基于类别一致性学习的稀疏邻域约束的联合聚类

蒋 超 许堉坤 张芮嘉 安佰龙
(国网上海市电力公司电力科学研究院 上海 200051)

摘 要 为了充分挖掘特征结构,提升聚类性能,提出一种基于类别一致性学习的稀疏邻域约束的联合聚类方法。将联合聚类问题转化为附加对偶正则化子的非负矩阵三因式分解,在非负矩阵分解的基础上,增加两个正则化子,使数据关联性与标签分配一致;提出一种目标优化的乘法交替方案,从理论上证明了算法的收敛性和正确性。利用三种评价方法在六个数据集上进行验证,并对其参数敏感性进行分析。实验结果表明,该算法具有较优的聚类性能。

关键词 联合聚类 稀疏邻域约束 非负矩阵分解 一致性学习

中图分类号 TP391.41 **文献标志码** A DOI:10.3969/j.issn.1000-386x.2024.12.045

JOINT CLUSTERING OF SPARSE NEIGHBORHOOD CONSTRAINTS BASED ON CLASS CONSISTENCY LEARNING

Jiang Chao Xu Yukun Zhang Ruijia An Bailong
(*Electric Power Research Institute*, *State Grid Shanghai Municipal Electric Power Company*, *Shanghai 200051*, *China*)

Abstract In order to fully mine the feature structure and improve the clustering performance, a joint clustering method with sparse neighborhood constraints based on category consistency learning is proposed. The joint clustering problem was transformed into a tri-factorization of nonnegative matrix with dual regularizer. Based on the nonnegative matrix decomposition, two regularizers were added to make the data relevance consistent with the label assignment. A multiplication alternation scheme for objective optimization was introduced, and the convergence and correctness of the algorithm were proved theoretically. The three evaluation methods were verified on six data sets, and their parameter sensitivity was analyzed. Experiments show that the proposed algorithm has better performance.

Keywords Joint clustering Sparse neighborhood constraint Nonnegative matrix decomposition Consistent learning

0 引 言

聚类作为机器学习中的一项基本而重要的任务,已经得到了极大的发展,其应用领域包括数据挖掘、基因工程和计算机视觉等[1-2]。具体而言,聚类是在不知道标签的情况下对数据进行分组,对于现如今大数据时代,数据种类繁多,如何实现高效精确的数据聚类已经越来越引起学者们的研究。

常用的聚类算法分为三类:基于图的聚类、基于密度的聚类和基于距离的聚类[3]。在基于图的方法中,比例切割、最小最大切割和归一化切割方法是应用较为广泛的三种方法[4]。基于密度的方法主要包括基于密度的噪声应用空间聚类(DBSCAN)、密度峰值聚类(DPC)以及最大分离概率聚类(MSPC)等方法[5]。对于基于距离的算法,除了广泛使用的欧氏距离外,还采用了 Minkowski 距离和马氏距离方法[6]。作为一种经典的基于距离的聚类,K-means 聚类更是得到了充分的研究发展,如 K-means ++ 聚类、多核 K-means 聚类和模糊 K-means 聚类等[7-9]。虽然上述方法在不同的应用领域均取得了较好的聚类效果,但是上述单边聚类方法仍然不足以探索文本和基因等数据的上下文信息,存在一定的应用局限性。

近年来,发展联合聚类技术越来越受到人们的重

收稿日期:2021-08-01。国网上海市电力公司项目(SGSHJY00GPJS1800310)。**蒋超**,工程师,主研领域:电力营销计量,线损分析与治理。**许堉坤**,工程师。**张芮嘉**,工程师。**安佰龙**,助理工程师。

视,这些联合聚类算法不仅对样本进行分组聚类,而且同步进行特征划分。由于样本与特征之间的二元性,联合聚类被广泛应用于结构化数据。马欣野等[10]提出了一种基于模糊化器-2 型区间集的模糊化子集联合聚类方法,利用模糊理论有效提升了联合聚类的非线性描述能力。王继奎等[11]提出一种基于鲁棒细粒度分解的 K 均值联合聚类算法,既能处理噪声环境中的重叠聚类,又能处理粗糙集在聚类定义中的不确定性。余炳光等[12]提出了结合 KNN 和图标签传播的密度峰值联合聚类算法,该算法考虑了各数据点间的相关性,可以有效地对各种形状和密度差异性较大的数据进行聚类。虽然取得了一系列卓有成效的研究,但仍存在许多问题有待解决,例如聚类准确度还存在较大提升空间,容易产生闭点的误分类,另外该方法难以挖掘特征结构,忽略了数据关联性和标签分配之间的一致性。

为解决上述问题,本文提出了一种基于类别一致性学习的稀疏邻域约束的联合聚类算法。将联合聚类转化为具有对偶正则化子的非负矩阵三因子分解学习类别一致性,提出并证明了一种求解目标优化问题的乘法迭代更新算法,通过理论和实验验证了该算法的优越性。

1　邻域约束联合聚类

1.1　问题表述

给定一个数据矩阵 $\boldsymbol{X}\in\mathbf{R}^{d\times n}$,其中 d 表示特征数,n 表示样本数,特征空间和样本空间表示为 $\boldsymbol{F}=\{f_1,f_2,\cdots,f_d\}$,$\chi=\{x_1,x_2,\cdots,x_n\}$。其中 $\boldsymbol{X}=(x_1,x_2,\cdots,x_n)=(f_1,f_2,\cdots,f_d)^{\mathrm{T}}$。

传统的聚类方法倾向于将样本分成 c 个簇。如果本文把离散的指标矩阵转换成连续的非负矩阵,该问题可以近似地转化为如下优化目标:

$$\min_{S,G}\frac{1}{2}\|\boldsymbol{X}-\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\|_F^2\quad S\geqslant 0,G\geqslant 0 \tag{1}$$

式中:系数矩阵为 $\boldsymbol{S}\in\mathbf{R}^{d\times c}$,指标矩阵为 $\boldsymbol{G}\in\mathbf{R}^{n\times c}$。

类似地,联合聚类试图将特征划分为 c_1 集群,将样本划分为 c_2 集群。如果将两个离散指标矩阵简化为连续非负矩阵,则该问题可以近似地转化为如下优化目标,

$$\min_{\boldsymbol{F},\boldsymbol{S},\boldsymbol{G}}\frac{1}{2}\|\boldsymbol{X}-\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\|_F^2\quad \boldsymbol{F}\geqslant 0,\boldsymbol{S}\geqslant 0,\boldsymbol{G}\geqslant 0 \tag{2}$$

式中 $\boldsymbol{S}\in\mathbf{R}^{c_1\times c_2}$ 是由聚类指标 $\boldsymbol{F}\in\mathbf{R}^{d\times c_2}$ 和 $\boldsymbol{G}\in\mathbf{R}^{n\times c_2}$ 共享的稀疏矩阵。

样本越相似,其类别标签就越一致。因此,本文附加了两个正则化器来保持数据相关性和标签分配之间的一致性:

$$\min_{\boldsymbol{F},\boldsymbol{S},\boldsymbol{G},\boldsymbol{Z}_1,\boldsymbol{Z}_2}\frac{1}{2}\|\boldsymbol{X}-\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\|_F^2+\frac{\alpha}{2}\|\boldsymbol{W}_1-\boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}}\|_F^2+\frac{\beta}{2}\|\boldsymbol{W}_2-\boldsymbol{F}\boldsymbol{Z}_2^{\mathrm{T}}\|_F^2$$
$$\text{s.t.}\quad \boldsymbol{F}\geqslant 0,\boldsymbol{S}\geqslant 0,\boldsymbol{G}\geqslant 0 \tag{3}$$

式中:$\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$ 分别表示特征和样本的相似性,$\boldsymbol{Z}_1$、$\boldsymbol{Z}_2$ 是系数矩阵。

1.2　优化算法

由于式(3)表示的函数非凸性,无法从式(3)中得到一个闭式解。当固定其他变量,目标函数在某个变量中是凸的。因此,本文提出了一种乘法迭代优化方案。

1.2.1　构造 L

首先,原始问题通过以下目标函数最小化:

$$J=\frac{1}{2}\|\boldsymbol{X}-\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\|_F^2+\frac{\alpha}{2}\|\boldsymbol{W}_1-\boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}}\|_F^2+\frac{\beta}{2}\|\boldsymbol{W}_2-\boldsymbol{F}\boldsymbol{Z}_2^{\mathrm{T}}\|_F^2$$
$$\text{s.t.}\quad \boldsymbol{F}\geqslant 0,\boldsymbol{S}\geqslant 0,\boldsymbol{G}\geqslant 0 \tag{4}$$

式中:$\alpha,\beta\geqslant 0$ 用于平衡重建误差。如果 $\alpha=\beta=0$,则该目标将退化为传统目标。此外,关于特征的第二个项求数据相关性和标签分配之间的一致性,第三种方法倾向于在样本方面保持数据相似性和标签分配之间的一致性。

为了求解约束,本文引入了拉格朗日乘子 $\boldsymbol{\Phi}$、$\boldsymbol{\Psi}$ 和 $\boldsymbol{\Omega}$,将拉格朗日函数构造为:

$$L=\frac{1}{2}\|\boldsymbol{X}-\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\|_F^2+\frac{\alpha}{2}\|\boldsymbol{W}_1-\boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}}\|_F^2+\frac{\beta}{2}\|\boldsymbol{W}_2-\boldsymbol{G}\boldsymbol{Z}_2^{\mathrm{T}}\|_F^2-$$
$$Tr(\boldsymbol{\Phi}\boldsymbol{F}^{\mathrm{T}})-Tr(\boldsymbol{\Psi}\boldsymbol{S}^{\mathrm{T}})-Tr(\boldsymbol{\Omega}\boldsymbol{G}^{\mathrm{T}}) \tag{5}$$

式中:

$$\|\boldsymbol{X}-\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\|_F^2=Tr((\boldsymbol{X}-\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{X}-\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}))=$$
$$Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})-2Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})+Tr(\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}) \tag{6}$$

$$\|\boldsymbol{W}_1-\boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}}\|_F^2=Tr((\boldsymbol{W}_1-\boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{W}_1-\boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}}))=$$
$$Tr(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_1)-2Tr(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}})+Tr(\boldsymbol{Z}_1\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}}) \tag{7}$$

$$\|\boldsymbol{W}_2-\boldsymbol{G}\boldsymbol{Z}_2^{\mathrm{T}}\|_F^2=Tr((\boldsymbol{W}_2-\boldsymbol{G}\boldsymbol{Z}_1^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{W}_2-\boldsymbol{G}\boldsymbol{Z}_1^{\mathrm{T}}))=$$
$$Tr(\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{W}_2)-2Tr(\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{G}\boldsymbol{Z}_2^{\mathrm{T}})+Tr(\boldsymbol{Z}_2\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{Z}_2^{\mathrm{T}}) \tag{8}$$

因此,拉格朗日函数可以改写为:

$$L=\frac{1}{2}Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})-Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})+\frac{1}{2}Tr(\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})+$$
$$\frac{\alpha}{2}Tr(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{W}_1)-\alpha Tr(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}})+\frac{\alpha}{2}Tr(\boldsymbol{Z}_1\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}})+$$
$$\frac{\beta}{2}Tr(\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{W}_2)-\beta Tr(\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{G}\boldsymbol{Z}_2^{\mathrm{T}})+\frac{\beta}{2}(\boldsymbol{Z}_2\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{Z}_2^{\mathrm{T}})-$$
$$Tr(\boldsymbol{\Phi}\boldsymbol{F}^{\mathrm{T}})-Tr(\boldsymbol{\Psi}\boldsymbol{S}^{\mathrm{T}})-Tr(\boldsymbol{\Omega}\boldsymbol{G}^{\mathrm{T}}) \tag{9}$$

1.2.2　$\boldsymbol{W}_1$和 $\boldsymbol{W}_2$的计算过程

对于稀疏性,考虑 k-近邻而不是考虑所有节点。因此,将特征关联矩阵 $\boldsymbol{W}_1$ 构造如下:

$$W_{ij}^1=\begin{cases}1 & f_j\in N(f_i)\\ 0 & \text{其他}\end{cases} \tag{10}$$

式中：W_{ij}^1表示f_j与f_i的距离，$N(f_i)$表示f_i的 k-近邻。样本关联矩阵 $\boldsymbol{W}_2$的结构如下：

$$W_{ij}^2=\begin{cases}1 & x_j\in N(x_i)\\ 0 & \text{其他}\end{cases} \tag{11}$$

式中：W_{ij}^2测量 x_j与 x_i和 $N(x_i)$的距离是指 x_i附近的 k_2最近的邻域。此外，一些核函数可以用来区分不同邻域之间的差异，但会引入额外的参数。

1.2.3 $\boldsymbol{Z}_1$和 $\boldsymbol{Z}_1$的计算过程

通过求 $\boldsymbol{L}$ 对 $\boldsymbol{Z}_1$和 $\boldsymbol{Z}_2$的导数，可以得到：

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{Z}_1}=-\alpha \boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{F}+\alpha \boldsymbol{Z}_1\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F} \tag{12}$$

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{Z}_2}=-\beta \boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{G}+\beta \boldsymbol{Z}_2\boldsymbol{G}^{\mathrm{T}}G \tag{13}$$

此外，令$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{Z}_1}=0$ 和$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{Z}_2}=0$，可以得到：

$$\boldsymbol{Z}_1=\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{F}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F})^{-1} \tag{14}$$

$$\boldsymbol{Z}_2=\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1} \tag{15}$$

1.2.4 更新 $\boldsymbol{F}$、$\boldsymbol{S}$ 和 $\boldsymbol{G}$

通过求 $\boldsymbol{L}$ 对 $\boldsymbol{F}$、$\boldsymbol{S}$ 和 $\boldsymbol{G}$ 的导数，可以得到：

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{F}}=-\boldsymbol{S}\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}+\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}-\alpha \boldsymbol{W}_1\boldsymbol{Z}_1+\alpha \boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}}\boldsymbol{Z}_1-\boldsymbol{\Phi} \tag{16}$$

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{S}}=-\boldsymbol{F}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G}+\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}-\boldsymbol{\Psi} \tag{17}$$

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{G}}=-\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}+\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}-\beta \boldsymbol{W}_2\boldsymbol{Z}_2+\beta \boldsymbol{G}\boldsymbol{Z}_2^{\mathrm{T}}\boldsymbol{Z}_2-\boldsymbol{\Omega} \tag{18}$$

结合 KKT 条件 $\boldsymbol{\Phi}_{ij}\boldsymbol{F}_{ij}=0$、$\boldsymbol{\Psi}_{ij}\boldsymbol{S}_{ij}=0$ 和 $\boldsymbol{\Omega}_{ij}\boldsymbol{G}_{ij}=0$，令$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{F}}=0$、$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{S}}=0$ 和$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{G}}=0$，可以得到：

$$(-\boldsymbol{X}\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}+\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}-\alpha \boldsymbol{W}_1\boldsymbol{Z}_1+\alpha \boldsymbol{F}\boldsymbol{Z}_1^{\mathrm{T}}\boldsymbol{Z}_1)_{ij}\boldsymbol{F}_{ij}=0 \tag{19}$$

$$(-\boldsymbol{F}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G}+\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})_{ij}\boldsymbol{S}_{ij}=0 \tag{20}$$

$$(-\boldsymbol{X}\boldsymbol{F}\boldsymbol{S}^{\mathrm{T}}+\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}-\beta \boldsymbol{W}_2\boldsymbol{Z}_2+\beta \boldsymbol{G}\boldsymbol{Z}_2^{\mathrm{T}}\boldsymbol{Z}_2)_{ij}\boldsymbol{G}_{ij}=0 \tag{21}$$

通过引入 $\boldsymbol{M}=\boldsymbol{W}_1\boldsymbol{Z}_1=\boldsymbol{M}^+-\boldsymbol{M}^-$，$\boldsymbol{N}=\boldsymbol{Z}_1^{\mathrm{T}}\boldsymbol{Z}_1=\boldsymbol{N}^+-\boldsymbol{N}^-$，$\boldsymbol{P}=\boldsymbol{W}_2\boldsymbol{Z}_2=\boldsymbol{P}^+-\boldsymbol{P}^-$ 以及 $\boldsymbol{Q}=\boldsymbol{Z}_2^{\mathrm{T}}\boldsymbol{Z}_2=Q^+-Q^-$，式(19)和式(21)可以改写为：

$$(-\boldsymbol{X}\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}+\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}-\alpha \boldsymbol{M}^++\alpha \boldsymbol{M}^-+\alpha \boldsymbol{F}\boldsymbol{N}^+-\alpha \boldsymbol{F}\boldsymbol{N}^-)_{ij}\boldsymbol{F}_{ij}=0 \tag{22}$$

$$(-\boldsymbol{X}\boldsymbol{F}\boldsymbol{S}^{\mathrm{T}}+\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}-\beta \boldsymbol{P}^++\beta \boldsymbol{P}^-+\beta \boldsymbol{G}\boldsymbol{Q}^+-\beta \boldsymbol{G}\boldsymbol{Q}^-)_{ij}\boldsymbol{G}_{ij}=0 \tag{23}$$

对于任意矩阵$\boldsymbol{A}$，满足$\boldsymbol{A}^+=\frac{|\boldsymbol{A}|+\boldsymbol{A}}{2}$和$\boldsymbol{A}^-=\frac{|\boldsymbol{A}|-\boldsymbol{A}}{2}$。

根据非负二次问题的优化框架，式(20)、式(22)和式(23)推出以下更新规则：

$$\boldsymbol{F}_{ij}=\left[\frac{(\boldsymbol{X}\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}+\alpha \boldsymbol{M}^++\alpha \boldsymbol{F}\boldsymbol{N}^-)_{ij}}{(\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}+\alpha \boldsymbol{M}^-+\alpha \boldsymbol{F}\boldsymbol{N}^+)_{ij}}\right]^{\frac{1}{2}} \tag{24}$$

$$\boldsymbol{S}_{ij}=\left[\frac{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G})_{ij}}{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})_{ij}}\right]^{\frac{1}{2}} \tag{25}$$

$$\boldsymbol{G}_{ij}=\left[\frac{(\boldsymbol{X}\boldsymbol{F}\boldsymbol{S}^{\mathrm{T}}+\beta \boldsymbol{P}^++\beta \boldsymbol{G}\boldsymbol{Q}^-)_{ij}}{(\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}+\beta \boldsymbol{P}^-+\beta \boldsymbol{G}\boldsymbol{Q}^+)_{ij}}\right]^{\frac{1}{2}} \tag{26}$$

1.2.5 标签分配

完成更新后，最终得到有效的集群指标矩阵 $\tilde{\boldsymbol{F}}$ 和 $\tilde{\boldsymbol{G}}$。此外，第 i 个特征的标签由式(27)进行分配。

$$l(f_i)=\arg\max_j \tilde{\boldsymbol{F}}_{ij} \tag{27}$$

第 i 个样本的标签由式(28)分配。

$$l(x_i)=\arg\max_j \tilde{\boldsymbol{G}}_{ij} \tag{28}$$

基于上述分析，整个优化过程见算法 1，简称 SNCC。

算法 1 稀疏邻域约束联合聚类

输入：数据矩阵 $\boldsymbol{X}\in\mathbf{R}^{d\times n}$，特征聚类数 c_1，样本聚类数 c_2，参数 α 和 β。

输出：特征标签$\{l(f_i)\}_{i=1}^{c_1}$，样本标签$\{l(x_i)\}_{i=1}^{c_2}$。

1. 初始化 $\boldsymbol{F}$、$\boldsymbol{S}$ 和 $\boldsymbol{G}$
2. 计算 $\boldsymbol{W}_1$和 $\boldsymbol{W}_2$
3. **while** 不收敛 **do**
4. 计算 $\boldsymbol{Z}_1=\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{F}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F})^{-1}$
5. 计算 $\boldsymbol{Z}_2=\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})^{-1}$
6. 更新 $\boldsymbol{F}_{ij}\leftarrow\boldsymbol{F}_{ij}\left[\frac{(\boldsymbol{X}\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}+\alpha \boldsymbol{M}^++\alpha \boldsymbol{F}\boldsymbol{N}^-)_{ij}}{(\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}+\alpha \boldsymbol{M}^-+\alpha \boldsymbol{F}\boldsymbol{N}^+)_{ij}}\right]^{\frac{1}{2}}$
7. 更新 $\boldsymbol{S}_{ij}\leftarrow\boldsymbol{S}_{ij}\left[\frac{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G})_{ij}}{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})_{ij}}\right]^{\frac{1}{2}}$
8. 更新 $\boldsymbol{G}_{ij}\leftarrow\boldsymbol{G}_{ij}\left[\frac{(\boldsymbol{X}\boldsymbol{F}\boldsymbol{S}^{\mathrm{T}}+\beta \boldsymbol{P}^++\beta \boldsymbol{G}\boldsymbol{Q}^-)_{ij}}{(\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}+\beta \boldsymbol{P}^-+\beta \boldsymbol{G}\boldsymbol{Q}^+)_{ij}}\right]^{\frac{1}{2}}$
9. **end while**
10. 分配特征标签 $l(f_i)$和样本标签 $l(x_i)$

1.3 收敛性分析

定义 1 如果满足下列条件，那么 $Z(h,h')$是 $F(h)$的辅助函数。

$$Z(h,h')\geqslant F(h),Z(h,h)=F(h) \tag{29}$$

引理 1[13] 如果 $Z(h,h')$是 $F(h)$的辅助函数，则在更新策略中 $F(h)$是非递增的。

$$h^{(t+1)}=\arg\max_h Z(h,h^{(t)}) \tag{30}$$

引理 2[14] $\forall \boldsymbol{A}\in\mathbf{R}_+^{n\times n},\boldsymbol{B}\in\mathbf{R}_+^{k\times k},\boldsymbol{S}\in\mathbf{R}_+^{n\times k},S'\in\mathbf{R}_+^{n\times k}$，且 $\boldsymbol{A}$、$\boldsymbol{B}$ 是对称的，满足以下不等式：

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\frac{(\boldsymbol{A}\boldsymbol{S}'\boldsymbol{B})_{ij}S_{ij}^2}{S'_{ij}}\geqslant Tr(\boldsymbol{S}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{S}\boldsymbol{B}) \tag{31}$$

定理 1 令：

$$J(F)=-Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})+\frac{1}{2}Tr(\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})-\alpha Tr(\boldsymbol{M}\boldsymbol{F}^{\mathrm{T}})+\frac{\alpha}{2}Tr(\boldsymbol{F}\boldsymbol{N}\boldsymbol{F}^{\mathrm{T}}) \tag{32}$$

其辅助函数为：

$$Z(\boldsymbol{F},\boldsymbol{F}') = -\sum_{i=1}^{d}\sum_{j=1}^{c_1}(\boldsymbol{XGS}^{\mathrm{T}})_{ij}F'_{ij}\left(1+\log\frac{F_{ij}}{F'_{ij}}\right)+\frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{c_1}\frac{(\boldsymbol{F}'\boldsymbol{SG}^{\mathrm{T}}\boldsymbol{GS}^{\mathrm{T}})_{ij}F_{ij}^2}{F'_{ij}}+\alpha\sum_{i=1}^{d}\sum_{j=1}^{c_1}M_{ij}^{-}\frac{F_{ij}^2+F_{ij}'^2}{2F'_{ij}}-\alpha\sum_{i=1}^{d}\sum_{j=1}^{c_1}M_{ij}^{+}F'_{ij}\left(1+\log\frac{F_{ij}}{F'_{ij}}\right)+\frac{\alpha}{2}\sum_{i=1}^{d}\sum_{j=1}^{c_1}\frac{(\boldsymbol{F}'\boldsymbol{N}^{+})_{ij}+F_{ij}^2}{F'_{ij}}-\frac{\alpha}{2}\sum_{i=1}^{d}\sum_{j=1}^{c_1}\sum_{k=1}^{c_1}N_{jk}^{-}F'_{ij}F'_{ik}\left(1+\log\frac{F_{ij}F_{ik}}{F'_{ij}F'_{ik}}\right)\tag{33}$$

式(33)在 $\boldsymbol{F}$ 中是凸的,而且它的全局极小值是:

$$F_{ij}=\arg\max_{F_{ij}}Z(\boldsymbol{F},\boldsymbol{F}')=F'_{ij}\left[\frac{(\boldsymbol{XGS}^{\mathrm{T}}+\alpha\boldsymbol{M}^{+}+\alpha\boldsymbol{FN}^{-})_{ij}}{(\boldsymbol{FSG}^{\mathrm{T}}\boldsymbol{GS}^{\mathrm{T}}+\alpha\boldsymbol{M}^{-}+\alpha\boldsymbol{FN}^{+})_{ij}}\right]^{\frac{1}{2}}\tag{34}$$

证明。首先,$J(F)$可以重写为:

$$J(\boldsymbol{F})=-Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}})+\frac{1}{2}Tr(\boldsymbol{GS}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}})-\alpha Tr(\boldsymbol{M}^{+}\boldsymbol{F}^{\mathrm{T}})+\alpha Tr(\boldsymbol{M}^{-}\boldsymbol{F}^{\mathrm{T}})+\frac{\alpha}{2}Tr(\boldsymbol{FN}^{+}\boldsymbol{F}^{\mathrm{T}})-\frac{\alpha}{2}Tr(\boldsymbol{FN}^{-}\boldsymbol{F}^{\mathrm{T}})\tag{35}$$

根据引理 2,可以得到:

$$Tr(\boldsymbol{GS}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}})=Tr(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}}\boldsymbol{GS}^{\mathrm{T}})\leqslant\sum_{i=1}^{d}\sum_{j=1}^{c_1}\frac{(\boldsymbol{F}'\boldsymbol{SG}^{\mathrm{T}}\boldsymbol{GS}^{\mathrm{T}})_{ij}F_{ij}^2}{F'_{ij}}\tag{36}$$

$$Tr(FN^{+}F^{\mathrm{T}})=Tr(FN^{+}F^{\mathrm{T}})\leqslant\sum_{i=1}^{d}\sum_{j=1}^{c_1}\frac{(F'N^{+})_{ij}F_{ij}^2}{F'_{ij}}\tag{37}$$

根据不等式:$\forall a,b>0,a\leqslant\frac{a^2+b^2}{2b}$,可以得到:

$$Tr(M^{-}F^{\mathrm{T}})=\sum_{i=1}^{d}\sum_{j=1}^{c_1}M_{ij}^{-}F_{ij}\leqslant\sum_{i=1}^{d}\sum_{j=1}^{c_1}M_{ij}^{-}\frac{F_{ij}^2+F_{ij}'^2}{2F'_{ij}}\tag{38}$$

根据不等式:$\forall z>0,z\geqslant 1+\log z$,得到:

$$Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}})=Tr(\boldsymbol{SG}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F})=\sum_{i=1}^{d}\sum_{j=1}^{c_1}(\boldsymbol{XGS}^{\mathrm{T}})_{ij}F_{ij}=\sum_{i=1}^{d}\sum_{j=1}^{c_1}(\boldsymbol{XGS}^{\mathrm{T}})_{ij}F'_{ij}\frac{F_{ij}}{F'_{ij}}\geqslant\sum_{i=1}^{d}\sum_{j=1}^{c_1}(\boldsymbol{XGS}^{\mathrm{T}})_{ij}F'_{ij}\left(1+\log\frac{F_{ij}}{F'_{ij}}\right)\tag{39}$$

$$Tr(\boldsymbol{M}^{+}\boldsymbol{F}^{\mathrm{T}})=\sum_{i=1}^{d}\sum_{j=1}^{c_1}M'_{ij}F_{ij}=\sum_{i=1}^{d}\sum_{j=1}^{c_1}M_{ij}^{+}F'_{ij}\frac{F_{ij}}{F'_{ij}}\geqslant\sum_{i=1}^{d}\sum_{j=1}^{c_1}M_{ij}^{+}F'_{ij}\left(1+\log\frac{F_{ij}}{F'_{ij}}\right)\tag{40}$$

$$Tr(\boldsymbol{FN}^{-}\boldsymbol{F}^{\mathrm{T}})=Tr(\boldsymbol{N}^{-}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F})=\sum_{i=1}^{d}\sum_{j=1}^{c_1}\sum_{k=1}^{c_1}N_{jk}^{-}F_{ij}F_{ik}=\sum_{i=1}^{d}\sum_{j=1}^{c_1}\sum_{k=1}^{c_1}N_{jk}^{-}F'_{ij}F'_{ik}\frac{F_{ij}F_{ik}}{F'_{ij}F'_{ik}}\geqslant\sum_{i=1}^{d}\sum_{j=1}^{c_1}\sum_{k=1}^{c_1}N_{jk}^{-}F'_{ij}F'_{ik}\left(1+\log\frac{F_{ij}F_{ik}}{F'_{ij}F'_{ik}}\right)\tag{41}$$

显然,$Z(\boldsymbol{F},\boldsymbol{F}')$由所有的边界组成的,满足定义 1。因此,$Z(\boldsymbol{F},\boldsymbol{F}')$是 $J(\boldsymbol{F})$的辅助函数。

为了最小化 $Z(\boldsymbol{F},\boldsymbol{F}')$,本文令:

$$\frac{\partial Z(\boldsymbol{F},\boldsymbol{F}')}{\partial F_{ij}}=-\frac{(\boldsymbol{XGS}^{\mathrm{T}})_{ij}F'_{ij}}{F_{ij}}+\frac{(\boldsymbol{F}'\boldsymbol{SG}^{\mathrm{T}}\boldsymbol{GS}^{\mathrm{T}})_{ij}F_{ij}}{F'_{ij}}-\alpha\frac{M_{ij}^{+}F'_{ij}}{F_{ij}}+\alpha\frac{M_{ij}^{-}F_{ij}}{F'_{ij}}+\alpha\frac{(\boldsymbol{F}'\boldsymbol{N}^{+})F_{ij}}{F'_{ij}}-\alpha\frac{(\boldsymbol{F}'\boldsymbol{N}^{-})F'_{ij}}{F_{ij}}\tag{42}$$

并得到其 Hessian 矩阵:

$$\frac{\partial^2 Z(\boldsymbol{F},\boldsymbol{F}')}{\partial F_{ij}\partial F_{kl}}=\delta_{ik}\delta_{ji}\left(\frac{(\boldsymbol{XGS}^{\mathrm{T}})_{ij}F'_{ij}}{F_{ij}^2}+\frac{(\boldsymbol{F}'\boldsymbol{SG}^{\mathrm{T}}\boldsymbol{GS}^{\mathrm{T}})_{ij}}{F'_{ij}}+\alpha\frac{M_{ij}^{+}F'_{ij}}{F_{ij}^2}\right)+\alpha\frac{M_{ij}^{-}}{F'_{ij}}+\alpha\frac{(\boldsymbol{F}'\boldsymbol{N}^{+})_{ij}}{F'_{ij}}+\alpha\frac{(\boldsymbol{F}'\boldsymbol{N}^{-})F'_{ij}}{F_{ij}^2}\tag{43}$$

当且仅当 $i=k$ 时,$\delta_{ik}=1$,否则 $\delta_{ik}=0$。此外,Hessian 矩阵是具有正值元素的对角矩阵,通过令$\frac{\partial Z(\boldsymbol{F},\boldsymbol{F}')}{\partial F_{ij}}=0$,可以得到 $Z(\boldsymbol{F},\boldsymbol{F}')$是凸的,且 $J(\boldsymbol{F})=\arg\min_{F_{ij}}Z(\boldsymbol{F},\boldsymbol{F}')$。

定理 2

$$J(\boldsymbol{S})=-Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}})+\frac{1}{2}Tr(\boldsymbol{GS}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}})\tag{44}$$

其辅助函数为:

$$Z(\boldsymbol{S},\boldsymbol{S}')=-\sum_{i=1}^{c_1}\sum_{j=1}^{c_2}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{XG})_{ij}S'_{ij}\left(\log\frac{S_{ij}}{S'_{ij}}\right)+\frac{1}{2}\sum_{i=1}^{c_1}\sum_{j=1}^{c_2}\frac{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{FS}'\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})_{ij}S_{ij}^2}{S'_{ij}}\tag{45}$$

式(45)在 $\boldsymbol{S}$ 中是凸的,其全局极小值为:

$$s_{ij}=\arg\max_{S_{ij}}Z(\boldsymbol{S},\boldsymbol{S}')=S'_{ij}\left[\frac{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{XG})_{ij}}{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}}\boldsymbol{G})_{ij}}\right]^{\frac{1}{2}}\tag{46}$$

证明。首先,$J(\boldsymbol{S})$可以重写为:

$$J(\boldsymbol{S})=-Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}})+\frac{1}{2}Tr(\boldsymbol{GS}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}})\tag{47}$$

根据引理 2,可知:

$$Tr(\boldsymbol{GS}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}})=Tr(\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}}\boldsymbol{G})\leqslant\sum_{i=1}^{c_1}\sum_{j=1}^{c_2}\frac{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{FS}'\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})_{ij}S_{ij}^2}{S'_{ij}}\tag{48}$$

根据不等式:$\forall z>0,z\geqslant 1+\log z$,得到:

$$Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{FSG}^{\mathrm{T}})=Tr(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{FS})\sum_{i=1}^{c_1}\sum_{j=1}^{c_2}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{XG})_{ij}S_{ij}=\sum_{i=1}^{c_1}\sum_{j=1}^{c_2}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{XG})_{ij}S'_{ij}\frac{S_{ij}}{S'_{ij}}\geqslant\sum_{i=1}^{c_1}\sum_{j=1}^{c_2}(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{XG})_{ij}S'_{ij}\left(1+\log\frac{S_{ij}}{S'_{ij}}\right)\tag{49}$$

显然,$Z(\boldsymbol{S},\boldsymbol{S}')$由所有的边界组成,满足定义 1。因此,$Z(\boldsymbol{S},\boldsymbol{S}')$是 $J(\boldsymbol{S})$的辅助函数。

为了最小化 $Z(\boldsymbol{S},\boldsymbol{S}')$,本文令:

$$\frac{\partial Z(\boldsymbol{S},\boldsymbol{S}')}{\partial S_{ij}}=-\frac{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{XG})_{ij}S'_{ij}}{S_{ij}}+\frac{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{FS}'\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})_{ij}S_{ij}}{S'_{ij}}\tag{50}$$

并得到其 Hessian 矩阵：

$$\frac{\partial^2 Z(\boldsymbol{S},\boldsymbol{S}')}{\partial S_{ij}\partial S_{kl}}=\delta_{ik}\delta_{ji}\left(\frac{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{G})_{ij}S'_{ij}}{S_{ij}^2}+\frac{(\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}'\boldsymbol{G}^{\mathrm{T}}\boldsymbol{G})_{ij}}{S'_{ij}}\right) \tag{51}$$

当且仅当 $i=k$ 时，$\delta_{ik}=1$，否则 $\delta_{ik}=0$。此外，Hessian 矩阵是具有正值元素的对角矩阵，通过令 $\frac{\partial Z(\boldsymbol{S},\boldsymbol{S}')}{\partial S_{ij}}=0$，可以得到 $Z(\boldsymbol{S},\boldsymbol{S}')$ 是凸的，并且 $J(\boldsymbol{S})=\arg\min_{S_{ij}} Z(\boldsymbol{S},\boldsymbol{S}')$。

定理 3

$$J(\boldsymbol{G})=-Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})+\frac{1}{2}Tr(\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})-\beta Tr(\boldsymbol{P}\boldsymbol{G}^{\mathrm{T}})+\frac{\beta}{2}Tr(\boldsymbol{G}\boldsymbol{Q}\boldsymbol{G}^{\mathrm{T}}) \tag{52}$$

它的辅助函数为：

$$Z(\boldsymbol{G},\boldsymbol{G}')=-\sum_{i=1}^{n}\sum_{j=1}^{c_2}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S})_{ij}G'_{ij}\left(1+\log\frac{G_{ij}}{G'_{ij}}\right)+\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{c_2}\frac{(\boldsymbol{G}'\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S})_{ij}G_{ij}^2}{G'_{ij}}+\beta\sum_{i=1}^{n}\sum_{j=1}^{c_2}P_{ij}^{-}\frac{G_{ij}^2+G_{ij}'^2}{2G'_{ij}}-\beta\sum_{i=1}^{n}\sum_{j=1}^{c_2}P_{ij}^{+}G'_{ij}\left(1+\log\frac{G_{ij}}{G'_{ij}}\right)+\frac{\beta}{2}\sum_{i=1}^{n}\sum_{j=1}^{c_2}\frac{(\boldsymbol{G}'\boldsymbol{Q}^{+})_{ij}G_{ij}^2}{G'_{ij}}-\frac{\beta}{2}\sum_{i=1}^{n}\sum_{j=1}^{c_2}\sum_{k=1}^{c_2}Q_{jk}^{-}G'_{ij}G'_{ik}\left(1+\log\frac{G_{ij}G_{ik}}{G'_{ij}G'_{ik}}\right) \tag{53}$$

式(53)在 G 中是凸的，其全局极小值为：

$$G_{ij}=\arg\max_{G_{ij}}G(F,F')=G'_{ij}\left[\frac{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}+\beta\boldsymbol{P}^{+}+\beta\boldsymbol{G}\boldsymbol{Q}^{-})_{ij}}{(\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}+\beta\boldsymbol{P}^{-}+\beta\boldsymbol{G}\boldsymbol{Q}^{+})_{ij}}\right]^{\frac{1}{2}} \tag{54}$$

证明。首先，$J(\boldsymbol{G})$可以重写为：

$$J(\boldsymbol{G})=-Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})+\frac{1}{2}Tr(\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})-\beta Tr(\boldsymbol{P}^{+}\boldsymbol{G}^{\mathrm{T}})+\beta Tr(\boldsymbol{P}^{-}\boldsymbol{G}^{\mathrm{T}})+\frac{\beta}{2}Tr(\boldsymbol{G}\boldsymbol{Q}^{+}\boldsymbol{G}^{\mathrm{T}})-\frac{\beta}{2}Tr(\boldsymbol{G}\boldsymbol{Q}^{-}\boldsymbol{G}^{\mathrm{T}}) \tag{55}$$

和定理 1 的证明一样，可以得到如下不等式：

$$Tr(\boldsymbol{G}\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})\leqslant\sum_{i=1}^{n}\sum_{j=1}^{c_2}\frac{(\boldsymbol{G}'\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S})_{ij}G_{ij}^2}{G'_{ij}} \tag{56}$$

$$Tr(\boldsymbol{G}\boldsymbol{Q}^{+}\boldsymbol{G}^{\mathrm{T}})\leqslant\sum_{i=1}^{n}\sum_{j=1}^{c_2}\frac{(\boldsymbol{G}\boldsymbol{Q}^{+})_{ij}G_{ij}^2}{G'_{ij}} \tag{57}$$

$$Tr(\boldsymbol{P}^{-}\boldsymbol{G}^{\mathrm{T}})\leqslant\sum_{i=1}^{n}\sum_{j=1}^{c_2}P_{ij}^{-}\frac{G_{ij}^2+G_{ij}'^2}{2G'_{ij}} \tag{58}$$

$$Tr(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^{\mathrm{T}})\geqslant\sum_{i=1}^{n}\sum_{j=1}^{c_2}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S})_{ij}G'_{ij}\left(1+\log\frac{G_{ij}}{G'_{ij}}\right) \tag{59}$$

$$Tr(\boldsymbol{P}^{+}\boldsymbol{G}^{\mathrm{T}})\geqslant\sum_{i=1}^{n}\sum_{j=1}^{c_2}P_{ij}^{+}G'_{ij}\left(1+\log\frac{G_{ij}}{G'_{ij}}\right) \tag{60}$$

$$Tr(\boldsymbol{G}\boldsymbol{Q}^{-}\boldsymbol{G}^{\mathrm{T}})\geqslant\sum_{i=1}^{n}\sum_{j=1}^{c_2}\sum_{k=1}^{c_2}Q_{jk}^{-}G'_{ij}G'_{ik}\left(1+\log\frac{G_{ij}G_{ik}}{G'_{ij}G'_{ik}}\right) \tag{61}$$

显然，$\partial Z(\boldsymbol{G},\boldsymbol{G}')$由所有的边界组成，满足定义 1。因此，$\partial Z(\boldsymbol{G},\boldsymbol{G}')$是 $J(\boldsymbol{G})$的辅助函数。此外，其一阶和二阶导数如下：

$$\frac{\partial Z(\boldsymbol{G},\boldsymbol{G}')}{\partial G_{ij}}=-\frac{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S})_{ij}G'_{ij}}{G_{ij}}+\frac{(\boldsymbol{G}'\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S})_{ij}G_{ij}}{G'_{ij}}-\beta\frac{P_{ij}^{+}G'_{ij}}{G_{ij}}+\beta\frac{P_{ij}^{-}G_{ij}}{G'_{ij}}+\beta\frac{(\boldsymbol{G}'\boldsymbol{Q}^{+})_{ij}G_{ij}}{G'_{ij}}-\beta\frac{(\boldsymbol{G}'\boldsymbol{Q}^{-})_{ij}G'_{ij}}{G_{ij}} \tag{62}$$

$$\frac{\partial^2 Z(\boldsymbol{G},\boldsymbol{G}')}{\partial G_{ij}\partial G_{kl}}=\delta_{ik}\delta_{ji}\left(\frac{(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S})_{ij}G'_{ij}}{G_{ij}}+\frac{(\boldsymbol{G}'\boldsymbol{S}^{\mathrm{T}}\boldsymbol{F}^{\mathrm{T}}\boldsymbol{F}\boldsymbol{S})_{ij}}{G'_{ij}}+\beta\frac{P_{ij}^{+}G'_{ij}}{G_{ik}^2}+\beta\frac{P_{ij}^{-}}{G'_{ij}}+\beta\frac{(\boldsymbol{G}'\boldsymbol{Q}^{+})_{ij}}{G'_{ij}}+\beta\frac{(\boldsymbol{G}'\boldsymbol{Q}^{-})_{ij}G'_{ij}}{G_{ij}^2}\right) \tag{63}$$

说明 $Z(\boldsymbol{G},\boldsymbol{G}')$是凸函数，令$\frac{\partial Z(\boldsymbol{G},\boldsymbol{G}')}{\partial G_{ij}}=0$，可以得到 $\arg\min_{G_{ij}}J(\boldsymbol{G})=\arg\min_{G_{ij}}Z(\boldsymbol{G},\boldsymbol{G}')$。

1.4 计算复杂度

对于一个数据集，其特征关联矩阵和样本关联矩阵的复杂度分别为 $O(d^2n)$和 $O(n^2d)$，其中 d 和 n 分别是特征数和样本数。此外，乘法变换过程的复杂度为 $O(dnct)$，其中 c 和 t 分别是聚类数和迭代数。总体而言，该算法的计算量为 $O(d^2n+n^2d+dnct)$。

2 实验与结果分析

2.1 评估指标

所采用的评价指标包括聚类精度、归一化互信息和调整后的 Rand 指数。聚类精度直接反映了数据点的错误分类，而归一化互信息和调整 Rand 指数则分别从组合数学和概率论的角度来判断属于同一类别的数据对的正确性。本质上，它们衡量的是实际分布和预测分布之间的差异。这些指标越高，表示性能就越好。

给定一个数据集 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$，设 $\boldsymbol{S}=\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\mid c_i=c_j,\tilde{c}_i=\tilde{c}_j,i<j\}$，$D=\{(\boldsymbol{x}_i,\boldsymbol{x}_j)\mid c_i\neq c_j,\tilde{c}_i\neq\tilde{c}_j,i<j\}$，其中 c_i和 $\tilde{c}_i$代表第 i 个样本的标签和预测。另外，用 $C=\{C_i\}_{i=1}^{k}$和 $\tilde{C}=\{\tilde{C}_i\}_{j=1}^{m}$分别表示实际聚类结果和预测聚类结果。互信息、信息熵和 Rand 指数的定义式如下：

$$I(C,\tilde{C})=\sum_{i=1}^{k}\sum_{j=1}^{m}|C_i\cap\tilde{C}_j|\log\frac{n|C_i\cap\tilde{C}_j|}{|C_i|\cdot|\tilde{C}_j|} \tag{64}$$

$$H(C)=\sum_{i=1}^{k}|C_i|\log\frac{n|C_i|}{n} \tag{65}$$

$$RI=\frac{2(|\boldsymbol{S}|+|\boldsymbol{D}|)}{n(n-1)} \tag{66}$$

此外，还计算了聚类精度、归一化互信息和调整后的 Rand 指数：

$$ACC = \frac{\sum_{i=1}^{n}\delta(c_i, map(\tilde{c}_i))}{n} \tag{67}$$

$$NMI = \frac{I(C,\tilde{C})}{\sqrt{H(C)H(\tilde{C})}} \tag{68}$$

$$ARI = \frac{RI = E(RI)}{\max(RI) - E(RI)} \tag{69}$$

式中:$map(\cdot)$和$\delta(\cdot)$分别表示置换映射函数和 Kronecker delta 函数。此外,利用 Kuhn-Munkres 算法可以得到最佳匹配。

2.2　数据集

为了全面验证算法,从基因表达、文本文档、数字图像和面部图像等多种数据类型中选取了 6 个公开可用的数据集,所采用数据集的基本信息见表 1。

表 1　数据集介绍

数据集	数据类型	实例个数	特征个数	类别
LUNG	基因表达	203	3 312	5
CLLSUB111	基因表达	111	11 340	3
BASEHOCK	文本文档	1 993	4 862	2
RELATHE	文本文档	1 427	4 322	2
MNISTSUB	数字图像	6 996	784	10
MSRA25	面部图像	1 799	256	12

2.3　实验设置

本节综合分析比较了一些经典的或最新的聚类方法,如:K-means,K 均值聚类算法作为一种典型的聚类算法得到了广泛的应用[8];NMF,非负矩阵分解算法也是一种常用算法[13];SNMF,结合奇异值分解(SVD)的对称非负矩阵分解方法[14];GNMF,图正则化的非负矩阵分解方法[15];DNMTF,双图正则非负矩阵三分解方法[16];DLLC,基于双图正则化的多尺度聚类[17];PNMF,投影非负矩阵分解方法[18];FNMTF,是一种快速的非负矩阵三因子分解算法,用于联合聚类[19];BKM,采用双边 K 均值算法进行快速联合聚类[20];SOBG,学习了一个结构最优二部图进行联合聚类[21]。具体介绍如下:

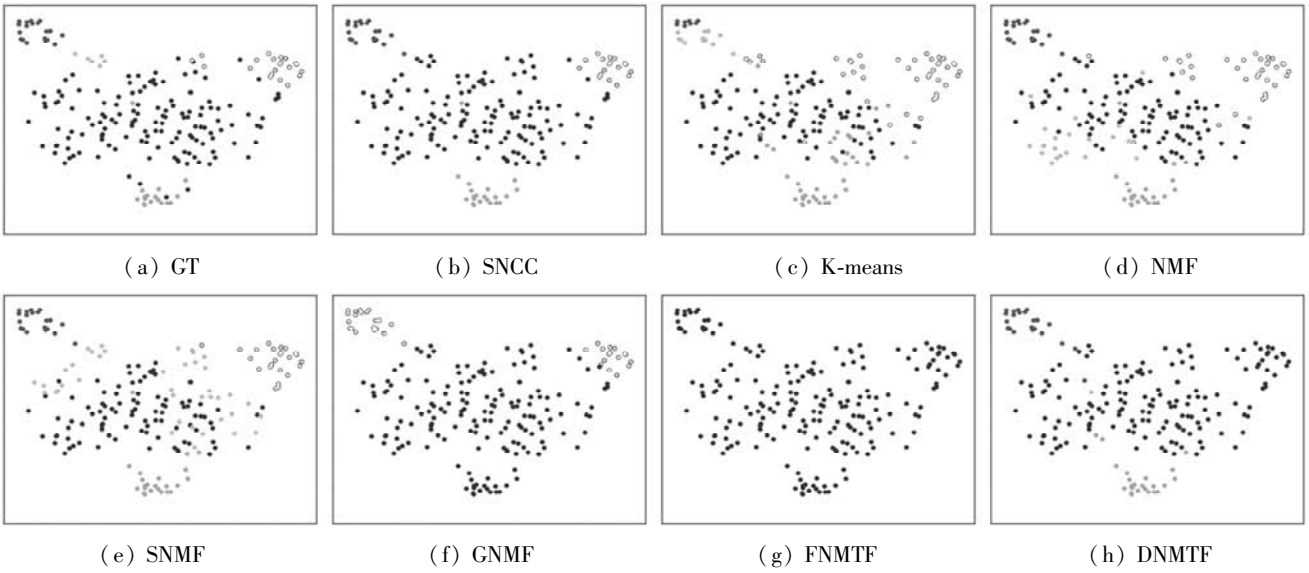
将 K-means 的参数设为默认值,对于 GNMF,其正则化参数 α 设置为 100,对于 DNMTF,其正则化参数 λ 和 μ 均设为 200,对于 SOBG,其参数 λ 设为 10,对于 DLLC,其正则化参数 α 和 β 均设为 1,PNMF 的参数 α、β 和 γ 设为 1,SNCC 参数 α 和 β 设为 0.1,对于基于 NMF 的方法,都是用 K-means 初始化的,迭代次数最多可达 20 次。

对于所有算法,它们的样本聚类数与实际类别数是一致的。对于联合聚类,由于特征聚类数目未知,将特征聚类数目设置为与样本聚类数目相同。对于图的构造,采用 k-近邻算法,权值为 0 或 1。此外,将特征和样本的最近邻数设为 10。考虑到某些方法的初始化敏感性,将所有实验重复进行 30 次,比较其均值和标准差。

所有的数据集都是在一台个人计算机上进行处理的,配置为 3.7 GHz 的 i7 中央处理器和 64 GB 的随机存取存储器。算法运行平台为 MATLAB 2016a,所有算法的程序均通过 MATLAB 脚本编写运行。

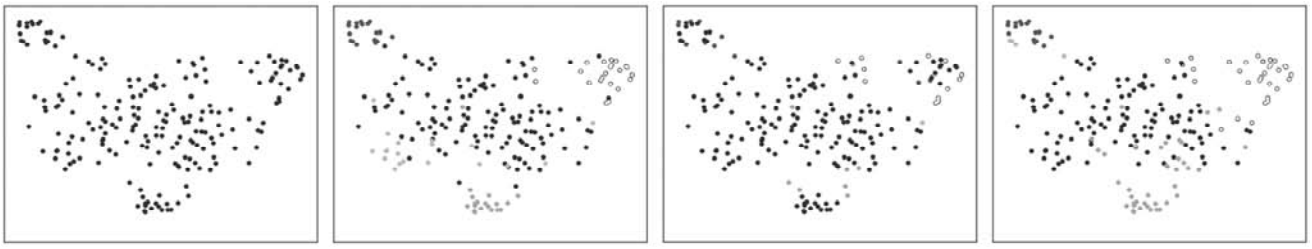
2.4　结果分析

LUNG 数据集的聚类结果如图 1 所示,可以看出,真实值(GT)包含不平衡的样本和难以分离的聚类。对于该数据集,即使某些方法只将所有样本分组到一个类别中,它们仍然比其他方法(如 FNMTF 和 BKM)的性能逊色。显然,也不希望将所有样本划分为两个或三个类,例如 GNMF 和 DNMTF。总的来说,SNCC 效果更好,因为它的结果更接近真实分布。



(a) GT　(b) SNCC　(c) K-means　(d) NMF

(e) SNMF　(f) GNMF　(g) FNMTF　(h) DNMTF



(i) BKM (j) SOBG (k) DLLC (l) PNMF

图 1 LUNG 数据集的聚类结果

如表 2 所示,SNCC 比其他算法获得了更高的聚类精度,相比之下,有些方法在 MNIST 和 MSRA25 数据集上聚类精度相对较低,如 BKM 和 SOBG。如表 3 所示,除 RELATHE 外,SNCC 在测试数据集上的表现优于其他数据集,但以归一化互信息度量,SNCC 仅次于 DNMTF。从表 4 可以看出,SNCC 在调整后的 Rand 指数中的性能仍较好。运行时间如图 2 所示。显然,SOBG 是耗时的,而 SNCC 的计算成本与 K-means 相差不大。面对图像数据集,大多数方法的时间消耗波动较大,而 SNCC 相对稳定。总体来看,SNCC 在聚类精度,时间成本上综合性能相对而言最佳。

表 2 多个数据集的结果准确度(ACC%)的比较

数据集	LUNG	CLL_SUB_111	BASE-HOCK	RELA-THE	MNIST_SUB	MSRA25
SNCC	84.5	54.7	69.0	57.3	53.9	52.6
K-means	69.3	44.1	62.8	56.2	52.4	47.3
NMF	74.6	45.8	64.4	56.4	50.5	47.2
SNMF	73.6	45.9	64.3	56.7	38.8	47.9
GNMF	75.8	47.8	51.3	54.9	23.5	39.8
FNMTF	68.1	46.0	53.7	52.1	41.8	18.3
DNMTF	79.5	50.9	65.2	56.9	37.1	46.0
BKM	68.5	46.0	52.0	56.7	11.3	10.3
SOBG	77.8	51.4	50.1	54.8	11.4	18.0
DLLC	56.3	48.3	60.0	55.2	47.5	41.5
PNMF	76.6	43.8	62.4	56.3	51.8	48.2

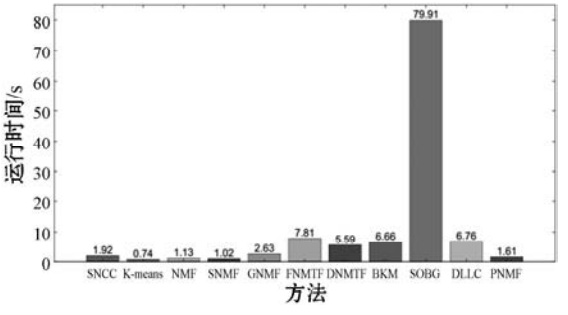
表 3 多个数据集的标准化互信息(NMI%)的比较

数据集	LUNG	CLL_SUB_111	BASE-HOCK	RELA-THE	MNIST_SUB	MSRA25
SNCC	59.8	21.9	11.4	1.7	52.8	60.4
K-means	52.0	10.4	5.2	1.2	51.1	54.4
NMF	55.5	14.2	6.2	1.1	46.2	53.8
SNMF	52.8	12.9	6.2	1.1	30.2	55.3
GNMF	35.6	10.2	0.6	0.4	15.5	44.1
FNMTF	8.0	5.7	0.4	0.2	36.3	14.7
DNMTF	56.8	17.9	10.1	1.9	31.5	55.0
BKM	4.8	2.9	0.4	0.7	0.1	0.6
SOBG	28.9	16.3	0.3	0.2	0.2	11.3
DLLC	23.4	14.0	3.1	0.8	42.4	44.8
PNMF	55.2	10.6	5.0	1.1	49.5	56.0
SNCC	59.8	21.9	11.4	1.7	52.8	60.4

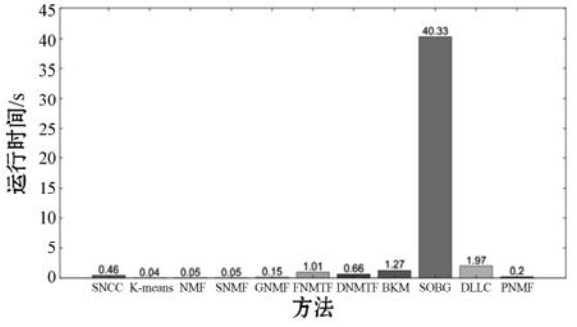
续表 3

表 4 多个数据集算法比较的调整后随机指数(ARI%)

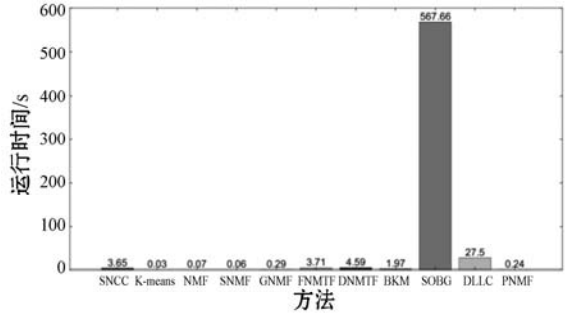
数据集	LUNG	CLL_SUB_111	BASE-HOCK	RELA-THE	MNIST_SUB	MSRA25
SNCC	63.0	10.9	14.6	2.2	39.5	33.4
K-means	41.8	2.4	6.8	1.6	37.7	31.0
NMF	48.7	4.7	8.3	1.6	34.0	30.2
SNMF	46.1	4.0	8.2	0.2	20.0	32.6
GNMF	38.6	5.1	0.4	0.2	6.8	22.8
FNMTF	3.9	-0.9	0.5	0.2	24.8	6.6
DNMTF	59.1	7.9	11.1	2.5	17.7	32.2
BKM	0.0	0.0	0.1	1.3	0.0	0.0
SOBG	33.4	16.0	0.0	0.1	0.0	1.3
DLLC	18.4	5.7	4.0	1.0	30.4	23.7
PNMF	50.5	2.5	6.4	1.6	36.6	33.2



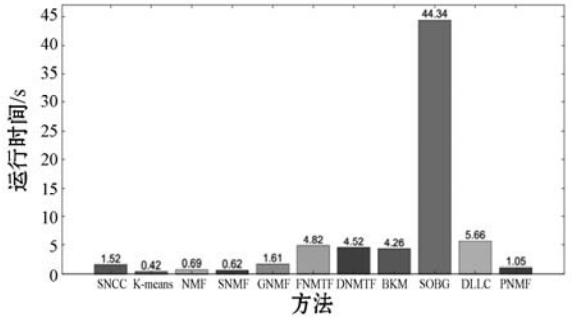
(a) BASEHOCK



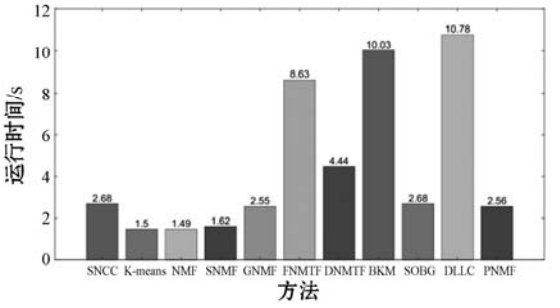
(b) LUNG



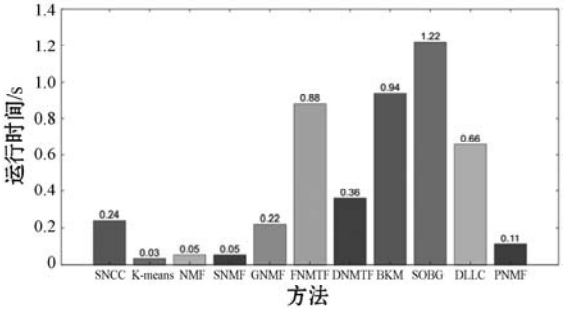
(c) CLL_SUB_111



(d) RELATHE



(e) MNIST_SUB

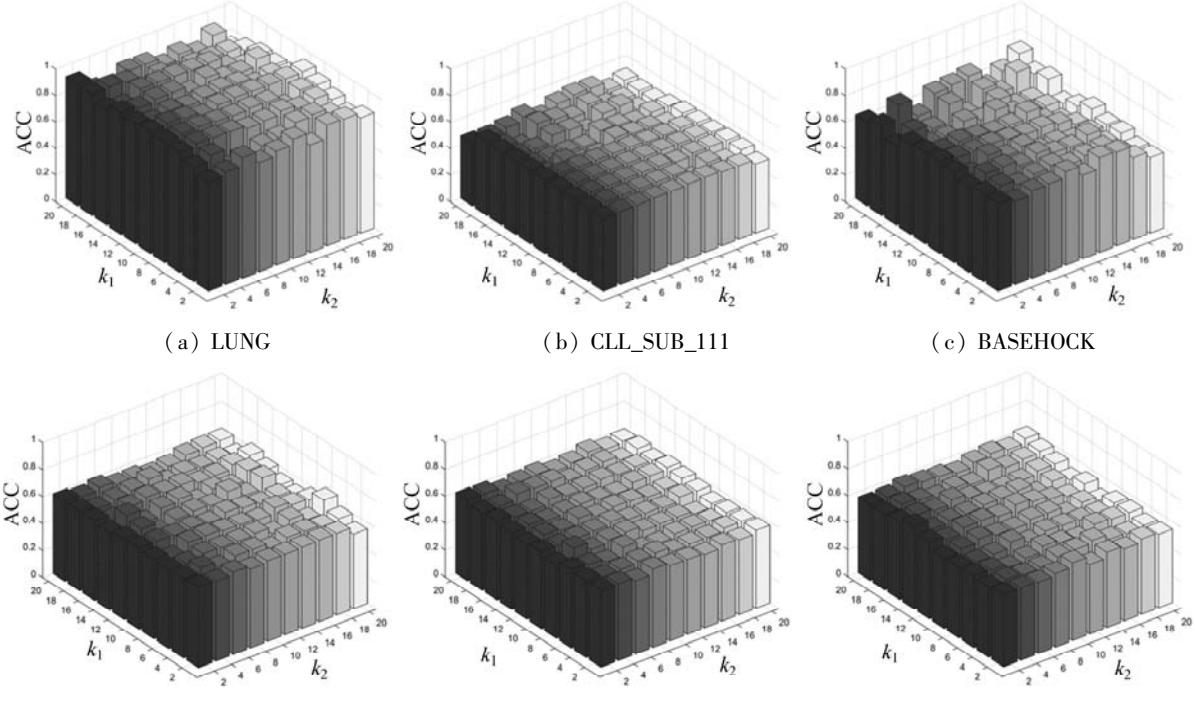


(f) MSRA25

图 2　不同数据集运行时间对比

2.5　参数灵敏度

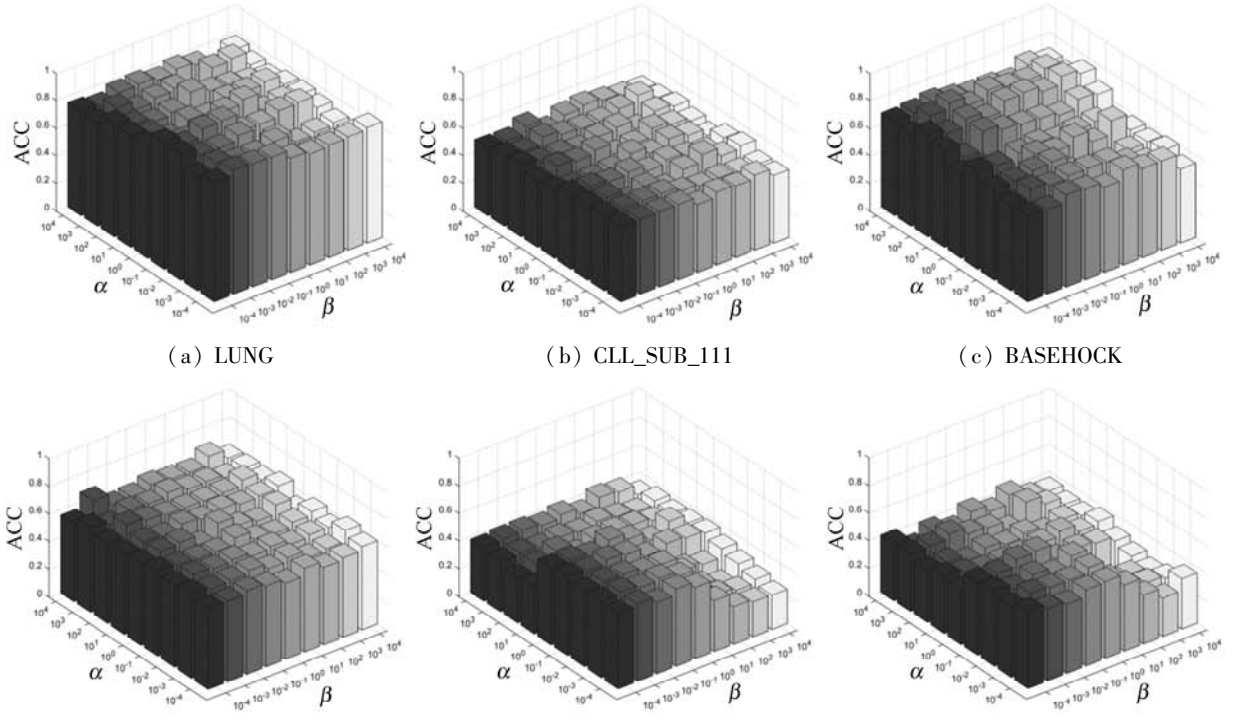
本节分析了算法的性能随参数的变化,图 3 和图 4 中说明了最近邻数和正则化的权重,并通过聚类精度测量。如果 k_1 略低于 k_2,性能会更好,另外权重 α 小于 β 也会产生更高的聚类精度。可以看出性能随着参数选择而波动,但是波动的范围均较小,说明提出算法对于参数的灵敏度较弱,进一步减弱了参数选择对性能的影响,提升了应用范围。



(a) LUNG　(b) CLL_SUB_111　(c) BASEHOCK

(d) RELATHE　(e) MNIST_SUB　(f) MSRA25

图 3　最近邻数的权重



(a) LUNG (b) CLL_SUB_111 (c) BASEHOCK

(d) RELATHE (e) MNIST_SUB (f) MSRA25

图 4 正则化参数的权重

2.6 相似性分析

本节讨论如何构造相似图,如表 5 所示,稀疏图的性能和密集图的性能类似,有时甚至更好,而高斯核和余弦相似性在图像和文本中取得了较好的结果,二进制加权作为一种常用的相似性度量,对各种数据的相似性均较好[17]。从表 5 中可知,相对于余弦和高斯相似性来说,二进制加权相似性在大部分数据集上的性能更好,计算量更少。

表 5 基于不同相似性和数据集下的性能

数据集	评价指标	二进制加权	余弦		高斯	
			密集	稀疏	密集	稀疏
LUNG	ACC	92.1	88.2	88.7	91.1	83.7
	NMI	72.3	64.1	68.3	73.4	58.7
	ARI	81.7	69.8	72.1	78.4	60.9
CLLSUB111	ACC	55.0	45.9	55.0	55.0	54.1
	NMI	26.3	12.9	23.9	23.9	21.5
	ARI	12.4	5.8	12.0	12.0	10.5
BASEHOCK	ACC	67.3	62.5	69.5	67.4	66.3
	NMI	11.3	4.8	11.4	9.1	8.3
	ARI	12.1	6.2	15.2	12.1	10.6
RELATHE	ACC	60.8	60.1	59.4	60.6	58.2
	NMI	3.6	2.9	2.6	3.1	2.0
	ARI	4.2	3.9	3.5	4.4	2.6
MNISTSUB	ACC	56.5	51.4	23.4	56.1	58.6
	NMI	53.6	48.3	12.7	53.5	54.8
	ARI	40.5	35.2	6.2	40.4	43.4
MSRA25	ACC	60.6	53.4	52.0	52.6	56.7
	NMI	64.9	59.9	62.5	60.0	64.2
	ARI	41.9	38.5	39.6	38.5	43.9

3 结 语

为了充分挖掘特征结构,提升聚类性能,提出了一种基于类别一致性学习的稀疏邻域约束的联合聚类方法。利用三种评价方法在六个数据集上进行了验证,分析结果可以得出如下结论:

(1) 提出方法由于保持了数据关联性和标签分配之间的一致性,在聚类精度,时间成本上相比较于其他方法具有优势。

(2) 提出算法对于参数的灵敏度较弱,进一步减弱了参数选择对性能的影响,提升了该方法的使用性能。

(3) 相对于余弦和高斯相似性来说,二进制加权相似性在大部分数据集上的性能更好,计算量更少。

参 考 文 献

[1] 李顺勇,张钰嘉,彭晓庆,等. 一种基于分层抽样的大数据快速聚类算法[J]. 计算机应用与软件,2020,37(10):256-261,277.

[2] 任昌鸿,安军. 改进 PSO 结合 DSA 技术的无线传感器网络均衡密度聚类方法[J]. 计算机应用与软件,2020,37(8):122-129.

[3] 彭春春,陈燕俐,荀艳梅. 支持本地化差分隐私保护的 K-modes 聚类方法[J]. 计算机科学,2021,48(2):105-113.

[4] 白璐,赵鑫,孔钰婷,等. 谱聚类算法研究综述[J]. 计算机工程与应用,2021,57(14):15-26.

[5] 张煜,陆亿红,黄德才. 基于密度峰值的加权犹豫模糊聚类算法[J]. 计算机科学,2021,48(1):145-151.

[6] 陈湘中,万烂军,李泓洋,等. 基于蚁群优化 K 均值聚类算法的滚轴故障预测[J]. 计算机工程与设计,2020,41(11):3218-3223.

[7] Park J H, Kang Y J. Evaluation index for sporty engine sound reflecting evaluators' tastes, developed using k-means cluster analysis[J]. International Journal of Automotive Technology,2020,21(6):1379-1389.

[8] Anderson N P, Gillman A, Yin Q, et al. Poster abstract 216: Identifying phenotypic subpopulations of chronic pain patients using K-means cluster analysis of body map data [J]. Pain Medicine,2020,21(6):1307-1310.

[9] Zhang B. Regional enterprise economic development dimensions based on K-means cluster analysis and nearest neighbor discriminant[J]. Journal of Intelligent and Fuzzy Systems, 2020,38(6):7365-7375.

[10] 马欣野,刘亚静,刘童. 基于欧式加权法的模糊 C 均值聚类算法[J]. 南方农机,2021,52(14):151-153.

[11] 王继奎,杨正国,易纪海,等. 稀疏约束的嵌入式模糊均值聚类算法[J]. 复旦学报(自然科学版),2020,59(6):725-733.

[12] 余炳光,刘冬梅. 特征逐减的可能性模糊聚类算法[J]. 计算机工程与应用,2019,55(19):58-65.

[13] 李向利,贾梦雪. 基于预处理的超图非负矩阵分解算法[J]. 计算机科学,2020,47(7):71-77.

[14] 王丽星. 基于 Huber 损失的非负矩阵分解算法在聚类中的研究[D]. 太原:山西大学,2020.

[15] Zhang C Q, Tao D C, Dong X. Generalized latent multi-view subspace clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,42(1):86-99.

[16] Jiao P F, Yu W, Wang W J. Exploring temporal community structure and constant evolutionary pattern hiding in dynamic networks[J]. Neurocomputing,2018,314(6):224-233.

[17] Mohanmmad K A, Nouman A, Yao J T. Variance based three-way clustering approaches for handling overlapping clustering[J]. International Journal of Approximate Reasoning,2020,118(5):47-63.

[18] Lv X C, Wang W H, Liu H F. Cluster-wise weighted NMF for hyperspectral images unmixing with imbalanced data[J]. Remote Sensing,2021,13(2):268.

[19] Venkatasubramanian M, Chetal K, Schnel D J, et al. Resolving single-cell heterogeneity from hundreds of thousands of cells through sequential hybrid clustering and NMF[J]. Bioinformatics,2020,36(12):3773-3780.

[20] Chen M L, Li X L. Robust matrix factorization with spectral embedding[J]. IEEE Transactions on Neural Networks and Learning Systems,2020,59(6):5698-5707.

[21] Wang C Y, Gao Y L, Kong X Z, et al. Unsupervised cluster analysis and gene marker extraction of scRNA-seq data based on non-negative matrix factorization[J]. IEEE Journal of Biomedical and Health Informatics,2021,26(8):458-467.

(上接第 280 页)

参 考 文 献

[1] 蔡晓晴,邓尧,张亮,等. 区块链原理及其核心技术[J]. 计算机学报,2021,44(1):84-131.

[2] 曾诗钦,霍如,黄韬,等. 区块链技术研究综述:原理、进展与应用[J]. 通信学报,2020,41(1):134-151.

[3] 张亮,刘百祥,张如意,等. 区块链技术综述[J]. 计算机工程,2019,45(5):1-12.

[4] 王君宇,吴清烈,曹卉宇. 国内区块链典型应用研究综述[J]. 科技与经济,2019,32(5):1-6.

[5] 于戈,聂铁铮,李晓华,等. 区块链系统中的分布式数据管理技术——挑战与展望[J]. 计算机学报,2021,44(1):28-53.

[6] 贺海武,延安,陈泽华. 基于区块链的智能合约技术与应用综述[J]. 计算机研究与发展,2018,55(11):2452-2466.

[7] Karger D, Lehman E, Leighton T, et al. Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the world wide web[C]//29th Annual ACM Symposium on the Theory of Computing,1997.

[8] A blockchain platform for the enterprise[EB/OL]. [2021-04-19]. https://hyperledger-fabric.readthedocs.io/en/release-2.3/index.html.

[9] FISCO BCOS 技术文档[EB/OL]. [2021-04-19]. https://fisco-bcos-documentation.readthedocs.io/zh_CN/latest/index.html.

[10] Corda release notes[EB/OL]. [2021-04-19]. https://docs.corda.net/docs/corda-os/4.7/release-notes.

[11] Benet J. IPFS-Content addressed, versioned, P2P file system[EB]. arXiv:1407.3561,2014.