

数据表示的相关性度量方法

刘吉元¹⁾ 刘新旺²⁾ 蔡志平²⁾ 祝 恩²⁾ 鞠儒生¹⁾

¹⁾ (国防科技大学系统工程学院 长沙 410072)

²⁾ (国防科技大学计算机学院 长沙 410072)

摘 要 数据表示之间的相关性度量是机器学习和人工智能技术的基石. 然而现有的度量方法要么数据表示的全局信息考虑不足, 要么复杂度较高, 限制了相关技术的进一步发展. 为解决上述问题, 本文提出一种数据表示的对齐度量方法, 称为表示对齐 (Representation Alignment, RA). 此度量方法能够全局性地衡量任意两个数据表示之间的相关性, 且其在样本数量和特征维度上的计算复杂度均为线性. 在此基础上, 我们将 RA 扩展到了对比学习领域, 进一步提出了基于对比的表示对齐 (Contrastive Representation Alignment, CRA) 度量方法. 上述两个度量方法可自然地用于多视图学习场景, 即可通过最大化不同视图数据之间的 RA 和 CRA 来融合各个视图之间的信息. 为验证这一点, 我们还提出了两个新颖多视图聚类算法, 并在七个基准数据集上取得了领先的聚类性能.

关键词 表示对齐; 对比学习; 多视图聚类

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2024.01568

On the Correlation Measurement of Data Representations

LIU Ji-Yuan¹⁾ LIU Xin-Wang²⁾ CAI Zhi-Ping²⁾ ZHU En²⁾ JU Ru-Sheng¹⁾

¹⁾ (College of Systems Engineering, National University of Defense Technology, Changsha 410072)

²⁾ (College of Computer, National University of Defense Technology, Changsha 410072)

Abstract The correlation measurement between data representations is the basis of machine learning and artificial intelligence techniques. However, existing methods either overlook the global information on the involved representations partially or scale poorly. To address the issue, this paper proposes the Representation Alignment (RA) to globally measure the correlation between two arbitrary data representations in linear time to the dimension and the number of samples. We further extend the representation alignment to the Contrastive Representation Alignment (CRA), where the concerned representations are two augmentations (in the scope of contrastive learning) of data samples. In addition, the two alignments can be naturally applied in multi-view learning systems by maximizing them on the latent representations of different data views so as to integrate their complementary information. To validate this, we develop two novel multi-view clustering algorithms and achieve state-of-the-art performance on seven benchmark datasets.

Keywords representation alignment; contrastive learning; multi-view clustering

收稿日期: 2023-08-21; 在线发布日期: 2024-04-17. 本课题得到国家自然科学基金青年科学基金项目 (No. 62306324)、国家自然科学基金面上项目 (No. 62376279, No. 62276271)、国家自然科学基金国家杰出青年科学基金项目 (No. 62325604) 资助. 刘吉元, 博士, 讲师, 中国计算机学会 (CCF) 会员, 主要研究领域为多视图聚类、多视图异常检测、联邦学习. E-mail: liujiyuan13@nudt.edu.cn. 刘新旺 (通信作者), 博士, 教授, 中国计算机学会 (CCF) 高级会员, 主要研究领域为多视图聚类、核学习. E-mail: xinwangliu@nudt.edu.cn. 蔡志平, 博士, 教授, 中国计算机学会 (CCF) 杰出会员, 主要研究领域为大数据分析、网络安全、分散计算. 祝 恩, 博士, 教授, 中国计算机学会 (CCF) 高级会员, 主要研究领域为聚类算法、异常检测、模式识别. 鞠儒生, 博士, 副教授, 主要研究领域为数据挖掘、计算机仿真.

1 引言

数据表示之间的相关性度量是机器学习和人工智能技术的基石,现已被广泛地应用在各个领域,如图像检索^[1, 2]、广告推荐^[3]、生物基因分析^[4]等.文献中常用的相关性度量方法可大致分为三类,即余弦相似度^[5]、核对齐^[6]和典型关联性分析(Canonical Correlation Analysis, CCA)^[7, 8].首先,余弦相似度的计算较为简单,应用最为广泛,常作为一个基本组件嵌入到各类机器学习算法中,如文献^[9, 10].其次,核对齐是用于计算两个核矩阵的相似度(即相关性)的方法,首次定义于文献^[6].在此基础上,Cortes 等人^[11]还额外对核矩阵进行了中心化操作,从而提出了中心核对齐的概念.现如今,核对齐已被广泛地应用于各类机器学习算法中,如核矩阵的选择^[12]、组合^[13, 14]、超参数微调^[15]等领域.其中,多核学习范式^[13]影响范围最广.第三类方法 CCA 首次提出于文献^[16],用于衡量人类运动天赋和学习天赋间的关系.经过几十年的发展,CCA 方法现如今已被广泛地应用于各类统计学习算法中.其中,较为典型的有多视图典型关联性分析方法^[17, 18]、深度典型关联性分析方法^[19-21].

虽然上述三类相关性度量方法被广泛地应用于相关学术研究和现实生活中,但仍存在不可忽视的缺陷.首先,数据的余弦相似度未能充分考虑数据表示的全局信息,即忽略了其余数据表示对目标表示相似度计算的修正作用.其次,核对齐度量方法不仅需根据数据表示计算核矩阵,还需计算核矩阵之间的 F 范数,故其计算至少需要平方复杂度.最后,CCA 方法需首先计算各个数据表示对应的映射向量,复杂度是数据表示维度的立方(计算过程可见文献^[7]).综合上述情况,如果机器学习算法中使用了核对齐或 CCA 任意一种度量方法,则其计算复杂度通常为平方甚至立方,如多核学习算法^[13, 22, 23]等.

为克服上述缺点,我们提出了一种线性的相关性度量方法,即表示对齐(Representation Alignment, RA).为说明该方法能更全面地考虑了数据的全局信息,我们分析了它与余弦相似度的计算过程,并进行了详细对比.此外,我们所提出的表示对齐方法还具有理论优势,即满足中心界理论,保证了其从训练到测试的泛化能力.再者,我们还考虑对比学习场景,并将表示对齐拓展为基于对比的表示对齐(Contrastive Representation Alignment,

CRA).在此基础上,我们详细分析了上述两个度量方法的性质以及它们之间的区别和联系.更进一步地,我们将上述两个度量应用于多视图聚类领域,即通过最大化不同视图数据表示之间的 RA 和 CRA,提出了两个新颖的算法.值得注意的是,所提出的算法均为线性复杂度,故适用于大规模聚类任务.最后,我们将上述两个算法与文献中常用的基准算法进行了对比.实验表明,前者具有更优的聚类性能.

总而言之,我们提出一个新颖的数据相关性度量方法 RA,并将之拓展到了对比学习场景,从而提出了 CRA 方法.不仅如此,我们还展示了它们在多视图聚类领域的应用,并进行了实验验证.值得注意的是,所提出的 RA 和 CRA 方法并非仅限于本文提出的两个多视图聚类算法,而是可以广泛地应用于表示学习、单视图/多视图分类/聚类/异常检测等机器学习算法领域.

2 相关研究

本节简要介绍了文中所涉及的关键概念及其相关研究,包括相关性度量方法、对比学习和多视图聚类三方面内容.

2.1 相关性度量方法

如引言所述,常用的相关性度量方法可大致分为三类.在相关研究中,最为常见和直接的度量方法便是计算各个数据表示之间的余弦相似度^[5],其表达式为

$$\hat{\rho}_a(\mathbf{H}, \mathbf{H}') = \frac{1}{n} \sum_{i=1}^n \frac{\langle \mathbf{h}_i, \mathbf{h}'_i \rangle}{\sqrt{\langle \mathbf{h}_i, \mathbf{h}_i \rangle \langle \mathbf{h}'_i, \mathbf{h}'_i \rangle}} \quad (1)$$

其中, \mathbf{H} 和 \mathbf{H}' 是 n 个数据样本在 d 维空间的不同表示,且 \mathbf{h}_i 和 \mathbf{h}'_i 分别为数据表示 \mathbf{H} 和 \mathbf{H}' 的第 i 行.再者,另一种普遍使用的方法是计算数据表示之间的核对齐^[6].令 \mathbf{K} 和 \mathbf{K}' 表示两个数据表示 \mathbf{H} 和 \mathbf{H}' 对应的核矩阵,则核对齐可计算为

$$\hat{\rho}_k(\mathbf{K}, \mathbf{K}') = \frac{\langle \mathbf{K}, \mathbf{K}' \rangle_F}{\sqrt{\langle \mathbf{K}, \mathbf{K} \rangle_F \langle \mathbf{K}', \mathbf{K}' \rangle_F}} \quad (2)$$

其中, $\langle \mathbf{K}, \mathbf{K}' \rangle_F = \text{Tr}(\mathbf{K}\mathbf{K}'^T)$.除此之外,相关文献还通过使用典型关联分析^[7, 8]来度量数据表示之间的相关性,即

$$\hat{\rho}_c(\mathbf{H}, \mathbf{H}') = \max_a \frac{\langle \mathbf{H}\mathbf{a}, \mathbf{H}'\mathbf{a}' \rangle}{\sqrt{\langle \mathbf{H}\mathbf{a}, \mathbf{H}\mathbf{a} \rangle \langle \mathbf{H}'\mathbf{a}', \mathbf{H}'\mathbf{a}' \rangle}} \quad (3)$$

其中, \mathbf{a} 和 \mathbf{a}' 分别为数据表示 \mathbf{H} 和 \mathbf{H}' 所对应的映射

向量,可通过矩阵的奇异值分解(Singular Vector Decomposition, SVD)进行求解.

2.2 对比学习

对比学习这一概念首先提出于文献[10],用于解决计算机视觉特征学习问题.因其能够大幅度提高深度网络的表示能力,这种学习方法迅速吸引了工业界和学术界的广泛关注^[24-26].简言之,对比学习通过最大化同类数据样本(正样本对)之间的相似度,并最小化非同类数据样本(负样本对)间的相似度,从而增强数据样本隐藏表示的判别性.记样本隐藏表示 \mathbf{z}_i 和 $\mathbf{z}_{j(i)}$ 为正样本对,且 \mathbf{z}_i 和 \mathbf{z}_k ($k \neq j(i)$) 为负样本对,那么第 i 个样本的损失函数可写为

$$\ell_{i,j(i)} = -\log \frac{\exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_{j(i)})/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{k \neq i} \exp(\text{sim}(\mathbf{z}_i, \mathbf{z}_k)/\tau)} \quad (4)$$

其中, $\mathbb{1}_{k \neq i} \in \{0, 1\}$ 为指示函数, τ 被称之为温度超参数,并且

$$\text{sim}(\mathbf{z}_i, \mathbf{z}_j) = \frac{\mathbf{z}_i^T \mathbf{z}_j}{\|\mathbf{z}_i\| \|\mathbf{z}_j\|} \quad (5)$$

在后续研究中, Yeh 等人^[27] 移除了上式分母中的正样本对,提高了算法的性能.此外, Khosla 等人^[26] 通过引入监督信号,将同类的样本视为正样本对,进一步提高了性能.本文在第 4 节中将对比学习的思想应用于表示对齐,提出了基于对比的表示对齐度量方法.

2.3 多视图聚类

在现实生活中,存在各式各样的多视图数据.例如,病人在医院进行肺部检查时,常会进行核酸检测、血液分析和核磁共振扫描;在构建人物画像时,互联网公司通常会收集用户的社交数据、购物数据、医疗数据等;人们在辨别小动物时,经常会从它的外形、声音、生活习性等多维度特征进行判断.上述这种由多个来源和特征共同构成的数据被称之为多视图数据^[28].而对上述多视图数据进行聚类分析的方法,被统称为多视图聚类方法^[28].在相关文献中,多视图聚类方法大多由传统的聚类方法发展而来,如核聚类、谱聚类等.常见的方法可总结为五类,即多核聚类^[29,30]、多视图谱聚类^[31]、多视图矩阵分解^[32]、多视图子空间聚类^[33]和深度聚类^[34].本文在第 5 节中正是基于多核聚类提出了两种高效的多视图聚类方法.

3 表示对齐

本节给出了表示对齐的定义,探索了它的性质,

并将之与常用的余弦相似度、核对齐和典型关联分析方法进行了对比,最后从理论上推导了表示对齐的中心界.

定义 1. 表示对齐(RA). 对于同一空间中的任意两个非零数据表示 \mathbf{H} 和 \mathbf{H}' , 其表示对齐可计算为

$$\hat{\rho}(\mathbf{H}, \mathbf{H}') = \frac{\langle \mathbf{H}, \mathbf{H}' \rangle_F}{\sqrt{\langle \mathbf{H}, \mathbf{H} \rangle_F \langle \mathbf{H}', \mathbf{H}' \rangle_F}} \quad (6)$$

由定义 1 可将表示对齐的期望记为

$$\rho = \mathbb{E}[\hat{\rho}(\mathbf{H}, \mathbf{H}')] \quad (7)$$

其中, \mathbf{H} 和 \mathbf{H}' 表示从同一空间两个任意数据分布中采样得到的数据表示. 同时, \mathbf{H} 和 \mathbf{H}' 是 n 个数据样本的 d 维表示,在后文中其维度大小均被默认为 $n \times d$. 此外,表示对齐的取值范围可见性质 1.

性质 1. 表示对齐 $\hat{\rho}$ 的取值范围为 $[-1, 1]$, 其中 1 表示两个数据表示之间最为相关, -1 表示最不相关.

证明. 令 \mathbf{h}_i 和 \mathbf{h}'_i 分别为矩阵 \mathbf{H} 和 \mathbf{H}' 的第 i 行,可直接应用柯西不等式,证明过程如下.

$$\begin{aligned} \langle \mathbf{H}, \mathbf{H}' \rangle_F &= \left(\sum_{i=1}^n \mathbf{h}_i \mathbf{h}'_i^T \right)^2 \\ &\leq \left(\sum_{i=1}^n \sqrt{\mathbf{h}_i \mathbf{h}_i^T} \sqrt{\mathbf{h}'_i \mathbf{h}'_i^T} \right)^2 \\ &\leq \left(\sum_{i=1}^n \mathbf{h}_i \mathbf{h}_i^T \right) \left(\sum_{i=1}^n \mathbf{h}'_i \mathbf{h}'_i^T \right) \\ &= \langle \mathbf{H}, \mathbf{H} \rangle_F \langle \mathbf{H}', \mathbf{H}' \rangle_F \quad (8) \end{aligned}$$

与余弦相似度 $\hat{\rho}_a$ 相比,表示对齐 $\hat{\rho}$ 的优势在于其更全面地考虑了数据的全局信息.为更好地说明这一点,可将余弦相似度 $\hat{\rho}_a$ 和表示对齐 $\hat{\rho}$ 改写为 n 个独立项之和,即

$$\hat{\rho}_a = \sum_{i=1}^n \frac{1}{n} \cdot \left\langle \frac{\mathbf{h}_i}{L_2(\mathbf{h}_i)}, \frac{\mathbf{h}'_i}{L_2(\mathbf{h}'_i)} \right\rangle \quad (9)$$

和

$$\hat{\rho} = \sum_{i=1}^n \left\langle \frac{\mathbf{h}_i}{L_F(\mathbf{H})}, \frac{\mathbf{h}'_i}{L_F(\mathbf{H}')} \right\rangle \quad (10)$$

其中, $L_2(\cdot)$ 和 $L_F(\cdot)$ 分别表示向量的 L_2 范数和矩阵的 F 范数.对比上述两个式子,全局信息的引入在以下两个地方:1) n 个独立项的求和;2) 每个数据表示的标准化. $\hat{\rho}_a$ 虽然在第一处考虑了其余数据表示,但是仅使用数据表示自身的 L_2 范数来对其进行标准化.反观 $\hat{\rho}$, 其在第一处考虑全局信息的基础上,使用所有数据表示的 F 范数来进行标准化,即使用其他样本的信息对目标样本进行了标准化,全局信息考虑的更为全面.从单位球投影的角度来看, $\hat{\rho}_a$ 将数据样本的向量表示投影到单位球的表面.而

$\hat{\rho}$ 则将其投影到一个半径为 η ($\eta \leq 1$) 的球内,同时保留了向量的大小和它们之间的角度关系.为增强说服力,我们进行了实验对比,并将结果展示如图 1. 具体而言, $\hat{\rho}_a$ 和 $\hat{\rho}$ 可通过下列三个场景进行计算:

- (1) 数据表示 \mathbf{H} 和 \mathbf{H}' 均分别从第一个高斯分布中随机采样得到;
- (2) 数据表示 \mathbf{H} 和 \mathbf{H}' 均分别从第二个高斯分布中随机采样得到;
- (3) 数据表示 \mathbf{H} 从第一个高斯分布中随机采样

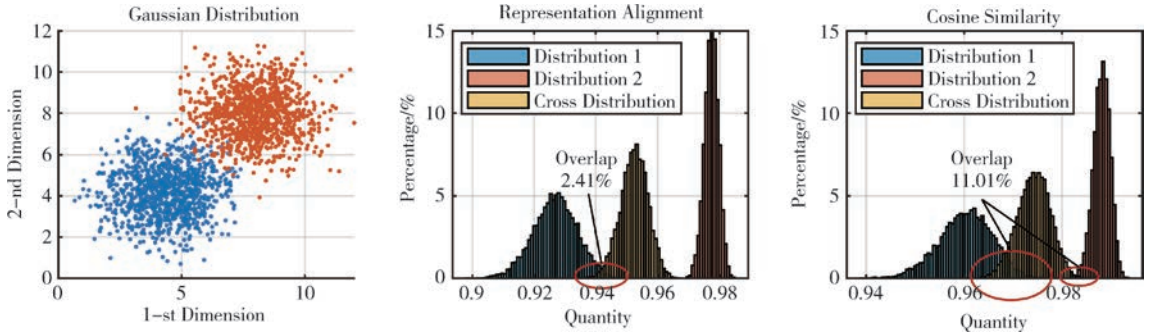


图 1 从左至右为二维空间中两个高斯分布(左)、表示对齐 $\hat{\rho}$ 的频率直方图(中)和余弦相似度均值 $\hat{\rho}_a$ 的频率直方图。在频率直方图中,左右两部分分别代表两个高斯分布内样本的相关性度量分布,中间部分代表两个高斯分布之间样本的相关性度量分布。

相较于核对齐和 CCA 度量方法,本文所提出的表示对齐具有更低的计算复杂度.具体而言,表示对齐在数据样本数量和表示维度上的计算复杂度均为线性.从公式(10)可以看出,计算数据表示的 F 范数和 n 个向量的内积只需 $O(dn)$ 的复杂度,即表示对齐的计算复杂度为 $O(dn)$.如使用核对齐来衡量数据表示的相关性,需 $O(dn^2)$ 的复杂度来计算与之对应的核矩阵,以及 $O(n^2)$ 的复杂度来计算公式(2).对于 CCA 方法而言,虽然其公式(3)的计算复杂度为 $O(dn)$,但是计算最优的投影矩阵 \mathbf{a} 和 \mathbf{a}' 需 $O(d^3n)$ 的复杂度[7].总而言之,相较于现有可选方案,数据对齐度量方法更适合处理大规模高维数据.

定理 1. 给定两个采样于 η 球和 η' 球的任意数据表示 \mathbf{H} 和 \mathbf{H}' ,它们的表示对齐经验估计分布在期望值的邻域,即以 $1-\sigma$ 的概率,下列不等式成立.

$$|\hat{\rho}(\mathbf{H}, \mathbf{H}') - \rho| \leq C(\mathbf{H}, \mathbf{H}', \eta, \eta') \sqrt{\frac{\ln(2/\sigma)}{2n}} \quad (11)$$

其中, C 为与变量 n 和 σ 无关的函数项.

定理 1 给出了表示对齐的中心界,说明了表示对齐的经验估计与其期望值之间差值的上界是一个

得到,而数据表示 \mathbf{H}' 从第二个高斯分布中随机采样得到.

在多次重复采样后,我们计算出上述三个场景下的 $\hat{\rho}_a$ 和 $\hat{\rho}$ 值,并将其展示在图 1 中.可以观察到,相较于表示对齐 $\hat{\rho}$, $\hat{\rho}_a$ 在三个场景下的概率直方图呈现出更大的重叠.这说明仅使用余弦相似度来衡量数据相关性有更大的概率误分类样本点,即判别性较差.上述现象清晰地说明了充分考虑数据表示全局信息的优势.

指数下降函数.这意味着表示对齐这一个度量方法的泛化性较好,即其在测试集上的表现接近于在训练集上的表现,使得其应用更具可预测性[6].详细证明过程请见附录.

4 基于对比的表示对齐

本节将表示对比这一度量方法拓展到了对比学习领域,提出了基于对比的表示对齐(CRA)度量方法.给定 n 个数据样本 $\{\mathbf{x}_i\}_{i=1}^n$,对比学习方法将产生两组对照样本,即

$$\mathbf{h}_i = t(\mathbf{x}_i) \text{ 和 } \mathbf{h}'_i = t'(\mathbf{x}_i) \quad (12)$$

其中, $t \in \mathcal{T}$ 和 $t' \in \mathcal{T}'$ 为两组变换函数.之后,我们将两组样本进行一一对应,并将对应的样本对视为正样本对,而非对应的样本对视为负样本对.在算法的训练过程中,需最大化正样本对的相似度,且最小化负样本对的相似度.在此设定下,我们将上述对照样本的数据表示 \mathbf{h}_i 和 \mathbf{h}'_i 约束在一个 d 维空间中.如此一来,便可将 n 个数据样本写为矩阵形式,即 $\mathbf{H} \in \mathbb{R}^{n \times d}$ 和 $\mathbf{H}' \in \mathbb{R}^{n \times d}$.对应的相似度矩阵可定义如下.

定义 2. 相似度矩阵.对照样本 \mathbf{H} 和 \mathbf{H}' 的相

似度矩阵可计算为它们乘积的归一化,即

$$\begin{aligned} \mathbf{C} &= \frac{\mathbf{H}\mathbf{H}'^T}{\sqrt{\langle \mathbf{H}, \mathbf{H} \rangle_F \langle \mathbf{H}', \mathbf{H}' \rangle_F}} \\ &= \begin{bmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,n} \\ C_{2,1} & C_{2,2} & \cdots & C_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n,1} & C_{n,2} & \cdots & C_{n,n} \end{bmatrix} \end{aligned} \quad (13)$$

其中, $C_{i,j}$ 度量了数据向量 \mathbf{h}_i 和 \mathbf{h}'_i 的相似度,其表达式为

$$C_{i,j} = \frac{\langle \mathbf{h}_i, \mathbf{h}'_j \rangle}{\sqrt{\langle \mathbf{H}, \mathbf{H} \rangle_F \langle \mathbf{H}', \mathbf{H}' \rangle_F}} \quad (14)$$

不难看出,每个相似度 $C_{i,j}$ 均包含一个统一的分母 $\sqrt{\langle \mathbf{H}, \mathbf{H} \rangle_F \langle \mathbf{H}', \mathbf{H}' \rangle_F}$, 而其取值取决于不同的分子 $\langle \mathbf{h}_i, \mathbf{h}'_j \rangle$.

结合定义 1 和定义 2, 不难看出上述两组对照样本的表示对齐等于其相似度矩阵的迹,即

$$\hat{\rho}(\mathbf{H}, \mathbf{H}') = \text{Tr}(\mathbf{C}) \quad (15)$$

换言之,所提出的表示对齐仅取决于对应相似度矩阵的对角线元素,而与其余值无关. 接下来,我们将其扩展为基于对比的表示对齐来解决这一问题,详见定义 3.

定义 3. 基于对比的表示对齐 (CRA). 将数据表示 \mathbf{H} 和 \mathbf{H}' 视为两组对照样本,基于对比的表示对齐可计算为

$$\hat{\rho}(\mathbf{H}, \mathbf{H}') = \text{Tr}(\mathbf{C}\mathbf{W}) \quad (16)$$

其中, \mathbf{C} 为定义 2 中的相似度矩阵, \mathbf{W} 为正负样本对的指示矩阵. 记 ω_p 和 ω_n 为两个正实数,那么

$$\mathbf{W}_{i,j} = \begin{cases} \omega_p \\ -\omega_n \end{cases} \quad (17)$$

在无监督学习场景下,最常见的方式便是将同一个数据的不同对照样本视为正样本对,而不同数据的对照样本视为负样本对^[10]. 在有监督学习场景下,同类数据的不同对照样本被视为正样本对,而其余对照样本被视为负样本对^[26]. 在此设定下,基于对比的表示对齐的指示矩阵在无监督和有监督场景下分别为

$$\begin{aligned} \mathbf{W}_{in} &= (\omega_p + \omega_n)\mathbf{I} - \omega_n\mathbf{1}, \\ \mathbf{W}_{sup} &= (\omega_p + \omega_n)\mathbf{Y}\mathbf{Y}^T - \omega_n\mathbf{1}. \end{aligned} \quad (18)$$

其中, \mathbf{I} 为单位矩阵, $\mathbf{1}$ 为全 1 矩阵,且 \mathbf{Y} 是大小为 $n \times k$ (k 为类别数量)的 one-hot 标签矩阵.

当在实际应用中使用 CRA 度量方法时,可预先给定指示矩阵 \mathbf{W} . 另一种可选方案是将 ω_p 和 ω_n 设置为变量,并添加相应的约束条件,然后在算法的优化过程中进行求解. 此外,我们在性质 2 中给出了

CRA 的界.

性质 2. 基于对比的表示对齐的取值范围为 $[-\delta, \delta]$, 其中 $\delta = L_F(\mathbf{W})$.

证明. 由柯西不等式,可得

$$\begin{aligned} [\text{Tr}(\mathbf{C}\mathbf{W})]^2 &= \left(\sum_{i,j=1}^n \omega_{i,j} C_{i,j} \right)^2 \\ &\leq \sum_{i,j=1}^n \omega_{i,j}^2 \sum_{i,j=1}^n C_{i,j}^2. \end{aligned} \quad (19)$$

上述不等式的右侧为 n 个独立项之和,且

$$\begin{aligned} \sum_{i,j=1}^n C_{i,j}^2 &= \sum_{i,j=1}^n \frac{(\mathbf{h}_i \mathbf{h}'_j{}^T)^2}{\text{Tr}(\mathbf{H}\mathbf{H}^T)\text{Tr}(\mathbf{H}'\mathbf{H}'^T)} \\ &= \frac{\sum_{i,j=1}^n (\mathbf{h}_i \mathbf{h}'_j{}^T)^2}{\left(\sum_{i=1}^n \mathbf{h}_i \mathbf{h}_i^T \right) \left(\sum_{i=1}^n \mathbf{h}'_i \mathbf{h}'_i{}^T \right)} \\ &= \frac{\sum_{i,j=1}^n (\mathbf{h}_i \mathbf{h}'_j{}^T)^2}{\sum_{i,j=1}^n (\mathbf{h}_i \mathbf{h}_i^T) (\mathbf{h}'_j \mathbf{h}'_j{}^T)} \leq 1. \end{aligned} \quad (20)$$

其中,上式的最后一步因下列柯西不等式而成立.

$$(\mathbf{h}_i \mathbf{h}'_j{}^T)^2 \leq (\mathbf{h}_i \mathbf{h}_i^T) (\mathbf{h}'_j \mathbf{h}'_j{}^T). \quad (21)$$

与上节中的表示对齐类似,基于对比的表示对齐的计算复杂度在数据样本数量和表示维度上均为线性. 具体而言,由定义 3 可得

$$\begin{aligned} \hat{\rho}(\mathbf{H}, \mathbf{H}') &= \text{Tr}(\mathbf{C}\mathbf{W}) \\ &= \frac{\text{Tr}(\mathbf{H}\mathbf{H}'^T\mathbf{W})}{\sqrt{\langle \mathbf{H}, \mathbf{H} \rangle_F \langle \mathbf{H}', \mathbf{H}' \rangle_F}}. \end{aligned} \quad (22)$$

在上式中,分母的计算只需 $O(dn)$ 复杂度. 同时,根据公式(18),上式的分子可计算为

$$\begin{aligned} \text{Tr}(\mathbf{H}\mathbf{H}'^T\mathbf{W}_{in}) &= (\omega_p + \omega_n)\text{Tr}(\mathbf{H}\mathbf{H}'^T) \\ &\quad - \omega_n \text{Tr}[(\mathbf{1}\mathbf{H})\mathbf{H}'^T] \end{aligned} \quad (23)$$

和

$$\begin{aligned} \text{Tr}(\mathbf{H}\mathbf{H}'^T\mathbf{W}_{sup}) &= (\omega_p + \omega_n)\text{Tr}(\mathbf{H}\mathbf{H}'^T\mathbf{Y}\mathbf{Y}^T) \\ &\quad - \omega_n \text{Tr}[(\mathbf{1}\mathbf{H})\mathbf{H}'^T] \end{aligned} \quad (24)$$

其中, $\text{Tr}(\mathbf{H}\mathbf{H}'^T)$ 的计算复杂度为 $O(dn)$. 同时,矩阵 $\mathbf{1}\mathbf{H}$ 的每行元素等于矩阵 \mathbf{H} 的列和,所以计算 $\text{Tr}[(\mathbf{1}\mathbf{H})\mathbf{H}'^T]$ 也仅需 $O(dn)$ 复杂度. 此外,

$$\text{Tr}(\mathbf{H}\mathbf{H}'^T\mathbf{Y}\mathbf{Y}^T) = \text{Tr}(\mathbf{B}^T\mathbf{B}'). \quad (25)$$

其中,

$$\mathbf{B} = \mathbf{H}^T\mathbf{Y}', \mathbf{B}' = \mathbf{H}'^T\mathbf{Y}. \quad (26)$$

不难看出, \mathbf{B} 和 \mathbf{B}' 可以通过索引和加和运算替代. 再者, $\text{Tr}(\mathbf{B}^T\mathbf{B}')$ 的计算只需 $O(kd)$ 的复杂度. 总而言之,公式(23)和公式(24)的计算复杂度均为 $O(kd)$, 意味着基于对比的表示对齐的计算复杂度为 $O(kd)$.

表示对齐和基于对比的表示对齐两个度量方法

之间的关系可表达为

$$\hat{Q}(\mathbf{H}, \mathbf{H}') = \omega_{\rho} \hat{\rho}(\mathbf{H}, \mathbf{H}') + \text{Tr}(\mathbf{C}\mathbf{W}'). \quad (27)$$

其中, $\mathbf{W}' = \mathbf{W} - \omega_{\rho} \mathbf{I}$. 结合公式(18), \mathbf{W}' 可改写为

$$\begin{aligned} \mathbf{W}'_{in} &= -\omega_n (\mathbf{1} - \mathbf{I}), \\ \mathbf{W}'_{sup} &= -\omega_p (\mathbf{I} - \mathbf{Y}\mathbf{Y}^T) - \omega_n (\mathbf{1} - \mathbf{Y}\mathbf{Y}^T). \end{aligned} \quad (28)$$

由公式(28), 易得 \mathbf{W}' 的对角线元素为 0. 这说明了相较于表示对齐, 基于对比的表示对齐包含一个额外项 $\text{Tr}(\mathbf{C}\mathbf{W}')$. 换言之, 后者保留了相似度矩阵 \mathbf{C} 中非对角线元素的信息. 我们在后文的实验中证明了此额外项对描述数据表示之间的相关性是有益的.

5 多视图领域的应用

在此节中, 我们展示了上述两种表示对齐度量方法的应用. 具体而言, 它们均可通过度量不同视图数据之间隐表示的相关性, 从而应用于多视图学习领域. 在此基础上, 我们分别提出了两种新颖的多视图聚类算法.

5.1 多视图场景的适配

在多视图场景中, 一般将数据视图个数记为 V , 且第 v 个视图的数据记为 $\mathbf{X}_v \in \mathbb{R}^{n \times d_v}$, 其中 n 和 d_v 代表数据样本的数量和第 v 个视图数据的特征维度. 在此设置下, 多视图学习算法常使用显式的函数 \mathcal{F} 将不同视图的数据映射为隐表示^[35, 36], 即

$$\mathbf{H}_v = \mathcal{F}(\mathbf{X}_v, \boldsymbol{\Theta}_v) \quad (29)$$

其中, $\boldsymbol{\Theta}_v$ 表示函数 \mathcal{F} 中的未知变量. 另一个常用的可选方案是通过约束隐藏表示 \mathbf{H}_v 与数据 \mathbf{X}_v 之间的关系, 而非显式地给出映射函数^[31, 37, 38], 即

$$\min_{\boldsymbol{\Theta}_v} \mathcal{G}(\mathbf{H}_v, \mathbf{X}_v, \boldsymbol{\Theta}_v). \quad (30)$$

其中, \mathcal{G} 和 $\boldsymbol{\Theta}_v$ 分别代表约束关系和待优化的变量.

在上述场景下, 可十分自然地使用表示对齐来度量隐表示 \mathbf{H}_v 和 $\mathbf{H}_{v'}$ 之间的相关性. 此外, 学术界已达成共识, 认为一个数据视图的隐表示可视为另一个视图隐表示的对照样本^[39-41]. 如此, 便可将基于对比的表示对齐应用于多视图数据的隐表示. 所以, 我们提出两个约束项来最大化隐表示 $\{\mathbf{H}_v\}_{v=1}^V$ 两两之间的相关性, 即

$$\begin{aligned} & \max_{\{\mathbf{H}_v\}_{v=1}^V} \hat{\mathcal{R}}_{\rho}(\{\mathbf{H}_v\}_{v=1}^V) \\ &= \max_{\{\mathbf{H}_v\}_{v=1}^V} \frac{1}{V(V-1)} \sum_{v_1, v_2=1}^V \mathbb{1}_{v_1 \neq v_2} \hat{\rho}(\mathbf{H}_{v_1}, \mathbf{H}_{v_2}) \end{aligned} \quad (31)$$

和

$$\begin{aligned} & \max_{\{\mathbf{H}_v\}_{v=1}^V} \hat{\mathcal{R}}_{\rho}(\{\mathbf{H}_v\}_{v=1}^V) \\ &= \max_{\{\mathbf{H}_v\}_{v=1}^V} \frac{1}{V(V-1)} \sum_{v_1, v_2=1}^V \mathbb{1}_{v_1 \neq v_2} \hat{Q}(\mathbf{H}_{v_1}, \mathbf{H}_{v_2}) \end{aligned} \quad (32)$$

其中, 当且仅当 $v_1 \neq v_2$ 时, 指示变量 $\mathbb{1}_{v_1 \neq v_2}$ 取值为 1, 否则为 0. 在接下来的章节中, 我们将通过实验验证最大化公式(31)和公式(32)能够有效地融合不同数据视图的互补信息, 从而提升多视图聚类算法的性能.

5.2 多视图聚类算法

本文采用了公式(30)所述的方式来约束多视图数据 $\{\mathbf{X}_v\}_{v=1}^V$ 和隐表示 $\{\mathbf{H}_v\}_{v=1}^V$ 之间的关系. 具体而言, 我们使用线性核函数构造下列两个核矩阵.

$$\mathbf{K}_v = \mathbf{X}_v \mathbf{X}_v^T, \mathbf{K}'_v = \mathbf{H}_v \mathbf{H}_v^T \quad (33)$$

使用公式(2)将核矩阵 \mathbf{K}'_v 和 \mathbf{K}_v 进行对齐, 可得

$$\begin{aligned} & \max_{\{\mathbf{H}_v\}_{v=1}^V} \hat{\rho}_k(\mathbf{K}_v, \mathbf{K}'_v) \\ & \doteq \max_{\{\mathbf{H}_v\}_{v=1}^V} \frac{\text{Tr}(\mathbf{X}_v \mathbf{X}_v^T \mathbf{H}_v \mathbf{H}_v^T)}{\sqrt{\text{Tr}(\mathbf{X}_v \mathbf{X}_v^T \mathbf{X}_v \mathbf{X}_v^T) \text{Tr}(\mathbf{H}_v \mathbf{H}_v^T \mathbf{H}_v \mathbf{H}_v^T)}}. \end{aligned} \quad (34)$$

此外, 使用公式(31)中的 $\hat{\mathcal{R}}_{\rho}$ 和公式(32)中的 $\hat{\mathcal{R}}_Q$ 对公式(34)进行约束, 上述公式可进一步转换为

$$\max_{\{\mathbf{H}_v\}_{v=1}^V} \frac{1}{V} \sum_{v=1}^V \hat{\rho}_k(\mathbf{K}_v, \mathbf{K}'_v) + \lambda \hat{\mathcal{R}}_{\rho}(\{\mathbf{H}_v\}_{v=1}^V) \quad (35)$$

和

$$\max_{\{\mathbf{H}_v\}_{v=1}^V} \frac{1}{V} \sum_{v=1}^V \hat{\rho}_k(\mathbf{K}_v, \mathbf{K}'_v) + \lambda \hat{\mathcal{R}}_Q(\{\mathbf{H}_v\}_{v=1}^V). \quad (36)$$

其中, λ 是一个用于平衡前后两项权重的超参数. 为便于表达, 我们将公式(35)中的目标式记为 MCRA (Multi-view Clustering with Representation Alignment) 算法, 同时将公式(36)中的目标式记为 MCCRA (Multi-view Clustering with Contrastive Representation Alignment) 算法. 此外, 我们还注意到对于聚类任务而言, 隐表示 \mathbf{H}_v 常被加以正交约束, 以增强不同样本之间的判别性^[31, 42]. 而且, 它们的规模一般被设置为 $n \times k$ (k 表示数据类别的数量). 总之, 上述两个多视图聚类算法的目标式可以统一记作

$$\begin{aligned} & \max_{\{\mathbf{H}_v\}_{v=1}^V} \frac{1}{V} \sum_{v=1}^V \frac{1}{\sqrt{k\mu_v}} \text{Tr}(\mathbf{H}_v^T \mathbf{X}_v \mathbf{X}_v^T \mathbf{H}_v) \\ & + \lambda \frac{1}{V(V-1)k} \sum_{v_1, v_2=1}^V \mathbb{1}_{v_1 \neq v_2} \text{Tr}(\mathbf{H}_{v_1} \mathbf{H}_{v_2}^T \hat{\mathbf{W}}) \\ & \text{s.t. } \mathbf{H}_v^T \mathbf{H}_v = \mathbf{I}_k. \end{aligned} \quad (37)$$

对于 MCRA 算法而言, 上述式子中 $\hat{\mathbf{W}} = \mathbf{I}$; 对于 MCCRA 算法而言, 上述式子中 $\hat{\mathbf{W}} = \mathbf{W}_{in}$. 此外, $\mu_v = \text{Tr}(\mathbf{K}_v \mathbf{K}_v^T)$, 且 \mathbf{I}_k 表示 k 阶单位矩阵. 为增强实

用性,我们将正样本对的权重 ω_p 和负样本对的权重 ω_n 分别设置为 $1/\sqrt{2n}$ 和 $1/\sqrt{2n(n-1)}$.

5.2.1 优化算法

为求解公式(37)中的目标式,我们提出了一个循环优化策略,即在每次循环中只求解一个变量,而固定其余变量^[32]. 具体而言,当求解矩阵 \mathbf{H}_v 时,目标式可转换为

$$\begin{aligned} \max_{\{\mathbf{H}_v\}_{v=1}^V} & \text{Tr}(\mathbf{H}_v^T \mathbf{G}_v \mathbf{H}_v) + \lambda \frac{1}{(V-1)k} \text{Tr}(\mathbf{H}_v \hat{\mathbf{H}}_v^T \hat{\mathbf{W}}) \\ \text{s. t.} & \mathbf{H}_v^T \mathbf{H}_v = \mathbf{I}_k. \end{aligned} \quad (38)$$

其中,

$$\begin{aligned} \mathbf{G}_v &= \frac{\mathbf{X}_v \mathbf{X}_v^T}{\sqrt{k\mu_v}}, \\ \hat{\mathbf{H}}_v^T &= \sum_{v'=1}^V \mathbb{1}_{v \neq v'} \mathbf{H}_{v'}. \end{aligned} \quad (39)$$

显然,公式(38)是一个 Stiefel 流形上的二次优化问题^[43],可求解如下:

- 1) 随机初始化 \mathbf{H}_v , 使得 $\mathbf{H}_v^T \mathbf{H}_v = \mathbf{I}_k$;
- 2) 更新 $\mathbf{M} = \mathbf{G}_v \mathbf{H}_v + \lambda \frac{1}{(V-1)k} \hat{\mathbf{W}}^T \hat{\mathbf{H}}_v$;
- 3) 对 \mathbf{M} 进行奇异值分解,即 $\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \mathbf{M}$, 其中 $\mathbf{U} \in \mathbb{R}^{n \times k}$, 且 $\mathbf{V} \in \mathbb{R}^{k \times k}$;
- 4) 更新 $\mathbf{H}_v = \mathbf{U} \mathbf{V}^T$;
- 5) 重复步骤 2) - 4) 直至收敛.

此外,我们还将优化的全过程展示如算法 1.

算法 1. MCRA 算法和 MCCRA 算法.

输入: 多视图数据 $\{\mathbf{X}_v\}_{v=1}^V$

输出: 数据类别划分

1. 随机初始化 $\{\mathbf{H}_v\}_{v=1}^V$, 使得 $\mathbf{H}_v^T \mathbf{H}_v = \mathbf{I}_k$;
2. $t = 0$;
3. WHILE $t < T$ DO
4. FOR $v = 1$ to V DO
5. 通过解决 Stiefel 流形上的二次规划问题(详见第 5.2.1 节),更新 \mathbf{H}_v ;
6. END FOR
7. 计算公式(37)的目标值;
8. $t = t + 1$
9. END WHILE

10. 将隐表示 $\{\mathbf{H}_v\}_{v=1}^V$ 进行拼接,并使用 k -均值算法计算出数据类别划分.

5.2.2 计算技巧和复杂度分析

接下来,我们将介绍相关的计算技巧来加速上述优化过程,并分析了 MCRA 和 MCCRA 两种算法的计算复杂度. 对于上节步骤 2) 中矩阵 \mathbf{M} 的计

算,可通过下式计算其第一项 $\mathbf{G}_v \mathbf{H}_v$.

$$\mathbf{G}_v \mathbf{H}_v = \frac{1}{\sqrt{k\mu_v}} \mathbf{X}_v (\mathbf{X}_v^T \mathbf{H}_v) \quad (40)$$

且 MCCRA 算法中的 $\hat{\mathbf{W}}^T \hat{\mathbf{H}}_v$ 可计算为

$$\hat{\mathbf{W}}^T \hat{\mathbf{H}}_v = \frac{n}{n-1} \hat{\mathbf{H}}_v - \frac{1}{n-1} \mathbf{1} \hat{\mathbf{H}}_v \quad (41)$$

其中, $\mathbf{1}$ 表示一个大小为 $n \times n$ 的全一矩阵. 不难看出,矩阵 $\mathbf{1} \hat{\mathbf{H}}_v$ 的每行元素等于矩阵 $\hat{\mathbf{H}}_v$ 的列和. 如此一来,公式(40)和公式(41)的计算复杂度仅为 $O(n)$. 此外,步骤 3) 中计算 $\mathbf{M} \in \mathbb{R}^{n \times k}$ 的奇异值分解只需 $O(n)$ 的复杂度. 再者,将隐表示 $\{\mathbf{H}_v\}_{v=1}^V$ 水平拼接,并使用 k -均值聚类进行类别划分,此过程也只需 $O(n)$ 的计算复杂度. 总而言之,MCRA 算法和 MCCRA 算法的计算复杂度在数据样本数量上均为线性.

6 实 验

6.1 实验设定

在下面的实验中,我们在七个常用的基准数据集上测试了 MCRA 算法和 MCCRA 算法的性能. 数据集为 Flower17^[44]、Caltech101^[45]、AwA^[46]、MNIST^[47]、YtVideo^[48]、AndroidM^[49] 和 DryBean^[50]. 此外,我们还将上述两个算法与文献中的六个代表性多视图聚类算法进行了对比(为公平起见,仅考虑了线性复杂度的算法),包括 MNMF^[51]、RMKMC^[52]、BMVC^[37]、LMSuC^[53] 和 OPMC^[32]. 再者,上述对比算法的执行代码均来自其作者发布的版本. 在执行过程中为确保对比的公平性,我们测试了上述算法在所有参数下的性能,并汇报最优的性能. MCRA 和 MCCRA 算法也是如此,其参数取值范围为 $\lambda \in [10^{-5}, 10^{-4}, \dots, 10^5]$. 注意,在后文中汇报的性能均为多次执行下的平均值,且聚类的性能由准确度 (ACC)、归一化互信息 (NMI) 和纯度 (Purity) 三方面指标进行衡量. 更详细的实验细节可见附录.

6.2 消融实验

为验证约束项 $\hat{\mathcal{R}}_o$ 和 $\hat{\mathcal{R}}_c$ 在多视图数据集上的有效性,我们设计了一个消融实验,且将使用的对比算法记为 MCKA (Kernel Aligned Multi-view Clustering with Kernel Alignment),目标式为

$$\begin{aligned} \max_{\{\mathbf{H}_v\}_{v=1}^V} & \frac{1}{V} \sum_{v=1}^V \text{Tr}(\mathbf{H}_v^T \mathbf{X}_v \mathbf{X}_v^T \mathbf{H}_v) \\ \text{s. t.} & \mathbf{H}_v^T \mathbf{H}_v = \mathbf{I}_k. \end{aligned} \quad (42)$$

这样一来,是否含有约束项 $\hat{\mathcal{R}}_0$ 和 $\hat{\mathcal{R}}_e$ 便成了 MCKA 算法与 MCRA 算法和 MCCRA 算法的唯一区别. 我们在七个标准数据集上测试了上述三个算法,相应的实验结果已展示在表 1 中. 可以看出,相较于 MCKA 算法,MCRA 算法和 MCCRA 算法均大幅提高了聚类的准确率. 这充分说明了最大化表示对齐和基于对比的表示对齐能够有效的提高多视图聚类算法的性能. 此外,MCCRA 算法在大多数数据集上超越了 MCRA 算法,表明了相较于表示对齐,基于对比的表示对齐能更有效地捕捉数据隐含表示的相关性,进一步验证了公式(27)中考虑非对角部分元素能带来有效的增益.

表 1 MCKA、MCRA 和 MCCRA 算法的准确度(ACC)对比

数据集	MCKA	MCRA	MCCRA
Flower17	37.33	<u>44.82</u>	46.50
Caltech101	28.30	<u>28.98</u>	29.62
AwA	9.81	10.33	<u>10.29</u>
MNIST	<u>98.79</u>	<u>98.79</u>	98.80
YtVideo	20.71	<u>25.25</u>	25.44
AndroidM	27.52	<u>31.66</u>	32.01
DryBean	66.53	<u>74.44</u>	75.09

注:最优和次优结果已加粗或用下划线标明.

6.3 对比实验

为进一步验证 MCRA 算法和 MCCRA 算法的有效性,我们还将之与现有文献中的六种多视图聚类算法进行了实验对比. 相应的实验结果展示如表 2. 可以看出,MCCRA 算法在每种设定下取得了最好的性能. 具体而言,相较于对比算法,该算法在 Flower17、Caltech101、MNIST、YtVideo、AndroidM 和 DryBean 六个数据集上准确率分别提高了 10.78%、1.91%、0.01%、2.10%、3.62% 和 4.49%, 归一化互信息分别提高了 11.01%、2.36%、0.07%、3.98%、3.21% 和 2.80%, 纯度分别提高了 10.44%、2.95%、0.01%、0.91%、1.25% 和 3.09%. 还可观察到,MCRA 算法的性能稍差于 MCCRA 算法,但相较于对比算法在 Flower17 数据集上的准确率提高了 9.10%, 在 Caltech101 数据集上提高了 1.27%, 在 YtVideo 数据集上提高了 1.91%. 不难看出,归一化互信息和纯度两个性能指标也表现出了相同的趋势. 虽然 MCCRA 算法和 MCRA 算法相较于 BMVC 算法在 AwA 数据集上的准确度分别下降了 0.16% 和 0.12%, 但应注意到 BMVC 算法包含了 6 个超参数,而前两个算法只包含 1 个超参数,更具实用性.

表 2 MCRA 算法和 MCCRA 算法以及文献中代表算法的准确度(ACC)、归一化互信息(NMI)和纯度(Purity)对比

指标	数据集	MNMF	RMKMC	BMVC	LMSuC	OPMC	MCRA	MCCRA
ACC	Flower17	35.72	23.31	27.57	31.40	29.56	<u>44.82</u>	46.50
	Caltech101	20.76	15.48	27.71	21.17	25.18	<u>28.98</u>	29.62
	AwA	6.74	9.13	10.45	8.18	9.49	<u>10.33</u>	10.29
	MNIST	97.96	91.55	85.63	98.75	98.79	<u>98.79</u>	98.80
	YtVideo	10.17	12.38	19.41	17.25	23.34	<u>25.25</u>	25.44
	AndroidM	26.80	23.29	26.90	27.90	28.39	<u>31.66</u>	32.01
	DryBean	36.09	55.24	50.45	70.60	58.24	<u>74.44</u>	75.09
NMI	Flower17	34.54	20.87	24.88	29.34	29.22	<u>45.00</u>	45.55
	Caltech101	41.69	23.78	45.33	43.49	46.41	<u>48.22</u>	48.77
	AwA	7.59	11.26	12.30	9.03	11.71	11.94	<u>12.06</u>
	MNIST	96.12	88.88	90.59	96.27	96.30	96.37	96.37
	YtVideo	8.24	10.17	15.80	14.08	20.74	<u>24.72</u>	24.72
	AndroidM	9.61	1.13	2.62	11.30	9.10	<u>14.23</u>	14.51
	DryBean	20.49	47.16	37.30	57.00	40.36	<u>59.23</u>	59.80
Purity	Flower17	36.87	24.68	28.90	33.16	31.47	<u>45.98</u>	47.31
	Caltech101	41.17	26.24	44.13	42.05	44.59	<u>47.11</u>	47.54
	AwA	8.57	11.24	12.19	10.03	11.23	<u>12.14</u>	12.19
	MNIST	97.96	92.59	89.90	98.75	98.79	<u>98.79</u>	98.80
	YtVideo	26.68	26.87	30.78	32.25	31.78	<u>33.25</u>	33.16
	AndroidM	32.28	23.29	26.90	34.74	30.88	<u>35.82</u>	35.99
	DryBean	42.30	59.56	57.17	72.00	58.24	<u>74.44</u>	75.09

注:最优和次优结果已加粗或用下划线标明.

6.4 参数与收敛性分析

在实验中,我们分析了超参数 λ 对聚类性能的影响. 以 Flower17 数据集为例,图 2 中左边两个子图展示了超参数 λ 在 $[10^{-5}, 10^{-4}, \dots, 10^5]$ 区间内取不同值时对应的性能变化. 不难看出,本文所提出的 MCRA 算法和 MCCRA 算法在绝大部分参数设置

下均超越了文献中的对比方法. 其次,随着超参数 λ 的增大,聚类性能有所提高,且在 10^{-1} 和 10^1 之间呈现出巨大的跃升,随后保持相对稳定. 这种正相关性说明了增加表示对齐和基于对比的表示对齐两项约束在目标式中的权重能明显地提高聚类性能,从而验证了这两项约束的有效性.

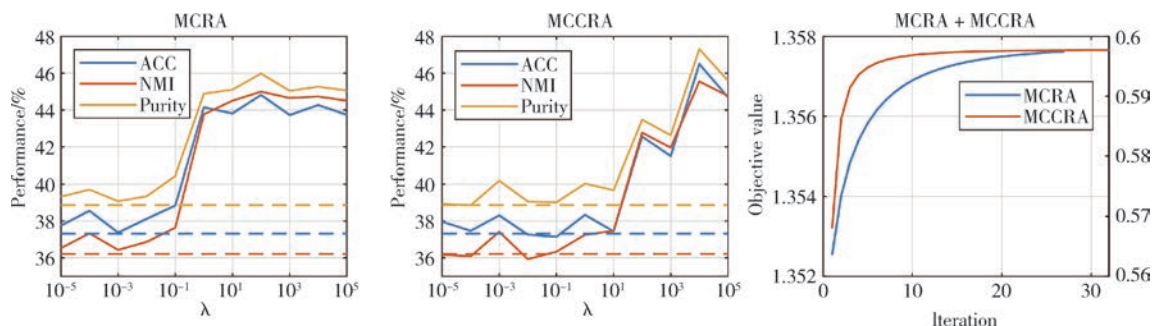


图 2 MCRA 和 MCCRA 算法在不同超参数 λ 取值下的性能变化曲线,及其在优化时目标式值的变化曲线.

注意,上述结果均来源于 Flower17 数据集上的实验,且虚线表示相关文献中对比算法的最优值.

此外,我们通过实验的方式探究了所提出算法的收敛性. 具体而言,我们收集了 MCRA 算法和 MCCRA 算法在每个优化轮次中目标式的值,并将 Flower17 数据集上的结果展示在图 2 中的第三幅子图. 可以看出,随着优化的不断进行,上述两个算法的目标值单调上升,且趋于上限,清晰地展现了其收敛性. 其他数据集上的结果可见附录.

7 讨论

本文所提出的表示对齐 RA 和基于对比的表示对齐 CRA 两种度量方法均受数据表示的特征顺序影响. 以两个数据表示 \mathbf{H} 和 \mathbf{H}' 为例,可通过将 \mathbf{H}' 的第一列(第一个特征)和第二列(第二个特征)调换得到新的数据表示 \mathbf{H}'' . 尽管数据表示 \mathbf{H}' 和 \mathbf{H}'' 在语义上未发生变化(仅是特征顺序发生改动),但是可观察到 $\hat{\rho}(\mathbf{H}, \mathbf{H}')$ 和 $\hat{\rho}(\mathbf{H}, \mathbf{H}'')$ 的值将出现较大差异. 因此,我们不建议将本文提出的 RA 和 CRA 两种度量方法应用在固定的数据表示上,如数据的原始表示,而建议将之应用于算法学习过程中的隐藏数据表示之间,如公式(29)和(30). 在后面一种应用场景下,最大化这两种度量可有效且自动地调整数据特征顺序,进而避免上述问题. 此外,此类应用场景可见于数据表示学习算法,尤其是多视图算法,如文献[22,36,37].

8 结论

在本文中,我们首次定义了数据表示的对齐的度量方法,并将之与三个现阶段常用的方法进行了对比,以凸显其优势. 在此基础上,通过引入对比学习的概念,我们提出了基于对比的表示对齐. 此外,我们在多视图学习场景中最大化上述两个表示对齐,从而构造了两个全新的多视图聚类算法. 最后,我们在七个标准数据集上测试了上述两个算法,并取得了领先的聚类性能.

致谢 感谢国家自然科学基金青年科学基金项目(No. 62306324)、国家自然科学基金面上项目(No. 62376279、No. 62276271)和国家自然科学基金国家杰出青年科学基金项目(No. 62325604)的资助.

参 考 文 献

- [1] Zhimin Gao, Lei Wang, and Luping Zhou. A probabilistic approach to cross-region matching-based image retrieval. *IEEE Transactions on Image Processing*, 2019,28(3): 1191-1204
- [2] Chenggang Yan, Biao Gong, Yuxuan Wei, and Yue Gao. Deep multi-view enhancement hashing for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021,43(4): 1445-1451

- [3] Sang-Min Choi, Sang-Ki Ko, and Yo-Sub Han. A movie recommendation algorithm based on genre correlations. *Expert Systems with Applications*, 2012, 39(9): 8079-8085
- [4] John Novembre, Toby Johnson, Katarzyna Bryc, Zolta'n Kutalik, Adam R. Boyko, Adam Auton, Amit Indap, Karen S. King, Sven Bergmann, Matthew R. Nelson, Matthew Stephens, and Carlos D. Bustamante. Genes mirror geography within Europe. *Nature*, 2008, 456(7218): 98-101
- [5] Peipei Xia, Li Zhang, and Fanzhang Li. Learning similarity with cosine similarity ensemble. *Information Sciences*, 2015, 307: 39-52
- [6] Nello Cristianini, John Shawe-Taylor, Andre' Elisseeff, and Jaz S. Kandola. On kernel-target alignment//Proceedings of the Annual Conference on Neural Information Processing Systems. Vancouver, Canada, 2001: 367-373
- [7] David R. Hardoon, Sa'ndor Szedma'k, and John Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural Computation*, 2004, 16(12): 2639-2664
- [8] Xinghao Yang, Weifeng Liu, Wei Liu, and Dacheng Tao. A survey on canonical correlation analysis. *IEEE Transactions on Knowledge and Data Engineering*, 2021, 33(6): 2349-2368
- [9] Kwang-Il Park, June Seok Hong, and Wooju Kim. A methodology combining cosine similarity with classifier for text classification. *Applied Artificial Intelligence*, 2020, 34(5): 396-411
- [10] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. A simple framework for contrastive learning of visual representations//Proceedings of the International Conference on Machine Learning. Virtual, 2020: 1597-1607
- [11] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. Algorithms for learning kernels based on centered alignment. *Journal of Machine Learning Research*, 2012, 13: 795-828
- [12] Lizhong Ding, Shizhong Liao, Yong Liu, Li Liu, Fan Zhu, Yazhou Yao, Ling Shao, and Xin Gao. Approximate kernel selection via matrix approximation. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(11): 4881-4891
- [13] Mehmet Gönen and Ethem Alpaydin. Multiple kernel learning algorithms. *Journal of Machine Learning Research*, 2011, 12: 2211-2268
- [14] Marius Kloft, Ulf Brefeld, Sören Sonnenburg, and Alexander Zien. lp-norm multiple kernel learning. *Journal of Machine Learning Research*, 2011, 12: 953-997
- [15] Charles A. Micchelli and Massimiliano Pontil. Learning the kernel function via regularization. *Journal of Machine Learning Research*, 2005, 6: 1099-1125
- [16] Harold Hotelling. The most predictable criterion. *Journal of Educational Psychology*, 1935, 26(2): 139-142
- [17] Kamalika Chaudhuri, Sham M. Kakade, Karen Livescu, and Karthik Sridharan. Multi-view clustering via canonical correlation analysis//Proceedings of the International Conference on Machine Learning. Montreal, Quebec, Canada, 2009: 129-136
- [18] Hok Shing Wong, Li Wang, Raymond Chan and Tiejong Zeng. Deep tensor CCA for multi-view learning. *IEEE Transactions on Big Data*, 2022, 8(6): 1664-1677
- [19] Xinghao Yang, Weifeng Liu, Dapeng Tao, and Jun Cheng. Canonical correlation analysis networks for two-view image recognition. *Information Sciences*, 2017, 385: 338-352
- [20] Hengrui Zhang, Qitian Wu, Junchi Yan, David Wipf, and Philip S. Yu. From canonical correlation analysis to self-supervised graph neural networks//Proceedings of the Annual Conference on Neural Information Processing Systems. Virtual, 2021: 76-89
- [21] Sagar Shrestha and Xiao Fu. Communication-efficient federated linear and deep generalized canonical correlation analysis. *IEEE Transactions on Signal Processing*, 2023, 71: 1379-1394
- [22] Hsin-Chien Huang, Yung-Yu Chuang, and Chu-Song Chen. Multiple kernel fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, 2012, 20(1): 120-134
- [23] Zhenwen Ren and Quansun Sun. Simultaneous global and local graph structure preserving for multiple kernel clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(5): 1839-1851
- [24] Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. COMPLETER: incomplete multi-view clustering via contrastive prediction//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Virtual, 2021: 11174-11183
- [25] Yunfan Li, Peng Hu, Jerry Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. Contrastive clustering//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual, 2021: 8547-8555
- [26] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning//Proceedings of the Annual Conference on Neural Information Processing Systems. Virtual, 2020: 18661-18673
- [27] Chun-Hsiao Yeh, Cheng-Yao Hong, Yen-Chi Hsu, Tyng-Luh Liu, Yubei Chen, and Yann LeCun. Decoupled contrastive learning//Proceedings of the European Conference on Computer Vision. Tel Aviv, Israel, 2022: 668-684
- [28] Yan Yang and Hao Wang. Multi-view clustering: A survey. *Big Data Mining and Analytics*, 2018, 1(2): 83-107.
- [29] Jiyuan Liu, Xinwang Liu, Jian Xiong, Qing Liao, Sihang Zhou, Siwei Wang, and Yuexiang Yang. Optimal neighborhood multiple kernel clustering with adaptive local kernels. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(6): 2872-2885
- [30] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Qing Liao, and Yuanqing Xia. Contrastive multi-view kernel learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

- 2023,45(8): 9552-9566
- [31] Erlin Pan and Zhao Kang. Multi-view contrastive graph clustering//Proceedings of the Annual Conference on Neural Information Processing Systems. Virtual, 2021: 2148-2159
- [32] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Li Liu, Siqi Wang, Weixuan Liang, and Jianguyong Shi. One-pass multi-view clustering for large-scale data//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada, 2021: 12324-12333
- [33] Jiyuan Liu, Xinwang Liu, Yuexiang Yang, Xifeng Guo, Marius Kloft, and Liangzhong He. Multiview subspace clustering via co-training robust data representation. *IEEE Transactions on Neural Networks and Learning Systems*, 2022,33(10): 5177-5189
- [34] Zhenyu Huang, Guocheng Niu, Xiao Liu, Wenbiao Ding, Xinyan Xiao, Hua Wu, and Xi Peng. Learning with noisy correspondence for cross-modal matching//Proceedings of the Annual Conference on Neural Information Processing Systems. Virtual, 2021: 29406-29419
- [35] Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu. Generalized latent multi-view subspace clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(1): 86-99
- [36] Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification//Proceedings of the International Conference on Learning Representations. Virtual, 2021
- [37] Zheng Zhang, Li Liu, Fumin Shen, Heng Tao Shen, and Ling Shao. Binary multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(7): 1774-1782
- [38] Shudong Huang, Ivor W. Tsang, Zenglin Xu, Jiancheng Lv, and Quanhui Liu. CDD: multi-view subspace clustering via cross-view diversity detection//Proceedings of the ACM Multimedia Conference. Virtual, 2021: 2308-2316
- [39] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding//Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020: 776-794
- [40] Han Zhang, Jing Yu Koh, Jason Baldrige, Honglak Lee, and Yinfei Yang. Cross-modal contrastive learning for text-to-image generation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Virtual, 2021: 833-842
- [41] Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. Contrastive clustering//Proceedings of the AAAI Conference on Artificial Intelligence. Virtual, 2021: 8547-8555
- [42] Zhenyu Huang, Peng Hu, Joey Tianyi Zhou, Jiancheng Lv, and Xi Peng. Partially view-aligned clustering//Proceedings of the Annual Conference on Neural Information Processing Systems. Virtual, 2020: 2892-2902
- [43] Feiping Nie, Rui Zhang, and Xuelong Li. A generalized power iteration method for solving quadratic problem on the stiefel manifold. *Science China Information Sciences*, 2017, 60(11): 112101;1-112101;10
- [44] Maria-Elena Nilsback and Andrew Zisserman. A visual vocabulary for flower classification//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York, USA, 2006: 1447-1454
- [45] Fei-Fei Li, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington, USA, 2004: 178-186
- [46] Christoph H. Lampert, Hannes Nickisch, and Stefan Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014,36(3): 453-465
- [47] Yann LeCun, L'eon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998,86(11): 2278-2324
- [48] Omid Madani, Manfred Georg, and David A. Ross. On using nearly-independent feature families for high precision and confidence. *Machine Learning*, 2013,92(2-3): 457-477
- [49] Babayeva Kamila. Android mischief dataset. Mendeley Data: <https://www.stratosphereips.org/android-mischief-dataset>, 2021
- [50] Murat Koklu and Ilker Ali Özkan. Multiclass classification of dry beans using computer vision and machine learning techniques. *Computers and Electronics in Agriculture*, 2020,174: 105507
- [51] Jing Gao, Jiawei Han, Jialu Liu, and Chi Wang. Multi-view clustering via joint nonnegative matrix factorization // Proceedings of the SIAM International Conference on Data Mining. Austin, USA, 2013: 252-260
- [52] Xiao Cai, Feiping Nie, and Heng Huang. Multi-view k-means clustering on big data//Proceedings of the International Joint Conference on Artificial Intelligence. Beijing, China, 2013: 2598-2604
- [53] Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. Large-scale multi-view subspace clustering in linear time//Proceedings of the AAAI Conference on Artificial Intelligence. New York, USA, 2020: 4412-4419
- [54] Colin McDiarmid et al. On the method of bounded differences. *Surveys in Combinatorics*, 1989,141(1): 148-188

附录

附 1. 定理证明

令 D 表示 $\{\mathbf{H}, \mathbf{H}'\}$, 且定义

$$\begin{aligned}\hat{\rho}_1(D) &= \frac{1}{n} \sum_{i=1}^n \mathbf{h}_i^\top \mathbf{h}'_i, \\ \hat{\rho}_2(D) &= \frac{1}{n} \sum_{i=1}^n \mathbf{h}_i \mathbf{h}_i, \\ \hat{\rho}_3(D) &= \frac{1}{n} \sum_{i=1}^n \mathbf{h}'_i{}^\top \mathbf{h}'_i, \\ \rho_1 &= \mathbb{E}[\hat{\rho}_1(D)], \\ \rho_2 &= \mathbb{E}[\hat{\rho}_2(D)], \\ \rho_3 &= \mathbb{E}[\hat{\rho}_3(D)]\end{aligned}\quad (43)$$

可得

$$\hat{\rho}(D) = \frac{\hat{\rho}_1(D)}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}}, \quad \rho = \frac{\rho_1}{\sqrt{\rho_2\rho_3}} \quad (44)$$

$$\begin{aligned}|\hat{\rho}(D) - \rho| &= \left| \frac{\hat{\rho}_1(D)}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}} - \frac{\rho_1}{\sqrt{\rho_2\rho_3}} \right| = \left| \frac{\hat{\rho}_1(D) \sqrt{\rho_2\rho_3} - \rho_1 \sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)\rho_2\rho_3}} \right| \\ &= \left| \frac{(\hat{\rho}_1(D) - \rho_1) \sqrt{\rho_2\rho_3} - \rho_1 (\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)} - \sqrt{\rho_2\rho_3})}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)\rho_2\rho_3}} \right| \\ &= \left| \frac{\hat{\rho}_1(D) - \rho_1}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}} - \rho \frac{\hat{\rho}_2(D)\hat{\rho}_3(D) - \rho_2\rho_3}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}(\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)} + \sqrt{\rho_2\rho_3})} \right|\end{aligned}\quad (47)$$

由法则可知 $-1 \leq \rho \leq 1$, 故

$$\begin{aligned}|\hat{\rho}(D) - \rho| &\leq \frac{|\hat{\rho}_1(D) - \rho_1|}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}} + \frac{|\hat{\rho}_2(D)\hat{\rho}_3(D) - \rho_2\rho_3|}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}(\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)} + \sqrt{\rho_2\rho_3})} \\ &= \frac{|\hat{\rho}_1(D) - \rho_1|}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}} + \frac{|\hat{\rho}_2(D)(\hat{\rho}_3(D) - \rho_3) + \rho_3(\hat{\rho}_2(D) - \rho_2)|}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}(\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)} + \sqrt{\rho_2\rho_3})} \\ &\leq \sqrt{\frac{\ln(2/\sigma)}{2n}} \frac{1}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}} \left[2\eta\eta' + \frac{\eta'^2\hat{\rho}_2(D) + \eta^2\rho_3}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)} + \sqrt{\rho_2\rho_3}} \right] \\ &\leq \sqrt{\frac{\ln(2/\sigma)}{2n}} \frac{1}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}} \left[2\eta\eta' + \eta'^2 \sqrt{\frac{\rho_2(D)}{\rho_3(D)}} + \eta^2 \sqrt{\frac{\rho_3}{\rho_2}} \right].\end{aligned}\quad (48)$$

因为 \mathbf{H} 和 \mathbf{H}' 分别是在半径为 η 和 η' 的球内采样得到, 即 $\rho_3/\rho_2 = \eta'/\eta$, 所以公式(47)可转化为

$$|\hat{\rho}(D) - \rho| \leq \sqrt{\frac{\ln(2/\sigma)}{2n}} \frac{1}{\sqrt{\hat{\rho}_2(D)\hat{\rho}_3(D)}} \left[\eta(2\eta' + \sqrt{\eta\eta'}) + \eta'^2 \sqrt{\frac{\rho_2(D)}{\rho_3(D)}} \right] = C(D, \eta, \eta') \sqrt{\frac{\ln(2/\sigma)}{2n}}. \quad (49)$$

附 2. 大规模问题

在计算基于对比的表示对齐时需将之分解为两部分, 即相似矩阵 \mathbf{C} 的构建以及矩阵 \mathbf{C} 与 \mathbf{W} 的乘法. 前者的复杂度为 $O(n^2)$, 同时因可采用公式(41)相似的加速技巧, 后者只需 $O(n)$ 的复杂度. 为使得基于对比的表示对齐能够处理大规模问题, 可将数据分为多个批次并分别进行计算, 如定义 4. 显然, 定义 4 中的方法只需 $O(bn)$ 的复杂度.

定义 4. 分 m 个批次给定两组数据表示, 其中批大小 $b \ll n$, 基于对比的表示对齐可近似估计为 $\hat{\mathbf{Q}}_c(\mathbf{H}, \mathbf{H}') =$

$$\sum_{i=1}^m \text{Tr}(\mathbf{C}_i \mathbf{W}_i).$$

其中, \mathbf{C}_i 和 \mathbf{W}_i 分别为第 i 批次样本的相似矩阵和预定义的指示矩阵.

通过构造另一数据集 $\tilde{D} = D \setminus \{(\mathbf{h}_i, \mathbf{h}'_i{}^\top)\} \cup \{(\tilde{\mathbf{h}}_i, \tilde{\mathbf{h}}'_i{}^\top)\}$, 可得下列误差界.

$$\begin{aligned}|\hat{\rho}_1(D) - \hat{\rho}_1(\tilde{D})| &\leq \frac{2\eta\eta'}{n}, \\ |\hat{\rho}_2(D) - \hat{\rho}_2(\tilde{D})| &\leq \frac{\eta^2}{n}, \\ |\hat{\rho}_3(D) - \hat{\rho}_3(\tilde{D})| &\leq \frac{\eta'^2}{n}.\end{aligned}\quad (45)$$

由 McDiarmid 不等式^[54]可知, 对于任意 $\sigma > 0$, 均有至少 $1 - \sigma$ 的概率使得下式成立.

$$\begin{aligned}|\hat{\rho}_1(D) - \hat{\rho}_1(\tilde{D})| &\leq \eta\eta' \sqrt{\frac{2\ln(2/\sigma)}{n}}, \\ |\hat{\rho}_2(D) - \hat{\rho}_2(\tilde{D})| &\leq \eta^2 \sqrt{\frac{\ln(2/\sigma)}{2n}}, \\ |\hat{\rho}_3(D) - \hat{\rho}_3(\tilde{D})| &\leq \eta'^2 \sqrt{\frac{\ln(2/\sigma)}{2n}}\end{aligned}\quad (46)$$

上式可推导为

Datasets/Caltech101/.

(3)AwA^[46]:该数据集收集了 50 类 30475 张动物的图片,包含了 6 个常见的图片特征,即 2688-D Color Histogram、2000-D Local Self-Similarity、252-D PHOG、2000-D SIFT、2000-D Color SIFT 和 2000-D SURF. 数据集网址为 <https://cvml.ist.ac.at/AwA/>.

(4)MNIST^[47]:该数据集收集了 60000 张 0 到 9 的手写数字图片,包含了 3 个深度网络特征,维度分别为 342、1024 和 64. 数据集网址为 <http://yann.lecun.com/exdb/mnist/>.

(5)YtVideo^[48]:该数据集收集了 101499 段 Youtube 视频,并从中提取出了 5 个数据特征,包括 64-D Audio Volume、512-D Vision Cuboids Histogram、64-D Vision HIST、647-D Vision HOG 和 838-D Vision MISC. 数据集网址为 <http://archive.ics.uci.edu/ml/datasets/YouTube+Multiview+Video+Games+Dataset>.

(6)AndroidM^[49]:该数据集收集了感染特洛伊病毒的安卓手机所产生的流量.经过梳理清洗,我们提取出了 803 条 TLS 网络流量的 3 部分特征,分别为 27-D 的流量包特征、786-D 的 TLS 协议特征和 42-D 的证书特征.数据集的网址为 <https://www.stratosphereips.org/android-mischief-dataset>.

(7)DryBean^[50]:该数据集收了 13611 个豆类果实的两部分,包括 12-D 的统计特征(如硬度、导电性等)和 4-D 的形状特征.数据集的网址为 <https://archive.ics.uci.edu/dataset/602/dry+bean+dataset>.

同时,上述基准数据集的样本数量、视图数量等信息总结如表 3.

此外,我们还简要介绍了文中所使用对比算法如下:

(1)MNMF^[51]:该算法首先通过在各个数据视图上应用 NMF 算法得出数据分类的指示矩阵,然后再约束它们向一个公共的指示矩阵靠齐,从而得出最后的聚类结果.

(2)RMKMC^[52]:该算法将标准的 k -均质聚类算法拓

展到多视图场景.

(3)BMVC^[37]:该算法首先在各个视图彼此约束的情况下将多视图数据编码为紧致的离散表示,然后通过离散矩阵分解求得最终的聚类结果.

(4)LMSuC^[53]:该算法通过在数据中寻找锚点的方法,将子空间聚类算法拓展到了大规模多视图数据上.

(5)OPMC^[32]:该算法提出了矩阵三分解方法,并使用公共的离散指示矩阵来融合多个数据视图的信息,从而最终计算出数据的类别划分.

表 3 基准数据集详情

数据集	样本数量	视图数量	类别数量
Flower17	1360	7	17
Caltech101	9144	5	101
AwA	30475	6	50
MNIST	60000	3	10
YtVideo	101499	5	31
AndroidM	803	3	8
DryBean	13611	2	7

附 4. 补充实验

表 4 展示了 MCKA 算法、MCRA 算法和 MCCRA 算法在七个基准数据集上标准化互信息(NMI)和纯度(Purity)的对比.不难看出,表中的数据与正文中表 1 展现的结果类似.其次,我们也在表 5 比较了上述算法的运行时间对比.可以看出,本文所提出的 MCRA 算法和 MCCRA 算法耗时为中游水平.值得注意的是,本文使用的对比算法均为线性复杂度.故 MCRA 算法和 MCCRA 算法的时间消耗在可接受范围内.再者,我们将上述两个算法在 Caltech101、AwA、MNIST、YtVideo、AndroidM 和 DryBean 六个数据集上的收敛性分析展示如图 3.

附 5. 实验环境

本文涉及的实验室均是在台式计算机上完成,CPU 型号为 Intel(R) Core(TM) i9-10900X@3.70GHz.

表 4 MCKA、MCRA 和 MCCRA 算法在七个基准数据集上标准化互信息(NMI)和纯度(Purity)的对比

指标	数据集	Flower17	Caltech101	AwA	MNIST	YtVideo	AndroidM	DryBean
NMI	MCKA	36.21	43.89	11.48	<u>96.37</u>	19.81	8.89	58.69
	MCRA	<u>45.00</u>	<u>48.22</u>	<u>11.94</u>	<u>96.37</u>	24.96	<u>14.23</u>	<u>59.23</u>
	MCCRA	45.55	48.77	12.06	96.37	<u>24.72</u>	14.51	59.79
Purity	MCKA	38.86	42.91	11.58	<u>98.79</u>	29.34	28.14	66.53
	MCRA	<u>45.98</u>	<u>47.11</u>	<u>12.14</u>	<u>98.79</u>	33.25	<u>35.82</u>	<u>74.44</u>
	MCCRA	47.31	47.54	12.19	98.80	33.16	35.99	75.09

注:其中最优和次优结果已用粗体和下划线标明.

表 5 MCRA、MCCRA 算法以及文献中的代表性算法在七个基准数据集上的耗时对比

(单位:s)

数据集	MNMF	RMKMC	BMVC	LMSuC	OPMC	MCRA	MCCRA
Flower17	1846.98	7.49	1.43	24.66	33.52	35.03	46.66
Caltech101	1736.73	14155.70	4.69	113.53	65.69	513.59	495.58
AwA	15952.07	6785.31	68.77	2379.84	1364.77	1873.53	1816.53
MNIST	2914.92	312.73	30.53	211.83	40.56	52.48	71.25
YtVideo	9553.65	59.69	151.24	1422.19	794.74	2277.29	2199.22
AndroidM	10.76	181.13	1.06	2.06	0.26	0.39	1.21
DryBean	23.30	2.57	3.24	22.24	0.49	4.99	5.61

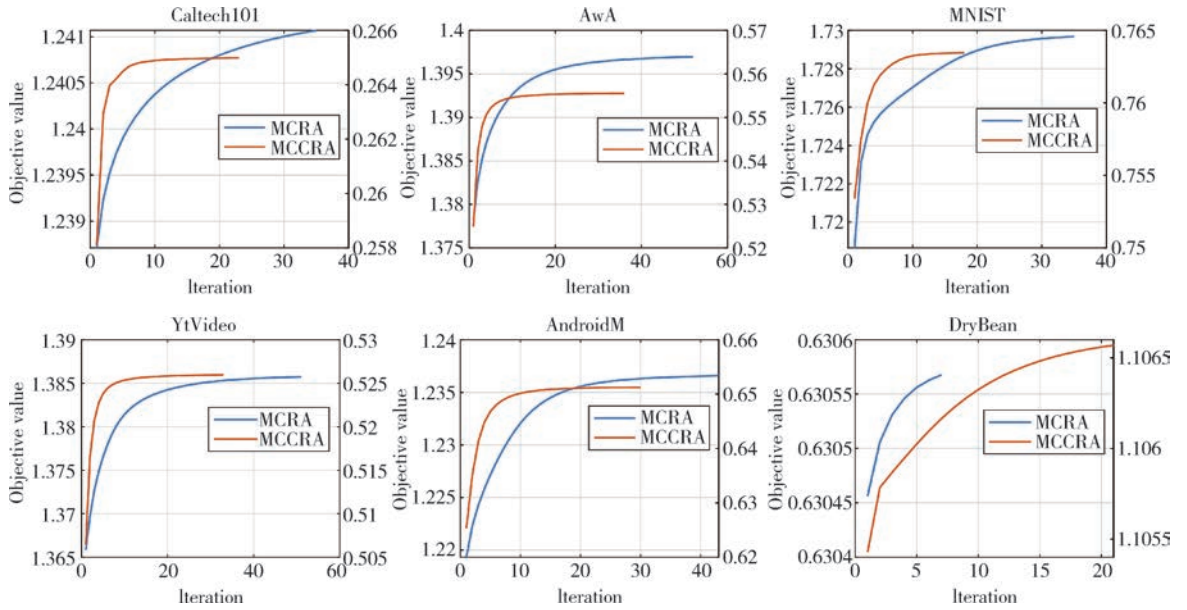


图3 MCRA算法和MCCRA算法在优化时目标式值的变化曲线



LIU Ji-Yuan, Ph. D., lecturer.

His research interests include multi-view clustering, multi-view anomaly detection and federated learning.

LIU Xin-Wang, Ph. D., professor. His research interests include multi-view clustering and kernel learning.

CAI Zhi-Ping, Ph. D., professor. His research interests include big data analysis, cyber-security and discrete computing.

ZHU En, Ph. D., professor. His research interests include clustering algorithm, anomaly detection and pattern recognition.

JU Ru-Sheng, Ph. D., professor. His research interests include data mining and computer simulation.

Background

This paper focuses on the correlation measurement between data representations. It is the basis of machine learning and artificial intelligence techniques, and has been widely applied in various fields, such as image retrieval, advertising recommendation, biological gene analysis, etc.

In literature, existing correlation measurement methods can be roughly divided into three categories, namely cosine similarity, kernel alignment and canonical correlation analysis. However, they either overlook the global information on the involved representations partially or scale poorly.

To address the issue, this paper proposes the Representation Alignment (RA) to globally measure the correlation between two arbitrary data representations in linear time to the dimension and the number of samples. As well, we further extend the representation alignment to the Contrastive Representation Alignment (CRA), where the concerned representations are two augmentations (in the scope of contrastive learning) of data samples. In addition,

the two alignments can be naturally applied in multi-view learning systems by maximizing them on the latent representations of different data views so as to integrate their complementary information. On the basis, we develop two novel multi-view clustering algorithms. They achieve state-of-the-art performance on seven benchmark datasets.

It is worth to note that the two proposed alignments are not only restricted to the two proposed multi-view clustering algorithms, but can be widely applied on representation learning, single-view/multi-view classification/clustering/anomaly detection systems, etc.

This research is supported by the National Natural Science Foundation of China (No. 6230632, 62276271, 62376279 and 62325604). The correlation measurement methods proposed in this paper are the key component and research hotspot of multi-view clustering algorithms. The authors always concentrate on the design and optimization of multi-view clustering algorithms and have published hundreds of highly-regarded papers during the last decade.