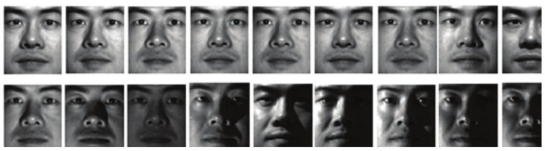


图 3　YaleB 数据集示例

Fig. 3　YaleB dataset samples

图 4　AR 数据集示例

Fig. 4　AR dataset samples

图 5　Yale 数据集示例

Fig. 5　Yale dataset samples

5.2　性能评价指标

利用准确度（Accuracy，ACC），标准化互信息（Normalized Mutual Information，NMI）和聚类纯度（Purity）3 个评价指标来评价实验性能。

ACC 的定义如下：

$$ACC=\frac{\sum_{i=1}^{n}\delta(map(r_i),l_i)}{n} \tag{25}$$

其中，$map(r_i)$是一个置换映射函数，使簇标签 r_i 与数据集中的等价标签匹配。n 为数据点个数，r_i 是 x_i 的预测标签，l_i 是对应的真实簇标签。$\delta(a,b)$是一个脉冲函数，如果 $a=b$，则函数值为 1，其他为 0。

NMI 的定义如下：

$$NMI(Y,C)=\frac{MI(Y,C)}{\sqrt{H(Y)H(C)}} \tag{26}$$

$$H(X)=\sum_{i=1}^{|X|}p(i)\log p(i) \tag{27}$$

其中，$p(i)=|X|/N$ 是从 X 中随机选择的一个对象落在第 X_i 类中的概率。

$$MI=\sum_{i=1}^{|Y|}\sum_{j=1}^{|C|}P(i,j)\log\frac{P(i,j)}{P(i)P(j)} \tag{28}$$

Purity 的定义如下：

$$Purity=\frac{1}{n}\sum_{i=1}^{c}\max_{j}(n_i^j) \tag{29}$$

其中，n_i^j 表示分配给集群 C_j 的 i 输入类的数据点数。c 是簇的个数。

5.3　基准方法

为了评价 SCGFAG 算法的聚类性能，本文将其与以下 7 种聚类算法进行了比较。

（1）K-means 聚类[4]。

（2）非负矩阵分解（NMF）[26]。

（3）鲁棒流形非负矩阵分解（Robust Manifold Non-negative Matrix Factorization，RMNMF）[14]。

（4）鲁棒图正则化非负矩阵分解（Robust Graph Regularized Nonnegative Matrix Factorization，RGNMF）[27]。

（5）基于自适应图正则化的低秩矩阵分解（Low-Rank Matrix Factorization With Adaptive Graph Regularizer，LMFAGR）[19]。

（6）半监督学习的非负低秩稀疏图[16]。

（7）判别半监督学习的非负稀疏编码（Sparse Probability Graph，SPG）[28]。

（8）半监督非负矩阵分解与不相似和相似性正则化（Semi-supervised Non-negative Matrix Factorization with Dissimilarity and Similarity Regularization，SNMFDSR）[29]。

5.4　半监督实验设计

对于每个数据集，本文从每一类中随机选取不同的样本作为标记样本，剩下的样本作为未标记样本。本节使用分类精度表示聚类结果。

对于 COIL20 数据集，每个人随机选出 2，4，6，8，10 张图像作为标记样本，其余的作为未标记样本。实验结果如表 3 所列。

表 3　在 COIL20 数据集上的聚类结果

Table 3　Clustering results on COIL20 dataset

（单位：%）

标记样本个数	NNLRS	SCGFAG
2	74.09	76.40
5	83.29	85.53
8	84.50	86.90
11	88.86	90.12
14	89.43	91.80

对于 YaleB 数据集，每个人随机选出 4，7，10，13，16 张图

像作为标记样本，其余的作为未标记样本。实验结果如表 4 所列。

表 4 在 YaleB 数据集上的聚类结果

Table 4 Clustering resultson YaleB dataset

（单位：%）

标记样本个数	SPG	NNLRS	SCGFAG
4	47.10	75.52	76.40
7	81.84	84.87	86.63
10	86.49	89.01	89.85
13	88.13	90.90	92.45
16	90.76	92.71	94.81

对于 AR 数据集，每个人随机选出 2,5,8,11,14 张图像作为标记样本，其余的作为未标记样本。实验结果如表 5 所列。

表 5 在 AR 数据集上的聚类结果

Table 5 Clustering result on AR dataset

（单位：%）

标记样本个数	SPG	NNLRS	SCGFAG
2	59.38	82.25	86.40
5	77.39	93.34	95.63
8	86.81	95.95	96.90
11	91.46	97.20	97.40
14	95.00	97.69	97.80

对于 Yale 数据集，每个人随机选出 2～6 张图像作为标记样本，其余的作为未标记样本。实验结果如表 6 所列。

表 6 在 Yale 数据集上的聚类结果

Table 6 Clustering result on Yale dataset

（单位：%）

标记样本个数	SPG	NNLRS	SCGFAG
2	51.25	56.66	59.25
3	55.16	67.10	68.23
4	58.58	71.83	74.15
5	66.45	76.89	79.35
6	71.46	81.54	84.44

由表 3－表 6 可以看出，SCGFAG 算法的聚类结果总体上明显优于其他算法。随着标记样本数量的增加，聚类结果也有明显的提升，这说明 GFHF 引入标签信息来指导相似矩阵的构建能够提高聚类的性能。

5.5 实验结果分析

对于 NNLRS，SPG 和本文的 SCGFAG 方法，本文实验是对数据集进行统一的设置，每个数据集的一个类选取 5 张图像作为标记样本，其余作为未标记样本。使用 ACC，NMI 和 Purity 作为评价指标，并对表 7－表 9 的实验结果进行了分析。

表 7 不同算法在不同数据集中的准确度

Table 7 ACC of different algorithms on different datasets

（单位：%）

Dataset	*K*-means	NMF	RMNMF	RGNMF	SPG	NNLRS	LMFAGR	SNMFDSR	SCGFAG
COIL20	68.24	43.67	64.20	63.46	60.57	66.08	70.89	84.52	**86.63**
YaleB	11.78	11.28	20.05	22.78	23.63	22.75	24.56	**37.42**	33.04
AR	27.24	15.60	33.26	28.10	27.23	29.10	30.01	31.34	**34.24**
Yale	48.88	35.07	46.95	60.12	55.15	65.45	63.69	50.23	**70.38**

表 8 不同数据集中的标准化互信息

Table 8 NMI of different methods on different datasets

（单位：%）

Dataset	*K*-means	NMF	RMNMF	RGNMF	SPG	NNLRS	LMFAGR	SNMFDSR	SCGFAG
COIL20	72.97	43.67	66.83	66.61	64.25	68.35	70.58	**92.45**	84.72
YaleB	6.19	5.54	21.41	20.88	22.56	21.31	25.59	**52.67**	38.03
AR	37.24	18.79	40.63	33.81	31.90	33.11	34.26	34.58	**35.11**
Yale	39.31	23.09	45.38	51.40	53.17	52.59	53.62	54.34	**55.03**

表 9 不同数据集中的纯度

Table 9 Purity of different methods on different datasets

（单位：%）

Dataset	*K*-means	NMF	RMNMF	RGNMF	SPG	NNLRS	LMFAGR	SNMFDSR	SCGFAG
COIL20	70.60	47.28	65.72	65.64	67.85	65.38	66.84	85.59	**87.92**
YaleB	12.46	11.88	20.01	24.84	25.27	25.62	27.45	37.60	**38.86**
AR	29.06	16.70	34.54	29.32	32.90	32.11	31.65	30.32	**38.15**
Yale	49.57	38.11	51.54	60.83	56.69	55.59	49.96	59.68	**66.73**

首先，这些基于图的 NMF 方法，如 RMNMF，RGNMF，LMFAGR 和本文提出的 SCGFAG，在大多数数据集中取得了比 *K*-means 和 NMF 更好的聚类性能。这是因为基于图的 NMF 利用相似图来挖掘给定数据的位置几何结构，可以显著提高 NMF 的聚类性能。

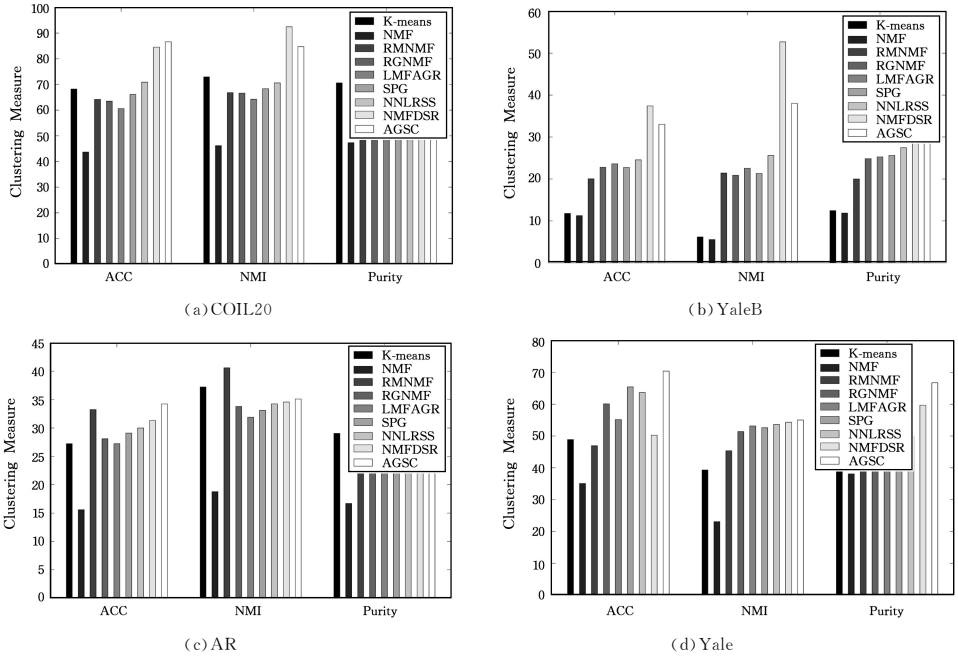
其次，RMNMF 和本文提出的 SCGFAG 采用 $l_{2,1}$ 范数来提高 NMF 的鲁棒性，降低了异常值对聚类结果的影响。而且本文提出的 SCGFAG 还使用了 l_1 范数，减小了噪声对聚类结果的影响。这说明鲁棒的 NMF 可以提高 NMF 的聚类性能。

再次，本文提出的 SCGFAG 相比其他的图正则化 NMF，比如 RMNMF 和 RGNMF，有一定的改进。虽然以上 4 种方法都使用了图拉普拉斯，但是 RMNMF 和 RGNMF 的输入图在矩阵分解的整个过程中是固定的。原始的图可能不能很好地正则化 NMF，从而限制了 NMF 的聚类性能。本文提出的 SCGFAG 通过学习自适应图很好地正则化 NMF，从而达到

了较好的聚类效果。

最后，本文提出的 SCGFAG 算法使用了 GFHF，引入半监督信息进行相似矩阵的构造，因此相较于 K-means，NMF，RMNMF，RGNMF 和 LMFAGR 算法，聚类性能有明显的提升。相比 SPG，NNLRS 和 SNMFDSR 算法，SCGFAG 算法使用了自适应图正则，自适应地学习图，并且引入了 l_1 范数和 $l_{2,1}$ 范数来减小噪声和异常值的影响。因此，SCGFAG 算法的聚类效果有所提升。虽然 SCGFAG 算法在某些情况下比 SNMFDSR 算法的效果稍差，但是其整体效果较好。

从图 6 中能更直观地观察到，本文的 SCGFAG 算法在 4 个数据集中取得了较好的聚类结果，表明本文提出的聚类算法能较好地改进 NMF 算法，是一种较好的聚类方法。



(a) COIL20　(b) YaleB　(c) AR　(d) Yale

图 6　在 COIL20，YaleB，AR 和 Yale 数据集上的聚类结果

Fig. 6　Clustering results on COIL20，YaleB，AR and Yale datasets

SCGFAG 算法的优势体现在以下三个方面。

(1)引入了 GFHF 半监督学习方法，使用标签信息来指导相似矩阵的构建。使用 GFHF 半监督学习方法所需的时间、精力和专业知识较少。

(2)自适应图正则化自适应地学习图，可以提高聚类精度。自适应图正则化具有对参数不敏感、尺度不变、操作简单等优点。

(3)目标函数采用 $l_{2,1}$ 范数作为度量，解决了其他聚类方法中常见的异常值问题，并对稀疏误差矩阵应用 l_1 范数来减小稀疏噪声对聚类的影响，使得聚类性能有明显提升。

结束语　本文提出了一种基于高斯场和自适应正则化的半监督聚类模型。该模型不仅能解决稀疏噪声和异常值问题，而且通过自适应正则化提高了聚类性能；另外还使用了 GFHF 半监督学习方法，通过标签信息来指导相似矩阵的构建，极大地提高了聚类性能；引入稀疏误差矩阵 S 和 l_1 范数来解决稀疏噪声问题；利用稀疏误差矩阵重构大量数据，得到鲁棒的分解结果；此外，将 $l_{2,1}$ 范数应用于矩阵分解，解决了异常值主导目标函数的问题。因此，SCGFAG 对由稀疏异常值重构的干净数据进行近似，通过 $l_{2,1}$ 范数对异常值进行约束，通过 l_1 范数对稀疏噪声进行约束以实现鲁棒性。最后，通过本文提出了一种迭代更新的优化方法，并证明了该方法的收敛性。最后，通过实验验证了 SCGFAG 的有效性。

未来的工作包括：1)对原始数据进行预处理，将其分为干净数据和损坏数据；2)将原始数据集进行投影操作。

参 考 文 献

[1] RAO S R，TRON R，VIDAL R，et al. Motion segmentation in the presence of outlying，incomplete，or corrupted trajectories[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2010，32(10)：1832-1845.

[2] SHI J，MALIK J. Normalized cuts and image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2000，22(8)：888-905.

[3] LIU G，LIN Z，YAN S，et al. Robust Recovery of Subspace Structures by Low-Rank Representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2013，35(1)：171-184.

[4] HARTIGAN J A，WONG M A. A K-means Clustering Algorithm：Algorithm AS 136[J]. Applied Statistics，1979，28(1)：100-108.

[5] VON LUXBURG U. A tutorial on spectral clustering[J]. Statistics and Computing，2007，17(4)：395-416.

[6] ZHOU S H，ZHU E，LIU X W，et al. Subspace segmentation-based robust multiple kernel clustering[J]. Information Fusion，2020，53：145-154

[7] ZHOU S，LIU X，ZHU C，et al. Spectral clustering-based local and global structure preservation for feature selection[C]// International Joint Conference on Neural Networks. IEEE，2014.

[8] DING C，LI T，JORDAN M. Convex and semi-nonnegative matrix factorizations[J]. IEEE Transactions on Software Enginee-

ring, 2010, 32(1): 45-55.
[9] ZHANG Y, KONG X W, WANG Z F, et al. Cluster analysis based on multi-view matrix Decomposition[J]. Acta Automata Sinica, 2018, 44(12): 2160-2169.
[10] DING Y, LI Y Z. Intrusion detection Algorithm based on PCA and Semi-supervised clustering[J]. Journal of Shandong University (Engineering Science), 2012, 42(5): 41-46.
[11] BASU S, BILENKO M, MOONEY R J, et al. A probabilistic framework for semi-supervised clustering[C]// Knowledge Discovery and Data Mining. 2004: 59-68.
[12] LIU H. Constrained nonnegative matrix factorization for image representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1299-1311.
[13] CAI D, HE X, HAN J, et al. Graph regularized non-negative matrix factorization for data representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1548-1560.
[14] HUANG J, NIE F P, HUANG H, et al. Robust Manifold Nonnegative Matrix Factorization[J]. ACM Transactions on Knowledge Discovery from Data, 2014, 8(3): 1-21.
[15] ZENG K, YU J, LI C, et al. Image clustering by hyper-graph regularized non-negative matrix factorization[J]. Neurocomputing, 2014, 138: 209-217.
[16] ZHANG X. Non-negative low rank and sparse graph for semi-supervised learning[C]// Computer Vision & Pattern Recognition. IEEE, 2012.
[17] NIE F, XU D, TSANG W H, et al. Flexible manifold embedding: a framework for semi-Supervised and unsupervised dimension reduction[J]. IEEE Transactions on Image Processing, 2010, 19(7): 1921-1932.
[18] ZHU X, GHAHRAMANI Z, LAFFERTY J D. Semi-supervised learning using Gaussian fields and Harmonic functions[C]// Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003). Washington DC, USA, 2003.
[19] LU G, WANG Y, ZOU J. Low-rank matrix factorization with adaptive graph regularizer [J]. IEEE Transactions on Image Processing, 2016, 25(5): 2196-2205.
[20] ZHANG L, ZHANG Q, DU B, et al. Adaptive manifold regularized matrix factorization for data clustering[C]// Twenty-sixth International Joint Conference on Artificial Intelligence. 2017.
[21] HE F, NIE F, WANG R, et al. Fast Semi-supervised learning with bipartite graph for large-scale data[J]. IEEE Transactions on Neural Networks, 2020, 31(2): 626-638.
[22] LI Z, LIU J, TANG J. Robust Structured Subspace Learning for Data Representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(10): 2085-2098.
[23] BHARDWAJ A, RAMAN S. Robust PCA-based solution to image composition using augmented Lagrange multiplier (ALM) [J]. Visual Computer, 2016, 32(5): 591-600.
[24] TOH K C, YUN S. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems[J]. Pacific Journal of Optimization, 2010, 6(3): 615-640.
[25] YANG J, YUAN X. Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization [J]. Mathematics of Computation, 2012, 82(281): 301-329.
[26] LEE D D, SEUNG H S. Algorithms for Non-negative Matrix Factorization [C] // Neural Information Processing Systems. 2000: 556-562.
[27] PENG C, KANG Z, HU Y, et al. Robust graph regularized nonnegative matrix factorization for clustering[J]. ACM Transactions on Knowledge Discovery from Data, 2017, 11(3): 33.
[28] HE R, ZHENG W S, HU B G, et al. Nonnegative sparse coding for discriminative semi-supervised learning[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2011.
[29] JIA Y H, KWONG S, HOU J H, et al. Semi-Supervised Non-Negative Matrix Factorization with Dissimilarity and Similarity Regularization [J]. IEEE Transaction on Neural Networks and Learning Systems, 2020, 31(7): 2510-2521.

ZHAO Min, born in 1995, postgraduate. Her main research interests include semi-supervised clustering and so on.

LIU Jing-lei, born in 1970, Ph.D, professor, master supervisor. His main research interests include artificial intelligent and theoretical computer science.