

基于自适应权重融合的深度多视子空间聚类

刘 静, 孙艳丰, 胡永利
(北京工业大学信息学部, 北京 100124)

摘要: 针对深度多视子空间聚类网络在进行数据融合时不能区分各视图可靠性, 以及缺乏对多视数据间一致性信息与互补性信息的利用等问题, 提出一种基于自适应的权重融合深度多视子空间聚类 (deep multi-view subspace clustering based on adaptive weight fusion, DMSC-AWF) 方法。首先, 通过使各视图共享同一个自表示层学习一个公共的表示矩阵, 同时为各视图分别构建自表示层来学习各视图特定的表示矩阵, 以此确保多视数据的一致性信息和互补性信息得以有效利用。然后, 在共享自表示层基础上引入注意力模块来量化不同视图的重要性, 注意力模块自适应地为每个视图数据分配权重。最后, 在4个公开数据集上进行聚类实验, 该方法的聚类结果相比于对比方法有明显的提升, 并且, 通过退化实验验证了注意力模块学习视权重有效性和重要性。

关键词: 深度子空间聚类; 表示矩阵; 多视; 权重自适应; 注意力模块; 权重分配

中图分类号: TP 183

文献标志码: A

文章编号: 0254-0037(2023)07-0758-11

doi: 10.11936/bjtxb2021070010

Deep Multi-view Subspace Clustering Based on Adaptive Weight Fusion

LIU Jing, SUN Yanfeng, HU Yongli

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

Abstract: In view of the inability of deep multi-view subspace clustering network to distinguish the reliability of each view when data fusion, and the lack of utilization of the consistent and complementary information between multi-view data, a deep multi-view subspace clustering method based on adaptive weight fusion (DMSC-AWF) was proposed. First, a common representation matrix was studied by making each view of sharing the same self-representation layer, and a self-representation layer was built for each visual to learn a specific representation matrix. The efficient use of consistent and complementary information that depends on the data was ensured. Second, the importance of different views was quantified by introducing attention modules based on the shared self-representation layer, which adaptively assigned weights to each visual data. Finally, clustering experiments were conducted on four public datasets, and the clustering results of this method were significantly improved compared with the comparison method. Moreover, the validity and importance of the attention module learning visual weight were verified by the degradation experiment.

Key words: deep subspace clustering; representation matrix; multi-view; weight adaption; attention module; weight distribution

子空间聚类方法^[1-5]作为一种有效的无监督学习方法, 在数据挖掘、计算机视觉等领域引起了广泛

关注。该方法假设将高维数据嵌入在低维子空间中, 并可由其描述表示, 因此, 子空间聚类的主要目

收稿日期: 2021-07-16; 修回日期: 2021-10-10

基金项目: 国家自然科学基金资助项目(61772048)

作者简介: 刘 静(1994—), 女, 博士研究生, 主要从事模式识别和深度学习方面的研究, E-mail: JingLiu@emails.bjut.edu.cn

通信作者: 孙艳丰(1964—), 女, 教授, 主要从事机器学习、模式识别和深度学习方面的研究, E-mail: yfsun@bjut.edu.cn

标是将一系列样本分割成互不相干的簇,使每个簇内的样本属于同一个子空间^[6-7]。近年来,基于自表示的子空间聚类方法^[8-10]因其优异的聚类性能备受关注。该方法的主要思想是将数据集自身作为字典使用,利用数据的表示矩阵构建相似矩阵,从而进行聚类任务。基于自表示的子空间聚类方法在图像表示^[11]、人脸聚类^[12-13]、运动分割^[14]、社交网络中的社区聚类、生物信息学^[15]、医学图像分析^[16]等方面具有广泛应用。

随着多媒体技术迅猛发展以及各种功能摄像机的普及,数据可以以不同形式被收集,从而产生多特征、多摄像角度、多模态等数据,这些数据被统称为多视数据。相较于单一视图采样的数据,对同一事物从不同角度或按不同方式所得到的多视数据拥有更丰富的信息。多视数据不同视图间存在一致性和互补性,一致性可理解为多视数据各视图间具有一致的语义信息,也就是多个视图的表示共享一套标签。互补性是指多视数据的多个视图之间信息互补,这些信息是每个视图独有的特征,这些特征从不同的角度描述了该数据。因此,结合一致性和互补性可提升人们对多视数据的理解。有效利用这些信息可以克服光照变化、遮挡等问题,有助于提高聚类性能。

多视子空间聚类旨在合并多视数据的特征以将相似的多视样本分到相同的簇中,而将差异较大的多视样本分到不同簇中。Zhang等^[13]通过对多视特征的张量表示施加低秩约束来获取不同视图的一致性信息,将张量的低秩约束分解为多个矩阵的低秩约束。Zhang等^[17]在多视子空间聚类方法中引入级联的思想,提出潜在多视子空间聚类方法。首先,用多个投影矩阵将不同视图数据投影到相同的低维空间以获得多视数据的公共表示,并处理样本中存在的噪声;然后,将具有公共信息的表示矩阵用于聚类任务。Wang等^[18]提出一种能够融合多视间的一致性信息和互补性信息的多视子空间聚类模型,利用互补约束项挖掘不同视图的互补信息,与此同时,利用一致约束项挖掘不同视图的一致信息。虽然上述方法存在显著差异,但是它们均运用一致性原则或互补性原则来确保多视子空间聚类任务的成功。总体而言,通过探索不同视图之间的一致性和互补性,使得多视聚类比单视聚类更加有效且更具泛化能力,这使得多视聚类^[19-21]在聚类任务中占据一席之地。

近几年,关于子空间聚类的研究主要关注线性

子空间的聚类,然而,实际中的数据不一定符合线性子空间聚类模型假设。例如:人脸图像会受到反射率以及被拍摄者姿态等影响,在这些外部因素影响下,面部图像往往位于一个非线性子空间中。因此,利用深度神经网络结构以一种无监督的方式来学习数据的非线性映射,从而使其更好地适用于子空间聚类是有必要的。深度学习的发展将多视子空间聚类带入了新的阶段,Abavisani等^[22]除了提出基于空间融合的各种方法之外,还提出一种基于相似性融合的网络(将与不同模态相对应的自表示层强制设置为相同),使得聚类性能获得显著提升。Zhu等^[23]提出了一种多视深度子空间聚类网络,该网络由2个子网组成,即分集网络和通用网络。分集网络学习视图特有的表示矩阵,而通用网络则为所有视图学习一个公共的表示矩阵。通过分集正则项利用多视图表示的互补性,可以捕获非线性和高阶视图间的关系。不同的视图共享相同的标签空间,因此,每个视图的表示矩阵都通过通用性正则项与公共视图对齐。

然而,上述方法在为各视图学习一个公共的表示矩阵时均未能甄别不同视图的可靠性。当存在一个不佳的视图数据时,聚类结果可能因此毁坏。因此,在聚类过程中为每一个视图学习一个权重是有必要的。同时,深度多视子空间聚类对将一致性信息与互补性信息相结合的研究还相对匮乏。如何在保证各视图之间一致性的情况下,充分利用不同视图的互补性信息来提高多视聚类性能,是多视聚类的一个重要问题。

综上所述,本文考虑到多视间共享一致性信息与不同视图所具有的内在个性化信息,提出了在获得所有视图的一致性信息基础上强调互补性(个性化)信息的模型,即基于自适应的权重融合深度多视子空间聚类模型(deep multi-view subspace clustering based on adaptive weight fusion, DMSC-AWF)。该模型通过各视图共享同一个自表示层学习一个共享的子空间表示矩阵,同时,又为各视图分别学习各自的自表示层来确保互补性信息。同时,为了避免某视图出现不可靠的信息从而影响最终聚类效果,引入注意力模块来量化不同视图的重要性,注意力模块自适应地为每种模态分配权重。注意力模块的使用为相似度融合质量提供了鲁棒性保障,从而提升了聚类性能。

1 方法

本文模型的网络体系结构包括自编码器、注

注意力模块和自表示模块。以2个视图为例,网络框架如图1所示。其中:蓝色虚线框表示整个自表示模块,自表示模块由一致性模块和互补性模块2个部分组成,分别对应虚线框中的蓝色和黄色部分; \oplus 为相加符号;紫色虚线框表示注意力模块,其输出为权重 w 。本文定义输入数据有 V 个视图,每个视图有 N 个样本,记为 $\mathbf{X}^v = \{\mathbf{x}_1^v, \mathbf{x}_2^v, \dots, \mathbf{x}_N^v\} \in \mathbb{R}^{N \times d_v} (v = 1, 2, \dots, V)$,其中 d_v 表示第 v 个视图的

样本维度。本模型采用的是卷积自编码器,相对于由全连接层构建的简单自编码器,卷积自编码器参数更少且对图像学习能力更强。本文不同视图数据对应各自特定的自编码器。编码器的输出为原始输入数据的潜在表示 $\mathbf{F}^v = \{f_1^v, f_2^v, \dots, f_N^v\}$,解码器的输出为各视图的重构数据 $\hat{\mathbf{X}}^v = \{\hat{\mathbf{x}}_1^v, \hat{\mathbf{x}}_2^v, \dots, \hat{\mathbf{x}}_N^v\}$ 。各视图的编码器参数记为 Θ_e^v ,解码器参数记为 Θ_d^v 。

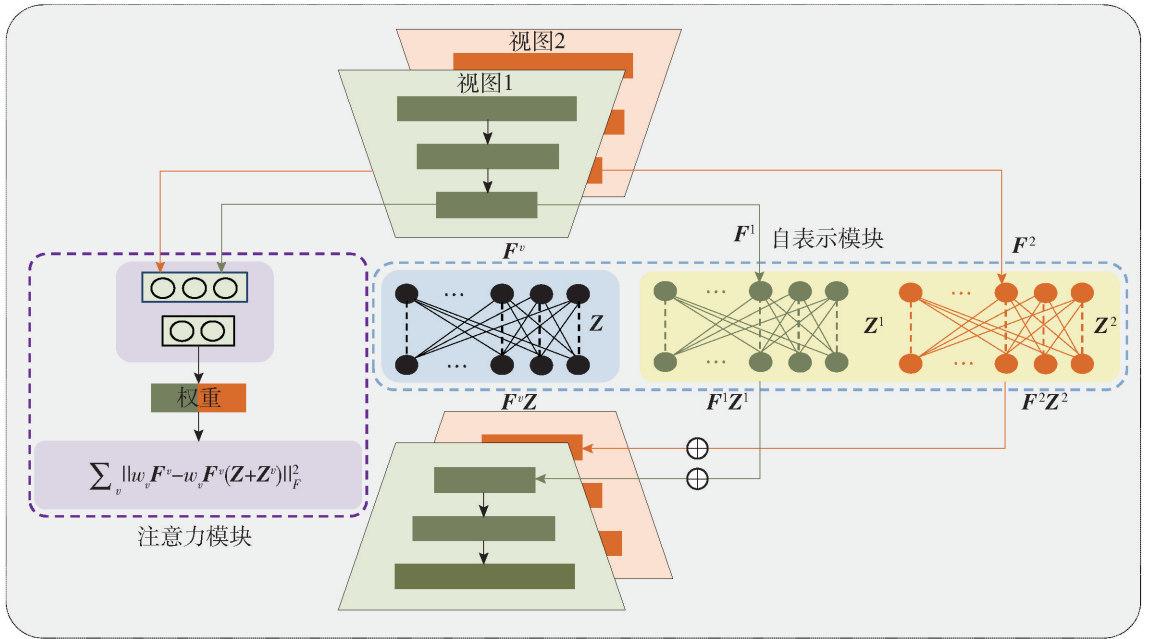


图1 DMSC-AWF网络框架

Fig. 1 Framework of DMSC-AWF

1.1 注意力模块

深度多模态聚类作为基于多视学习的无监督子空间聚类方法,具有优异的聚类性能。然而,对某些视图数据进行采集时可能存在噪声干扰,会出现数据不可靠的情况,直接进行视图间融合会降低生成的相似度矩阵的有效性。针对这一问题,通过引入注意力模块为各视图自适应地分配融合权重。对于信息量损坏的视图,网络并不撤销对这些信息的注意力,而是降低对其分配的权重。相反,网络也能在训练过程中增加对干净视图的注意力。注意力模块的权重均由网络学习得出,不需要人为指定权重大小,这也符合本文为不同视图自适应匹配权重的特点。

注意力模块以学习不同视图的权重为目标,本模块的输入为潜在表示 $\mathbf{F}^v (v = 1, 2, \dots, V)$ 拼接而成的连接特征 \mathbf{F} ,其输出是 V 维向量 w 。通常,它由3个全连接层和1个Softmax层组成,通过求平均函数获得输出权重 w 。注意力模块详细网络结构见图2。

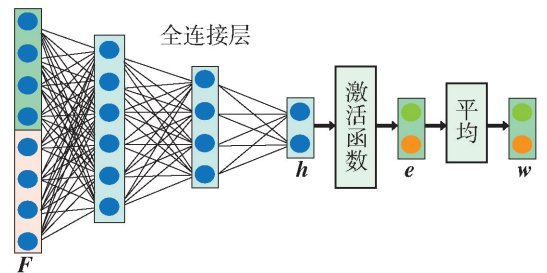


图2 注意力模块结构

Fig. 2 Structure of attention module

本文使用公式

$$\mathbf{F} = \text{concat}(\mathbf{F}^1, \mathbf{F}^2, \dots, \mathbf{F}^V) \quad (1)$$

$$\mathbf{h} = \text{FCs}(\mathbf{F}) \quad (2)$$

$$\mathbf{e} = \text{Softmax}(\text{Sigmoid}(\mathbf{h})/\tau) \quad (3)$$

$$\mathbf{w} = \text{Mean}(\mathbf{e}, \text{dim} = 0) \quad (4)$$

描述该过程。式中: $\text{concat}(\cdot)$ 表示连接运算符; $\text{FCs}(\cdot)$ 表示3个全连接的层; \mathbf{h} 、 \mathbf{e} 为隐含特征; τ 为校准因子; Sigmoid 函数与校准因子 τ 可以看作是一

种技巧,用来避免为质量最优的视图分配接近1的分数; $\text{Mean}(\mathbf{e}, \text{dim} = 0)$ 表示对 \mathbf{e} 每一列求平均值。为了简单起见,将该模块中的参数表示为 Θ_a 。

1.2 自表示模块

因为对于深度多模态聚类缺乏对互补性信息的研究,所以本文在深度子空间网络的自表示层进行改进。传统的模型假设各视图共享学习一个公共的自表示层,自表示模块的输入为各视图的潜在表示 \mathbf{F}^v , 输出为经自表示的潜在表示 $\mathbf{F}^v \mathbf{Z}$, \mathbf{Z} 为公共自表示层系数,即公共表示矩阵,同时 $\mathbf{F}^v \mathbf{Z}$ 也将作为后续解码操作的输入,这种传统自表示模型可以用

$$\begin{aligned} \min_{\mathbf{Z}, \Theta_c^v, \Theta_d^v} \sum_v \|\mathbf{F}^v - \mathbf{F}^v \mathbf{Z}\|_F^2 + \frac{\lambda}{2} \|\mathbf{Z}\|_p \\ \text{s. t. } \text{diag}(\mathbf{Z}) = 0 \end{aligned} \quad (5)$$

进行约束。式中: \mathbf{F}^v 为各视图输入数据经编码器非线性映射得到的潜在表示; $\|\cdot\|_p$ 可以根据需求进行选择,例如常用的稀疏 $\|\cdot\|_1$ 、低秩 $\|\cdot\|_*$ 或者最小二乘 $\|\cdot\|_2$ 等; 要求 \mathbf{Z} 对角线元素为0是为了避免出现单位矩阵等平凡解。

本模块为传统自表示模型增加了视图特定的自表示层,结构上本模块的输入仍是各视图的潜在表示 \mathbf{F}^v , 而输出除了经公共自表示层自表示的潜在表示 $\mathbf{F}^v \mathbf{Z}$, 还增加了 $\mathbf{F}^v \mathbf{Z}^v$ 。因此,本模块可以分为一致性模块和互补性模块。解码器的输入变成了 $\mathbf{F}^v \mathbf{Z}$ 与 $\mathbf{F}^v \mathbf{Z}^v$ 之和,在数学上,模型可以表示为

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{Z}^v, \Theta_c^v, \Theta_d^v} \frac{\lambda_1}{2} \sum_v \|\mathbf{F}^v - \mathbf{F}^v (\mathbf{Z} + \mathbf{Z}^v)\|_F^2 + \\ \frac{\lambda_2}{2} \left(\|\mathbf{Z}\|_p + \sum_v \|\mathbf{Z}^v\|_p \right) \\ \text{s. t. } \text{diag}(\mathbf{Z}) = 0, \text{diag}(\mathbf{Z}^v) = 0 \end{aligned} \quad (6)$$

式中 \mathbf{Z}^v 为各视图特定的自表示层系数,即个性表示矩阵,此处同样要求 \mathbf{Z}^v 对角线元素为0。

最后,本文用网络收敛后学习到的公共表示矩阵 \mathbf{Z} 构建相似度矩阵 $\mathbf{A} (\mathbf{A} = (|\mathbf{Z}| + |\mathbf{Z}^T|)/2)$, 对 \mathbf{A} 执行谱聚类。

1.3 损失函数

由于多视数据各视图之间既有一致性又有互补性,通常用约束各视图的个性表示矩阵与公共表示矩阵距离相近来约束各视图一致性。Cao 等^[12] 考虑到视图间信息多样性,引入希尔伯特-施密特独立性准则 (Hilbert-Schmidt independence criterion, HSIC)^[12,24] 来学习多视图间的互补性信息,如 RGB 信息和深度信息,主要目的是衡量 2 个变量的分布

差异,因此,可以度量非线性和高阶相关性。在此基础上,将互补性正则化 R_s 定义为

$$R_s = \sum_{i \neq j} \text{HSIC}(\mathbf{Z}^i, \mathbf{Z}^j) \quad (7)$$

式中 \mathbf{Z}^i 和 \mathbf{Z}^j 分别表示第 i 个和第 j 个视图对应的视特定表示矩阵。

式(7)利用 HSIC 确保学到的各视图的个性表示矩阵具有足够的差异性,从而达到有效利用来自多个视图的互补信息的目的。

所有视图共享相同标签,因此,各视图特定表示矩阵应该与公共表示矩阵对齐。本文将一致性正则化 R_c 定义为

$$R_c = \sum_v \|\mathbf{Z} - \mathbf{Z}^v\|_F^2 \quad (8)$$

结合注意力模块和自表示模块的权重和约束,基于权重融合的多视子空间聚类模型,最终可以表示为

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{Z}^v, \Theta_c^v, \Theta_d^v} \sum_v \|\mathbf{X}^v - \hat{\mathbf{X}}^v\|_F^2 + \\ \frac{\lambda_1}{2} \sum_v \|\mathbf{w}_v \mathbf{F}^v - \mathbf{w}_v \mathbf{F}^v (\mathbf{Z} + \mathbf{Z}^v)\|_F^2 + \\ \frac{\lambda_2}{2} \left(\|\mathbf{Z}\|_p + \sum_v \|\mathbf{Z}^v\|_p \right) + \frac{\lambda_3}{2} \sum_v \|\mathbf{Z} - \mathbf{Z}^v\|_F^2 + \\ \frac{\lambda_4}{2} \sum_{i \neq j} \text{HSIC}(\mathbf{Z}^i, \mathbf{Z}^j) \\ \text{s. t. } \text{diag}(\mathbf{Z}) = 0, \text{diag}(\mathbf{Z}^v) = 0 \end{aligned} \quad (9)$$

式中: \mathbf{X}^v 为模型输入数据, v 代表第 v 个视图; $\hat{\mathbf{X}}^v$ 为重构数据; $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 为正则化项系数。式(9)第1项表示重构损失;第2~4项表示带权重的自表示损失及其正则项;第5项表示一致性正则项,约束各视图个性表示矩阵向公共表示矩阵逼近;第6项表示互补性正则项保证视特定表示矩阵相互独立,确保模型可以充分挖掘多视信息的多样性。DMSC-AWF 模型的详细学习过程在算法1中给出。

算法1 DMSC-AWF 模型学习过程

输入: 无标签多视数据 $\mathbf{X}^v = \{\mathbf{x}_1^v, \mathbf{x}_2^v, \dots, \mathbf{x}_N^v\} (v = 1, 2, \dots, V)$; 超参数 $\lambda_1, \lambda_2, \lambda_3, \lambda_4$; 校准因子 τ ; 迭代次数 T ; 学习率 l ; 用预训练的自编码器参数初始化模型自编码器参数 Θ_c^v, Θ_d^v ; 随机初始化 \mathbf{Z}, Θ_a 。

输出: 聚类结果 Y

1. for $t = 1$ to T do
2. 将多视数据集输入到编码器, 获得潜在表示 \mathbf{F}^v 。
3. 根据式(1)~(4)对潜在表示进行处理, 计算权重 \mathbf{w} 。
4. 将权重 \mathbf{w} 用于式(9)来优化网络。

5. 使用随机梯度下降算法来更新模型相关参数 Θ_e^v 、 Θ_d^v 、 Z 、 Z^v 、 Θ_a 。
6. end for
7. 对完成迭代后获得的 Z 进行 $A = (|Z| + |Z^T|)/2$ 操作。
8. 将 A 用于聚类以获得聚类结果。

2 实验结果

为验证本模型的有效性,本文进行了大量实验,包括聚类实验、注意力模块、互补性模块退化实验和具有噪声数据的实验等,并将其与近年来的几种方法进行比较,以验证所提出模型的性能。此外,对实验中使用的数据集、对比方法、参数设置进行了详细说明。本文实验将相似度矩阵 $A = (|Z| + |Z^T|)/2$ 用于 4 个多视数据集以实现聚类任务。

本文方法使用 Python 3.7 环境及采用 Tensorflow 框架编码,操作系统为 Linux,显卡为 RTX 2080S $\times 2$,显存为 8 GB $\times 2$ 。

2.1 对比方法与数据集

在聚类实验中,将本文提出的 DMSC-AWF 与近年来性能较为优异的子空间聚类算法进行比较。对比方法包括单视传统子空间聚类算法和深度子空间聚类算法,共 4 种。

1) 基于低秩表示的子空间结构鲁棒恢复 (robust recovery of subspace structures by low-rank representation, LRR)^[11]。在所有候选中寻求最低秩表示,这些候选可以将数据样本表示为给定字典中具有最佳单视图的基的线性组合。

2) 深度子空间聚类网络 (deep subspace clustering networks, DSCN)^[25]。该算法将通过原始数据计算出的稀疏表示矩阵作为先验知识,再结合自编码器网络将原始数据映射到潜在空间。该模型首次将子空间聚类用于深度自编码器框架。

3) 对子空间聚类的深度认知 (deep cognitive subspace clustering, DCSC)^[26]。该算法对损失施加一个自定速正则化,是一种鲁棒的深度子空间聚类算法。

4) 无监督学习视觉特征的深度聚类 (deep clustering for unsupervised learning of visual features, DC)^[27]。该算法使用 k -means 在聚类之间迭代,并将聚类分配预测为伪标签来更新其权值,从而产生

判别损失,是一种可扩展的无监督聚类方法。

对比方法包括多视传统子空间聚类算法和深度子空间聚类算法,共 5 种。

1) 共同规范化多视谱聚类 (co-regularized multi-view spectral clustering, Co-Reg SPC)^[28]。该算法通过谱聚类目标函数隐式结合多个视图的图形,并进行聚类,以实现更好的聚类结果。

2) 通过低秩稀疏分解实现鲁棒的多视谱聚类 (robust multi-view spectral clustering via low-rank and sparse decomposition, RMSC)^[29]。为解决无甄别结合含有噪声的多视数据影响聚类结果的问题,该模型为单视图构造相应转移概率矩阵,并通过这些矩阵获得一个共享的低秩转移概率矩阵,并将其作为标准马尔可夫链聚类方法的关键输入,该模型具有低秩和稀疏性。

3) 通过深度矩阵分解进行多视聚类 (multi-view clustering via deep matrix factorization, DMF)^[30]。该算法通过半非负矩阵分解学习多视数据的多层次语义信息。

4) 潜在的多视子空间聚类 (latent multi-view subspace clustering, LMSC)^[17]。该算法寻找原始数据的潜在表示,同时,根据学习到的潜在表示进行数据重构。

5) 深度多模态子空间聚类 (deep multimodal subspace clustering networks, MSCN)^[22]。研究发现,该模型在深度多模态子空间聚类任务中的空间融合方法依赖于模式之间的空间对应。

为评估提出的基于权重融合的多视子空间聚类恢复算法性能,本研究在 4 个公共数据集上进行了大量实验,数据集简介如下。

1) ORL 人脸数据集^[31]。该数据集包含 40 个人的 10 张不同的图像。对于每个受试者,图像采集时的控制光照条件、面部表情和面部细节不同。对于 ORL 原始数据集,本实验调整图像大小为 48×48 ,提取 3 种类型的特征,即强度 (4 096 维)、局部二值模式 (local binary pattern, LBP) (3 304 维) 和 Gabor (6 750 维)。标准的 LBP 特征是从 72×80 像素的图像中提取的,直方图大小为 $59 \sim 910$ 个像素块。Gabor 特征是在 4 个方向 $\theta = 0^\circ, 45^\circ, 90^\circ, 135^\circ$ 的情况下以同一个尺度 $\lambda = 4$ 提取的,其分辨率为 25×30 像素的松散脸部裁剪。除了强度之外,所有描述符都缩放为单位范数。

2) Still DB 数据集^[23]。该数据集由 467 张行为动作图像构成,主要包括 6 类:跑步、行走、接球、

投掷、蹲下、踢腿。由于行为动作姿势的相似和图片有杂乱的背景,该数据集是具有挑战性的行为图像数据库。

3) BBC Sport 数据集^[32]。该数据集包含 544 份来自 BBC 体育网站的体育新闻文章,这些文章在 2004、2005 年涉及 5 个主题领域的 2 个视图。对于每个示例,第 1 个视图有 3 183 个特征,第 2 个视图有 3 203 个特征。BBC Sport 的部分子集如图 3 所示。

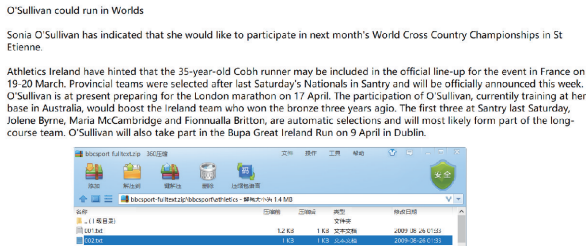


图3 BBC Sport 文本数据集的部分原始样本

Fig. 3 Part of the original sample of the BBC Sport text dataset

4) RGB-D 对象数据集^[33]。该数据集包含从多个视图中获取的 300 个物理上(传感器距离物体的实际距离)截然不同的对象的视觉和深度图像,这

些对象按照 WordNet 上下名关系(类似于 ImageNet)被组织成 51 个类别。本实验中随机选择 50 个类,每类 10 个样本。所有视觉图像和深度图像大小均为 64 × 64 像素。实验中使用递归中值滤波,直到所有缺失的值都被填满。RGB 图像和深度图像的子集如图 4 所示。

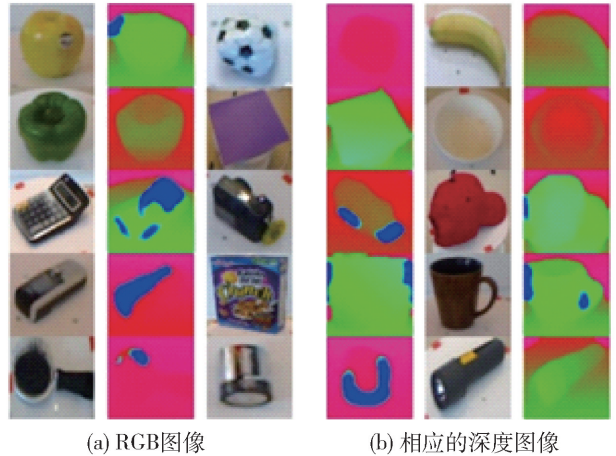


图4 RGB-D 对象数据集的部分样本

Fig. 4 Part of sample from the RGB-D object dataset

4 个公开数据集 ORL、Still DB、BBC Sport 和 RGB-D 在本文实验中的使用情况如表 1 所示。

表 1 数据集情况统计

Table 1 Statistics of datasets

数据集	类别	视图 个数	每个视图 样本数	维度大小		
				视图 1	视图 2	视图 3
ORL	40	3	400	4 096	3 304	6 750
Still DB	6	3	467	200	200	200
BBC Sport	5	2	544	3 183	3 203	
RGB-D	50	2	500	64 × 64 × 3	64 × 64 × 1	

2.2 参数设置与评估指标

本文模型使用 [64, 32, 16] 通道的 3 层编码器,相应地使用 [16, 32, 64] 通道的 3 层解码器。卷积层的卷积核大小都设置为 3 × 3,而非线性激活函数为整流线性单元。然后,通过反卷积层将潜在特征返回到与输入大小相同的空间。同时,在所有实验中将 l_r 设置为 0.001。以 RGB-D 数据集为例,其输入为 2 个视图数据:一个视图数据为彩色视觉图像(有 3 个通道),另一个视图数据为深度图像(有 1 个通道)。2 个视图的通道分别为 3—64—32—16—32—64—3 和 1—64—32—16—32—64—1。横向和纵向平移步长为 2,填充方式为 same,迭代次数为 40, $\lambda_1 = 1.00$, $\lambda_2 = 1.00$, $\lambda_3 = 0.10$, $\lambda_4 = 0.10$, $\tau = 0.50$ 。为确定上述关键

参数($\lambda_1, \lambda_2, \lambda_3, \lambda_4, \tau$)取值对聚类性能的影响,本文固定其他 4 个参数,依次分析各参数对聚类性能的影响。不同参数在 RGB-D 数据集的 DMSC-AWF 性能见图 5。图 5(a) ~ (d) 中 $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 的取值范围设置为 [0.01, 0.10, 1.00, 10.00, 100.00]。结果表明,当 λ_1, λ_2 被设为 1.00 时,本文方法在不同的性能指标上均有良好的表现。同样,可以观察到,当 λ_3, λ_4 设为 0.10 时,本文提出的 DMSC-AWF 具有更好的性能。当 τ 分别取 0.1、0.5、1.0、10.0、20.0 时进行实验,当 $\tau = 0.5$ 时效果最优。

本实验使用 4 个流行的评价指标来评估聚类效果,包括准确度 (accuracy, ACC) A_{CC} 、归一化互信息 (normalized mutual information, NMI) N_{MI} 、精确率和

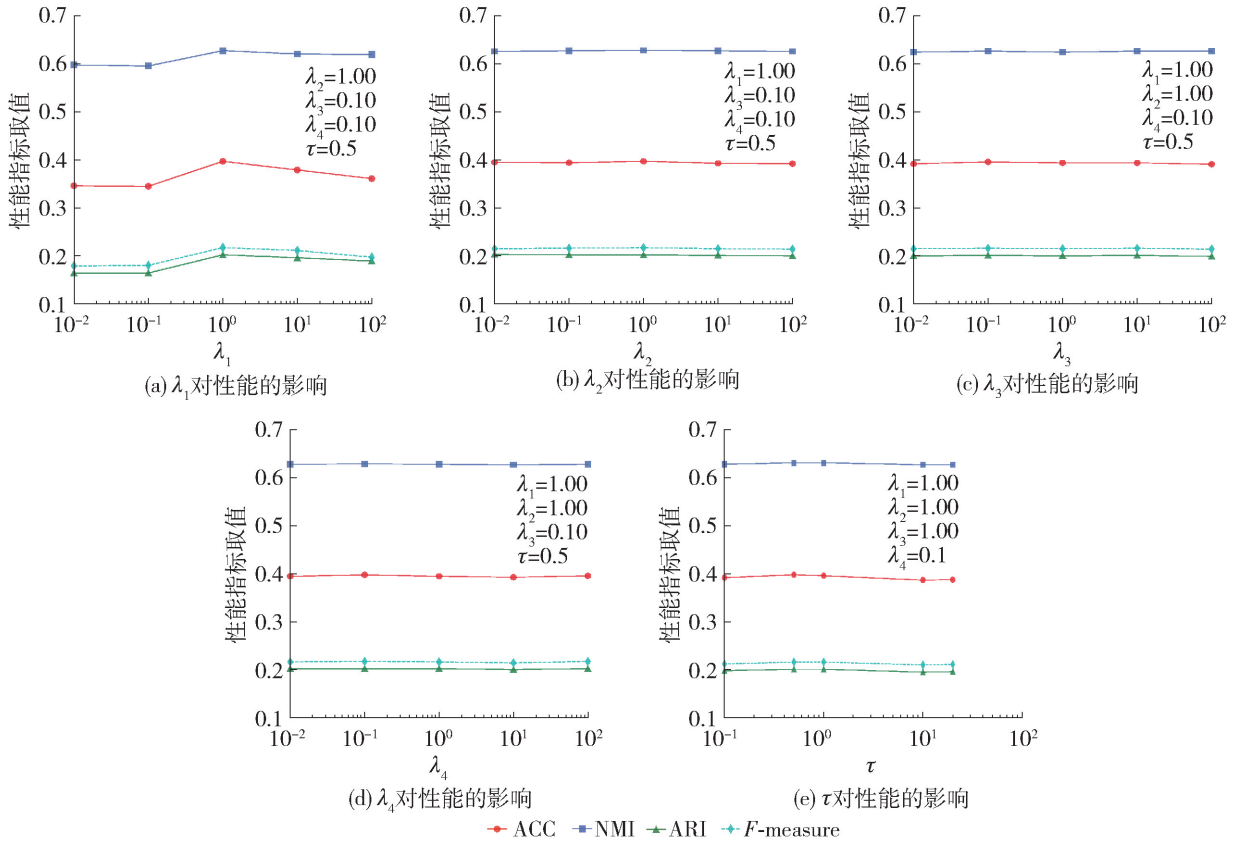


图5 不同参数对 DMSC-AWF 性能的影响

Fig. 5 Effect of different parameters on DMSC-AWF performance

召回率的加权调和平均 F_{measure} 和可以综合评价性能的调整兰德指数 (adjusted Rand index, ARI) I_{AR} 。其中,分数越高,表示聚类性能越好。

A_{CC} 的计算公式为

$$A_{\text{CC}} = \max_m \frac{\sum_{i=1}^n \{l_i = m(y_i)\}}{n} \quad (10)$$

式中: l_i 为真实标签; y_i 为模型产生的聚类结果; $m(y_i)$ 为置换映射函数,将聚类标签映射到类标签; n 为样本数。置换映射函数的最佳映射可通过匈牙利算法获得。

N_{MI} 是数据真实标签 l 和聚类标签 k 之间相似度的归一化度量,计算公式为

$$N_{\text{MI}} = \frac{I(l; y)}{\max\{H(l), H(y)\}} \quad (11)$$

式中: $I(l; y)$ 表示 l 和 y 之间的互信息; $H(l)$ 、 $H(y)$ 分别表示 l 和 y 的熵。NMI 的结果不因簇(类)的排列而改变,它们归一化为 $[0, 1]$, 0 表示没有相关性, 1 表示完全相关。

I_{AR} 反映聚类标签与数据真实标签的重叠程度,具体计算公式见文献[34]。

F_{measure} 常用于衡量模型的有效性,取值范围为 $[0, 1]$, 公式为

$$F_{\text{measure}} = \frac{2 \times P \times R}{P + R} \quad (12)$$

式中: P 为精确率; R 为召回率。

2.3 实验结果对比与分析

实验时,将 MSCN 与本模型 DMSC-AWF 的迭代步数保持一致,并以最后一次迭代的结果作为最终结果。其他对比实验结果取自文献[23]。在实验设置中,所有方法都使用 k -means 聚类方法来得到最终的聚类结果。解的初始化严重影响 k -means 的性能,因此,实验时把每个代码均运行 30 次,并给出了平均性能和标准偏差。为验证本模型及本模型提出的各个模块的有效性,分别进行了聚类实验、退化实验以及具有噪声实验。

2.3.1 聚类实验

为验证本文模型的有效性,本实验将不同方法在 ORL、Still DB、BBC Sport 和 RGB-D 数据集上的聚类结果进行对比,实验结果如表 2~5 所示。表中黑体数字为最佳结果。

根据聚类结果显示,本文提出的方法在 ORL、

表 2 不同方法在 ORL 下的聚类性能对比

Table 2 Comparison of the clustering performance of different methods under the ORL dataset

方法	ACC	NMI	ARI	F_{measure}
LRR	0.773 ± 0.003	0.895 ± 0.006	0.724 ± 0.002	0.731 ± 0.004
DSCN	0.801 ± 0.009	0.883 ± 0.005	0.704 ± 0.012	0.711 ± 0.011
DCSC	0.811 ± 0.003	0.893 ± 0.003	0.709 ± 0.021	0.718 ± 0.004
DC	0.788 ± 0.002	0.865 ± 0.011	0.684 ± 0.007	0.701 ± 0.008
Co-Reg SPC	0.715 ± 0.000	0.853 ± 0.003	0.602 ± 0.004	0.615 ± 0.000
RMSC	0.723 ± 0.025	0.872 ± 0.012	0.645 ± 0.029	0.654 ± 0.028
DMF	0.823 ± 0.021	0.933 ± 0.010	0.783 ± 0.001	0.773 ± 0.002
LMSC	0.819 ± 0.017	0.931 ± 0.011	0.769 ± 0.044	0.758 ± 0.009
MSCN	0.833 ± 0.008	0.928 ± 0.001	0.790 ± 0.005	0.787 ± 0.001
DMSC-AWF (本文)	0.848 ± 0.007	0.937 ± 0.004	0.802 ± 0.003	0.819 ± 0.007

表 3 不同方法在 Still DB 下的聚类性能对比

Table 3 Comparison of the clustering performance of different methods under the Still DB dataset

方法	ACC	NMI	ARI	F_{measure}
LRR	0.306 ± 0.039	0.109 ± 0.030	0.066 ± 0.002	0.240 ± 0.052
DSCN	0.323 ± 0.006	0.216 ± 0.011	0.145 ± 0.002	0.293 ± 0.019
DCSC	0.325 ± 0.007	0.222 ± 0.008	0.148 ± 0.003	0.301 ± 0.002
DC	0.315 ± 0.001	0.199 ± 0.003	0.131 ± 0.001	0.280 ± 0.011
Co-Reg SPC	0.263 ± 0.024	0.093 ± 0.016	0.092 ± 0.004	0.226 ± 0.035
RMSC	0.285 ± 0.020	0.106 ± 0.056	0.113 ± 0.063	0.232 ± 0.021
DMF	0.336 ± 0.017	0.154 ± 0.010	0.124 ± 0.001	0.265 ± 0.005
LMSC	0.328 ± 0.029	0.137 ± 0.032	0.088 ± 0.007	0.269 ± 0.055
MSCN	0.312 ± 0.008	0.168 ± 0.001	0.133 ± 0.005	0.261 ± 0.001
DMSC-AWF (本文)	0.353 ± 0.006	0.210 ± 0.007	0.152 ± 0.003	0.304 ± 0.012

表 4 不同方法在 BBC Sport 下的聚类性能对比

Table 4 Comparison of the clustering performance of different methods under the BBC Sport dataset

方法	ACC	NMI	ARI	F_{measure}
LRR	0.832 ± 0.026	0.690 ± 0.019	0.667 ± 0.008	0.774 ± 0.023
DSCN	0.821 ± 0.000	0.652 ± 0.000	0.856 ± 0.001	0.683 ± 0.001
DCSC	0.843 ± 0.000	0.683 ± 0.001	0.864 ± 0.012	0.712 ± 0.002
DC	0.724 ± 0.000	0.556 ± 0.001	0.781 ± 0.000	0.492 ± 0.000
Co-Reg SPC	0.564 ± 0.000	0.718 ± 0.003	0.696 ± 0.001	0.766 ± 0.002
RMSC	0.737 ± 0.003	0.608 ± 0.007	0.723 ± 0.025	0.655 ± 0.002
DMF	0.890 ± 0.031	0.821 ± 0.003	0.883 ± 0.012	0.889 ± 0.001
LMSC	0.900 ± 0.044	0.826 ± 0.006	0.893 ± 0.012	0.887 ± 0.071
MSCN	0.888 ± 0.003	0.813 ± 0.002	0.859 ± 0.001	0.854 ± 0.002
DMSC-AWF (本文)	0.920 ± 0.004	0.831 ± 0.003	0.902 ± 0.003	0.857 ± 0.001

表5 不同方法在RGB-D下的聚类性能对比

Table 5 Comparison of the clustering performance of different methods under the RGB-D dataset

方法	ACC	NMI	ARI	F_{measure}
LRR	0.299 ± 0.010	0.589 ± 0.002	0.143 ± 0.002	0.156 ± 0.001
DSCN	0.339 ± 0.006	0.589 ± 0.004	0.163 ± 0.004	0.179 ± 0.004
DCSC	0.340 ± 0.002	0.591 ± 0.002	0.170 ± 0.001	0.182 ± 0.003
DC	0.340 ± 0.002	0.594 ± 0.003	0.177 ± 0.004	0.184 ± 0.004
Co-Reg SPC	0.268 ± 0.003	0.602 ± 0.007	0.155 ± 0.020	0.175 ± 0.018
RMSC	0.341 ± 0.015	0.603 ± 0.006	0.162 ± 0.010	0.178 ± 0.010
DMF	0.286 ± 0.006	0.549 ± 0.004	0.107 ± 0.002	0.123 ± 0.001
LMSC	0.335 ± 0.028	0.593 ± 0.030	0.151 ± 0.035	0.167 ± 0.034
MSCN	0.354 ± 0.003	0.608 ± 0.001	0.190 ± 0.002	0.203 ± 0.004
DMSC-AWF (本文)	0.397 ± 0.006	0.628 ± 0.003	0.202 ± 0.004	0.217 ± 0.004

Still DB、BBC Sport 和 RGB-D 数据集上明显优于其他对比方法的聚类性能。例如:对于 RGB-D 来说,本文提出的模型 ACC 提高了 4.3%, NMI 提高了 2.0%, ARI 提高了 0.8%, F_{measure} 提高了 1.4%。这证明了所提出的 DMSC-AWF 模型在多视子空间聚类任务中的有效性。同时,在聚类结果中可以看出:1) 无论是单视还是多视,使用深度网络的方法相比于传统方法具有较高的准确度,这表明深度网络具有更好地提取非线性数据特征的能力,有助于提高聚类效果。2) 在同样使用深度网络的情况下,多视聚类方法相较于单视聚类方法取得了更好的聚类性能,这表明有效利用多视角数据,在大多数情况下可以获得较为不错的聚类性能。

综上可知,性能的提升主要体现在 3 个方面:一是在网络中以端到端的方式学习相似度矩阵;二是将多视数据的一致性信息和互补性信息结合,并应用到自表示中;三是做相似度融合时将注意力模块应用到网络中,为不同质量的视图数据分配合适的权重。

2.3.2 注意力模块及互补性模块退化实验

为验证本文提出的模块的有效性,以 RGB-D 数据集为例,对包含添加了互补性信息的自表示模块和注意力模块的模型 DMSC-AWF、在 DMSC-AWF 模型基础上去掉注意力模块的模型 DMSC-AWF/att 以及在 DMSC-AWF/att 模型基础上去掉互补性模块的 MSCN 模型进行实验,最终聚类结果见表 6。

表6 退化实验的聚类性能对比

Table 6 Clustering performance of degraded experiments

方法	ACC	NMI	ARI	F_{measure}
DMSC-AWF (本文)	0.397 ± 0.006	0.628 ± 0.003	0.202 ± 0.004	0.217 ± 0.004
DMSC-AWF/att	0.382 ± 0.006	0.634 ± 0.004	0.200 ± 0.007	0.215 ± 0.007
MSCN	0.354 ± 0.003	0.608 ± 0.001	0.190 ± 0.002	0.203 ± 0.004

由表 6 可以看出,DMSC-AWF 相较 MSCN 和 DMSC-AWF/att 具有较为优异的性能优势。因为 DMSC-AWF/att 与 MSCN 的区别在于 DMSC-AWF/att 的自表示模块中包含互补性模块,所以由 DMSC-AWF/att 与 MSCN 的实验结果可知,在自表示模块中添加互补性模块是非常有必要的。DMSC-AWF 在聚类性能上大致高于没有添加注意力模块的 DMSC-AWF/att,同时,DMSC-AWF 聚类结果的标准差小于 DMSC-AWF/att,这说明添加注意力模块有助于提高网络的鲁棒性,从而提升聚类稳定性。本退化实验结

果证实了注意力模块及互补性模块的有效性。

2.3.3 具有噪声数据的聚类实验

以 RGB-D 数据集为例,在本数据集的某一视图上添加高斯噪声,从而获得一组新的噪声数据。将这组噪声数据作为一个视图同另一个视图一起送入网络中,并将本模型 DMSC-AWF 与 MSCN 的聚类结果进行对比。将高斯噪声均值设置为 0,标准差设置为 0.2。在本部分预训练实验中一个视图的输入数据为噪声数据。具有噪声数据的聚类结果如表 7 所示。表中黑体数字为最佳结果。

表 7 标准差为 0.2 的噪声数据的聚类性能
Table 7 Clustering performance of noise data with standard deviation of 0.2

方法	ACC	NMI	ARI	$F_{measure}$
MSCN	0.345 ± 0.004	0.609 ± 0.003	0.167 ± 0.002	0.183 ± 0.002
DMSC-AWF (本文)	0.388 ± 0.007	0.627 ± 0.004	0.200 ± 0.005	0.215 ± 0.005

根据表 7 中展示的添加噪声实验的聚类结果可以看出,如果原始图像包含噪声,则会降低聚类性能,但本文提出的方法依旧可以保持一个不错的聚类效果。相比于 MSCN 来说,本文模型 ACC 提高了 3.8%,NMI 提高了 0.5%,ARI 提高了 1.7%, $F_{measure}$ 提高了 1.7%。为进一步直观体现本模型面对噪声数据的鲁棒性,本文对不同视图的权重进行了可视化,展示了有噪声的视图和无噪声的视图之间的权重差异,实验结果如图 6 所示。本噪声实验验证了在面对不可靠数据时,本文提出的基于权重融合的方法相对于各视图直接融合的传统方法具有更强的鲁棒性,从而证实了融合时对各视图分配不同权重是有必要的。

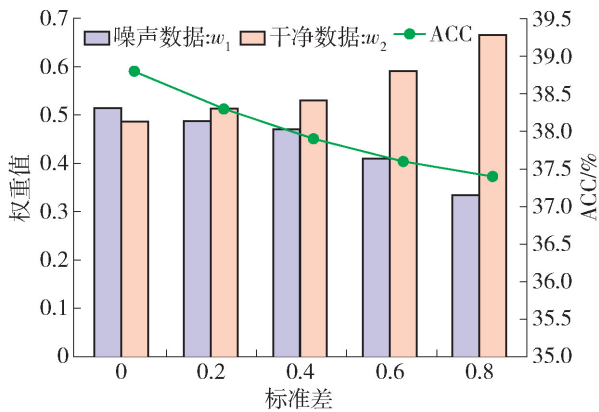


图 6 对噪声实验权重可视化

Fig.6 Visualization of experimental weight of noise

3 结论

1) DMSC-AWF 方法强制各视图共享自表示层(相似度融合)以确保一致性,并在此基础上强调了各视图独立的互补性信息,即为各视图分别学习一个自表示层。同时,在多视信息融合时利用注意力模块学习到各视图的权重分配,为相似度融合质量提供了鲁棒性保障。

2) DMSC-AWF 算法可以提升数据融合质量,从而提升聚类性能。

3) DMSC-AWF 的注意力模块有助于提高网络的鲁棒性,使网络具有一定抗噪声能力,从而提升聚类稳定性。

参考文献:

[1] MASCI J, MEIER U, CIRESAN D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction [C] // International Conference on Artificial Neural Networks. Berlin: Springer, 2011: 52-59.

[2] VIDAL R, MA Y, SASTRY S. Generalized principal component analysis (GPCA) [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(12): 1945-1959.

[3] CHEN G, LERMAN G. Spectral curvature clustering (SCC) [J]. International Journal of Computer Vision, 2009, 81(3): 317-330.

[4] 许凯, 吴小俊, 尹贺峰. 基于分布式低秩表示的子空间聚类算法 [J]. 计算机研究与发展, 2016, 53(7): 1605-1611.

XU K, WU X J, YIN H F. Distributed low rank representation-based subspace clustering algorithm [J]. Journal of Computer Research and Development, 2016, 53(7): 1605-1611. (in Chinese)

[5] 王卫卫, 李小平, 冯象初, 等. 稀疏子空间聚类综述 [J]. 自动化学报, 2015, 41(8): 1373-1384.

WANG W W, LI X P, FENG X C, et al. A survey on sparse subspace clustering [J]. Acta Automatica Sinica, 2015, 41(8): 1373-1384. (in Chinese)

[6] PARSONS L, HAQUE E, LIU H, et al. Subspace clustering for high dimensional data: a review [J]. ACM SIGKDD Explorations Newsletter, 2004, 6(1): 90-105.

[7] AGRAWAL R, GEHRKE J, GUNOPULOS D, et al. Automatic subspace clustering of high dimensional data for data mining applications [J]. Data Mining and Knowledge Discovery, 1998, 27(2): 94-105.

[8] ELHAMIFAR E, VIDAL R. Sparse subspace clustering: algorithm, theory, and applications [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 35(11): 2765-2781.

[9] LIU G C, LIN Z C, YU Y. Robust subspace segmentation by low-rank representation [C] // International Conference on Machine Learning. New York: ACM, 2010: 663-670.

[10] VIDAL R, FAVARO P. Low rank subspace clustering (LRSC) [J]. Pattern Recognition Letters, 2014, 43(7): 47-61.

- [11] LIU G C, LIN Z C, YAN S C, et al. Robust recovery of subspace structures by low-rank representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 171-184.
- [12] CAO X C, ZHANG C Q, FU H Z, et al. Diversity-induced multi-view subspace clustering [C] // *IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2015: 586-594.
- [13] ZHANG C Q, FU H Z, LIU S, et al. Low-rank tensor constrained multiview subspace clustering [C] // *International Conference on Computer Vision*. Piscataway: IEEE, 2015: 1582-1590.
- [14] ALDROUBI A, SEKMEN A. Nearness to local subspace algorithm for subspace and motion segmentation [J]. *IEEE Signal Processing Letters*, 2012, 19(10): 704-707.
- [15] TARI L, BARAL C, KIM S. Fuzzy c-means clustering with prior biological knowledge[J]. *Journal of Biomedical Informatics*, 2009, 42(1): 74-81.
- [16] MILITELLO C, VITABILE S, RUNDO L, et al. A fully automatic 2D segmentation method for uterine fibroid in MRgFUS treatment evaluation[J]. *Computers in Biology and Medicine*, 2015, 62(7): 277-292.
- [17] ZHANG C Q, HU Q H, FU H Z, et al. Latent multi-view subspace clustering [C] // *IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2017: 4279-4287.
- [18] WANG X B, GUO X J, LEI Z, et al. Exclusivity-consistency regularized multi-view subspace clustering [C] // *IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2017: 923-931.
- [19] DANG X, BAILEY J. A framework to uncover multiple alternative clusterings [J]. *Machine Learning*, 2015, 98(1/2): 7-30.
- [20] REN P Z, XIAO Y, XU P F, et al. Robust auto-weighted multi-view clustering [C] // *International Joint Conference on Artificial Intelligence*. San Francisco: Morgan Kaufmann, 2018: 2644-2650.
- [21] LIU Y, HE L F, CAO B K, et al. Multi-view multi-graph embedding for brain network clustering analysis[C] // *AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI, 2018: 117-124.
- [22] ABAVISANI M, PATEL V M. Deep multimodal subspace clustering networks[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2018, 12(6): 1601-1614.
- [23] ZHU P F, HUI B Y, ZHANG C Q, et al. Multi-view deep subspace clustering networks[EB/OL]. [2021-05-05]. <https://arxiv.org/abs/1908.01978>.
- [24] GRETTON A, BOUSQUET O, SMOLA A, et al. Measuring statistical dependence with Hilbert-Schmidt norms [C] // *International Conference on Algorithmic Learning Theory*. Berlin: Springer, 2005: 63-77.
- [25] JI P, ZHANG T, LI H D, et al. Deep subspace clustering networks[C] // *Proceedings of the Conference and Workshop on Neural Information Processing Systems*. La Jolla, California: NIPS, 2017: 24-33.
- [26] JIANG Y B Y, YANG Z Y, XU Q Q, et al. When to learn what: deep cognitive subspace clustering [C] // *ACM International Conference on Multimedia*. New York: ACM, 2018: 718-726.
- [27] CARON M, BOJANOWSKI P, JOULIN A, et al. Deep clustering for unsupervised learning of visual features[C] // *European Conference on Computer Vision*. Berlin: Springer, 2018: 132-149.
- [28] KUMAR A, RAI P, DAUME H. Co-regularized multi-view spectral clustering [C] // *Neural Information Processing Systems*. Cambridge: MIT Press, 2011: 1413-1421.
- [29] XIA R K, PAN Y, DU L, et al. Robust multi-view spectral clustering via low-rank and sparse decomposition [C] // *AAAI Conference on Artificial Intelligence*. Menlo Park: AAAI, 2014: 2149-2155.
- [30] PENG X, ZHU H Y, FENG J S, et al. Deep clustering with sample-assignment invariance prior [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2019, 31(11): 4857-4868.
- [31] FELZENSZWALB P, MCALLESTER D, RAMANAN D. A discriminatively trained, multiscale, deformable part model[C] // *IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2008: 1-8.
- [32] GREENE D, CUNNINGHAM P. Practical solutions to the problem of diagonal dominance in kernel document clustering [C] // *International Conference on Machine Learning*. New York: ACM, 2006: 377-384.
- [33] LAI K, BO L Y, REN X F, et al. A large-scale hierarchical multiview RGB-D object dataset[C] // *IEEE International Conference on Robotics and Automation*. Piscataway: IEEE, 2011: 1817-1824.
- [34] RAND W M. Objective criteria for the evaluation of clustering methods [J]. *Journal of the American Statistical Association*, 1971, 66(336): 846-850.

(责任编辑 梁洁)