

文章编号: 2096-1618(2023)05-0526-09

基于多视角自适应图正则的非负矩阵分解聚类

林虹燕¹, 杜元花¹, 周楠², 田永强³

(1. 成都信息工程大学应用数学学院, 四川 成都 610225; 2. 成都大学电子信息与电气工程学院, 四川 成都 610106; 3. 华为技术有限公司, 云南 昆明 650011)

摘要:为充分利用各个视角数据内在几何结构关系,提出一种新的基于自适应图正则非负矩阵分解的多视角聚类。该算法在一个统一的框架内,通过各视角亲和矩阵自适应学习提取共识的亲和矩阵进行图嵌入来提取多视角数据共识局部结构信息。另外,通过非负矩阵分解来提取多视角数据全局重构信息。最终使各个视角的共识表达,既保持了数据多视角共识全局重构信息,也保持了数据多视角局部结构信息。该优化问题在考虑了所有数据的一致性和每个视角之间互补性的同时,引入了各个视角数据的局部结构信息,达到数据表达和聚类的效果。通过4组真实数据集的实验,结果表明所提出的方法与已有多视角聚类方法相比具有一定的优越性。

关键词:多视角学习;拉普拉斯秩约束;图嵌入;非负矩阵分解;聚类

中图分类号:TP181

文献标志码:A

doi:10.16836/j.cnki.jcuit.2023.05.006

0 引言

多视角学习在近年成为机器学习的热门话题之一,在图像处理、卫星通信、模式识别等领域有广泛应用。多视角学习能够整合多视角数据的互补性和异构性,并将其重构为一个共识表达,在越来越多的领域受到欢迎。

随着多视角数据在实际应用中的日益增多,提出了多种多视角聚类方法^[1-4]。Nie等^[5]通过探索拉普拉斯秩约束图来构建多视角聚类,可近似地为每个具有不同置信度的视角构建共识亲和图。Zhan等^[6]提出一种多视图一致性聚类方法,通过学习一致性图,直接获得聚类标签,并且无需进行任何后处理。Wen等^[7]提出一种简单有效的不完备多视角聚类框架,该框架同时考虑了这些不完备多视角观测的局部几何信息并具备不平衡的判别能力。Luo等^[8]提出一种多视角双聚类方法,该方法在统一的框架中同时探索共识表达和双聚类结构。Liu等^[9]提出一种基于潜在空间近邻学习的新型多视角聚类方法。其中,多视角非负矩阵分解通过对多个视角数据进行矩阵分解达到原数据矩阵的低秩逼近,从而可以发现每个视角的内在结构表达成一个基,并将互补信息融合成一个低维空间的共识表达。上述大多数方法只重视原始空间数据的一致性,忽略了各个视角的独特性,得不到满意的数据表达。同时,低维空间的数据表达跟聚类结果也

有很强的联系,共识表达可以展现数据的内在结构,另一个矩阵则为低维空间的系数表达矩阵。因此,在数据表达的同时可以对数据进行聚类。最后,通过对拉普拉斯图进行秩约(constraint Laplacian rank, CLR)^[10]来构造多视角图,该多视角图能很好地学习到数据内在结构。受到该思想的启发,本文提出一种多视角自适应图非负矩阵分解(multi-view adaptive graph non-negative matrix factorization, MAGNMF)的聚类方法。该方将多个视角共识图嵌入框架^[11]与非负矩阵分解特征学习相结合,不仅考虑了多视角数据的全局重构信息,也考虑了多视角的局部结构信息,以此达到数据表达和聚类的效果。本文主要贡献如下:

(1)提出了一种结合非负矩阵分解、CLR和图嵌入的多视角特征学习聚类方法,使低维表达既能保持高维数据的内在局部结构信息,也能提取出多个视角的全局重构信息;

(2)提出了一个基于快速块坐标更新(block coordinate update, BCU)^[12]迭代更新算法来求解所提出的优化模型;

(3)通过在4个真实数据集上进行大量验证,结果表明本文所提出的多视角数据表达聚类方法MAGNMF在多数情况下优于现有的方法。

1 相关工作

1.1 符号

本文用到的符号如表1所示。

收稿日期:2022-08-31

基金项目:国家自然科学基金资助项目(11901063);四川省自然科学基金资助项目(23NSFSC2919)

表 1 符号

符号	说明
n	样本数
c	聚类数
m	视角数
$X = [X^{(1)}, X^{(2)}, \dots, X^{(m)}]$	数据集
S	共识相似度矩阵
$A^{(v)}$	第 v 个视角的亲矩阵
$U^{(v)}$	第 v 个视角的基矩阵
F	共识低维表达矩阵
L	拉普拉斯矩阵
D	度矩阵
$\text{Tr}(\cdot)$	矩阵的迹
$\ \cdot\ _F$	矩阵 Frobenius 范数

1.2 拉普拉斯秩约束算法

现有的多视角聚类方法多基于多图聚类,如协同训练^[13],多视角谱聚类^[14]等。这些方法都是在固定输入的数据图上进行聚类,如果输入的图质量较差,则聚类结果也较差,其次需要进行后处理才能完成聚类。拉普拉斯秩约束定理通过限制拉普拉斯矩阵秩的个数($n-k$)来约束数据的相似图的连通分量个数(k),给定亲和矩阵 A ,寻找相似矩阵 S 。通过数据的原始亲和矩阵 A ,可以学习到相似矩阵 S 以及相应的拉普拉斯矩阵 $L = D - (S^T + S)/2$,它的秩为 $n-k$,其中对角矩阵 D 的第 i 个元素为 $\sum_j (s_{ij} + s_{ji})/2$ 。在这个限制之下,学习到的 S 是有合适排列的块对角矩阵,这样就可以直接将数据点划分到 k 簇。其模型表示为

$$\min_{\sum_j s_{ij} = 1, s_{ij} \geq 0, \text{rank}(L_S) = n-k} \|S - A\|_F^2 \quad (1)$$

其中 $A_{i,j} = \begin{cases} \exp(-\|X_i - X_j\|_2^2/t), & \|X_i - X_j\|_2^2 \leq \varepsilon; \\ 0, & \text{otherwise.} \end{cases}$, 限制 S 的每一行之和为 1,以防每个元素为 0。

1.3 图正则化的非负矩阵分解算法

现实中的数据基本都是高维的,非负矩阵分解(nonnegative matrix factorization, NMF)^[15]是将高维空间的数据矩阵近似表达成两个低维矩阵, U 是基矩阵, F 是低维表达矩阵。 X 是数据矩阵,其中 X 每一列表示一个样本点,每一行表示一个特征。

$$X \approx UF^T \quad (2)$$

为了寻找一个高质量的近似,可以将 NMF 的目标函数变成以下优化问题:

$$\min_{U, F \geq 0} \|X - UF^T\|_F^2 \quad (3)$$

这样 NMF 就可以学到局部表达。为更好地学习数

据空间的内部几何和判别结构,同时维持高维数据在低维空间的流型结构,避免过拟合,加入了图正则项。于是,图正则化的非负矩阵分解^[16]模型表示如下:

$$\min_{U, F \geq 0} \|X - UF^T\|_F^2 + \lambda \text{Tr}(F^T L F) \quad (4)$$

其中 $U \geq 0, F \geq 0$ 表示矩阵 U 与 F 中所有元素都是非负的,正则项参数 $\lambda \geq 0$,用于控制低维表达对于高维数据局部结构的保持性, $L = D - (S^T + S)/2$ 为图拉普拉斯矩阵,其中 S 为数据矩阵的相似矩阵。将图正则项与原始非负矩阵分解结合被称为图正则化非负矩阵分解。

2 自适应图正则的非负矩阵分解

本文提出一种新的多视角自适应图非负矩阵分解模型。根据拉普拉斯秩约束需要找到共识的拉普拉斯约束图,该图可以近似地作为每个视角的相似度图的质心。图非负矩阵分解将数据从嵌入的高维环境空间中的低维流形中采样,得到在低维空间的表达。本文模型还加入了惩罚项以防止平凡解的出现。

给定一个多视角数据集 $X = [X^{(1)}, X^{(2)}, \dots, X^{(m)}]$,本文模型目标是找到共识相似矩阵 S ,共识低维表达矩阵 F 和每个视角的基矩阵 $U^{(v)}$ 。因此,目标函数被表示成:

$$\min_{S, F, U} \frac{\lambda}{2} \sum_{v=1}^m \|S - A^{(v)}\|_F^2 + \frac{1}{2} \sum_{v=1}^m \|X^{(v)} - U^{(v)} F^T\|_F^2 +$$

$$\frac{\lambda}{2} \text{Tr}(F^T L_S F) + \sum_{v=1}^m \|U^{(v)}\|_F^2$$

$$\text{s. t. } S \geq 0, s_i 1_n = 1, F \geq 0, f_i 1_n = 1, U \geq 0 \quad (5)$$

这里, $\lambda > 0$ 是 CLR 和图正则项的平衡系数。限制 $S, F, U^{(v)}$ 为非负矩阵,且 S 与 F 的每一行和为 1,有效地避免了平凡解和减小计算复杂度。最后,将 K-means 算法^[17]应用到共低维表达矩阵,得到聚类结果。

3 求解算法

3.1 加速 BCU 迭代方法

本文采用加速 BCU 策略来推导模型(5)的求解算法。在每次迭代中,更新目标函数的一个变量,将其他两个变量固定在其最近更新的值。具体来说,每个变量通过迭代地执行以下的表达式来进行更新。

$$S^{k+1} = \arg \min_{S \geq 0, s_i 1_n = 1} (\nabla_S f(S^k, F^k, (U^{(v)})^k), (S - S^k) - \frac{L_S^k}{2} \|S - S^k\|_F^2) \quad (6)$$

$$\mathbf{F}^{k+1} = \arg \min_{\mathbf{F} \geq 0, \sum_{j=1}^n f_j = 1} (\nabla_S f(\mathbf{S}^{k+1}, \hat{\mathbf{F}}^k, (\mathbf{U}^{(v)})^k), \mathbf{F} - \hat{\mathbf{F}}^k) - \frac{\mathbf{L}_F^k}{2} \|\mathbf{F} - \hat{\mathbf{F}}^k\|_F^2 \quad (7)$$

$$(\mathbf{U}^{(v)})^{k+1} = \arg \min_{\mathbf{U}^{(v)} \geq 0} f(\mathbf{S}^{k+1}, \mathbf{F}^{k+1}, (\mathbf{U}^{(v)})^k) \quad (8)$$

其中 \mathbf{L}_S^k 和 \mathbf{L}_F^k 分别是 $\nabla_S f(\cdot)$ 和 $\nabla_F f(\cdot)$ 的利普希茨 (Lipchitz) 常数, 并且插值点

$$\hat{\mathbf{S}}^k = \mathbf{S}^k + \omega_S^k (\mathbf{S}^k - \mathbf{S}^{k-1}) \quad (9)$$

$$\hat{\mathbf{F}}^k = \mathbf{F}^k + \omega_F^k (\mathbf{F}^k - \mathbf{F}^{k-1}) \quad (10)$$

这里 $\omega_S^k, \omega_F^k \in [0, 1]$ 分别为插值点 $\hat{\mathbf{S}}^k$ 和 $\hat{\mathbf{F}}^k$ 的权重, 文献 [18-19] 认为, 通过该插值的办法可以明显提高 BCU 的求解效率。

3.2 更新 \mathbf{S}

令 $\mathbf{C} = \nabla_S f(\hat{\mathbf{S}}^k, \hat{\mathbf{F}}^k, (\mathbf{U}^{(v)})^k)$ 且 $\mathbf{L}_1 = \mathbf{L}_S^k$, 于是式(6)可以表示成

$$\begin{aligned} & \text{Tr}((\mathbf{S} - \hat{\mathbf{S}}^k)^T \mathbf{B}) - \frac{\mathbf{L}_1}{2} \text{Tr}((\mathbf{S} - \hat{\mathbf{S}}^k)^T (\mathbf{S} - \hat{\mathbf{S}}^k)) \\ &= \text{Tr}(\mathbf{S}^T \mathbf{B} - (\hat{\mathbf{S}}^k)^T \mathbf{B}) - \frac{\mathbf{L}_1}{2} \text{Tr}(\mathbf{S}^T \mathbf{S} - 2\mathbf{S}^T \hat{\mathbf{S}}^k + (\hat{\mathbf{S}}^k)^T \hat{\mathbf{S}}^k) \end{aligned} \quad (11)$$

消除式(5)中与 \mathbf{S} 无关的项, 式(6)可以被重新表示为

$$\begin{aligned} & \arg \max_{\mathbf{S} \geq 0, \sum_{j=1}^n s_j = 1} \text{Tr}(\mathbf{S}^T \mathbf{B} - \frac{\mathbf{L}_1}{2} \mathbf{S}^T \mathbf{S} + \mathbf{L}_1 \mathbf{S}^T \hat{\mathbf{S}}^k) \\ & \Leftrightarrow \arg \max_{\mathbf{S} \geq 0, \sum_{j=1}^n s_j = 1} -\frac{\mathbf{L}_1}{2} \text{Tr}(\mathbf{S}^T \mathbf{S} - \frac{2}{\mathbf{L}_1} \mathbf{S}^T \mathbf{B} - 2\mathbf{S}^T \hat{\mathbf{S}}^k) \\ & \Leftrightarrow \arg \min_{\mathbf{S} \geq 0, \sum_{j=1}^n s_j = 1} \text{Tr}(\mathbf{S}^T \mathbf{S} - \frac{2}{\mathbf{L}_1} \mathbf{S}^T \mathbf{B} - 2\mathbf{S}^T \hat{\mathbf{S}}^k) \\ & \Leftrightarrow \arg \min_{\mathbf{S} \geq 0, \sum_{j=1}^n s_j = 1} \text{Tr}(\mathbf{S}^T \mathbf{S} - 2\mathbf{S}^T (\frac{1}{\mathbf{L}_1} \mathbf{B} + \hat{\mathbf{S}}^k)) \end{aligned}$$

最后, 式(6)等价于:

$$\arg \max_{\mathbf{S} \geq 0, \sum_{j=1}^n s_j = 1} \frac{1}{2} \left\| \mathbf{S} - (\hat{\mathbf{S}}^k + \frac{1}{\mathbf{L}_1} \mathbf{B}) \right\|_F^2 \quad (12)$$

令 $\mathbf{C} = \hat{\mathbf{S}}^k + (1/\mathbf{L}_1) \nabla_S f(\hat{\mathbf{S}}^k, \hat{\mathbf{F}}^k, (\mathbf{U}^{(v)})^k)$, 式(12)可以等价表示为

$$\min_{\mathbf{S} \geq 0, \sum_{j=1}^n s_j = 1} \frac{1}{2} \left\| \mathbf{S} - \mathbf{C} \right\|_F^2 \quad (13)$$

式(13)按照列可以分解为 n 个相互独立的子问题:

$$\min_{s \geq 0, \sum_{j=1}^n s_j = 1} \frac{1}{2} \left\| s - c \right\|_F^2 \quad (14)$$

如此, 这个问题就变成了单纯形上的投影。根据文献 [20] 中的方法可以将式(14)很好地求解, 其求解方法

如下。

算法1 单纯形上的投影: $s = \text{Proj-Sim}(c)$

第1步 按上升顺序排列 c 形如 $c_{(1)} \leq \dots \leq c_{(p)}$, 令 $i = p-1$;

第2步 计算 $t_i = \frac{\sum_{j=i+1}^p c_j - 1}{p-i}$, 如果 $t = t_i$ 且 $t_i \geq p_{(i)}$, 直接到第4步, 如果 $i \geq 1$, 令 $i \leftarrow i-1$ 并且重复第2步, 如果 $i=0$, 直接到第3步;

第3步 令 $\hat{t} = \frac{\sum_{j=1}^p c_j - 1}{n}$;

第4步 将 c 的投影 $s = (c - \hat{t})_+$ 投到 Δ^p 上。

3.3 更新 \mathbf{F}

与 \mathbf{S} 子问题求解方法类似, 令 $\mathbf{D} = \nabla_F f(\mathbf{S}^{k+1}, \hat{\mathbf{F}}^k, (\mathbf{U}^{(v)})^k)$ 且 $\mathbf{L}_2 = \mathbf{L}_F^k$, 于是式(7)可以表示成

$$\begin{aligned} & \text{Tr}((\mathbf{F} - \hat{\mathbf{F}}^k)^T \mathbf{D}) - \frac{\mathbf{L}_2}{2} \text{Tr}((\mathbf{F} - \hat{\mathbf{F}}^k)^T (\mathbf{F} - \hat{\mathbf{F}}^k)) \\ &= \text{Tr}(\mathbf{F}^T \mathbf{D} - (\hat{\mathbf{F}}^k)^T \mathbf{D}) - \frac{\mathbf{L}_2}{2} \text{Tr}(\mathbf{F}^T \mathbf{F} - 2\mathbf{F}^T \hat{\mathbf{F}}^k + (\hat{\mathbf{F}}^k)^T \hat{\mathbf{F}}^k) \end{aligned} \quad (15)$$

消除式(5)中与 \mathbf{F} 无关的项, 式(7)可以被重新表示为

$$\begin{aligned} & \arg \max_{\mathbf{F} \geq 0} \text{Tr}(\mathbf{F}^T \mathbf{D} - \frac{\mathbf{L}_2}{2} \mathbf{F}^T \mathbf{F} + \mathbf{L}_2 \mathbf{F}^T \hat{\mathbf{F}}^k) \\ & \Leftrightarrow \arg \max_{\mathbf{F} \geq 0} -\frac{\mathbf{L}_2}{2} \text{Tr}(\mathbf{F}^T \mathbf{F} - \frac{2}{\mathbf{L}_2} \mathbf{F}^T \mathbf{D} - 2\mathbf{F}^T \hat{\mathbf{F}}^k) \\ & \Leftrightarrow \arg \min_{\mathbf{F} \geq 0} \text{Tr}(\mathbf{F}^T \mathbf{F} - \frac{2}{\mathbf{L}_2} \mathbf{F}^T \mathbf{D} - 2\mathbf{F}^T \hat{\mathbf{F}}^k) \\ & \Leftrightarrow \arg \min_{\mathbf{F} \geq 0} \text{Tr}(\mathbf{F}^T \mathbf{F} - 2\mathbf{F}^T (\frac{1}{\mathbf{L}_2} \mathbf{D} + \hat{\mathbf{F}}^k)) \end{aligned}$$

最后, 式(7)等价于:

$$\arg \max_{\mathbf{F} \geq 0} \frac{1}{2} \left\| \mathbf{F} - (\hat{\mathbf{F}}^k + \frac{1}{\mathbf{L}_2} \mathbf{D}) \right\|_F^2 \quad (16)$$

令 $\mathbf{E} = \hat{\mathbf{F}}^k + (1/\mathbf{L}_2) \nabla_F f(\mathbf{S}^{k+1}, \hat{\mathbf{F}}^k, (\mathbf{U}^{(v)})^k)$, 式(16)可以等价表示为

$$\min_{\mathbf{F} \geq 0, \sum_{j=1}^n f_j = 1} \frac{1}{2} \left\| \mathbf{F} - \mathbf{E} \right\|_F^2 \quad (17)$$

式(17)也可以通过算法1来求解。

3.4 更新 $\mathbf{U}^{(v)}$

由于式(8)具有闭形式解, 且在每个视角中都是独立的, 因此可以将这个问题的每个视角分开处理。先求出关于 $\mathbf{U}^{(v)}$ 式子的梯度, 去除目标函数式(5)中与 $\mathbf{U}^{(v)}$ 无关的项, 得到:

$$\begin{aligned}
\text{(i) 令 } f_1 &= \left\| \mathbf{X}^{(v)} - \mathbf{U}^{(v)} \mathbf{F}^T \right\|_F^2 \\
&= \text{Tr}((\mathbf{X}^{(v)} - \mathbf{U}^{(v)} \mathbf{F}^T)^T (\mathbf{X}^{(v)} - \mathbf{U}^{(v)} \mathbf{F}^T)) \\
&= \text{Tr}((\mathbf{X}^{(v)})^T \mathbf{X}^{(v)} - (\mathbf{X}^{(v)})^T \mathbf{U}^{(v)} \mathbf{F}^T - \\
&\quad \mathbf{F} (\mathbf{U}^{(v)})^T \mathbf{X}^{(v)} + \mathbf{F} (\mathbf{U}^{(v)})^T \mathbf{U}^{(v)} \mathbf{F}^T) \\
&= \text{Tr}((\mathbf{X}^{(v)})^T \mathbf{X}^{(v)} - 2 (\mathbf{X}^{(v)})^T \mathbf{U}^{(v)} \mathbf{F}^T + \\
&\quad \mathbf{F} (\mathbf{U}^{(v)})^T \mathbf{U}^{(v)} \mathbf{F}^T)
\end{aligned}$$

$$\text{于是, } \frac{\nabla f_1}{\nabla \mathbf{U}^{(v)}} = 2\mathbf{U}^{(v)} \mathbf{F}^T \mathbf{F} - 2\mathbf{X}^{(v)} \mathbf{F}.$$

那么可以得到

$$\frac{\nabla f_1'}{\nabla \mathbf{U}^{(v)}} = \frac{1}{2} \sum_{v=1}^m \frac{\nabla f_1}{\nabla \mathbf{U}^{(v)}} = \sum_{v=1}^m (\mathbf{U}^{(v)} \mathbf{F}^T \mathbf{F} - \mathbf{X}^{(v)} \mathbf{F})$$

$$\text{(ii) 令 } f_2 = \frac{1}{2} \sum_{v=1}^m \left\| \mathbf{U}^{(v)} \right\|_F^2 = \frac{1}{2} \sum_{v=1}^m \text{Tr}((\mathbf{U}^{(v)})^T \mathbf{U}^{(v)}),$$

$$\text{得到 } \frac{\nabla f_2}{\nabla \mathbf{U}^{(v)}} = \sum_{v=1}^m \mathbf{U}^{(v)}$$

结合(i)和(ii)得到

$$\frac{\nabla f}{\nabla \mathbf{U}^{(v)}} = \sum_{v=1}^m (\mathbf{U}^{(v)} \mathbf{F}^T \mathbf{F} - \mathbf{X}^{(v)} \mathbf{F} - \mathbf{U}^{(v)}) \quad (18)$$

令梯度为零,于是得到

$$\frac{\nabla f}{\nabla \mathbf{U}^{(v)}} = \mathbf{U}^{(v)} \mathbf{F}^T \mathbf{F} - \mathbf{X}^{(v)} \mathbf{F} - \mathbf{U}^{(v)} = 0 \quad (19)$$

得到

$$\mathbf{U}^{(v)} \mathbf{F}^T \mathbf{F} - \mathbf{X}^{(v)} \mathbf{F} - \mathbf{U}^{(v)} = 0 \quad (20)$$

因此 $\mathbf{U}^{(v)}$ 的封闭解为

$$\mathbf{U}^{(v)} = \mathbf{X}^{(v)} \mathbf{F} (\mathbf{F}^T \mathbf{F} - \mathbf{I})^{-1} \quad (21)$$

3.5 参数设置

算法共有5个参数:输入平滑参数 $\lambda \in [10^{-3}, 10^3]$, 以及前文提到的更新变量 \mathbf{S} 和 \mathbf{F} 以更新相应的利普希茨 (Lipchitz) 常数 L_S^k 和 L_F^k 和权重 ω_S^k 和 ω_F^k 。

3.5.1 推导 L_S^k

去除目标函数式(5)中与 \mathbf{S} 无关的项,得到

$$\min_{S \geq 0, s_{i1}, \dots, s_{in}} \frac{\lambda}{2} \sum_{v=1}^m \left\| \mathbf{S} - \mathbf{A}^{(v)} \right\|_F^2 - \frac{\lambda}{2} \text{Tr}(\mathbf{F}^T \mathbf{L} \mathbf{F})$$

$$\Leftrightarrow \min_{S \geq 0, s_{i1}, \dots, s_{in}} \frac{\lambda}{2} \sum_{v=1}^m \left\| \mathbf{S} - \mathbf{A}^{(v)} \right\|_F^2 -$$

$$\frac{\lambda}{4} \sum_{i,j} \|f_i - f_j\|_2^2 \left(\frac{s_{ij} + s_{ji}}{2} \right)$$

$$\Leftrightarrow \min_{S \geq 0, s_{i1}, \dots, s_{in}} \frac{\lambda}{2} \sum_{v=1}^m \left\| \mathbf{S} - \mathbf{A}^{(v)} \right\|_F^2 - \frac{\lambda}{4} \sum_{i,j} \|f_i - f_j\|_2^2 s_{ij}$$

其中 f_i 是 \mathbf{F} 的第 i 行。

$$\text{(iii) 令 } f_v = \left\| \mathbf{S} - \mathbf{A}^{(v)} \right\|_F^2 = \text{Tr}((\mathbf{S} - \mathbf{A}^{(v)})^T (\mathbf{S} - \mathbf{A}^{(v)})) \\ = \text{Tr}(\mathbf{S}^T \mathbf{S} - 2\mathbf{S}^T \mathbf{A}^{(v)} + (\mathbf{A}^{(v)})^T \mathbf{A}^{(v)})$$

$$\text{于是, } \frac{\nabla f_v}{\nabla \mathbf{S}} = 2\mathbf{S} - 2\mathbf{A}^{(v)} = 2(\mathbf{S} - \mathbf{A}^{(v)}).$$

$$\text{同时令 } \bar{f} = \frac{\lambda}{2} \sum_{v=1}^m \left\| \mathbf{S} - \mathbf{A}^{(v)} \right\|_F^2 = \frac{\lambda}{2} \sum_{v=1}^m f_v, \text{ 那么可以得到}$$

$$\frac{\nabla \bar{f}}{\nabla \mathbf{S}} = \frac{\lambda}{2} \sum_{v=1}^m \frac{\nabla f_v}{\nabla \mathbf{S}} = \lambda \sum_{v=1}^m (\mathbf{S} - \mathbf{A}^{(v)}).$$

(iv) 令 $\mathbf{H}_{ij} = \|f_i - f_j\|_2^2, h_i, s_i$ 分别是 \mathbf{H} 和 \mathbf{S} 的第 i 行。

$$\text{那么 } \hat{f} = \frac{\lambda}{4} \|f_i - f_j\|_2^2 s_{ij} = \frac{\lambda}{4} \text{Tr}(\mathbf{H}^T \mathbf{S}),$$

$$\text{所以, } \frac{\nabla \hat{f}}{\nabla \mathbf{S}} = \frac{\lambda}{4} \mathbf{H}$$

$$\text{结合(iii)和(iv)得到 } \frac{\nabla f}{\nabla \mathbf{S}} = \lambda \sum_{v=1}^m (\mathbf{S} - \mathbf{A}^{(v)}) - \frac{\lambda}{4} \mathbf{H}$$

则任何 $\bar{\mathbf{S}}$ 和 $\hat{\mathbf{S}}$ 有

$$\left\| \nabla_{\mathbf{S}} f(\bar{\mathbf{S}}) - \nabla_{\mathbf{S}} f(\hat{\mathbf{S}}) \right\|_F$$

$$= \lambda \left\| \sum_{v=1}^m (\bar{\mathbf{S}} - \mathbf{A}^{(v)}) - \sum_{v=1}^m (\hat{\mathbf{S}} - \mathbf{A}^{(v)}) \right\|_F$$

$$= \lambda \left\| \sum_{v=1}^m (\hat{\mathbf{S}} - \bar{\mathbf{S}}) \right\|_F$$

$$\leq \lambda \sum_{v=1}^m \|I\|_2 \left\| \hat{\mathbf{S}} - \bar{\mathbf{S}} \right\|_F$$

因此, Lipschitz 常数为

$$L_S^k = \lambda \sum_{v=1}^m \|I\|_2 \quad (22)$$

定义

$$L_S^k = \lambda m \quad (23)$$

3.5.2 推导 L_F^k

去除目标函数式(5)中与 \mathbf{F} 无关的项,得到:

$$\begin{aligned}
\text{(v) 令 } f_1 &= \left\| \mathbf{X}^{(v)} - \mathbf{U}^{(v)} \mathbf{F}^T \right\|_F^2 \\
&= \text{Tr}((\mathbf{X}^{(v)} - \mathbf{U}^{(v)} \mathbf{F}^T)^T (\mathbf{X}^{(v)} - \mathbf{U}^{(v)} \mathbf{F}^T)) \\
&= \text{Tr}((\mathbf{X}^{(v)})^T \mathbf{X}^{(v)} - (\mathbf{X}^{(v)})^T \mathbf{U}^{(v)} \mathbf{F}^T - \\
&\quad \mathbf{F} (\mathbf{U}^{(v)})^T \mathbf{X}^{(v)} + \mathbf{F} (\mathbf{U}^{(v)})^T \mathbf{U}^{(v)} \mathbf{F}^T) \\
&= \text{Tr}((\mathbf{X}^{(v)})^T \mathbf{X}^{(v)} - 2 (\mathbf{X}^{(v)})^T \mathbf{U}^{(v)} \mathbf{F}^T + \\
&\quad \mathbf{F} (\mathbf{U}^{(v)})^T \mathbf{U}^{(v)} \mathbf{F}^T)
\end{aligned}$$

$$\text{于是, } \frac{\nabla f_1}{\nabla \mathbf{F}} = 2\mathbf{F} (\mathbf{U}^{(v)})^T \mathbf{U}^{(v)} - 2 (\mathbf{X}^{(v)})^T \mathbf{U}^{(v)}.$$

那么可以得到

$$\frac{\nabla f_1'}{\nabla \mathbf{F}} = \frac{1}{2} \sum_{v=1}^m \frac{\nabla f_1}{\nabla \mathbf{F}} = \sum_{v=1}^m (\mathbf{F} (\mathbf{U}^{(v)})^T \mathbf{U}^{(v)} - (\mathbf{X}^{(v)})^T \mathbf{U}^{(v)})$$

$$\text{(vi) 令 } f_2 = \frac{\lambda}{2} \text{Tr}(\mathbf{F}^T \mathbf{L} \mathbf{F}), \text{ 得到, } \frac{\nabla f_2}{\nabla \mathbf{F}} = \frac{\lambda}{2} (\mathbf{L} + \mathbf{L}^T) \mathbf{F}$$

结合(v)和(vi)得到

$$\frac{\nabla f}{\nabla \mathbf{F}} = \sum_{v=1}^m [\mathbf{F} (\mathbf{U}^{(v)})^T \mathbf{U}^{(v)} - (\mathbf{X}^{(v)})^T \mathbf{U}^{(v)}] - \frac{\lambda}{2} (\mathbf{L} + \mathbf{L}^T) \mathbf{F} \quad (24)$$

则任何 $\bar{\mathbf{F}}$ 和 $\hat{\mathbf{F}}$ 有

$$\begin{aligned}
&\left\| \nabla_{\mathbf{F}} f(\bar{\mathbf{F}}) - \nabla_{\mathbf{F}} f(\hat{\mathbf{F}}) \right\|_F \\
&= \left\| \sum_{v=1}^m (\hat{\mathbf{F}} - \bar{\mathbf{F}}) (\mathbf{U}^{(v)})^T \mathbf{U}^{(v)} - \frac{\lambda}{2} (\mathbf{L} + \mathbf{L}^T) (\hat{\mathbf{F}} - \bar{\mathbf{F}}) \right\|_F
\end{aligned}$$

$$\begin{aligned} &\leq \left\| \sum_{v=1}^m (\hat{\mathbf{F}} - \tilde{\mathbf{F}}) (\mathbf{U}^{(v)})^T \mathbf{U}^{(v)} \right\|_F + \frac{\lambda}{2} \left\| (\mathbf{L} + \mathbf{L}^T) (\hat{\mathbf{F}} - \tilde{\mathbf{F}}) \right\|_F \\ &\leq \sum_{v=1}^m \left\| (\mathbf{U}^{(v)})^T \mathbf{U}^{(v)} \right\|_2 \cdot \left\| \hat{\mathbf{F}} - \tilde{\mathbf{F}} \right\|_F + \frac{\lambda}{2} \left\| \mathbf{L} + \mathbf{L}^T \right\|_2 \cdot \\ &\quad \left\| \hat{\mathbf{F}} - \tilde{\mathbf{F}} \right\|_F \\ &= \left(\sum_{v=1}^m \left\| (\mathbf{U}^{(v)})^T \mathbf{U}^{(v)} \right\|_2 + \frac{\lambda}{2} \left\| \mathbf{L} + \mathbf{L}^T \right\|_2 \right) \left\| \hat{\mathbf{F}} - \tilde{\mathbf{F}} \right\|_F \end{aligned}$$

因此, Lipschitz 常数为

$$\mathbf{L}_F^k = \sum_{v=1}^m \left\| ((\mathbf{U}^{(v)})^{k+1})^T (\mathbf{U}^{(v)})^{k+1} \right\|_2 + \frac{\lambda}{2} \left\| \mathbf{L}^{k+1} + (\mathbf{L}^{k+1})^T \right\|_2 \quad (25)$$

定义

$$\mathbf{L}_F^k = \sum_{v=1}^m \left\| ((\mathbf{U}^{(v)})^{k+1})^T (\mathbf{U}^{(v)})^{k+1} \right\|_2 + \lambda \left\| \mathbf{L}^{k+1} \right\|_2 \quad (26)$$

3.5.3 推导权重 ω_S^k 及 ω_F^k

根据文献[12]提到的办法,定义插值权重如下:

$$\omega_S^k = \min(\hat{\omega}_S^k, \delta_\omega \sqrt{\frac{\mathbf{L}_S^{k-1}}{\mathbf{L}_S^k}}) \quad (27)$$

$$\omega_F^k = \min(\hat{\omega}_F^k, \delta_\omega \sqrt{\frac{\mathbf{L}_F^{k-1}}{\mathbf{L}_F^k}}) \quad (28)$$

这里 $\delta_\omega < 1$, $\hat{\omega}_i^l = (\tau_{l-1} - 1) / \tau_l$ 且 $\tau_0 = 1$, $\tau_l = \frac{1}{2} (1 + \sqrt{1 + 4\tau_{l-1}^2})$ 。

结合 3.1 ~ 3.5, 整体求解模型式(5)的算法如下:

算法2 多视角自适应图非负矩阵分解(MAGNMF)

输入:亲和矩阵 $\mathbf{A} = \{A^{(1)}, A^{(2)}, \dots, A^{(m)}\} \in \mathbf{R}^{n \times n}$,

多视角非负矩阵

$\mathbf{X} = \{X^{(1)}, X^{(2)}, \dots, X^{(m)}\} \in \mathbf{R}^{d \times n}$, 参数 λ

输出:共识相似矩阵 $\mathbf{S} \in \mathbf{R}^{n \times n}$ 、每个视角的基矩阵

$\mathbf{U}^{(v)} = \{\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \dots, \mathbf{U}^{(m)}\} \in \mathbf{R}^{d \times c}$ 以及共识系数

矩阵 $\mathbf{F} \in \mathbf{R}^{n \times c}$;

初始化:计算拉普拉斯矩阵 $\mathbf{L}_S = \mathbf{D}_A - \frac{\mathbf{S}^T + \mathbf{S}}{2} \in \mathbf{R}^{n \times n}$,

将 \mathbf{L}_S 前 c 个最小向量作为共识系数矩阵 $\mathbf{F} \in \mathbf{R}^{n \times c}$;

步骤1 通过式(17) ~ (20)计算 $\mathbf{L}_S^k, \mathbf{L}_F^k, \omega_S^k$ 和 ω_F^k ;

步骤2 令 $\hat{\mathbf{S}}^k = \mathbf{S}^k + \omega_S^k (\mathbf{S}^k - \mathbf{S}^{k-1})$;

步骤3 令 $\hat{\mathbf{F}}^k = \mathbf{F}^k + \omega_F^k (\mathbf{F}^k - \mathbf{F}^{k-1})$;

步骤4 利用算法1求解式(11)更新 \mathbf{S}^{k+1} ;

步骤5 利用算法1求解式(13)更新 \mathbf{F}^{k+1} ;

步骤6 通过式(16)更新 $(\mathbf{U}^{(v)})^{k+1}$;

步骤7 如果 $f(\hat{\mathbf{S}}^{k+1}, \hat{\mathbf{F}}^{k+1}, (\mathbf{U}^{(v)})^{k+1}) \leq f(\mathbf{S}^k, \mathbf{F}^k,$

$(\mathbf{U}^{(v)})^k$), 令 $\hat{\mathbf{S}}^k = \mathbf{S}^k, \hat{\mathbf{F}}^k = \mathbf{F}^k$ 回到步骤(4);

步骤8 令 $k \leftarrow k+1$;

步骤9 重复步骤(1) ~ (8)直至收敛。

3.6 收敛性分析

定理1 算法2的迭代更新策略可使式(5)中的目标函数值单调递减,并收敛到最小值点。

为简单起见,假设 $\omega_S^k = 0, \omega_F^k = 0$, 即不进行外推。 $\omega_S^k = 0, \omega_F^k = 0$ 的情况较复杂,但可以按文献[12]类似地处理。

根据式(5),令

$$\varphi(\mathbf{S}, \mathbf{F}, \mathbf{U}^{(v)}) = \frac{\lambda}{2} \sum_{v=1}^m \left\| \mathbf{S} - \mathbf{A}^{(v)} \right\|_F^2 + \sum_{v=1}^m \left\| \mathbf{U}^{(v)} \right\|_F^2 +$$

$$\frac{\lambda}{2} \text{Tr}(\mathbf{F}^T \mathbf{L}_S \mathbf{F}) + \frac{1}{2} \sum_{v=1}^m \left\| \mathbf{X}^{(v)} - \mathbf{U}^{(v)} \mathbf{F}^T \right\|_F^2$$

s. t. $\mathbf{S} \geq 0, s_{i_1 n} = 1, \mathbf{F} \geq 0, f_{i_1 n} = 1, \mathbf{U} \geq 0$ 于是 $\varphi(\mathbf{S}, \mathbf{F}, \mathbf{U}^{(v)}) \geq 0$, 因此函数目标值下界为0。根据算法2的更新策略,由文献[15]可得:

$$\begin{aligned} \varphi(\mathbf{S}^k, \mathbf{F}^{k+1}, (\mathbf{U}^{(v)})^{k+1}) - \varphi(\mathbf{S}^{k+1}, \mathbf{F}^{k+1}, (\mathbf{U}^{(v)})^{k+1}) &\geq \\ \frac{\mathbf{L}_S^{k+1}}{2} \cdot \left\| \mathbf{S}^k - \mathbf{S}^{k+1} \right\|_F^2 &\geq 0 \end{aligned}$$

同理可得:

$$\varphi(\mathbf{S}^k, \mathbf{F}^{k+1}, (\mathbf{U}^{(v)})^{k+1}) \geq \varphi(\mathbf{S}^{k+1}, \mathbf{F}^{k+1}, (\mathbf{U}^{(v)})^{k+1})$$

于是

$$\begin{aligned} \varphi(\mathbf{S}^k, \mathbf{F}^k, (\mathbf{U}^{(v)})^k) &\geq \varphi(\mathbf{S}^{k+1}, \mathbf{F}^k, (\mathbf{U}^{(v)})^k) \\ &\geq \varphi(\mathbf{S}^{k+1}, \mathbf{F}^{k+1}, (\mathbf{U}^{(v)})^k) \\ &\geq \varphi(\mathbf{S}^{k+1}, \mathbf{F}^{k+1}, (\mathbf{U}^{(v)})^{k+1}) \end{aligned}$$

目标值的单调递减。又由于目标函数存在下界,于是算法2的更新策略收敛到一个极小值点。

通过算法2的更新规则,可使目标函数单调递减并收敛到一个极小值点,完成证明。

图1展示了MAGNMF算法在ORL数据集上,不同迭代次数下的目标函数值。由图1可知,随着迭代次数的增加,目标函数值呈单调递减趋势,并最终达到收敛。

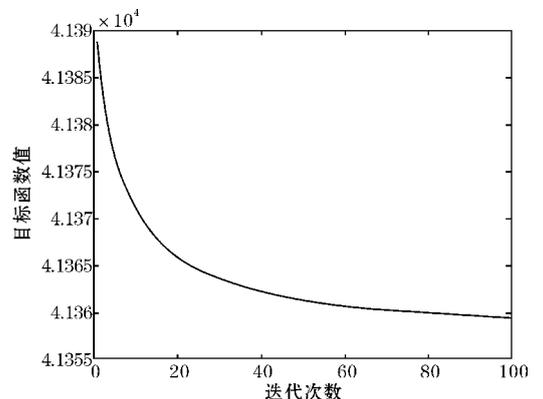


图1 收敛曲线

4 实验

将 MAGNMF 算法与目前先进的算法在聚类效果上进行比较。为保证算法比较的公平性,对于需要 k 近邻构建亲和矩阵的算法,统一将近邻数 k 设置为 5,运行算法得到的低维表达之后,再使用 K-means 算法进行聚类得到聚类结果,其中 K-means 算法直接得到结果。全部的算法都在 MATLAB R2018b 上进行。

4.1 数据集

实验用到 4 个数据集,其中包含 1 个文本数据集和 3 个图像数据集。

BBC:该数据来源于 BBC 新闻网页,由 685 个文件构成,每个文档被分成 4 个部分(4 个视角),包含 5 个类别的主题标签。

COIL20:该数据集的图片含有 20 个类别,共有 1440 张,这些图片为 32×32 的灰度照片并从 3 个不同的角度描述了样本。

Digit:该数据集由 2000 张手写数字图片(0~9)组成,视角 1 是 76 个傅里叶系数,视角 2 是 240 像素的像素特征。

ORL:该数据集共 400 张人脸图片,由 40 个不同的人的人脸图片构成,每个人有 10 张图片,每张图片由 4 个不同的特征描述。

4 个数据集的主要信息如表 2 所示,其中 d_i 表示第 i 个视角数据的维度。

表 2 数据集的主要信息统计

指标	BBC	COIL20	Digit	ORL
样本数	685	1440	2000	400
聚类数	5	20	10	40
视角	4	3	6	4
d_1	4659	1024	216	512
d_2	4633	3304	76	59
d_3	4665	3304	64	864
d_4	4684	-	6	254
d_5	-	-	240	-
d_6	-	-	47	-

4.2 对比方法

选取 8 种聚类算法来对比本文所提出的多视角聚类算法,各聚类算法如下。

K 均值聚类(K-means):通过迭代更新,将数据点分给相应的聚类中心。

共正则谱聚类(coregSC)^[21]:在谱聚类框架下,通过共同规范聚类假设,使所有视角一致,最终得到聚类结果。

非负矩阵分解(NMF):通过基矩阵和系数矩阵近似表达原数据,得到低维空间的表达。

基于非负矩阵分解的多视角聚类(multiview non-negative matrix factorization, MultiNMF)^[21]:跨多个视角的 NMF。

基于图正则项的多视角聚类(graph nonnegative matrix factorization, GMNMF)^[22]:通过有局部图正则项的多视角 NMF 进行特征提取,考虑了数据之间内部视图的相关性。

自加权多视图学习(auto-weighted multiple graph learning, AMGL)^[23]:模型没有额外的参数,能够自动学习每个视图的最优权值。

多图自加权多视角聚类(self-weighted multiview clustering, SwMC)^[24]:通过融合不同的权重图来构造相似图,然后利用相似图构造一个具有明确块对角结构的拉普拉斯图。

基于一致相似度矩阵的图非负矩阵分解(similarity graph nonnegative matrix factorization, SGNMF)^[25]:通过多图自加权多视角聚类学习相似度矩阵,来构造拉普拉斯图正则项,最后将该正则项加入原始的非负矩阵分解模型中。

4.3 评价指标

采用精确度(ACC)^[26]和归一化互信息(NMI)^[26]两种指标来度量聚类性能。指标数值越大,聚类性能越好,4 个真实数据集和 ORL 数据集的不同视角组合的 ACC 和 NMI 如表 3~5 所示,其中加粗的数值为最高数值,加下划线的数值为次高数值。

表 3 4 个数据集的 ACC 实验结果

算法	BBC	COIL20	Digit	ORL
K-means	38.20(6.21)	63.82(0.00)	70.20(8.31)	63.80(2.43)
NMF	45.36(5.43)	69.04(1.63)	52.39(0.06)	23.50(0.75)
MultiNMF	48.53(4.04)	71.42(3.65)	83.40(0.01)	78.64(3.25)
GMNMF	44.99(4.18)	<u>75.30(0.00)</u>	86.70(0.00)	55.65(1.71)
coregSC	47.01(0.00)	65.52(0.20)	75.56(5.96)	<u>77.12(0.00)</u>
AMGL	<u>53.96(11.05)</u>	56.80(5.00)	74.40(8.19)	54.60(5.00)
SwMC	37.87(0.31)	68.60(1.90)	74.00(0.21)	77.10(2.60)
SCGNMF	50.73(2.36)	60.23(2.65)	72.11(0.00)	69.63(3.10)
Our method	70.77(0.00)	75.76(0.00)	<u>85.25(0.00)</u>	85.50(0.00)

表4 4个数据集的NMI实验结果

算法	BBC	COIL20	Digit	ORL
K-means	8.93(9.96)	77.94(0.01)	71.50(3.66)	82.33(1.25)
NMF	30.203.03	84.04(4.98)	45.22(0.00)	45.26(1.11)
MultiNMF	33.52(2.36)	76.46(0.36)	80.40(0.01)	72.40(0.00)
GMNMF	24.28(5.32)	<u>87.50(0.00)</u>	<u>84.20(0.00)</u>	75.91(0.70)
coregSC	28.63(0.00)	77.91(0.15)	74.21(3.27)	<u>90.01(0.00)</u>
AMGL	<u>36.97(18.97)</u>	76.50(3.10)	82.64(4.73)	77.40(3.40)
SwMC	4.90(2.61)	72.40(0.70)	84.82(0.57)	88.81(1.20)
SCGNMF	32.77(0.84)	63.68(0.32)	74.35(1.20)	85.89(1.10)
Our method	64.50(0.00)	89.00(0.00)	85.52(0.00)	92.94(0.00)

表5 ORL数据集各视角在本方法的实验结果

视角	ACC	NMI
view1	76.75	90.25
view2	47.25	70.42
view3	39.25	56.96
view4	37.75	62.03
view1+view2	81.75	90.25
view1+view3	77.00	90.13
view1+view4	81.50	88.25
view2+view3	51.50	71.67
view2+view4	56.00	72.73
view3+view4	61.00	75.77
view1+view2+view3	74.50	90.96
view1+view2+view4	66.50	81.96
view1+view3+view4	<u>81.00</u>	<u>91.99</u>
view2+view3+view4	67.75	83.79
view1+view2+view3+view4	85.50	92.94

4.4 多视角聚类实验

将本文所提出的 MAGNMF 算法和其他 8 种对比算法在 4 个真实数据集下的聚类效果 (ACC、NMI) 总结在表 3 和表 4。通过观察和对比实验结果,得到如下结论:

K-means 和 NMF 表现的效果不是很好,SwMC、MultiNMF、GMNMF、coregSC、AMGL 以及 SCGNMF 等算法的效果不相上下,本文的 MAGNMF 算法效果在大多数情况下优于所比较的其他算法。

在 BBC、COIL20 和 ORL 等数据集上,MAGNMF 算法对比其他算法效果有较大提升。特别是在 BBC 数据集上,对比排第二的算法,MAGNMF 在 ACC 度量下

提高了 16.04%,在 NMI 度量下提高了 31.73%。在 ORL 数据集中,对比排第二的算法,MAGNMF 在 ACC 度量下提高了 8.37%,在 NMI 度量下提高了 2.93%。

在 Digit 数据集的实验中,MAGNMF 方法在 ACC 度量下取得了第二好的聚类结果,仅与最好的方法差 1.45%,在 NMI 度量下取得了最高的聚类结果,优于第二好的结果 1.32%。

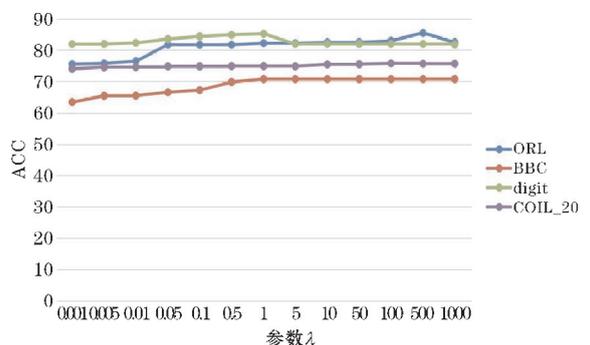
从实验结果来看,本文的算法效果相对于其他的算法在聚类上有一定的优越性,说明拉普拉斯秩约束的共识相似度矩阵学习和图正则化的非负矩阵的结合,使 MAGNMF 方法可以让原始空间的数据在低维空间中更好地保持原始数据的全局重构信息与局部结构信息,使低维表达更具有判别性,从而提升样本的聚类精度。

4.5 多视角聚类有效性验证

为验证本文所提出的 MAGNMF 算法在多视角聚类问题的有效性,在 ORL 数据集的不同视角组合上进行实验,表 5 展示了不同视角组合下,采用 MAGNMF 算法在 ACC 和 NMI 度量下的聚类结果。实验结果表明,完整的多视角数据集比单视角或者不完整视角数据集的聚类性能更好。

4.6 参数敏感性分析

算法中的 λ 是拉普拉斯秩算法和图正则项的参数,它是促进所有视角一致性和权衡这两部分的平滑度。为使实验具有公平性,将对对比算法中所有实验计算亲和矩阵的近邻值设置为 $K=5$ 。参数 λ 在集合 $\{0.001, 0.05, 0.01, 0.5, 1, 5, 10, 50, 100, 500, 1000\}$ 中变动。图 2 和图 3 中的 4 条折线分别表示了 4 个数据集在不同 λ 数值时对应 ACC 和 NMI 度量下的聚类结果。由此可知,在 4 个数据集上,聚类结果对于参数 λ 的变化并不敏感。

图2 参数 λ 对ACC的影响

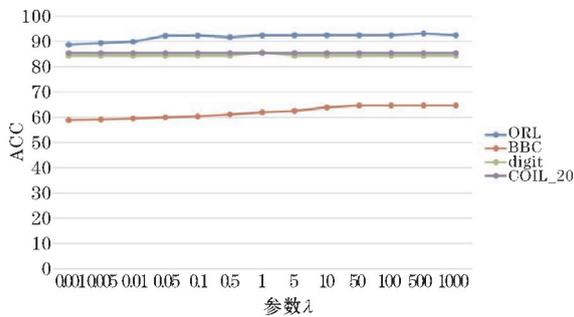


图 3 参数 λ 对 NMI 的影响

5 结束语

提出了一种新的基于多视角图自适应非负矩阵分解 (MAGNMF) 方法来处理多视角数据的聚类问题。该方法的优点包括 3 个方面:(1) 利用 CLR 和图嵌入的结合,同时考虑了所有数据的一致性和每个视角之间的互补性,达到数据表达和聚类的效果;(2) 该方法有且仅有一个参数需要调优;(3) 一个有效的 BCU 迭代更新算法被提出。通过 4 个现实数据集的实验,结果表明,本文所提出的 MAGNMF 方法在大部分情况下都优于目前的方法。

参考文献:

- [1] LIU J, CHI W, JING G, et al. Multi-view clustering via joint nonnegative matrix factorization [EB/OL]. <http://www.docin.com/p-1774193084.html>, 2016-11-03.
- [2] Sankaran P, Asari V K. A Multi-View Approach on Modular PCA for Illumination and Pose Invariant Face Recognition [C]. Applied Imagery Pattern Recognition Workshop. IEEE Computer Society, 2004.
- [3] Richard H, Ablin P, Hyvärinen A, et al. Adaptive Multi-View ICA: Estimation of noise levels for optimal inference [J]. arXiv preprint arXiv: 2102.10964, 2021.
- [4] ZHAN K, Shi J, Jing W, et al. Adaptive Structure Concept Factorization for Multiview Clustering [J]. Neural Computation, 2018, 30(4): 1-24.
- [5] Nie F, Li J, Li X. Self-weighted Multiview Clustering with Multiple Graphs [J]. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017: 2564-2570.
- [6] Zhan K, Nie F, Wang J, et al. Multiview consensus graph clustering [J]. IEEE Transactions on Image Processing, 2018, 28(3): 1261-1270.
- [7] Wen J, Zhang Z, Zhang Z, et al. Generalized Incomplete Multiview Clustering With Flexible Locality Structure Diffusion [J]. IEEE Transactions on Cybernetics, 2020(99): 1-14.
- [8] Luo S, Cao X. Multiview Subspace Dual Clustering [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021: 1-13.
- [9] Liu B Y, Huang L, Wang C D, et al. Multiview clustering via proximity learning in latent representation space [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021: 1-14.
- [10] Nie F, Wang X, Jordan M, et al. The constrained laplacian rank algorithm for graph-based clustering [J]. Proceedings of the AAAI conference on artificial intelligence. 2016, 30(1): 1969-1976.
- [11] Yan S, Xu D, Zhang B, et al. Graph embedding and extensions: A general framework for dimensionality reduction [J]. IEEE transactions on pattern analysis and machine intelligence, 2006, 29(1): 40-51.
- [12] Xu Y, Yin W. A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Applications to Nonnegative Tensor Factorization and Completion [J]. SIAM Journal on Imaging Sciences, 2015, 6(3): 1758-1789.
- [13] Blum A, Mitchell T. Combining Labeled and Unlabeled Data with Co-Training [C]. Proceedings of the 11th Annual Conference on Computational Learning Theory, 1998.
- [14] Xiao C, Nie F, Huang H, et al. Heterogeneous image feature integration via multi-modal spectral clustering [C]. CVPR 2011. IEEE, 2011.
- [15] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization [J]. Nature, 1999, 401(6755): 788-791.
- [16] Deng C, 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.
- [17] Wong J A H A. Algorithm AS 136: A K-Means Clustering Algorithm [J]. Journal of the Royal

- Statistical Society, 1979, 28(1):100–108.
- [18] Xu Y, Yin W. A globally convergent algorithm for nonconvex optimization based on block coordinate update [J]. *Journal of Scientific Computing*, 2017, 72(2):700–734.
- [19] Yangang, Xu. Alternating proximal gradient method for sparse nonnegative Tucker decomposition [J]. *Mathematical Programming Computation: A Publication of the Mathematical Programming Society*, 2015, 7(1):39–70.
- [20] Chen Y, Ye X. Projection onto a simplex [J]. *arXiv preprint arXiv*, 2011:1101–6081.
- [21] Kumar A, Rai P, Daume H. Co-regularized Multi-view Spectral Clustering [C]. *Neural Information Processing Systems*. Curran Associates Inc. 2011.
- [22] Wang Z, Kong X, Fu H, et al. Feature extraction via multi-view non-negative matrix factorization with local graph regularization [C]. *IEEE International Conference on Image Processing*. IEEE, 2015.
- [23] Nie F, Li J, Li X. Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification [J]. *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*. 2016:1881–1887.
- [24] Nie F, Jing L, Li X. Self-weighted multiview clustering with Multiple Graphs [C]. *Twenty-Sixth International Joint Conference on Artificial Intelligence*. 2017.
- [25] 李向利, 逯喜燕, 范学珍. 学习一致相似度矩阵的图非负矩阵分解 [J]. *广西大学学报(自然科学版)*, 2022, 47(1):262–273.
- [26] Shi S, Nie F, Wang R, et al. Self-weighting multi-view spectral clustering based on nuclear norm [J]. *Pattern Recognition*, 2022, 124:108429.

Non-negative Matrix Factorization Clustering Algorithm based on Multi-view Adaptive Graph Regularization

LIN Hongyan¹, DU Yuanhua¹, ZHOU Nan², TIAN Yongqiang³

(1. College of Applied Mathematics, Chengdu University of Information Technology, Chengdu 610225, China; 2. College of electronic information and electrical engineering, Chengdu University, Chengdu 610106, China; 3. Huawei Technologies Company Limited, Kunming 650011, China)

Abstract: In order to make full use of the inherent geometric structure relationship of each view data, this paper proposes a new multi-view clustering method based on adaptive graph regular non-negative matrix factorization. In a unified framework, the algorithm adaptively learns the consensus affinity matrix to extract the consensus local structure information of multi-view data for graph embedding. The global reconstruction information of multi-view data is extracted by non-negative matrix factorization. Finally, the consensus expression of each perspective not only achieves the effect of data expression and clustering. Experiments on four real datasets show that the proposed method is superior to the multi-view clustering method.

Keywords: multi-view learning; constraint Laplacian rank; graph embedding; non-negative matrix factorization; clustering