

# 基于独立自表达学习的不完全多视图聚类

诸葛文章<sup>1,2†</sup>, 范瑞东<sup>1†</sup>, 罗廷金<sup>1</sup>, 陶红<sup>1</sup>, 侯臣平<sup>1\*</sup>

1. 国防科技大学理学院, 长沙 410073

2. 海军研究院, 北京 100161

\* 通信作者. E-mail: hcpnudt@hotmail.com

† 同等贡献

收稿日期: 2021-02-26; 修回日期: 2021-04-26; 接受日期: 2021-08-19; 网络出版日期: 2022-07-13

国家自然科学基金 (批准号: 61922087, 61906201, 62006238) 和湖南省杰出青年自然科学基金 (批准号: 2019JJ20020) 资助

**摘要** 不完全多视图聚类是通过结合多视图数据的异构不完全特征来获得数据本征结构, 从而提高聚类性能的一种学习范式. 在实际应用中, 各个视图除了缺失某些完整样本外, 还会受到缺失值与异常值的影响, 使得大部分传统的不完全多视图聚类方法失效. 为解决上述问题, 本文提出一种基于独立自表达学习的不完全多视图聚类方法. 该方法通过自表达重构, 补充缺失的特征的同时学习视图独有的自表达矩阵, 然后为自表达矩阵添加低秩约束, 更好地挖掘本征结构, 并通过引入希尔伯特-施密特独立性准则来衡量不同视图间的差异性. 多个数据集上的实验结果表明, 所提方法在大多数情况下能取得较对比方法更优的聚类结果.

**关键词** 不完全多视图聚类, 特征任意缺失, 自表达, 差异性

## 1 引言

随着数据分析技术的发展, 在很多应用中, 数据由不同来源或性质的多种特征组成, 称为多视图数据<sup>[1~3]</sup>. 多视图数据中各个视图的特征通常来源于不同的空间, 其统计性质也存在较大差异, 因此如何更好地利用多个视图的信息提高数据分析效果引起了广泛的研究兴趣, 由此产生的研究方向被命名为多视图学习 (multi-view learning, MVL). 多视图聚类<sup>[4~11]</sup> (multi-view clustering, MVC) 作为多视图学习的范式之一, 由于其免除了高昂的数据标注代价, 更好地利用多个视图的信息提高数据分析效果, 成为近几十年的一个重要研究课题. 然而, 在某些应用场景中, 并不是每个视图都能完整地采集到所有样本的特征. 例如, 在跨语言文档数据中, 一些文档仅有部分语言的译本; 在采集网页数据时, 部分网页可能屏蔽了不同的敏感词汇, 导致采集的数据特征出现缺失值. 这种部分视图缺失样本特征信息的聚类问题<sup>[11~18]</sup> 称为不完全多视图聚类 (incomplete multi-view clustering, IMC).

**引用格式:** 诸葛文章, 范瑞东, 罗廷金, 等. 基于独立自表达学习的不完全多视图聚类. 中国科学: 信息科学, 2022, 52: 1186–1203, doi: 10.1360/SSI-2021-0070  
Zhuge W Z, Fan R D, Luo T J, et al. Incomplete multi-view clustering via independent self-representation learning (in Chinese). Sci Sin Inform, 2022, 52: 1186–1203, doi: 10.1360/SSI-2021-0070

根据多视图数据缺失方式的不同, 现有的不完全多视图聚类问题可以分为以下两类: (1) 基于特征整体缺失的不完全多视图聚类问题; (2) 基于特征任意缺失的不完全多视图聚类问题. 基于特征整体缺失的不完全多视图聚类问题假设各个视图的不同样本特征是完整缺失的, 近年来研究者提出了许多方法来解决此类问题. Li 等<sup>[19]</sup> 基于非负矩阵分解提出了部分多视角聚类方法, 通过将不同视图的数据矩阵分解到同一个潜在的子空间, 在该子空间内不同视图共有部分的数据样本对应相同的低维表示, 进而获取所有数据的低维表示. Hu 等<sup>[20]</sup> 提出双对齐不完全多视图聚类, 通过引入区分存在样本与缺失样本的 0-1 对角矩阵作为半监督矩阵分解的权重矩阵来学习共有的低维表示, 同时使用  $\ell_{2,1}$  范数降低缺失数据样本的影响. Wang 等<sup>[21]</sup> 提出了扰动定向的不完全多视图聚类算法, 首先在每个视图上构造不完全图矩阵, 使用图矩阵每个位置确定元素的均值对该位置缺失元素进行填补, 然后通过分析扰动风险界和各图矩阵融合结果之间的关系学习一个视图共有的图矩阵.

特征任意缺失是指除特征整体缺失情形外, 由于遮挡或信号故障等问题, 各个视图的部分特征缺失. 作为比特征整体缺失更一般的缺失情形, 研究者对基于特征任意缺失的不完全多视图聚类问题展开深入的探索. Xu 等<sup>[22]</sup> 提出了基于不完整视图的多视图学习方法 (multi-view learning with incomplete views, MVL-IV), MVL-IV 在每个视图上引入 0-1 指示矩阵区分该视图存在与缺失的特征, 基于各视图现有的特征学习视图共有的低维表示. Tao 等<sup>[23]</sup> 在 MVL-IV 的基础上, 基于无监督情形下类间最大化准则, 要求所学的共有表示矩阵具有相应的聚类特性. Tao 等<sup>[24]</sup> 将基于矩阵分解的不完全多视图子空间学习方法归纳到同一个学习框架, 并加以改进, 使其能处理视图特征任意缺失的多视图数据. 尽管传统的方法取得了优良的效果, 但是仍然存在以下局限性: (1) 现有基于特征任意缺失的多视图聚类方法大多通过矩阵分解来获得一个共有的子空间表示, 然而多视图数据很少能被一个子空间很好的描述, 因此传统方法很难学得较好的子空间表示. (2) 现有基于特征任意缺失的多视图聚类方法通常只考虑到不同视图的一致性, 没有考虑视图间的差异性, 即没有考虑捕捉使不同视图互补的信息.

针对上述问题, 本文提出基于独立自表达学习的不完全多视图聚类方法 (incomplete multi-view clustering via independent self-representation learning, ISRL). 为了能够处理特征任意缺失的数据且更清晰地捕捉数据的全局结构, ISRL 基于自表达准则, 同时进行各个视图的自表达矩阵学习以及数据矩阵补全, 并假定自表达矩阵由低秩的表示矩阵构造而成, 即考虑数据位于多个子空间附近, 因此可以直接基于表示矩阵获取聚类结果; 为了利用不同视图的一致性及差异性信息, ISRL 结合各个视图独有的自表达矩阵, 学习所有视图共有的自表达矩阵, 同时基于希尔伯特-施密特独立性准则 (Hilbert-Schmidt independence criterion, HSIC), 设计了基于互补性信息的视图权重学习机制. 此外, 本文针对模型对应的优化问题提出了一个高效的交替迭代求解算法. 相比于传统的多视图子空间聚类方法, 该方法的计算复杂度显著降低, 且在多个数据集上的实验结果表明该方法在大多数情况下能取得较对比方法更优或相当的聚类结果.

本文在第 2 节介绍现有的不完全多视图聚类方法以及多视图自表达聚类方法. 第 3 节介绍本文提出的 ISRL 方法并给出算法的求解流程. 第 4 节给出算法的收敛性分析及计算复杂度分析. 第 5 节通过在多个数据集上进行实验来验证 ISRL 方法的有效性. 最后总结全文.

## 2 相关工作回顾

假设一个多视图数据集的第  $v$  视图的样本所组成的数据矩阵表示为  $\mathbf{X}^{(v)} = [\mathbf{x}_1^{(v)}; \dots; \mathbf{x}_n^{(v)}] \in \mathbb{R}^{n \times d^{(v)}}$ . 当多视图数据面临视图特征任意缺失的问题时, 为了区分各视图的数据矩阵中存在和缺失的

表 1 常用符号说明  
Table 1 Description of common symbols

Symbol	Description	Symbol	Description
$\mathbf{X}^{(v)} \in \mathbb{R}^{n \times d^{(v)}}$	The missing data matrix of the $v$ -th view	$d_v$	The feature dimensions of the $v$ -th view
$\mathbf{M}^{(v)} \in \mathbb{R}^{n \times d^{(v)}}$	The learned complete data matrix of the $v$ -th view	$n$	Data size
$\mathbf{Z}^{(v)} \in \mathbb{R}^{n \times n}$	The self-representation matrix of the $v$ -th view	$V$	The numbers of views
$\mathbf{\Omega}^{(v)} \in \mathbb{R}^{n \times d^{(v)}}$	The missing indicating matrix of the $v$ -th view	$C$	The numbers of classes
$\mathbf{F}^{(v)} \in \mathbb{R}^{n \times r}$	The representation matrix of the $v$ -th view	$r$	The representation dimensions
$\mathbf{F} \in \mathbb{R}^{n \times r}$	The common representation matrix	$\alpha$	The view weight vector

特征, 针对每个  $\mathbf{X}^{(v)}$  引入一个指示矩阵  $\mathbf{\Omega}^{(v)} \in \{0, 1\}^{n \times d^{(v)}}$ , 其第  $(i, j)$  个元素  $\Omega_{ij}^{(v)}$  定义如下:

$$\Omega_{ij}^{(v)} = \begin{cases} 1, & \text{如果 } x_{ij}^{(v)} \text{ 存在于第 } v \text{ 个视图,} \\ 0, & \text{否则.} \end{cases} \quad (1)$$

视图特征任意缺失情形下的多视图聚类的目标为将具有特征缺失的  $n$  个多视图样本数据聚集成  $C$  个语义类, 实现不同组的样本分割. 此外, 我们将文中常用的符号表示总结于表 1 中.

在本文中, 我们规定  $\mathbf{M} \in \mathbb{R}^{p \times q}$  的  $\ell_{2,1}$  范数为

$$\|\mathbf{M}\|_{2,1} = \sum_{i=1}^p \left( \left( \sum_{j=1}^q m_{ij}^2 \right)^{\frac{1}{2}} \right).$$

在介绍了视图特征任意缺失情形和常用符号说明之后, 在下文中首先回顾现有的处理特征任意缺失情形下的基于矩阵分解的多视图子空间聚类方法, 然后回顾针对特征完整的数据设计的基于自表达的多视图子空间聚类算法.

## 2.1 特征任意缺失情形下的基于矩阵分解的多视图聚类方法

传统的特征任意缺失情形下的多视图聚类方法 [22~24] 大多基于矩阵分解, 同时进行各视图的数据矩阵的补全和潜在子空间的学习. 基于视图特征不完整的数据矩阵  $\{\mathbf{X}^{(v)}\}_{v=1}^V$ , 这类方法学习补全的数据矩阵  $\{\mathbf{M}^{(v)}\}_{v=1}^V$ , 其中  $\mathbf{M}^{(v)} \in \mathbb{R}^{n \times d^{(v)}}$  为补全后的  $\mathbf{X}^{(v)}$ . 同时, 在一个潜在的子空间里分解  $\{\mathbf{M}^{(v)}\}_{v=1}^V$ , 学习视图共有的低维表示矩阵  $\mathbf{F} \in \mathbb{R}^{n \times r}$ . 这类基于矩阵分解的子空间学习方法的优化问题可以归纳为如下统一的形式:

$$\begin{aligned} \min_{\{\mathbf{M}^{(v)}, \mathbf{U}^{(v)}\}_{v=1}^V, \mathbf{F}} & \sum_{v=1}^V \|\mathbf{M}^{(v)} - \mathbf{F}\mathbf{U}^{(v)}\|_F^2 + \mathcal{R}(\mathbf{F}, \{\mathbf{U}^{(v)}\}_{v=1}^V) \\ \text{s.t. } & \mathbf{\Omega}^{(v)} \odot \mathbf{M}^{(v)} = \mathbf{\Omega}^{(v)} \odot \mathbf{X}^{(v)}, \mathbf{U}^{(v)} \in \mathcal{C}_U^{(v)}, (\forall v), \mathbf{F} \in \mathcal{C}_F, \end{aligned} \quad (2)$$

其中,  $\mathbf{U}^{(v)} \in \mathbb{R}^{d^{(v)} \times r}$  是第  $v$  个视图的投影矩阵,  $\mathcal{R}(\mathbf{F}, \mathbf{U}^{(v)})$  是  $\mathbf{F}$  和  $\mathbf{U}^{(v)}$  的正则项,  $\mathcal{C}_F$  和  $\mathcal{C}_U^{(v)}$  为  $\mathbf{F}$  和  $\mathbf{U}^{(v)}$  的约束. 在获得  $\mathbf{F}$  之后, 运行  $K$ -均值等聚类方法获取数据的聚类结果. 基于框架 (2), 大多数基于矩阵分解的不完全多视图聚类方法可以扩展为特征任意缺失情形下的多视图子空间学习方法.

## 2.2 特征完全情形下的基于自表达的多视图聚类方法

现有的多视图子空间聚类方法<sup>[6,8,9,25,26]</sup>通常基于各视图的数据矩阵  $\{\mathbf{X}^{(v)}\}_{v=1}^V$  学习视图独有的自表达矩阵  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$ , 同时设计正则项使得  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$  之间存在交互, 现有方法的优化问题可以归纳为如下统一的形式<sup>[27]</sup>:

$$\begin{aligned} \min_{\{\mathbf{Z}^{(v)}\}_{v=1}^V} \sum_{v=1}^V \mathcal{L}(\mathbf{X}^{(v)}, \mathbf{Z}^{(v)} \mathbf{X}^{(v)}) + \mathcal{R}(\{\mathbf{Z}^{(v)}\}_{v=1}^V), \\ \text{s.t. } \mathbf{Z}^{(v)} \in \mathcal{C}_Z^{(v)}, (\forall v). \end{aligned} \quad (3)$$

其中,  $\mathcal{L}(\cdot, \cdot)$  为重构损失函数,  $\mathcal{R}(\cdot)$  为基于不同假设设计的反映  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$  之间关系的正则化项. 在获得  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$  之后, 这些方法通常基于  $\mathbf{S} = \sum_{v=1}^V (|\mathbf{Z}^{(v)}| + |\mathbf{Z}^{(v)}|^T)$  的方式构造视图共有的相似性矩阵  $\mathbf{S}$ .

## 3 基于独立自表达学习的不完全多视图聚类方法

### 3.1 目标函数

针对各视图缺失部分特征的数据矩阵  $\{\mathbf{X}^{(v)}\}_{v=1}^V$ , 不完全多视图子空间聚类方法在各视图上进行自表达重构同时基于自表达的结果对数据矩阵进行补全. 为实现这一目标, 这类方法应同时学习视图独有的自表达矩阵  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$  以及补全的数据矩阵  $\{\mathbf{M}^{(v)}\}_{v=1}^V$ , 其中  $\mathbf{M}^{(v)} \in \mathbb{R}^{n \times d^{(v)}}$  为补全后的  $\mathbf{X}^{(v)}$ . 由于各视图的信息可能是片面且不完整的, 为了结合多个视图的信息取得更好的学习效果, 这类方法同时学习视图共有的自表达矩阵  $\mathbf{Z} \in \mathbb{R}^{n \times n}$ . 综上, 基于不完整特征的多视图子空间聚类框架的目标函数形式如下:

$$\begin{aligned} \min_{\{\mathbf{M}^{(v)}, \mathbf{Z}^{(v)}\}_{v=1}^V, \mathbf{Z}} \sum_{v=1}^V \mathcal{L}(\mathbf{M}^{(v)}, \mathbf{Z}^{(v)} \mathbf{M}^{(v)}) + \mathcal{R}(\{\mathbf{Z}^{(v)}\}_{v=1}^V, \mathbf{Z}), \\ \text{s.t. } \mathbf{\Omega}^{(v)} \odot \mathbf{M}^{(v)} = \mathbf{\Omega}^{(v)} \odot \mathbf{X}^{(v)}, \mathbf{Z}^{(v)} \in \mathcal{C}_Z^{(v)}, (\forall v), \mathbf{Z} \in \mathcal{C}_Z, \end{aligned} \quad (4)$$

其中,  $\mathcal{C}_Z$  是  $\mathbf{Z}$  的约束, 引入正则化项  $\mathcal{R}(\{\mathbf{Z}^{(v)}\}_{v=1}^V, \mathbf{Z})$  的目的是使  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$  和  $\mathbf{Z}$  之间的信息可以合理的交互, 挖掘视图共有的结构同时修正  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$  中不准确的信息. 接下来, 本文基于式 (4) 介绍 ISRL 的具体模型.

以往的多视图子空间聚类方法<sup>[28,29]</sup>将最小化自表达矩阵的秩作为目标函数的一部分, 当平衡参数改变时, 这些方法所求得的自表达矩阵的秩往往也会发生改变. 与这些方法不同, 本文提出的 ISRL 方法基于工作<sup>[25,30]</sup>, 要求  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$  和  $\mathbf{Z}$  属于集合  $z_{n,n,r}$ , 集合  $z_{n,n,r}$  的定义如下:

$$z_{n,n,r} = \{\mathbf{F}\mathbf{F}^T : \mathbf{F} \in \mathbb{R}^{n \times r}, \mathbf{F}^T \mathbf{F} = \mathbf{I}_r\}. \quad (5)$$

由式 (5) 可知,  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$  和  $\mathbf{Z}$  的秩被固定为  $r$ , 通常情况下,  $r \ll \min(n, d)$ .

为了降低视图中难以自表达的特征的影响, ISRL 采用  $\ell_{2,1}$  范数度量视图的自表达损失. 假定  $\mathbf{Z}^{(v)} = \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T$  和  $\mathbf{Z} = \mathbf{F}\mathbf{F}^T$ ,  $\mathcal{L}(\mathbf{M}^{(v)}, \mathbf{Z}^{(v)} \mathbf{M}^{(v)})$  的形式如下:

$$\mathcal{L}(\mathbf{M}^{(v)}, \mathbf{Z}^{(v)} \mathbf{M}^{(v)}) = \|(\mathbf{M}^{(v)} - \mathbf{Z}^{(v)} \mathbf{M}^{(v)})^T\|_{2,1} = \|(\mathbf{M}^{(v)})^T [\mathbf{I}_n - \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T]\|_{2,1}. \quad (6)$$

相比于传统的 Frobenius 范数,  $\ell_{2,1}$  范数更倾向于使得  $(\mathbf{M}^{(v)})^T[\mathbf{I}_n - \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T]$  的行为为零. 因此本文的基本假设是特定于特征的, 即如果数据矩阵  $\mathbf{M}^{(v)}$  中的某些特征难以通过自表达矩阵恢复, 此时采用  $\ell_{2,1}$  范数使得模型更加侧重于恢复数据里可以采用自表达描述的特征.

为了设计  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$  和  $\mathbf{Z}$  之间的合理交互机制, ISRL 让视图独有的自表达矩阵  $\mathbf{Z}^{(v)}$  和视图共有自表达矩阵  $\mathbf{Z}$  之间以加权线性组合的方式交互信息, 因此,  $\mathcal{R}(\{\mathbf{Z}^{(v)}\}_{v=1}^V, \mathbf{Z})$  中包含如下部分:

$$\sum_{v=1}^V \alpha_v \|\mathbf{Z}^{(v)} - \mathbf{Z}\|_F^2 = \sum_{v=1}^V \alpha_v \|\mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T - \mathbf{F}\mathbf{F}^T\|_F^2, \quad (7)$$

其中  $\alpha_v$  衡量  $\mathbf{Z}^{(v)}$  和  $\mathbf{Z}$  之间交互的重要性. 为了确定  $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_V]^T$ , ISRL 考虑  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$  之间的互补性信息, 给互补信息较多的视图赋予较大的权重. 为实现上述目标, 基于希尔伯特 - 施密特独立性准则<sup>[31]</sup>, ISRL 使用  $\text{HSIC}(\mathbf{Z}^{(v)}, \mathbf{Z}^{(w)})$  来衡量  $\mathbf{Z}^{(v)}$  和  $\mathbf{Z}^{(w)}$  之间的独立性.  $\mathcal{R}(\{\mathbf{Z}^{(v)}\}_{v=1}^V, \mathbf{Z})$  包含如下部分:

$$\begin{aligned} & \sum_{v=1, w \neq v}^V \alpha_v \alpha_w \text{HSIC}(\mathbf{Z}^{(v)}, \mathbf{Z}^{(w)}) + \mu \|\boldsymbol{\alpha}\|_2^2, \\ & \text{s.t. } \boldsymbol{\alpha} \geq 0, \boldsymbol{\alpha}^T \mathbf{1}_V = 1, \end{aligned} \quad (8)$$

其中,  $\mu \geq 0$  为  $\boldsymbol{\alpha}$  分布的平滑参数, 式 (8) 的约束使得  $\boldsymbol{\alpha}$  为概率向量, 能够更好地衡量不同视图的重要程度. 本文使用内积核  $\mathbf{K}^{(v)} = \mathbf{Z}^{(v)}(\mathbf{Z}^{(v)})^T$  来计算 HSIC, 即

$$\text{HSIC}(\mathbf{Z}^{(v)}, \mathbf{Z}^{(w)}) = \text{tr}(\mathbf{H}\mathbf{K}^{(v)}\mathbf{H}\mathbf{K}^{(w)}), \quad (9)$$

其中,  $\mathbf{H} = \mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$  为中心化矩阵. 考虑到  $\mathbf{Z}^{(v)} = \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T$  及约束  $(\mathbf{F}^{(v)})^T\mathbf{F}^{(v)} = \mathbf{I}_r$ , 等式  $\mathbf{K}^{(v)} = \mathbf{Z}^{(v)}(\mathbf{Z}^{(v)})^T = \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T$  和  $\text{HSIC}(\mathbf{Z}^{(v)}, \mathbf{Z}^{(w)}) = \text{HSIC}(\mathbf{F}^{(v)}, \mathbf{F}^{(w)})$  成立.

通过结合  $\mathcal{L}(\mathbf{M}^{(v)}, \mathbf{Z}^{(v)}\mathbf{M}^{(v)})$  和  $\mathcal{R}(\{\mathbf{Z}^{(v)}\}_{v=1}^V, \mathbf{Z})$  的具体形式以及  $\mathbf{Z}^{(v)}$  和  $\mathbf{Z}$  的约束, ISRL 的目标函数形式如下:

$$\begin{aligned} & \min_{\boldsymbol{\Upsilon}} \sum_{v=1, w \neq v}^V \alpha_v \alpha_w \text{HSIC}(\mathbf{F}^{(v)}, \mathbf{F}^{(w)}) + \sum_{v=1}^V \alpha_v \|\mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T - \mathbf{F}\mathbf{F}^T\|_F^2 + \mu \|\boldsymbol{\alpha}\|_2^2 \\ & + \lambda \sum_{v=1}^V \|(\mathbf{M}^{(v)})^T[\mathbf{I}_n - \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T]\|_{2,1}, \\ & \text{s.t. } \boldsymbol{\Omega}^{(v)} \odot \mathbf{M}^{(v)} = \boldsymbol{\Omega}^{(v)} \odot \mathbf{X}^{(v)}, (\mathbf{F}^{(v)})^T\mathbf{F}^{(v)} = \mathbf{I}_r, \mathbf{F}^T\mathbf{F} = \mathbf{I}_r, \boldsymbol{\alpha} \geq 0, \boldsymbol{\alpha}^T \mathbf{1}_V = 1, \end{aligned} \quad (10)$$

其中,  $\boldsymbol{\Upsilon} = \{\{\mathbf{M}^{(v)}, \mathbf{F}^{(v)}\}_{v=1}^V, \boldsymbol{\alpha}, \mathbf{F}\}$  收集了 ISRL 的所有变量,  $\lambda \geq 0$  为自表达损失和正则损失的平衡参数. 不同于式 (4) 学习自表达矩阵  $\{\mathbf{Z}^{(v)}\}_{v=1}^V$  和  $\mathbf{Z}$ , ISRL 学习  $\{\mathbf{F}^{(v)}\}_{v=1}^V$  和  $\mathbf{F}$ , 因此 ISRL 不需要在计算共有的相似性矩阵后通过运行谱聚类获得聚类结果, 可以直接基于  $\mathbf{F}$  获取聚类结果.

### 3.2 求解算法

由于  $\ell_{2,1}$  范数的影响, ISRL 的优化问题 (10) 难以直接求解, 本文采用交替迭代的方式更新 (10) 中的 4 组变量  $\{\mathbf{M}^{(v)}\}_{v=1}^V, \{\mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F}$  和  $\boldsymbol{\alpha}$ . 基于重加权策略<sup>[32]</sup>, 在每轮迭代中求解如下替代问题:

$$\min_{\boldsymbol{\Upsilon}} \lambda \sum_{v=1}^V \text{tr}\{[\mathbf{I}_n - \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T]\mathbf{M}^{(v)}\mathbf{D}^{(v)}(\mathbf{M}^{(v)})^T[\mathbf{I}_n - \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T]\} + \mu \|\boldsymbol{\alpha}\|_2^2$$

$$\begin{aligned}
 & + \sum_{v=1, w \neq v}^V \alpha_v \alpha_w \text{tr}(\mathbf{H} \mathbf{F}^{(v)} (\mathbf{F}^{(v)})^T \mathbf{H} \mathbf{F}^{(w)} (\mathbf{F}^{(w)})^T) + \sum_{v=1}^V \alpha_v \|\mathbf{F}^{(v)} (\mathbf{F}^{(v)})^T - \mathbf{F} \mathbf{F}^T\|_F^2, \\
 \text{s.t. } & \mathbf{\Omega}^{(v)} \odot \mathbf{M}^{(v)} = \mathbf{\Omega}^{(v)} \odot \mathbf{X}^{(v)}, (\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_r, \mathbf{F}^T \mathbf{F} = \mathbf{I}_r, \boldsymbol{\alpha} \geq 0, \boldsymbol{\alpha}^T \mathbf{1}_V = 1,
 \end{aligned} \tag{11}$$

其中,  $\{\mathbf{D}^{(v)}\}_{v=1}^V$  为对角矩阵,  $\mathbf{D}^{(v)} \in \mathbb{R}^{d^{(v)} \times d^{(v)}}$ . 在变量  $\{\mathbf{M}^{(v)}\}_{v=1}^V, \{\mathbf{F}^{(v)}\}_{v=1}^V$  完成一轮更新后,  $\mathbf{D}^{(v)}$  的第  $j$  个 ( $j = 1, \dots, d^{(v)}$ ) 对角元素  $d^{(v)}$  按照如下准则更新:

$$d_{jj}^{(v)} = \frac{1}{2\|\mathbf{e}_{:j}^{(v)}\|_2}, \tag{12}$$

其中,  $\mathbf{e}_{:j}^{(v)}$  为第  $v$  个视图的自表达误差矩阵  $\mathbf{E}^{(v)} = [\mathbf{I}_n - \mathbf{F}^{(v)} (\mathbf{F}^{(v)})^T] \mathbf{M}^{(v)} \in \mathbb{R}^{n \times d^{(v)}}$  的第  $j$  列. 在更新变量  $\{\mathbf{D}^{(v)}\}_{v=1}^V$  后, 算法依次更新  $\{\mathbf{M}^{(v)}\}_{v=1}^V, \{\mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F}$  和  $\boldsymbol{\alpha}$ .

**固定其他变量, 更新  $\mathbf{F}^{(v)}$ .** 当  $\{\mathbf{F}^{(w)}\}_{w \neq v}, \{\mathbf{M}^{(v)}\}_{v=1}^V, \boldsymbol{\alpha}, \mathbf{F}$  和  $\{\mathbf{D}^{(v)}\}_{v=1}^V$  固定时, 问题 (11) 转化为如下关于  $\mathbf{F}^{(v)}$  的子问题:

$$\min_{(\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_r} \text{tr}[(\mathbf{F}^{(v)})^T \mathbf{L}^{(v)} \mathbf{F}^{(v)}], \tag{13}$$

其中,  $\mathbf{L}^{(v)} = 2\alpha_v \mathbf{H} [\sum_{w \neq v} \alpha_w \mathbf{F}^{(w)} (\mathbf{F}^{(w)})^T] \mathbf{H} - \lambda \mathbf{M}^{(v)} \mathbf{D}^{(v)} (\mathbf{M}^{(v)})^T - 2\alpha_v \mathbf{F} \mathbf{F}^T$ . 最小化问题 (13) 等价于如下最大化问题:

$$\max_{(\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_r} \text{tr}[(\mathbf{F}^{(v)})^T \mathbf{A}^{(v)} \mathbf{F}^{(v)}], \tag{14}$$

其中,  $\mathbf{A}^{(v)} = \rho^{(v)} \mathbf{I}_n - \mathbf{L}^{(v)}$ , 参数  $\rho^{(v)}$  为一个足够大的正实数, 使得  $\mathbf{A}^{(v)}$  为半正定矩阵. 为加快求解效率, 算法不直接求解问题 (13) 和 (14) 的最优解, 而是对如下问题进行求解:

$$\max_{(\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_r} \text{tr}[(\mathbf{F}^{(v)})^T \mathbf{C}^{(v)}], \tag{15}$$

其中,  $\mathbf{C}^{(v)} = 2\rho^{(v)} \mathbf{I}_n \mathbf{F}^{(v)} - 4\alpha_v \mathbf{H} [\sum_{w \neq v} \alpha_w \mathbf{F}^{(w)} (\mathbf{F}^{(w)})^T] \mathbf{H} \mathbf{F}^{(v)} + 4\alpha_v \mathbf{F} \mathbf{F}^T \mathbf{F}^{(v)} + 2\lambda \mathbf{M}^{(v)} \mathbf{D}^{(v)} \cdot (\mathbf{M}^{(v)})^T \mathbf{F}^{(v)}$ . 对于优化问题 (14), 假设  $\mathbf{C}^{(v)}$  的紧致奇异值分解为  $\mathbf{C}^{(v)} = \mathbf{U}_C^{(v)} \boldsymbol{\Sigma}_C^{(v)} (\mathbf{V}_C^{(v)})^T$ , 其中  $\mathbf{U}_C^{(v)} \in \mathbb{R}^{n \times r}, \boldsymbol{\Sigma}_C^{(v)} \in \mathbb{R}^{r \times r}$  且  $\mathbf{V}_C^{(v)} \in \mathbb{R}^{r \times r}$ . 借鉴文献 [33] 给出了形式相同的问题的优化方法, 我们的子问题 (15) 的最优解为

$$\mathbf{F}^{(v)} = \mathbf{U}_C^{(v)} (\mathbf{V}_C^{(v)})^T. \tag{16}$$

**固定其他变量, 更新  $\{\mathbf{M}^{(v)}\}_{v=1}^V$ .** 当  $\boldsymbol{\alpha}, \mathbf{F}, \{\mathbf{D}^{(v)}\}_{v=1}^V$  和  $\{\mathbf{F}^{(v)}\}_{v=1}^V$  固定时, 问题 (11) 中视图之间的关系是解耦的,  $\{\mathbf{M}^{(v)}\}_{v=1}^V$  可以通过独立地求解如下  $V$  个子问题更新:

$$\begin{aligned}
 & \min_{\mathbf{M}^{(v)}} \text{tr}\{[\mathbf{I}_n - \mathbf{F}^{(v)} (\mathbf{F}^{(v)})^T] \mathbf{M}^{(v)} \mathbf{D}^{(v)} (\mathbf{M}^{(v)})^T [\mathbf{I}_n - \mathbf{F}^{(v)} (\mathbf{F}^{(v)})^T]\}, \\
 \text{s.t. } & \mathbf{\Omega}^{(v)} \odot \mathbf{M}^{(v)} = \mathbf{\Omega}^{(v)} \odot \mathbf{X}^{(v)}.
 \end{aligned} \tag{17}$$

基于式 (17), 算法采用梯度下降的方式更新  $\mathbf{M}^{(v)}$ , 即

$$\mathbf{M}^{(v)} = \mathbf{M}^{(v)} - \zeta^{(v)} \Delta(\mathbf{M}^{(v)}), \tag{18}$$

其中,  $\Delta(\mathbf{M}^{(v)})$  为式 (18) 的梯度,  $\zeta^{(v)} \geq 0$  为步长.  $\Delta(\mathbf{M}^{(v)})$  的计算方式如下:

$$\Delta(\mathbf{M}^{(v)}) = 2(\mathbf{M}^{(v)}\mathbf{D}^{(v)} - \mathbf{F}^{(v)}[(\mathbf{F}^{(v)})^T\mathbf{M}^{(v)}\mathbf{D}^{(v)}]) \odot (\mathbf{1}_{n \times d^{(v)}} - \mathbf{\Omega}^{(v)}). \quad (19)$$

根据精确线性搜索,  $\zeta^{(v)}$  的值为

$$\zeta^{(v)} = \max\left(0, \frac{\text{tr}[\mathbf{E}^{(v)}\mathbf{D}^{(v)}(\mathbf{G}^{(v)})^T]}{\text{tr}[\mathbf{G}^{(v)}\mathbf{D}^{(v)}(\mathbf{G}^{(v)})^T]}\right), \quad (20)$$

其中,  $\mathbf{E}^{(v)} = \mathbf{M}^{(v)} - \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T\mathbf{M}^{(v)}$ ,  $\mathbf{G}^{(v)} = \Delta(\mathbf{M}^{(v)}) - \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T\Delta(\mathbf{M}^{(v)})$ . 由于采用精确线性搜索, 式 (20) 确定的步长确保了按照式 (18) 更新的  $\mathbf{M}^{(v)}$  使得式 (17) 的目标函数是不增的.

**固定其他变量, 更新  $\alpha$ .** 当  $\mathbf{F}$ ,  $\{\mathbf{D}^{(v)}\}_{v=1}^V$ ,  $\{\mathbf{F}^{(v)}\}_{v=1}^V$  和  $\{\mathbf{M}^{(v)}\}_{v=1}^V$  固定时, 问题 (11) 可以转化成如下关于  $\alpha$  的最小化问题:

$$\min_{\alpha \geq 0, \alpha^T \mathbf{1}_{V=1}} \alpha^T \mathbf{J} \alpha + \beta^T \alpha, \quad (21)$$

其中,  $\mathbf{J} \in \mathbb{R}^{V \times V}$  和  $\beta = [\beta_1, \dots, \beta_V]^T \in \mathbb{R}^V$  分别为基于固定变量计算的矩阵和向量. 矩阵  $\mathbf{J}$  的第  $(v, w)$  个元素  $j_{vw}$  的值为

$$j_{vw} = \begin{cases} \mu, & \text{如果 } v = w, \\ \text{HSIC}(\mathbf{F}^{(v)}, \mathbf{F}^{(w)}), & \text{其他情形.} \end{cases} \quad (22)$$

向量  $\beta$  中第  $v$  个元素为  $\beta_v = \|\mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T - \mathbf{F}\mathbf{F}^T\|_F^2 = 2r - 2\text{tr}[(\mathbf{F}^{(v)})^T\mathbf{F}\mathbf{F}^T\mathbf{F}^{(v)}]$ . 通过选择合理的参数  $\mu$ ,  $\mathbf{J}$  可以变成半正定矩阵或正定矩阵, 此时问题 (21) 变成了一个较易求解的二次凸规划问题.

**固定其他变量, 更新  $\mathbf{F}$ .** 当  $\{\mathbf{D}^{(v)}\}_{v=1}^V$ ,  $\{\mathbf{F}^{(v)}\}_{v=1}^V$ ,  $\{\mathbf{M}^{(v)}\}_{v=1}^V$  和  $\alpha$  固定时, 问题 (10) 可以转化成如下关于  $\mathbf{F}$  的最大化问题:

$$\max_{\mathbf{F}^T\mathbf{F}=\mathbf{I}_r} \text{tr} \left\{ \mathbf{F}^T \left[ \sum_{v=1}^V \alpha_v \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T \right] \mathbf{F} \right\}. \quad (23)$$

与问题 (14) 的求解相类似, 算法求解以下优化问题更新  $\mathbf{F}$ , 更新的  $\mathbf{F}$  使得式 (23) 目标函数是单调不减的.

$$\max_{\mathbf{F}^T\mathbf{F}=\mathbf{I}_r} \text{tr}(\mathbf{F}^T\mathbf{C}), \quad (24)$$

其中,  $\mathbf{C} = 2 \sum_{v=1}^V \alpha_v \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T \mathbf{F}$ . 假设  $\mathbf{C}$  的紧致奇异值分解为  $\mathbf{C} = \mathbf{U}_C \mathbf{\Sigma}_C (\mathbf{V}_C)^T$ , 其中  $\mathbf{U}_C \in \mathbb{R}^{n \times r}$ ,  $\mathbf{\Sigma}_C \in \mathbb{R}^{r \times r}$  和  $\mathbf{V}_C \in \mathbb{R}^{r \times r}$ , 问题 (24) 的解为

$$\mathbf{F} = \mathbf{U}_C \mathbf{V}_C^T. \quad (25)$$

基于上述 5 步, 5 组变量  $\{\mathbf{F}^{(v)}\}_{v=1}^V$ ,  $\{\mathbf{M}^{(v)}\}_{v=1}^V$ ,  $\{\mathbf{S}^{(v)}\}_{v=1}^V$ ,  $\alpha$ ,  $\mathbf{F}$  和  $\{\mathbf{D}^{(v)}\}_{v=1}^V$  交替的更新, 所提算法重复该更新过程直到式 (10) 的目标函数值收敛. 对于初始化问题, ISRL 算法平等地对待各个视图和特征, 将每个  $\alpha_v$  都初始化为  $1/V$  且将每个  $\mathbf{D}^{(v)}$  初始化为  $\mathbf{I}_{d^{(v)}}$ , 同时在满足约束的基础上随机地初始化其他变量. 优化问题 (10) 的求解主要步骤总结为算法 1.

**Algorithm 1** Incomplete multi-view clustering via independent self-representation learning (ISRL)

**Input:** Multi-view data with missing partial features  $\{\mathbf{X}^{(v)}\}_{v=1}^V$ , indicating matrix  $\{\Omega^{(v)}\}_{v=1}^V$ , clustering number  $C$ , parameters  $r$ ,  $\mu$ , and  $\lambda$ .

**Initialization:** Initialize  $\alpha_v = 1/V$ , initialize  $\mathbf{D}^{(v)} = \mathbf{I}_{d^{(v)}}$ , initialize  $\mathbf{F}$ ,  $\mathbf{F}^{(v)}$ , and  $\mathbf{M}^{(v)}$ ;

- 1: **while** not converged **do**
- 2:   Update the self-representation matrix  $\{\mathbf{F}^{(v)}\}_{v=1}^V$  by Eq. (16);
- 3:   Update the complete data matrix  $\{\mathbf{M}^{(v)}\}_{v=1}^V$  by Eq. (18);
- 4:   Update the weight vector  $\alpha$  by Eq. (21);
- 5:   Update the common representation matrix  $\mathbf{F}$  by Eq. (25);
- 6:   Update the reweighted matrix  $\mathbf{D}^{(v)}$  by Eq. (12);
- 7: **end while**

**Output:**  $\mathbf{F}$ ,  $\alpha$ ,  $\{\mathbf{F}^{(v)}\}_{v=1}^V$ ,  $\{\mathbf{M}^{(v)}\}_{v=1}^V$ .

## 4 理论分析

本节对 ISRL 算法进行系统的分析, 首先分析将自表达矩阵  $\mathbf{Z}$  分解为式 (5) 得到的一些性质, 其次对算法 1 的收敛性进行分析, 最后分析算法 1 的计算复杂度.

### 4.1 模型分析

传统的多视图自表达聚类方法通常假设  $\mathbf{X}^{(v)} = \mathbf{X}^{(v)} \mathbf{Z}^{(v)}$ , 其目标函数为

$$\min_{\mathbf{Z}^{(v)}} \sum_{v=1}^V \|\mathbf{Z}^{(v)}\|_* + \mathcal{R}(\{\mathbf{Z}^{(v)}\}_{v=1}^V, \mathbf{Z}), \text{ s.t. } \mathbf{X}^{(v)} = \mathbf{X}^{(v)} \mathbf{Z}^{(v)}. \quad (26)$$

然而现实应用的多视图数据通常含有一些异常值和噪声, 强制执行等式约束  $\mathbf{X}^{(v)} = \mathbf{X}^{(v)} \mathbf{Z}^{(v)}$  可能不会学得较好的  $\mathbf{Z}^{(v)}$ , 因此我们将等式约束松弛为正则化项. 为表现出子表达矩阵  $\mathbf{Z}^{(v)}$  的低秩性和对称性, 我们约束  $\mathbf{Z}^{(v)} = \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T$ , 其中  $\mathbf{F}^{(v)} \in \mathbb{R}^{n \times r}$ ,  $(\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_r$ . 即我们的有关变量  $\mathbf{F}^{(v)}$  的多视图模型为

$$\begin{cases} \min_{\mathbf{F}^{(v)}, \mathbf{F}} \sum_{v=1}^V \mathcal{L}(\mathbf{M}^{(v)}, \mathbf{M}^{(v)} \mathbf{F}^{(v)} (\mathbf{F}^{(v)})^T) + \mathcal{R}(\{\mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F}), \\ \text{s.t. } (\mathbf{F}^{(v)})^T \mathbf{F}^{(v)} = \mathbf{I}_r, \mathbf{F}^T \mathbf{F} = \mathbf{I}_r. \end{cases} \quad (27)$$

然而正则化项  $\mathcal{R}(\{\mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F})$  将视图的自表达矩阵耦合在一起, 为了更清晰地展示分解自表达矩阵  $\mathbf{Z}^{(v)}$  的性质, 我们令  $\mathcal{R}(\{\mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F}) = 0$ ,  $\mathcal{L}(\mathbf{M}^{(v)}, \mathbf{M}^{(v)} \mathbf{F}^{(v)} (\mathbf{F}^{(v)})^T) = \|\mathbf{M}^{(v)} - \mathbf{M}^{(v)} \mathbf{F}^{(v)} (\mathbf{F}^{(v)})^T\|_F^2$ . 即针对于每个视图上的  $\mathbf{F}^{(v)}$ , 其对应的优化问题为

$$\min_{\mathbf{F}^{(v)} \text{ s.t. } \mathbf{F}^{(v)} = \mathbf{I}_r} \|\mathbf{M}^{(v)} - \mathbf{M}^{(v)} \mathbf{F}^{(v)} (\mathbf{F}^{(v)})^T\|_F^2. \quad (28)$$

**定理1** 假设数据矩阵  $\mathbf{M}^{(v)}$  的奇异值分解 (singular value decomposition, SVD) 为  $\mathbf{M}^{(v)} = \mathbf{U} \Sigma \mathbf{V}^T$ , 则目标函数 (28) 有唯一的闭式解  $\mathbf{F}^{(v)} = \mathbf{V}_r$ , 对应的目标函数值为  $\sum_{i=r+1}^n \delta_i^2$ , 其中  $\delta_i$  为数据矩阵  $\mathbf{M}^{(v)}$  的第  $i$  个特征值.

**证明** 首先将式 (28) 进行转化

$$\min_{\mathbf{F}^{(v)} \text{ s.t. } \mathbf{F}^{(v)} = \mathbf{I}_r} \|\mathbf{M}^{(v)} - \mathbf{M}^{(v)} \mathbf{F}^{(v)} (\mathbf{F}^{(v)})^T\|_F^2$$



$$= \min_{\mathbf{F}^{(v)\top} \mathbf{F}^{(v)} = \mathbf{I}_r} \text{tr}((\mathbf{M}^{(v)})^\top \mathbf{M}^{(v)} - (\mathbf{F}^{(v)})^\top (\mathbf{M}^{(v)})^\top \mathbf{M}^{(v)} \mathbf{F}^{(v)}). \quad (29)$$

假设数据矩阵  $\mathbf{M}^{(v)}$  的奇异值分解为  $\mathbf{M}^{(v)} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^\top$ , 上述问题 (29) 的解等价于如下问题:

$$\max_{\mathbf{F}^{(v)\top} \mathbf{F}^{(v)} = \mathbf{I}_r} \text{tr}((\mathbf{F}^{(v)})^\top \mathbf{V} \mathbf{\Sigma}^\top \mathbf{\Sigma} \mathbf{V}^\top \mathbf{F}^{(v)}). \quad (30)$$

根据主成分分析理论 (principal component analysis, PCA), 问题 (29) 具有唯一的闭式解  $\mathbf{F}^{(v)} = \mathbf{V}_r$ . 从而问题 (29) 的目标函数值为

$$\text{tr}((\mathbf{M}^{(v)})^\top \mathbf{M}^{(v)} - (\mathbf{F}^{(v)})^\top (\mathbf{M}^{(v)})^\top \mathbf{M}^{(v)} \mathbf{F}^{(v)}) = \text{tr}(\mathbf{V} \mathbf{\Sigma}^\top \mathbf{\Sigma} \mathbf{V}^\top - \mathbf{V}_r^\top \mathbf{V} \mathbf{\Sigma}^\top \mathbf{\Sigma} \mathbf{V}^\top \mathbf{V}_r) = \sum_{i=r+1}^n \delta_i^2. \quad (31)$$

## 4.2 收敛性分析

为证明算法 1 的收敛性, 本文首先介绍如下引理 [34].

**引理 1** 对任意的非 0 向量  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$ , 有如下不等式成立:

$$\|\mathbf{a}\|_2 - \frac{\|\mathbf{a}\|_2^2}{2\|\mathbf{b}\|_2} \leq \|\mathbf{b}\|_2 - \frac{\|\mathbf{b}\|_2^2}{2\|\mathbf{b}\|_2}. \quad (32)$$

**定理 2** 按照算法 1 的迭代规则, 随着算法迭代次数的增加, ISRL 的目标函数 (10) 单调不减, 直至收敛. 此时算法 1 达到问题 (10) 的一个平稳点.

**证明** 将每次迭代中更新的  $\mathbf{F}^{(v)}$ ,  $\mathbf{M}^{(v)}$ ,  $\boldsymbol{\alpha}$  和  $\mathbf{F}$  分别表示为  $\tilde{\mathbf{F}}^{(v)}$ ,  $\tilde{\mathbf{M}}^{(v)}$ ,  $\tilde{\mathbf{F}}$  和  $\tilde{\boldsymbol{\alpha}}$ . 令  $\mathcal{G}(\mathbf{M}^{(v)}, \mathbf{F}^{(v)}) = \lambda \text{tr}\{[\mathbf{I}_n - \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^\top] \mathbf{M}^{(v)} \mathbf{D}^{(v)} (\mathbf{M}^{(v)})^\top [\mathbf{I}_n - \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^\top]\}$  及  $\mathcal{R}(\{\mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F}) = \sum_{v=1}^V \alpha_v \|\mathbf{F}^{(v)} - \mathbf{F}\|_F^2 + \sum_{v=1, w \neq v}^V \alpha_v \alpha_w \text{tr}(\mathbf{H} \mathbf{F}^{(v)} (\mathbf{F}^{(v)})^\top \mathbf{H} \mathbf{F}^{(w)} (\mathbf{F}^{(w)})^\top) + \mu \|\boldsymbol{\alpha}\|_2^2$ . 式 (11) 优化问题的目标函数可以写成  $\mathcal{R}(\{\mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F}, \boldsymbol{\alpha}) + \sum_{v=1}^V \mathcal{G}(\mathbf{M}^{(v)}, \mathbf{F}^{(v)})$ .

首先, 文献 [33] 中给出了与本文中形式一致的优化问题 (23) 的算法和详细的收敛性证明, 可以得到  $\{\mathbf{F}^{(v)}\}_{v=1}^V$  依次更新后, 式 (11) 的目标函数单调不减;  $\mathbf{M}^{(v)}$  基于精确线性搜索的梯度下降法更新, 因此式 (11) 的目标函数进一步降低; 更新的  $\boldsymbol{\alpha}$  为对应子问题的最优解, 目标函数单调不减;  $\mathbf{F}$  的更新与  $\{\mathbf{F}^{(v)}\}_{v=1}^V$  的更新类似. 综上可得出

$$\begin{aligned} & \sum_{v=1}^V \mathcal{G}(\mathbf{M}^{(v)}, \mathbf{F}^{(v)}) + \mathcal{R}(\{\mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F}, \boldsymbol{\alpha}) \geq \sum_{v=1}^V \mathcal{G}(\mathbf{M}^{(v)}, \tilde{\mathbf{F}}^{(v)}) + \mathcal{R}(\{\tilde{\mathbf{F}}^{(v)}\}_{v=1}^V, \mathbf{F}, \boldsymbol{\alpha}) \\ & \geq \sum_{v=1}^V \mathcal{G}(\tilde{\mathbf{M}}^{(v)}, \tilde{\mathbf{F}}^{(v)}) + \mathcal{R}(\{\tilde{\mathbf{F}}^{(v)}\}_{v=1}^V, \mathbf{F}, \boldsymbol{\alpha}) \geq \sum_{v=1}^V \mathcal{G}(\tilde{\mathbf{M}}^{(v)}, \tilde{\mathbf{F}}^{(v)}) + \mathcal{R}(\{\tilde{\mathbf{F}}^{(v)}\}_{v=1}^V, \mathbf{F}, \tilde{\boldsymbol{\alpha}}) \\ & \geq \sum_{v=1}^V \mathcal{G}(\tilde{\mathbf{M}}^{(v)}, \tilde{\mathbf{F}}^{(v)}) + \mathcal{R}(\{\tilde{\mathbf{F}}^{(v)}\}_{v=1}^V, \tilde{\mathbf{F}}, \tilde{\boldsymbol{\alpha}}). \end{aligned} \quad (33)$$

令  $\tilde{\mathbf{E}}^{(v)} = [\mathbf{I}_n - \tilde{\mathbf{F}}^{(v)}(\tilde{\mathbf{F}}^{(v)})^\top] \tilde{\mathbf{M}}^{(v)}$ , 结合  $\{\mathbf{D}^{(v)}\}_{v=1}^V$ , 其中  $d_{ii}^{(v)} = 1/(2\|\mathbf{e}_{:j}^{(v)}\|_2)$ , 由式 (33) 可以推导出

$$\lambda \sum_{v=1}^V \sum_{j=1}^{d^{(v)}} \frac{\|\tilde{\mathbf{e}}_{:j}^{(v)}\|_2^2}{2\|\mathbf{e}_{:j}^{(v)}\|_2} + \mathcal{R}(\{\tilde{\mathbf{F}}^{(v)}\}_{v=1}^V, \tilde{\mathbf{F}}, \tilde{\boldsymbol{\alpha}}) \leq \lambda \sum_{v=1}^V \sum_{j=1}^{d^{(v)}} \frac{\|\mathbf{e}_{:j}^{(v)}\|_2^2}{2\|\mathbf{e}_{:j}^{(v)}\|_2} + \mathcal{R}(\{\mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F}, \boldsymbol{\alpha}), \quad (34)$$

其中,  $\tilde{\mathbf{e}}_{:j}^{(v)}$  为  $\tilde{\mathbf{E}}^{(v)}$  的第  $j$  列. 由引理 1 可知

$$\lambda \sum_{v=1}^V \sum_{j=1}^{d^{(v)}} \|\tilde{\mathbf{e}}_{:j}^{(v)}\|_2 - \lambda \sum_{v=1}^V \sum_{j=1}^{d^{(v)}} \frac{\|\tilde{\mathbf{e}}_{:j}^{(v)}\|_2^2}{2\|\mathbf{e}_{:j}^{(v)}\|_2} \leq \lambda \sum_{v=1}^V \sum_{j=1}^{d^{(v)}} \|\mathbf{e}_{:j}^{(v)}\|_2 - \lambda \sum_{v=1}^V \sum_{j=1}^{d^{(v)}} \frac{\|\mathbf{e}_{:j}^{(v)}\|_2^2}{2\|\mathbf{e}_{:j}^{(v)}\|_2}. \quad (35)$$

将式 (34) 和 (35) 相加后, 可以得到

$$\sum_{v=1}^V \left( \|\tilde{\mathbf{E}}^{(v)}\|^T_{2,1} + \mathcal{R}(\{\tilde{\mathbf{F}}^{(v)}\}_{v=1}^V, \tilde{\mathbf{F}}, \tilde{\boldsymbol{\alpha}}) \leq \sum_{v=1}^V \left( \|\mathbf{E}^{(v)}\|^T_{2,1} + \mathcal{R}(\{\mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F}, \boldsymbol{\alpha}) \right). \quad (36)$$

式 (10) 可以简写为  $\sum_{v=1}^V \mathcal{L}(\mathbf{M}^{(v)}, \mathbf{F}^{(v)}) + \mathcal{R}(\{\mathbf{F}^{(v)}\}_{v=1}^V, \mathbf{F}, \boldsymbol{\alpha})$ , 其中函数  $\mathcal{L}(\mathbf{M}^{(v)}, \mathbf{F}^{(v)}) = \lambda \cdot \|\mathbf{M}^{(v)}\|^T_{2,1} + \|\mathbf{I}_n - \mathbf{F}^{(v)}(\mathbf{F}^{(v)})^T\|_{2,1} = \|\mathbf{E}^{(v)}\|^T_{2,1}$ . 将算法 1 收敛后的  $\mathbf{F}^{(v)}$ ,  $\mathbf{M}^{(v)}$ ,  $\boldsymbol{\alpha}$  和  $\mathbf{F}$  分别记为  $\hat{\mathbf{F}}^{(v)}$ ,  $\hat{\mathbf{M}}^{(v)}$ ,  $\hat{\boldsymbol{\alpha}}$  和  $\hat{\mathbf{F}}$ , 基于梯度下降法的性质可知,  $\{\{\hat{\mathbf{F}}^{(v)}, \hat{\mathbf{M}}^{(v)}\}_{v=1}^V, \hat{\boldsymbol{\alpha}}, \hat{\mathbf{F}}\}$  满足优化问题 (11) 的 KKT (Karush-Kuhn-Tucker) 条件. 又优化问题 (10) 和 (11) 的 KKT 条件相同, 因此  $\{\{\hat{\mathbf{F}}^{(v)}, \hat{\mathbf{M}}^{(v)}\}_{v=1}^V, \hat{\boldsymbol{\alpha}}, \hat{\mathbf{F}}\}$  满足 (10) 达到平稳点的 KKT 条件.

### 4.3 时间复杂度分析

算法 1 采用交替迭代的方式求解问题 (10), 在每次迭代中, 更新  $\{\mathbf{F}^{(v)}\}_{v=1}^V$  的计算复杂度为  $O(V^2nr^2 + V^2nr + Vnd + Vndr + Vnr^2 + Vr^3)$ ; 更新  $\{\mathbf{M}^{(v)}\}_{v=1}^V$  和  $\{\mathbf{D}^{(v)}\}_{v=1}^V$  的计算复杂度为  $O(Vnd + Vndr)$ ; 更新  $\boldsymbol{\alpha}$  的计算复杂度为  $O(V^3\tau)$ , 其中  $\tau$  为求解问题 (21) 的迭代次数; 更新  $\mathbf{F}$  的计算复杂度为  $O(Vnr^2 + r^3)$ . 通常  $r \ll \min(n, d)$  且  $V \ll \min(n, d)$ , 因此算法 1 的计算复杂度为  $O(VndrT)$ , 其中  $T$  是算法 1 的迭代次数.

## 5 实验

本节通过实验对 ISRL 算法进行系统的评估. 本文首先将 ISRL 与相关聚类方法在 6 个数据集上进行不同缺失特征比例下的聚类效果对比, 然后给出 ISRL 目标函数的收敛性曲线并研究参数对 ISRL 性能的影响, 最后给出 ISRL 的计算时间分析.

### 5.1 实验设定

**数据集.** 对比实验在 6 个数据集上进行, 分别为 MSRC-v1, Ionosphere, NNSpt, Protein, WebKB, 以及 Digits. 本文于表 2 中对这 6 个数据集的特征类型进行简要总结, 并对每个数据集进行具体描述.

(1) MSRC-v1<sup>1)</sup>数据集由属于 8 类的 240 张照片组成, 每类有 30 张照片. 背景类的图片被剔除, 剩余的自行车、人脸、汽车、牛、树、建筑和飞机这 7 类图片组成数据集, 仍命名为 MSRC-v1.

(2) Ionosphere<sup>2)</sup>数据集由 351 个电离层数据组成, 这些数据被分为两类: 225 个“好”示例和 126 个“坏”示例.

(3) NNSpt<sup>3)</sup>数据集由 840 张图片组成, 这些图片属于两类: 420 张 NBA 图片和 420 张 NASCAR 图片.

(4) Protein<sup>4)</sup>数据集由 629 个酵母蛋白质数据组成, 这些酵母蛋白质数据被分为 2 类: 497 个膜蛋白质和 132 核糖体蛋白质.

(5) WebKB<sup>5)</sup>数据集由从 4 所大学的网站收集的 1051 个网页文本数据组成, 这些网页文本数据被分为两类: 230 个课程网页数据以及 821 个非课程网页数据.

1) <https://www.microsoft.com/en-us/research/project/>.

2) <http://archive.ics.uci.edu/ml/datasets/Ionosphere>.

3) <http://www.cst.ecnu.edu.cn/~sjsun/software/MvLapSVMcode.zip>.

4) <https://noble.gs.washington.edu/proj/sdp-svm/>.

5) <http://www.cs.cmu.edu/afs/cs/project/theo-11/www/wwkb/>.

表 2 6 个真实多视图数据集的简要特征描述

Table 2 A brief feature description of six real multi-view datasets

View	MSRC-v1	Ionosphere	NNSpt	Protein	WebKB	Digits
1	LBP(256)	View1(34)	GRAY(1024)	FFT(4910)	View1(2949)	FOU(76)
2	HOG(100)	PCA(25)	TFIDF(296)	FAC(216)	View2(334)	GE(441)
3	CENTRIST(1302)			Pfam(3735)		KAR(64)
4	GIST(512)					PIX(240)
5	CMT(48)					ZER(47)
6	SIFT(200)					MOR(6)

(6) Digits<sup>6)</sup>数据集由 0~9 这 10 类数字手写体图片组成, 每类 200 张, 共计 2000 张图片.

上述 6 个数据集本身具有完整的特征, 为了模拟特征人员缺失情形下的学习场景, 实验删除一定比例的特征生成不完整的数据. 具体来说, 本文基于一个概率来移除每个  $x_{ij}^{(v)}$ , 该概率可以被视为数据特征的缺失比例 (incomplete ratio, IR), 在实验中, IR 从 0% 到 50% 以 10% 的步长进行变动. 在生成具有不同缺失特征比例的多视图数据后, 本次实验将各视图数据矩阵的每个特征的值归一化到  $[0, 1]$  之间.

**对比方法.** 在实验中, 将本文提出的 ISRL 方法与如下 5 种方法进行对比: 不完全多视图学习 (MVL-IV) [22], 多视图潜在子空间聚类 (latent multi-view subspace clustering, LMSC) [27], 基于块对角表示的不完全多视图学习 (incomplete multi-view learning with block diagonal representation, IML-BDR) [24], 双对齐不完全多视图聚类 (doubly aligned incomplete multi-view clustering, DAIMC) [20], 低秩张量约束的多视图子空间聚类 (low-rank tensor constrained multi-view subspace clustering, LTMSC) [6], 一致和独有的多视图子空间聚类 (consistent and specific multi-view subspace clustering, CSMSC) [29]. DAIMC 原本是针对不完全多视图数据设计的, 本实验基于式 (2) 将其扩展为能处理特征随机缺失的多视图数据的方法. LMSC, LTMSC 和 CSMSC 是针对特征完整的多视图数据设计的方法, 本次实验首先使用基于深度矩阵分解的补全方法 (matrix completion by deep matrix factorization, DMF) [35] 对各视图的数据矩阵进行补全. MVL-IV, DAIMC, IML-BDR 和 ISRL 直接学习数据的低维表示, LMSC, LTMSC 和 CSMSC 在构造自表达矩阵后, 通过谱嵌入学习低维表示, 最后在各方法的共有低维表示上运行 K-means 去获取数据的聚类结果. 方法代码在处理器为 Intel(R) Xeon(R) CPU E3-1245 v3 (3.4 GHz), 内存为 32 GB 以及操作系统为 Windows 10 的工作站上运行, 平台为 MATLAB R2017a.

**参数设定.** 在验证实验中, 各方法的参数都是通过网格搜索的方式确定的, 记录最优参数组合下的聚类结果. 对比方法的参数搜索范围是根据各自的论文确定的, 代码是从其作者主页下载的. 对于 ISRL,  $\lambda$  在  $\{10^{-1}, 1, 10^1, 10^2, 10^3, 10^4\}$  内调节,  $r$  在  $\{1C, 2C, 3C, 4C, 5C, 6C\}$  内调节,  $\mu$  在  $\{r, 10^{0.5}r, 10r\}$  内调节. 本次实验所有采用交替下降求解策略的方法的停止准则为

$$\frac{J(t-1) - J(t)}{J(t-1)} < 10^{-4}, \quad (37)$$

其中,  $J(t)$  是第  $t$  次迭代中的目标函数值.

**评价准则.** 为衡量聚类效果, 实验使用两种聚类评价准则, 分别为聚类准确率 (clustering accuracy, ACC) 和标准化互信息 (normalized mutual information, NMI). 在每个数据集上, 基于各缺失比例独立地制造 10 个特征缺失的多视图数据, 运行各方法后, 报告平均结果及标准差 (STD).

6) <https://archive.ics.uci.edu/ml/datasets/Multiple+Features>.

## 5.2 聚类结果对比

表 3 和 4 分别报告了 6 个数据集上不同特征缺失比例 IR 下各方法的聚类 ACC 和 NMI 结果. 最优结果加粗, 符号 ‘●/○/○’ 表示置信度 95% 情况下 ISRL 较对比方法 “优/平/差”.

从表 3 和 4 中的结果可以看出:

(1) 对比基于矩阵分解的多视图子空间学习方法和结合补全方法的多视图子空间聚类方法, 在一些数据集如 MSRC-v1 上, 3 个多视图子空间学习方法的聚类效果大多数情况下优于 3 个多视图子空间聚类方法; 在另一些数据集例如 Digits 上, 3 个多视图子空间聚类方法的聚类效果大多数情况下优于 3 个多视图子空间学习方法. 这可能是由于不同类别的方法适合不同的数据集导致的.

(2) 将 MVL-IV, DAIMC 和 IML-BDR 3 个方法进行对比, IML-BDR 在 MSRC-v1, NNSpt, Protein, 以及 WebKB 数据集上的性能优于其他两个方法, 而 DAIMC 在 Digits 数据集上的聚类效果更好. 类似的现象存在于 LMSC, CSMSC 和 LTMSC 方法的对比中, 例如, CSMSC 在 MSRC-v1, NNSpt 数据集上的取得了更好的聚类效果, LTMSC 在 Digits 数据集上性能更好, LMSC 在 Protein 和 WebKB 数据集上的某些缺失比例下实现了优于其他两个方法的结果. 这可能是因为这些方法虽然属于同类方法, 但是由于采用不同的范数、正则化项或约束, 这使得它们擅长处理某些特定的数据.

(3) 随着缺失比例 IR 的增加, 大多数情况下各方法取得的聚类效果是降低的, 这种下降可能是由于缺失信息的增多导致的. 在这些实验结果中, CSMSC 在 Protein 数据集上的结果是一个明显的例外, 随着缺失比例的提高, CSMSC 取得了更好的聚类结果, 这种现象可能是由于随着 IR 增加, DMF 在补全数据矩阵时可以更有效地去除数据噪声导致, 从而使得 CSMSC 在去噪的数据上展现更好的聚类性能. 此外, 在 NNSpt 数据集上, LTMSC 的效果随着缺失比例的提高先升后降. 对于 LTMSC 方法来说, 出现这种情形的原因可能是当 IR 较低时, DMF 补全起到了去噪的效果; 而当 IR 较高时, DMF 补全产生的偏差导致 LTMSC 性能再次降低.

(4) 在 ACC 和 NMI 两种聚类指标下, 本文提出的 ISRL 方法在大多数情况下实现了比其他方法更优或相当的聚类结果. 这可能是由于 ISRL 结合了数据矩阵补全与自表达学习的过程, 蕴含视图共有和独有的自表达矩阵共同学习机制, 使各视图的信息能更合理的交互.

## 5.3 收敛性实验

为了验证算法 1 的收敛性, 图 1 绘制了 IR = 20% 的 3 个数据集上 ISRL 目标函数 (10) 的收敛曲线. 对于 ISRL 算法,  $r$  被设定为  $3C$ ,  $\lambda$  被固定为  $10^3$ ,  $\beta$  被设定为  $10r$ . 如图 1 所示, 在 3 个数据集上, 随着迭代次数的增加, 式 (10) 的目标函数单调不增. 算法 1 在 50 步内使得式 (10) 的目标函数值收敛到一个固定的值, 这表明算法 1 的收敛速度还是较快的.

## 5.4 参数影响分析

本文接下来分析参数对 ISRL 性能的影响. ISRL 方法有 3 个参数,  $r$  在  $\{C, 2C, 3C, 4C, 5C, 5C\}$  范围内调节,  $\lambda$  在  $\{10^{-1}, 1, 10^1, 10^2, 10^3, 10^4\}$  范围内调节,  $\mu$  在  $\{r, 10^{0.5}r, 10r\}$  范围内调节. 实验选取的两个数据集为 MSRC-v1 和 Ionosphere, 缺失比例 IR 被固定为 20%. 图 2 绘制了两个数据集上不同参数组合下的 ISRL 聚类 ACC 结果. 从图 2 中的结果可以看出:  $\mu$  需要设定为合适的值, ISRL 才能通过调节  $r$  和  $\lambda$  实现较好的性能. 当  $\mu$  固定为合适的值后, 在 MSRC-v1 数据集上, ISRL 的性能主要受  $r$  的影响, 在 Ionosphere 数据集上, ISRL 的性能主要受  $\lambda$  的影响. ISRL 在两个数据集上的最优参数组合不同, 其原因是 ISRL 的最优参数组合是由数据特性决定的.

表 3 各算法在 6 个公开数据集上不同缺失特征比例下的聚类 ACC 对比 (均值 (标准差))

Table 3 ACC comparison (AVE(STD)) of all algorithms in six data sets at different missing feature ratios IR

	IR = 0%	IR = 10%	IR = 20%	IR = 30%	IR = 40%	IR = 50%
MSRC-v1						
Algorithm						
MVL-IV	84.0(10.5)●	86.1(9.7)●	89.3(8.3)●	85.6(8.7)●	86.0(6.7)●	89.3(6.0)●
DAIMC	91.4(4.5)●	90.3(3.8)●	90.3(4.3)●	88.7(4.0)●	87.5(4.8)●	86.8(3.0)●
IML-BDR	<b>98.0(0.9)⊙</b>	<b>98.2(0.6)⊙</b>	97.1(1.3)●	96.4(1.2)⊙	96.0(1.6)⊙	94. (1.5)●
LMSC	72.2(5.6)●	71.2(3.6)●	67.7(6.4)●	69.7(4.8)●	65.7(4.9)●	64.4(3.1)●
LTMSC	64.1(1.5)●	44.0(4.2)●	36.8(4.1)●	29.3(4.6)●	32.4(3.6)●	28.9(3.0)●
CSMSC	83.4(0.2)●	82.7(3.7)●	87.7(5.2)●	82.5(4.4)●	84.1(5.7)●	82.9(3.3)●
ISRL	<b>98.0(0.3)</b>	98.0(0.3)	<b>98.0(0.4)</b>	<b>97.3(0.5)</b>	<b>96.6(1.0)</b>	<b>95.6(1.5)</b>
Ionosphere						
Algorithm						
MVL-IV	81.5(6.1)●	79.0(5.6)●	76.7(7.4)●	75.0(7.0)●	71.3(4.7)●	70.5(3.3)●
DAIMC	73.2(9.0)●	70.5(8.3)●	66.3(4.0)●	66.7(2.8)●	66.5(3.6)●	65.7(2.0)●
IML-BDR	75.8(3.5)●	74.8(3.0)●	73.8(3.4)●	73.7(3.8)●	75.7(5.1)●	76.3(3.4)⊙
LMSC	78.3(3.7)●	75.2(4.4)●	79.3(4.9)●	80.1(6.9)●	77.3(3.7)●	74.8(3.3)●
LTMSC	53.3(0.0)●	79.6(1.7)●	77.9(2.9)●	82.1(4.7)●	78.7(4.9)⊙	<b>78.1(4.4)⊙</b>
CSMSC	72.6(0.0)●	75.0(1.2)●	75.2(1.4)●	75.3(1.1)●	74.2(0.9)●	72.4(1.1)●
ISRL	<b>100.0(0.0)</b>	<b>97.3(0.5)</b>	<b>93.4(1.3)</b>	<b>89.0(1.7)</b>	<b>82.5(2.9)</b>	77.7(1.7)
NNSpt						
Algorithm						
MVL-IV	83.2(12.4)●	78.9(17.1)●	88.3(12.7)●	84.8(13.8)●	88.5(8.8)●	91.6(1.3)●
DAIMC	69.9(4.0)●	68.3(1.0)●	68.2(1.3)●	70.0(3.1)●	70.4(1.6)●	72.4(8.5)●
IML-BDR	94.2(0.4)●	94.0(0.6)●	94.1(0.4)●	94.3(0.4)●	93.3(0.5)●	92.4(1.3)●
LMSC	67.0(2.9)●	71.4(1.4)●	72.3(1.7)●	70.5(1.0)●	69.6(1.7)●	67.2(3.9)●
LTMSC	50.1(0.0)●	54.9(15.1)●	93.1(15.1)⊙	83.4(23.0)⊙	59.8(20.2)●	50.1(0.0)●
CSMSC	<b>99.3(0.0)⊙</b>	<b>99.3(0.2)⊙</b>	<b>99.4(0.2)⊙</b>	<b>99.4(0.2)⊙</b>	<b>99.3(0.2)⊙</b>	<b>99.2(0.3)⊙</b>
ISRL	98.9(0.0)	99.0(0.1)	99.0(0.3)	99.1(0.2)	98.9(0.2)	99.0(0.2)
Protein						
Algorithm						
MVL-IV	94.5(1.2)●	94.7(1.4)●	93.6(2.3)●	90.2(4.9)●	88.5(5.7)●	87.3(6.2)●
DAIMC	87.9(1.6)●	87.2(1.8)●	87.5(1.6)●	88.4(3.6)●	87.7(3.0)●	89.3(4.3)●
IML-BDR	95.6(0.5)●	95.8(0.7)●	95.3(0.7)●	95.0(1.4)●	94.7(1.8)●	93.5(2.2)●
LMSC	84.5(1.7)●	83.5(3.6)●	83.7(2.8)●	83.2(3.2)●	86.5(4.7)●	86.5(3.2)●
LTMSC	78.9(0.0)●	78.9(0.1)●	79.0(0.2)●	79.0(0.2)●	79.0(0.2)●	79.0(0.2)●
CSMSC	76.5(0.0)●	76.5(0.0)●	76.6(1.5)●	87.9(1.2)●	94.6(1.1)●	<b>98.3(0.4)⊙</b>
ISRL	<b>99.1(0.3)</b>	<b>98.9(0.3)</b>	<b>98.6(0.4)</b>	<b>98.3(0.7)</b>	<b>98.2(0.5)</b>	98.1(0.4)
WebKB						
Algorithm						
MVL-IV	91.5(3.3)●	84.3(4.5)●	81.1(1.1)●	81.4(1.4)●	81.1(0.7)●	80.4(1.0)●
DAIMC	88.7(5.5)●	86.9(7.5)●	83.7(7.3)●	85.3(7.3)●	78.4(8.4)●	78.3(6.5)●
IML-BDR	91.5(1.9)●	92.0(1.8)●	91.4(1.9)●	91.2(1.5)●	91.4(1.3)●	89.6(1.7)●
LMSC	90.0(5.5)●	81.4(9.4)●	79.5(8.7)●	78.0(12.6)●	75.3(6.0)●	68.8(3.6)●
LTMSC	77.9(0.0)●	78.0(0.0)●	78.0(0.0)●	78.0(0.0)●	78.1(0.1)●	78.0(0.0)●
CSMSC	78.0(0.0)●	80.4(4.3)●	81.3(6.9)●	78.6(10.8)●	85.3(7.6)●	77.8(7.5)●
ISRL	<b>96.4(0.1)</b>	<b>96.0(0.3)</b>	<b>95.4(0.4)</b>	<b>94.9(0.6)</b>	<b>93.8(0.6)</b>	<b>92.5(0.5)</b>
Digits						
Algorithm						
MVL-IV	78.2(11.2)●	79.0(9.6)●	81.0(5.8)●	76.9(5.3)●	79.3(7.2)●	83.1(6.0)●
DAIMC	86.7(5.9)●	91.9(2.7)●	89.4(5.0)●	89.3(4.7)●	91.1(3.4)⊙	90.6(4.4)⊙
IML-BDR	76.9(4.1)●	80.3(2.5)●	77.9(1.7)●	77.2(2.6)●	74.9(2.5)●	70.0(3.2)●
LMSC	82.4(4.7)●	84.8(5.0)●	84.0(5.0)●	85.7(5.5)●	88.1(4.2)●	86.6(4.7)●
LTMSC	91.3(0.0)●	93.2(1.5)⊙	93.1(1.1)●	92.6(1.4)⊙	91.7(1.1)●	87.2(4.9)●
CSMSC	88.9(0.0)●	90.6(0.6)●	90.9(0.7)●	81.8(3.9)●	77.1(7.5)●	72.4(4.2)●
ISRL	<b>94.8(0.6)</b>	<b>94.2(0.5)</b>	<b>94.4(0.8)</b>	<b>93.6(0.9)</b>	<b>93.0(0.8)</b>	<b>92.8(0.7)</b>

表 4 各算法在 6 个数据集上不同缺失特征比例 IR 下的聚类 NMI 对比 (均值 (标准差))

Table 4 NMI comparison (AVE(STD)) of all algorithms in six data sets at different missing feature ratios IR

	IR = 0%	IR = 10%	IR = 20%	IR = 30%	IR = 40%	IR = 50%
MSRC-v1						
Algorithm						
MVL-IV	79.9(10.2)●	82.0(8.7)●	84.9(6.8)●	80.4(8.6)●	80.2(5.9)●	82.6(7.0)●
DAIMC	85.8(5.0)●	85.1(3.9)●	83.7(6.0)●	81.2(5.1)●	82.0(4.6)●	78.5(3.6)●
IML-BDR	<b>96.0(1.9)○</b>	<b>96.4(1.1)○</b>	94.4(2.2)○	93.0(2.0)○	92.3(2.8)○	89.5(2.6)○
LMSC	65.3(7.0)●	65.6(5.5)●	61.9(3.7)●	59.2(5.7)●	57.1(4.5)●	56.2(2.8)●
LTMSC	64.3(0.8)●	44.9(2.3)●	37.3(3.2)●	26.6(5.4)●	28.2(4.8)●	21.4(3.4)●
CSMSC	80.6(0.3)●	78.4(3.6)●	81.4(4.6)●	75.9(2.9)●	77.4(5.2)●	74.9(3.2)●
ISRL	95.9(0.5)	95.8(0.5)	<b>95.7(0.9)</b>	<b>94.1(1.0)</b>	<b>93.0(1.6)</b>	<b>91.1(2.6)</b>
Ionosphere						
Algorithm						
MVL-IV	30.9(13.0)●	26.5(14.4)●	21.6(12.0)●	22.4(16.2)●	13.3(5.8)●	11.5(5.2)●
DAIMC	19.0(17.3)●	15.3(12.2)●	8.6(6.8)●	8.1(2.5)●	7.6(3.7)●	6.1(1.9)●
IML-BDR	21.4(6.2)●	20.0(7.3)●	19.5(7.2)●	16.7(7.2)●	20.2(8.0)●	19.5(5.9)○
LMSC	35.4(3.8)●	24.8(9.8)●	32.8(8.8)●	35.5(9.6)●	28.2(5.2)○	25.1(8.7)○
LTMSC	15.9(0.0)●	37.8(2.8)●	36.4(4.2)●	42.9(8.2)○	<b>37.9(6.8)○</b>	<b>36.0(5.3)○</b>
CSMSC	16.7(0.0)●	22.9(3.2)●	23.7(4.0)●	23.8(3.1)●	21.1(2.0)●	17.0(2.4)●
ISRL	<b>99.7(0.8)</b>	<b>81.5(2.9)</b>	<b>63.8(5.6)</b>	<b>48.2(5.1)</b>	30.9(6.4)	21.5(3.0)
NNSpt						
Algorithm						
MVL-IV	42.0(24.4)●	37.5(30.7)●	55.1(19.9)●	46.9(25.4)●	52.5(16.8)●	58.6(4.4)●
DAIMC	12.5(5.5)●	10.1(1.2)●	10.0(1.7)●	13.2(6.5)●	13.3(3.5)●	19.5(17.4)●
IML-BDR	68.1(1.5)●	67.2(2.3)●	67.9(1.7)●	68.7(1.6)●	64.8(1.7)●	61.7(4.6)●
LMSC	11.3(2.2)●	14.3(1.8)●	16.2(3.5)●	12.8(1.3)●	11.9(2.5)●	10.0(4.1)●
LTMSC	1.0(0.0)●	9.5(26.7)●	77.4(26.9)○	59.8(40.7)●	18.3(36.1)●	1.1(0.2)●
CSMSC	<b>94.1(0.0)○</b>	<b>94.7(1.6)○</b>	<b>94.9(1.5)○</b>	<b>94.9(1.4)○</b>	<b>94.3(1.2)○</b>	<b>93.7(2.4)○</b>
ISRL	92.0(0.0)	92.2(0.5)	92.4(1.7)	93.1(1.3)	91.7(1.1)	92.6(1.0)
Protein						
Algorithm						
MVL-IV	62.2(3.1)●	63.8(4.3)●	61.9(5.9)●	51.2(17.7)●	45.1(22.9)●	43.8(23.9)●
DAIMC	42.0(4.4)●	42.0(5.3)●	41.7(6.1)●	46.1(11.3)●	42.1(11.1)●	46.2(13.4)●
IML-BDR	67.0(2.6)●	69.2(3.2)●	67.5(3.2)●	67.5(5.8)●	67.3(6.4)●	63.3(6.6)●
LMSC	34.9(4.2)●	35.4(4.9)●	38.4(4.6)●	36.1(3.7)●	44.7(10.0)●	44.7(9.2)●
LTMSC	0.5(0.0)●	0.5(0.1)●	1.4(1.3)●	1.8(1.4)●	1.8(1.4)●	1.3(1.3)●
CSMSC	2.5(0.0)●	2.5(0.0)●	34.9(1.7)●	50.1(2.0)●	67.6(4.3)●	<b>84.4(2.7)○</b>
ISRL	<b>91.3(2.2)</b>	<b>89.7(2.2)</b>	<b>88.0(3.3)</b>	<b>85.7(4.4)</b>	<b>85.1(3.5)</b>	83.9(3.2)
WebKB						
Algorithm						
MVL-IV	51.4(13.2)●	27.1(16.7)●	15.9(2.7)●	16.4(3.1)●	15.1(1.8)●	12.3(2.0)●
DAIMC	41.3(16.0)●	38.0(19.0)●	28.2(14.1)●	32.9(16.0)●	17.7(16.8)●	12.7(14.4)●
IML-BDR	50.5(7.4)●	52.6(6.9)●	50.2(7.4)●	49.2(6.1)●	49.6(5.0)●	40.8(6.7)●
LMSC	46.7(12.2)●	30.4(17.3)●	27.7(14.8)●	25.9(16.3)●	11.5(7.7)●	3.2(2.8)●
LTMSC	0.6(0.0)●	0.4(0.0)●	0.4(0.0)●	0.4(0.0)●	0.8(0.8)●	0.4(0.1)●
CSMSC	0.4(0.0)●	11.3(17.8)●	25.7(17.2)●	29.1(11.7)●	35.0(12.8)●	15.1(12.4)●
ISRL	<b>75.4(0.4)</b>	<b>72.9(1.3)</b>	<b>69.0(2.6)</b>	<b>66.6(3.2)</b>	<b>60.9(3.2)</b>	<b>54.6(2.8)</b>
Digits						
Algorithm						
MVL-IV	69.1(11.8)●	70.2(9.4)●	72.2(5.3)●	68.6(6.7)●	70.5(6.4)●	73.2(6.3)●
DAIMC	82.3(4.1)●	85.8(1.8)●	84.5(2.8)●	84.4(2.7)●	85.6(1.7)○	85.8(2.2)○
IML-BDR	70.9(3.2)●	72.8(1.9)●	70.6(1.7)●	69.0(2.1)●	65.9(3.0)●	60.1(2.4)●
LMSC	79.1(2.0)●	78.3(2.4)●	78.9(2.2)●	79.8(2.5)●	80.6(1.8)●	79.6(1.8)●
LTMSC	86.6(0.0)●	88.3(1.5)○	87.8(1.3)○	86.8(1.7)○	85.4(1.2)●	81.9(1.8)●
CSMSC	83.1(0.1)●	84.5(0.8)●	84.7(0.8)●	83.5(1.2)●	79.5(2.1)●	79.2(1.4)●
ISRL	<b>89.8(0.6)</b>	<b>88.9(0.6)</b>	<b>88.8(1.0)</b>	<b>87.8(1.0)</b>	<b>86.6(1.2)</b>	<b>85.9(0.8)</b>

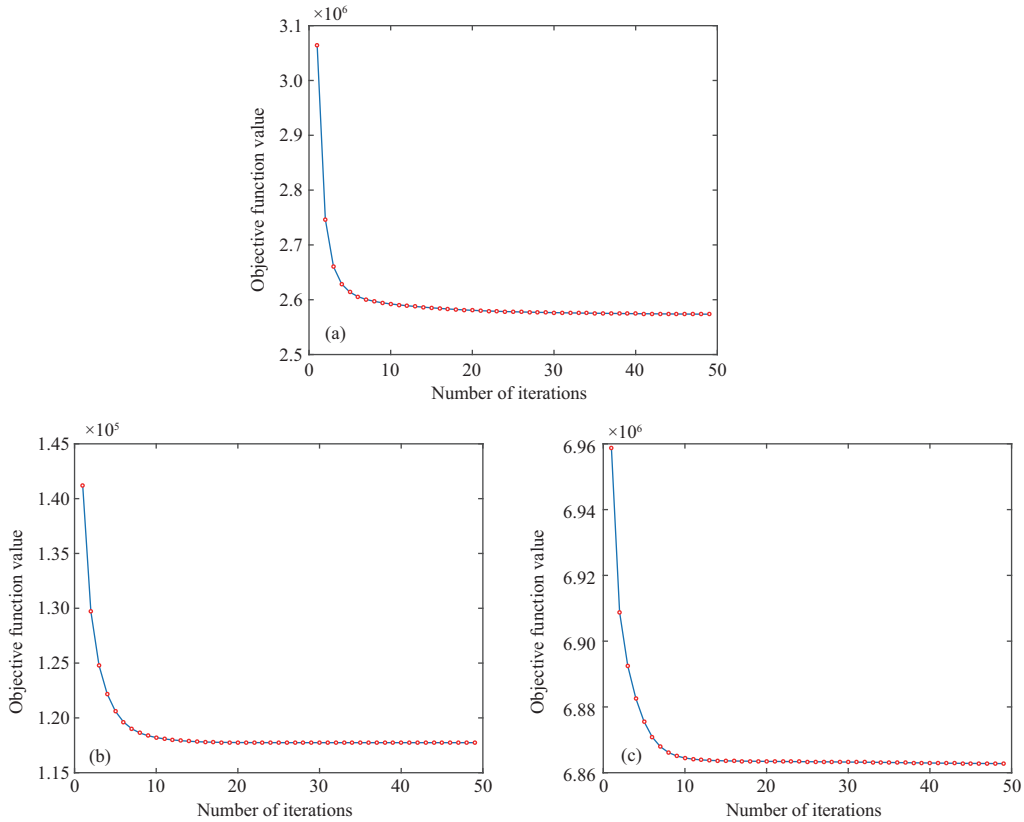


图 1 (网络版彩图) ISRL 的收敛曲线

Figure 1 (Color online) The convergence curve of ISRL. (a) MSRC-v1; (a) Ionosphere; (c) NNSpt

### 5.5 计算时间分析

为了验证 ISRL 的计算效率, 表 5 报告了 ISRL 及其对比方法在 6 个数据集上的运行时间. 实验设定缺失比例为 20%, 并基于预先确定的参数将每个方法运行 20 次后取平均时间. 此外, 由于对比方法 LMSC, LTMSC 与 CSMSC 不能直接处理缺失数据, 因此实验将矩阵补全所花费的时间也加入对应算法之中. 从表 5 可以看出, 在 Ionosphere 数据集上 ISRL 不如 MVL-IV 和 DAIMC, 这可能是由于 Ionosphere 为小数据集, 数据个数并不是影响计算复杂度的主要因素. 在较大的数据集上, 例如 Digits 上, 本文提出的 ISRL 方法计算时间方面优势更加明显, 这验证了 ISRL 的线性计算复杂度的有效性.

## 6 结束语

针对无标注且视图特征任意缺失情形下的多视图聚类问题, 本文研究如何同时进行数据补全和多视图子空间聚类, 设计了可行的改造方式, 提出了基于独立自表达学习的不完全多视图聚类方法 ISRL. ISRL 在每个视图上基于自表达重构对数据进行补全的同时, 学习一个视图共有的自表达矩阵使得各个视图的信息可以交互. 不同于传统方法, ISRL 并不显式地计算自表达矩阵, 而是基于低秩假设, 学习反映自表达矩阵潜在结构的表示矩阵. 为了区分视图的不同重要性, ISRL 设计了一个考虑视图互

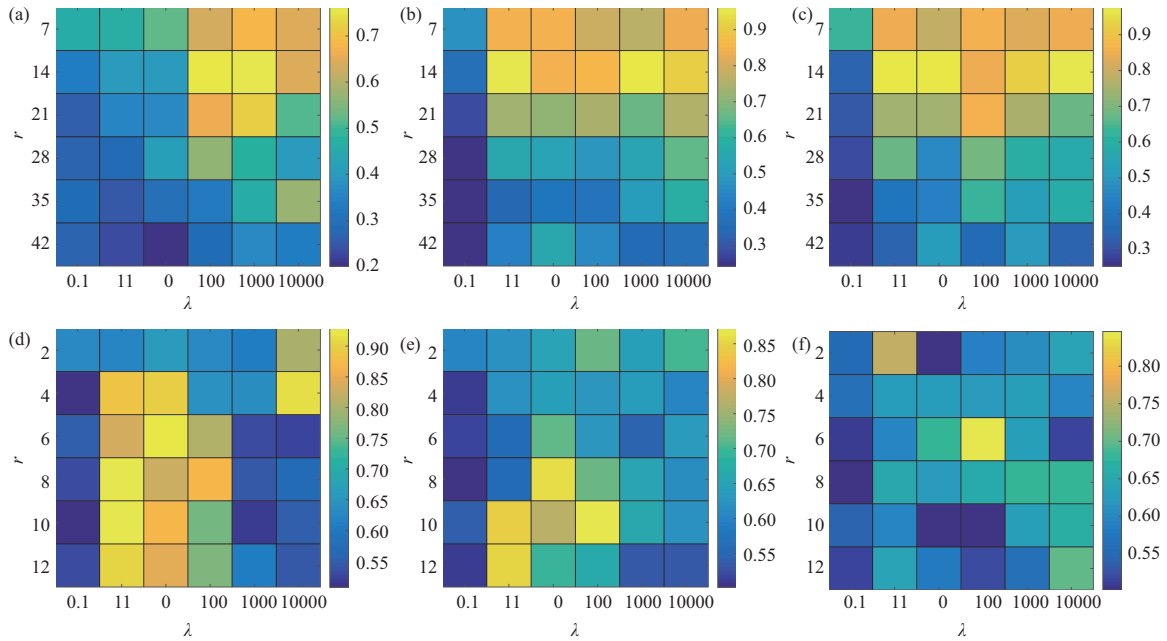


图 2 ISRL 对参数  $\mu$ ,  $r$  和  $\lambda$  的敏感度分析

Figure 2 The sensitivity analysis of ISRL with parameters  $\mu$ ,  $r$  and  $\lambda$ . (a) MSRC-v1:  $\mu = r$ ; (b) MSRC-v1:  $\mu = 10^{0.5}r$ ; (c) MSRC-v1:  $\mu = 10r$ ; (d) Ionosphere:  $\mu = r$ ; (e) Ionosphere:  $\mu = 10^{0.5}r$ ; (f) Ionosphere:  $\mu = 10r$

表 5 ISRL 及其对比方法在 6 个数据集上的运行时间 (s)

Table 5 Average runtime comparison (s) on 6 datasets

Dataset	MVL-IV	DAIMC	IML-BDR	LMSC	LTMSC	CSMSC	ISRL
MSRC-v1	1.1582	6.0947	0.7913	1.7175	3.6664	1.0469	0.3576
Ionosphere	0.0315	0.0812	0.6358	1.5895	1.5703	0.1344	0.1261
NNSpt	1.7509	3.1533	4.5344	9.5049	13.8337	1.2181	0.4306
Protein	4.3931	510.2782	6.4285	13.4356	27.5087	6.3576	2.0101
WebKB	3.8727	32.5769	7.7573	18.6160	31.1849	4.3133	1.2825
Digits	3.5635	3.3618	100.0725	93.4729	326.9522	5.2347	0.7072

补性的视图权重学习机制. 此外, 本文针对 ISRL 的优化问题提出了一个高效的求解算法, 设计了相应的实验对 ISRL 算法的有效性进行了验证. 在未来的研究中, 拟从以下两方面来继续深入本文的工作: (1) 在已有理论的基础上, 尝试直接分析 ISRL 算法的整体目标函数的解的性质; (2) 考虑深入分析  $l_{2,1}$  对 ISRL 算法的影响及其隐含的运行机理.

参考文献

- 1 Yang Y, Wang H. Multi-view clustering: a survey. *Big Data Min Anal*, 2018, 1: 83–107
- 2 Fan R D, Luo T J, Zhuge W Z, et al. Multi-view subspace learning via bidirectional sparsity. *Pattern Recognit*, 2020, 108: 107524
- 3 Tang C, Zhu X Z, Liu X W, et al. Cross-view local structure preserved diversity and consensus learning for multi-view unsupervised feature selection. In: *Proceedings of AAAI Conference on Artificial Intelligence*, 2019. 5101–5108
- 4 Tang C, Zhu X Z, Liu X W, et al. Learning a joint affinity graph for multiview subspace clustering. *IEEE Trans*



- Multimedia, 2019, 21: 1724–1736
- 5 Lu Y H, Zhen M M, Fang T. Multi-view based neural network for semantic segmentation on 3D scenes. *Sci China Inf Sci*, 2019, 62: 229101
  - 6 Zhang C Q, Fu H Z, Liu S, et al. Low-rank tensor constrained multi-view subspace clustering. In: *Proceedings of IEEE International Conference on Computer Vision*, Santiago, 2015. 1582–1590
  - 7 Wang P Y, Shui P P, Yu F G, et al. 3D shape classification based on convolutional neural networks fusing multi-view information. *Sci Sin Inform*, 2019, 49: 436–449 [王鹏宇, 水盼盼, 余锋根, 等. 基于多视角卷积神经网络的三维模型分类方法. *中国科学: 信息科学*, 2019, 49: 436–449]
  - 8 Gao H C, Nie F P, Li X L, et al. Multi-view subspace clustering. In: *Proceedings of IEEE International Conference on Computer Vision*, Santiago, 2015. 4238–4246
  - 9 Zhuge W Z, Hou C P, Jiao Y Y, et al. Robust auto-weighted multi-view subspace clustering with common subspace representation matrix. *PLoS ONE*, 2017, 12: 0176769
  - 10 Li X L, Chen M L, Wang Q. Multiview-based group behavior analysis in optical image sequence. *Sci Sin Inform*, 2018, 48: 1227–1241 [李学龙, 陈穆林, 王琦. 光学影像序列中基于多视角聚类的群组行为分析. *中国科学: 信息科学*, 2018, 48: 1227–1241]
  - 11 Tang C, Liu X W, Zhu X Z, et al. CGD: multi-view clustering via cross-view graph diffusion. In: *Proceedings of AAAI Conference on Artificial Intelligence*, 2020. 5924–5931
  - 12 Liu X L, Ye Z H. Partial multi-view clustering based on StarGAN and subspace learning. *J South China Univ Technol (Nat Sci Edit)*, 2020, 48: 87–98 [刘小兰, 叶泽慧. 基于 StarGAN 和子空间学习的缺失多视图聚类. *华南理工大学学报 (自然科学版)*, 2020, 48: 87–98]
  - 13 Liu X W, Gao W, Zhu X Z, et al. Multiple kernel k-means with incomplete kernels. *IEEE Trans Pattern Anal Mach Intell*, 2020, 42: 1191–1204
  - 14 Liu X W, Zhu X Z, Li M M, et al. Late fusion incomplete multi-view clustering. *IEEE Trans Pattern Anal Mach Intell*, 2019, 41: 2410–2423
  - 15 Xu N, Guo Y Q, Zheng X, et al. Partial multi-view subspace clustering. In: *Proceedings of the 26th ACM International Conference on Multimedia*, Seoul, 2018. 1794–1801
  - 16 Yin Q Y, Wu S, Wang L. Incomplete multi-view clustering via subspace learning. In: *Proceedings of Conference on Information and Knowledge Management*, Melbourne, 2015. 382–392
  - 17 Yin Q Y, Wu S, Wang L. Unified subspace learning for incomplete and unlabeled multi-view data. *Pattern Recognit*, 2017, 67: 313–327
  - 18 Zhao B Y, Zhang C Q, Chen L, et al. Generative model for partial multi-view clustering. *Acta Automat Sin*, 2020. doi:10.16383/j.aas.c200121 [赵博宇, 张长青, 陈蕾, 等. 生成式不完整多视图数据聚类. *自动化学报*, 2020. doi: 10.16383/j.aas.c200121]
  - 19 Li S Y, Jiang Y, Zhou Z H. Partial multi-view clustering. In: *Proceedings of AAAI Conference on Artificial Intelligence*, 2014. 1968–1974
  - 20 Hu M L, Chen S C. Doubly aligned incomplete multi-view clustering. In: *Proceedings of International Joint Conference on Artificial Intelligence*, Stockholm, 2018. 2262–2268
  - 21 Wang H, Zong L L, Liu B, et al. Spectral perturbation meets incomplete multiview data. In: *Proceedings of International Joint Conference on Artificial Intelligence*, Macao, 2019. 3677–3683
  - 22 Xu C, Tao D C, Xu C. Multi-view learning with incomplete views. *IEEE Trans Image Process*, 2015, 24: 5812–5825
  - 23 Tao H, Hou C P, Yi D Y, et al. Unsupervised maximum margin incomplete multi-view clustering. In: *Proceedings of the 1st CCF International Conference on Artificial Intelligence*, Jinan, 2018. 13–25
  - 24 Tao H, Hou C P, Yi D Y, et al. Joint embedding learning and low-rank approximation: a framework for incomplete multi-view learning. 2019. ArXiv:1812.10012
  - 25 Liu G C, Lin Z C, Yan S C, et al. Robust recovery of subspace structures by low-rank representation. *IEEE Trans Pattern Anal Mach Intell*, 2013, 35: 171–184
  - 26 Cao X C, Zhang C Q, Fu H Z, et al. Diversity-induced multi-view subspace clustering. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 2015. 586–594
  - 27 Zhang C Q, Fu H Z, Hu Q H, et al. Generalized latent multi-view subspace clustering. *IEEE Trans Pattern Anal Mach Intell*, 2020, 42: 86–99

- 28 Zhang C Q, Hu Q H, Fu H Z, et al. Latent multi-view subspace clustering. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017. 4333–4341
- 29 Luo S R, Zhang C Q, Zhang W, et al. Consistent and specific multi-view subspace clustering. In: Proceedings of AAAI Conference on Artificial Intelligence, New Orleans, 2018. 3730–3737
- 30 Wang Y X, Xu H, Leng C L. Provable subspace clustering: when LRR meets SSC. In: Proceedings of Conference on Neural Information Processing Systems, 2013. 64–72
- 31 Gretton A, Bousquet O, Smola A, et al. Measuring statistical dependence with Hilbert-Schmidt norms. In: Proceedings of International Conference on Algorithmic Learning Theory, Singapore, 2005. 63–78
- 32 Nie F P, Wang X Q, Huang H. Multiclass capped  $\ell_p$ -norm SVM for robust classifications. In: Proceedings of AAAI Conference on Artificial Intelligence, San Francisco, 2017. 2415–2421
- 33 Nie F P, Zhang R, Li X L. A generalized power iteration method for solving quadratic problem on the Stiefel manifold. *Sci China Inf Sci*, 2017, 60: 112101
- 34 Nie F P, Huang H, Cai X, et al. Efficient and robust feature selection via joint  $\ell_{2,1}$ -norms minimization. In: Proceedings of Conference on Neural Information Processing Systems, Vancouver, 2010. 1813–1821
- 35 Fan J C, Cheng J Y. Matrix completion by deep matrix factorization. *Neural Netw*, 2018, 98: 34–41

## Incomplete multi-view clustering via independent self-representation learning

Wenzhang ZHUGE<sup>1,2†</sup>, Ruidong FAN<sup>1†</sup>, Tingjin LUO<sup>1</sup>, Hong TAO<sup>1</sup> & Chenping HOU<sup>1\*</sup>

1. *College of Science, National University of Defense Technology, Changsha 410073, China;*

2. *Naval Research Institute, Beijing 100161, China*

\* Corresponding author. E-mail: hcpnudt@hotmail.com

† Equal contribution

**Abstract** Incomplete multi-view clustering is a learning paradigm that combines heterogeneous and incomplete characteristics of multi-view data to obtain data structure and improve clustering performance. In practical application, each view will also be affected by missing values and outliers, in addition to missing some complete samples, which makes most traditional incomplete multi-view clustering methods ineffective. An incomplete multi-view clustering method based on independent self-representation learning is proposed to solve the above issue. The proposed method fills in the missing features and learns the unique self-representation matrix of each view using self-representation reconstruction. Then, to better mine the structure, low-rank constraints are added to the self-representation matrix, and the diversity among different views is measured by introducing the Hilbert-Schmidt independence criterion. Experimental results on multiple data sets show that the proposed method can achieve better clustering than other advanced methods in most cases.

**Keywords** incomplete multi-view clustering, arbitrary missing feature, self-representation, diversity