计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(5): 1216−1225, 2025

消除互补性争议的多视图聚类算法

赵玉涵　陈松灿
（南京航空航天大学计算机科学与技术学院　南京　211106）
（模式分析与机器智能工业和信息化部重点实验室（南京航空航天大学）　南京　211106）
（zhaoyuhan7@nuaa.edu.cn）

Multi-View Clustering Algorithm via Eliminating the Complementarity Controversy

Zhao Yuhan and Chen Songcan
（*College of Computer Science and Technology*, *Nanjing University of Aeronautics and Astronautics*, *Nanjing* 211106）
（*Key Laboratory of Pattern Analysis and Machine Intelligence* (*Nanjing University of Aeronautics and Astronautics*), *Ministry of Industry and Information Technology*, *Nanjing* 211106）

Abstract　Multi-view clustering aims to use heterogeneous information from different views to discover the underlying data structure and divide the samples into clusters. Consistency and complementarity are two key elements that affect the performance of multi-view clustering. Consistency emphasizes the semantic similarity between different views. Complementarity emphasizes the mutual supplementation of specific information within each view. At present, the study of consistency has been relatively in-depth, but the study of complementarity is controversial, in which some methods believe that consistency and complementarity can assist each other, but constraining them to the same feature space actually causes a conflict between them. Other approaches accordingly argue that complementary information should be discarded, but this would result in a waste of information. Intuitively, complementarity should exist. The contribution of this paper is to find that existing methods do not have enough insight into the essence of complementarity, i.e., consistency and complementarity are not independent but entangled with each other, which results in conflict. Motivated by this finding, we realize the separation of the two kinds of information through disentangling, specifically making them located in different feature subspaces instead of the same feature space, thus developing a multi-view clustering algorithm that takes into account both consistency and complementarity, effectively extracting the complementary information while avoiding the conflict between consistency and complementarity. Comparative experiments on standard datasets demonstrate the effectiveness of the proposed algorithm.

Key words　multi-view clustering; consistency and complementarity; representation disentanglement; contrastive learning; pseudo-labeling

摘　要　多视图聚类旨在利用来自不同视图的异构信息发现底层数据结构，并划分样本所属类别. 一致性和互补性是影响多视图聚类性能的2个关键要素. 一致性强调不同视图间的语义相似性，互补性则强调每个视图内特有信息的相互补充. 目前对一致性研究已相对深入，但对互补性研究存在争议，其中一些方法认为一致性和互补性能互助，但仅将二者约束至同一特征空间中实际上造成了二者的冲突. 而另一些

收稿日期：2024-06-03；**修回日期**：2024-09-03
基金项目：航空发动机及燃气轮机重大专项基础研究项目(J2019-IV-0018-0086)；国家自然科学基金项目(62076124)
This work was supported by the National Science and Technology Major Project (J2019-IV-0018-0086) and the National Natural Science Foundation of China (62076124).
通信作者：陈松灿(s.chen@nuaa.edu.cn)

方法则据此认为应丢弃互补信息，但这又造成信息浪费. 直觉上互补性应该存在，贡献在于发现了现有方法没有足够洞悉并触及到互补性的本质，即一致性和互补性并非独立而是相互耦合，结果导致冲突. 受此启发，通过解耦实现了 2 种信息的分离，具体使它们位于不同的特征子空间而非现在的同一特征空间，从而发展出了一种兼顾一致性和互补性的多视图聚类算法，在有效提取出互补信息的同时避免二者冲突. 在标准数据集上的对比实验验证了所提算法的有效性.

关键词 多视图聚类；一致性和互补性；表示解耦；对比学习；伪标记

中图法分类号 TP391

DOI: 10.7544/issn1000-1239.202440426 **CSTR:** 32373.14.issn1000-1239.202440426

随着数据采集和存储技术的飞速发展，人们获取数据的来源越来越丰富，对研究对象的描述也逐渐从单一角度发展成为多个角度. 例如，采用多机位从不同视角拍摄物体，或者利用多种特征提取器获取图像的不同特征信息. 这些来自异构域，却用于描述同一对象的数据被称为多视图数据[1]. 针对这类数据的一个重要研究方向是多视图聚类(multi-view clustering, MVC)，其思想是利用来自不同视图的异构信息发现底层数据结构，并将样本划分至所属类别，实现比单一视图聚类更好的效果[2]. 影响多视图聚类性能的 2 个关键要素分别是一致性和互补性，图 1 是二者关系的示意图[3]. 具体来说，一致性对应视图间的公有部分，强调同一样本在不同视图上的语义相似性. 互补性则对应视图内的私有部分，强调每个视图具有自身独特的属性，能够相互补充促进，从而更全面地描述研究对象[4].

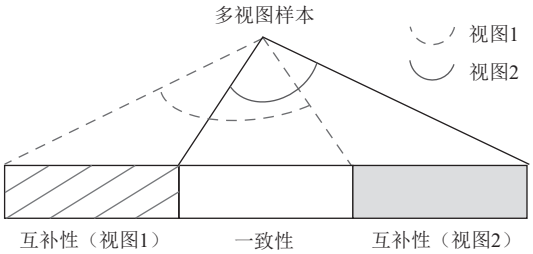


Fig. 1 Illustration of consistency and complementarity

图 1 一致性与互补性的示意图

目前，针对一致性的研究已相当深入，有大量工作围绕其开展[5-9]，但对互补性的研究仍存在争议. 其中一些方法认为一致性与互补性均有助于多视图聚类[10-12]，但却将二者对应的约束施加在同一特征空间，实际上造成了二者之间的对立和冲突，例如文献[11]中所提算法利用对比学习(contrastive learning, CL)来强调一致性，希望不同视图对应的表示更为相似，但与此同时该算法还利用重建误差损失来确保所学表示能够保留每个视图的特有信息，这又使得表示间的相似性减少. 相反，另一些方法则认为互补性会对一致性的学习造成干扰[13-14]，影响聚类性能，因此应当丢弃互补信息，例如文献 [13] 将低级特征空间中的互补信息过滤并丢弃，继而在高级特征空间中使用一致性约束进行样本表示和语义标记的学习. 这类方法通过放弃互补性来避免冲突，但却造成了多视图信息的浪费.

由上述分析可知，当前针对互补性的争议仍然停留在建模过程中是否应该使用互补性，换言之，互补性对于多视图聚类来说是锦上添花还是画蛇添足. 从互补性的定义来说，不同视图提供的特有信息相互补充，有利于全面描述样本. 从经验和直觉来说，除了使用一致性来关注所有视图兼有的对象主体(如动物形态)外，互补性能额外提供同样具有判别性的信息(如动物栖息地、皮毛颜色等)[15]. 因此，本文认为互补性本身具有价值，但由于现有方法没有足够洞悉并触及到互补性的本质，即一致性和互补性并非独立而是相互耦合，才会导致二者冲突并引发争议.

本文的这一发现为当前争议提供了一个出口，即，若能解开一致信息和互补信息之间的耦合，便可充分利用互补信息，并使一致性和互补性相互促进而非制约. 据此，本文通过表示解耦实现一致信息和互补信息的分离，使它们位于不同的特征子空间而非同一特征空间，提出了消除互补性争议的多视图聚类(multi-view clustering via eliminating the complementarity controversy, MCECC)算法，该算法能够兼顾一致性和互补性，在有效提取出互补信息的同时避免二者冲突. 除此之外，本文发现常用的样本层面对比损失只在视图间构建正样本对，而忽视了视图内的样本关系，导致视图内同类样本被强制构建为负样本对，造成大量假负样本对的产生. 为解决这一问题，本文从聚类层面设计对比损失，利用伪标记来重新定义视图内样本关系，去除假负样本对.

本文的主要贡献包括 3 个方面：

1)发现当前争议并未触及互补性的本质,由此出发进一步提出 MCECC 算法. 不同于现有方法在同一特征空间中同时学习一致性和互补性而导致二者冲突,MCECC 通过表示解耦将一致信息和互补信息分离至不同的特征子空间,在有效提取出信息的同时避免冲突.

2)设计聚类层面对比损失用于增强整体表示的判别性. 不同于现有多视图聚类方法仅关注视图间对比,MCECC 进一步关注视图内对比,利用置信度标准选择伪标记,将同类样本在视图内拉近,不同类样本在视图内和视图间推远,从而去除假负样本对并加强对比力度.

3)在多个标准数据集上的对比实验验证了本文所提 MCECC 算法的有效性.

1 相关工作

1.1 多视图聚类

多视图聚类旨在利用来自不同视图的异构信息来发现底层数据结构,并将样本划分至所属类别. 现有多视图聚类方法可大致分为 5 类[2],分别是基于非负矩阵分解的方法[16-17]、基于子空间学习的方法[18-19]、基于多核学习的方法[20-21]、基于图学习的方法[22-23]和基于深度学习的方法[9-13,24-27]. 其中前 4 类方法也被统称为传统方法,其表征能力有限且计算复杂度较高,在需要处理复杂数据的真实场景中性能受限[13]. 而深度神经网络通过非线性参数映射,能够挖掘数据背后丰富的信息,具有强大的表征能力,在多视图聚类领域中逐渐成为主流. 例如,文献 [10] 使用双层嵌套编码器网络,在统一框架内同时执行特定视图和整体样本的表示学习. Xu 等人[13]提出多层次特征学习框架,在低级特征空间中进行视图重建,在高级特征空间中学习表示和语义标记. Lin 等人[26]则从信息论的角度出发,通过最大化互信息和最小化条件熵,将多视图的表示学习和视图补全融合在一个整体深度框架中.

由于上述原因,本文算法基于深度学习开展,但所提思想并不局限于此,而是可以推广到传统方法中. 具体来说,本文从多视图聚类中的 2 个重要性质入手,分析当前针对互补性研究的争议,并强调争议的产生是源于现有方法对互补性本质的忽视,进而由本质出发,设计能够消除争议的算法.

1.2 对比学习

对比学习是无监督学习中的一个重要范式[28-30],其基本思想非常直观,即在特征空间中拉近相似样本(即正样本对)而推远不同样本(即负样本对). 如 Chen 等人[29]设计的对比学习框架将同一样本的不同增广视图视为正样本对. 而多视图数据中样本的不同视图恰好对应着对比学习中人为生成的增广视图,且多视图学习中的一致性强调同一样本在不同视图上的语义相似性,即将同一样本的不同视图视为正样本对予以拉近. 因此,对比学习的思想在多视图学习中天然存在.

由于这种天然的联系,对比学习被广泛运用在多视图聚类中[13,26-27,31-33]. 例如,文献 [26] 从样本层面进行对比学习,通过最大化同一样本不同视图表示间的互信息来保证多视图聚类中的一致性. Chen 等人[27]则提出针对聚类结果的跨视图对比学习算法,用于探索不同视图上一致的语义标签信息.

2 消除互补性争议的多视图聚类算法

本节详细介绍本文提出的 MCECC 算法,算法的模型框架如图 2 所示.

该模型主要由 3 个模块组成,分别是表示解耦模块、视图重建模块和聚类层面对比学习模块. 具体来说,表示解耦模块旨在利用互信息约束和正交约束来解开一致信息与互补信息间的耦合. 视图重建模块旨在综合利用一致信息和互补信息来准确复原出原始视图. 聚类层面对比学习模块则旨在利用置信度标准选择伪标记,将同类样本在视图内拉近,将不同类样本在视图内和视图间推远.

现将问题描述如下:给定具有 V 个视图和 N 个样本的多视图数据集 $\{\boldsymbol{X}^{(v)}\in\mathbb{R}^{D_v\times N}\}_{v=1}^{V}$,其中 D_v 表示视图 v 的维度,$\boldsymbol{X}^{(v)}=(\boldsymbol{X}_1^{(v)},\boldsymbol{X}_2^{(v)},\cdots,\boldsymbol{X}_N^{(v)})$ 表示视图 v 的数据,$\boldsymbol{X}_i^{(v)}$ 表示样本 i 的视图 v 数据. 多视图聚类的目标是将 N 个样本划分至 K 个聚类中.

2.1 表示解耦模块

由于一致性与互补性并非独立存在而是相互耦合,所以才会导致冲突. 受此启发,本文设计表示解耦模块,旨在解开一致信息和互补信息之间的耦合,将二者分离至不同的特征子空间,从而突破现有方法中一致性和互补性相互抑制的困境,在有效提取出互补信息的同时实现 2 种性质的相互促进.

具体来说,将 $\boldsymbol{X}^{(v)}$ 分别通过其对应的一致编码器 $f_{\mathrm{con}}^{(v)}(\boldsymbol{X}^{(v)})$ 和互补编码器 $f_{\mathrm{com}}^{(v)}(\boldsymbol{X}^{(v)})$,得到一致表示矩阵 $\boldsymbol{H}^{(v)}\in\mathbb{R}^{d\times N}$ 和互补表示矩阵 $\boldsymbol{S}^{(v)}\in\mathbb{R}^{d\times N}$,其中 d 是特征子空间的维度,$\boldsymbol{X}_i^{(v)}$ 的一致表示和互补表示分别记为 $\boldsymbol{H}_i^{(v)}$ 和 $\boldsymbol{S}_i^{(v)}$. 通过编码器网络进行的初步分离旨在保证

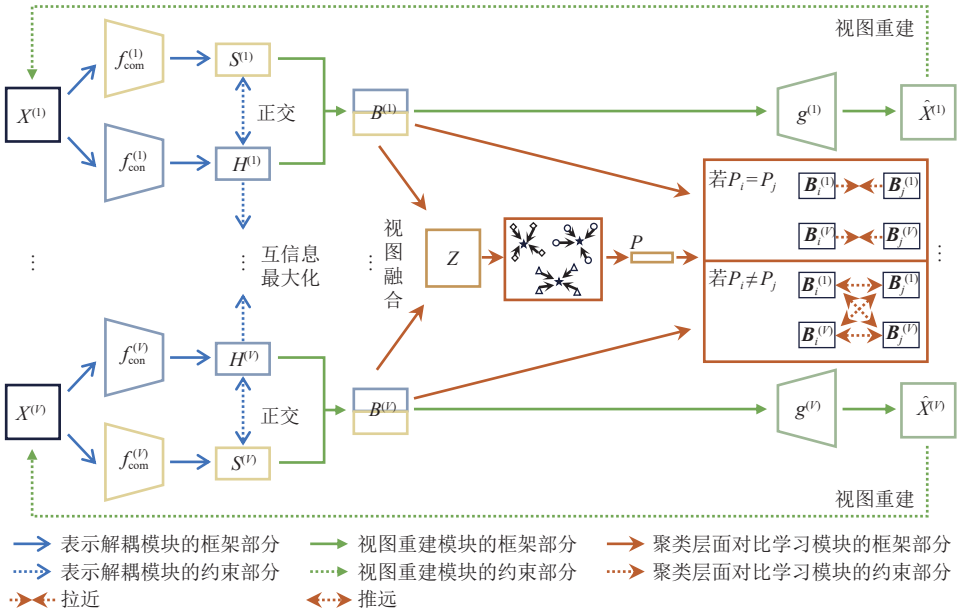


Fig. 2 The framework of MCECC

图 2 MCECC 模型框架

一致信息和互补信息能够被独立学习且不会彼此干扰. 进一步, MCECC 在表示解耦模块中同时使用互信息约束和正交约束来实现 2 种信息的深度分离, 其思想是先通过互信息最大化来固定一致表示所在的特征子空间, 再利用正交约束确定互补表示所在的特征子空间, 并使得 2 个特征子空间彼此正交, 从而有效解开一致信息和互补信息的耦合关系, 减少二者间的冗余.

首先, 对每个视图的一致表示进行样本层面对比学习. 具体来说, 将同一样本在不同视图上的一致表示视为正对, 通过最大化正对之间的互信息来提高一致表示的相似性[26,32]. 于是, 视图 t 和视图 k 间的样本层面对比损失被定义为

$$\ell_{\text{ins}}^{(tk)}=-\sum_{i=1}^{N}(I(\boldsymbol{H}_i^{(t)};\boldsymbol{H}_i^{(k)})+\alpha\sum_{v=t,k}H(\boldsymbol{H}_i^{(v)})),\tag{1}$$

其中 I 和 H 分别表示互信息函数和信息熵函数. α 是用于调节正则化项的平衡参数, 本文将其固定为 9, 与文献 [26] 保持一致. 模型中编码器网络的最后一层为 softmax 函数, 因此 $\boldsymbol{H}^{(v)}$ 中的每个元素都可被视为一个过聚类的类别概率, $\boldsymbol{H}^{(v)}$ 则可被看作离散随机变量 h 在 d 类上的分布[34]. 从该角度出发, 式(1)中的正则化项可以避免所有样本被分配到一个特定聚类中. 进一步, 将式(1)推广到多视图场景, 得到相应样本层面对比损失

$$\mathcal{L}_{\text{icl}}=\frac{1}{2N}\sum_{t=1}^{V}\sum_{k\neq t}\ell_{\text{ins}}^{(tk)}.\tag{2}$$

式(2)通过最大化视图间互信息来固定一致表示所在的特征子空间. 进一步, 再在视图内部使用正交约束来实现一致表示和互补表示的解耦. 正交损失被定义为

$$\mathcal{L}_{\text{ort}}=\frac{1}{N}\sum_{v=1}^{V}\sum_{i=1}^{N}((\boldsymbol{S}_i^{(v)})^{\mathrm{T}}\boldsymbol{H}_i^{(v)}-\beta H(\boldsymbol{S}_i^{(v)})),\tag{3}$$

其中 β 与式(1)中的 α 作用相似. 式(3)获得最优解时, 样本的一致表示与互补表示彼此正交, 从而能够确定互补表示所在的特征子空间, 实现一致信息和互补信息的解耦, 并最大程度减少表示间的冗余, 挖掘尽可能真实的互补信息. 而最大化 $H(\boldsymbol{S}_i^{(v)})$ 则能够避免模型的学习陷入平凡解.

2.2 视图重建模块

视图重建模块旨在综合利用一致表示和互补表示来恢复出原始视图, 保证模型能够学习到充分的视图信息. 具体来说, 将一致表示和互补表示进行拼接融合, 得到样本的视图表示 $\boldsymbol{B}_i^{(v)}=(\boldsymbol{H}_i^{(v)};\boldsymbol{S}_i^{(v)})$, 再将其通过视图 v 对应的解码器 $g^{(v)}(\boldsymbol{B}_i^{(v)})$ 来生成重建后的视图数据 $\hat{\boldsymbol{X}}_i^{(v)}$. 视图重建误差损失被定义为

$$\mathcal{L}_{\text{rec}}=\frac{1}{N}\sum_{v=1}^{V}\sum_{i=1}^{N}\left\|\boldsymbol{X}_i^{(v)}-\hat{\boldsymbol{X}}_i^{(v)}\right\|_2^2.\tag{4}$$

虽然重建误差损失被广泛运用在多视图聚类中, 但从模型全局来看, 本文的视图重建模块与现有方法存在不同. 主要体现在, 现有方法希望通过重建误差损失来保留视图表示中的互补信息, 而强调一致性的约束又会从视图表示中去除互补信息, 这就使

得二者的作用相互抵触, 产生了冲突. 然而本文在表示解耦模块中率先将一致表示和互补表示分离, 使得重建误差损失仍可用于保留充分的视图信息, 强调一致性的约束也可精确作用于一致表示, 进而消除冲突, 使得重建误差损失和模型中其他部分相互促进.

2.3 聚类层面对比学习模块

由于缺乏标记信息的引导, 样本层面的对比学习往往只将同一样本的不同视图视为正对, 而忽视了同类样本之间的关系, 甚至将这些本应拉近的样本彼此推远, 造成了假负样本对的产生. 为此, 本文利用学习过程中产生的伪标记, 设计了聚类层面的对比损失函数, 将同类样本在视图内拉近, 不同类样本在视图内和视图间推远, 从而增强所学表示的判别性.

在视图融合的过程中, 由于同一样本不同视图上的一致表示基本相同, 所以无需重复参与融合, 此处对它们进行平均, 得到多视图样本的一致表示为

$$\boldsymbol{H}_i = \frac{1}{V}\sum_{v=1}^{V}\boldsymbol{H}_i^{(v)}. \tag{5}$$

于是多视图样本表示 $\boldsymbol{Z}_i = (\boldsymbol{H}_i; \boldsymbol{S}_i^{(1)}; \cdots; \boldsymbol{S}_i^{(V)})$, 多视图样本表示矩阵 $\boldsymbol{Z}=(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \cdots, \boldsymbol{Z}_N) \in \mathbb{R}^{d(V+1)\times N}$. 使用 k-means 对 $\boldsymbol{Z}$ 进行聚类, 此处的目标是得到聚类中心 $\boldsymbol{U}=(\boldsymbol{U}_1, \boldsymbol{U}_2, \cdots, \boldsymbol{U}_K) \in \mathbb{R}^{d(V+1)\times K}$, 其中 $\boldsymbol{U}_c$ 表示第 c 个聚类的中心. 进而通过计算样本到每个聚类中心的距离[35-36]得到聚类软标记 $\boldsymbol{Q}=(\boldsymbol{Q}_1, \boldsymbol{Q}_2, \cdots, \boldsymbol{Q}_N) \in \mathbb{R}^{K\times N}$, 具体公式为

$$q_{ci} = \frac{(1+\|\boldsymbol{Z}_i-\boldsymbol{U}_c\|_2^2)^{-1}}{\sum_{c=1}^{K}(1+\|\boldsymbol{Z}_i-\boldsymbol{U}_c\|_2^2)^{-1}}, \tag{6}$$

其中 $\boldsymbol{Q}_i$ 是样本 i 的聚类软标记, q_{ci} 是 $\boldsymbol{Q}_i$ 的元素, 代表样本 i 划分到第 c 个聚类的概率.

进一步, 本文设置置信度阈值 γ, 以获取更为可靠的伪标记 $\boldsymbol{P} \in \mathbb{R}^n$ 用于聚类层面对比学习[37], 其中 n 是所选取出的可靠样本数. 具体来说, 若样本 i 的软标记中最大类概率值高于 γ, 则选择其作聚类层面对比学习, 并将最大类概率对应的聚类标记作为样本 i 的伪标记 P_i. 该过程可表示为

$$P_i = \begin{cases} \arg\max(\boldsymbol{Q}_i), & \text{if } \max(\boldsymbol{Q}_i) \geqslant \gamma, \\ -1, & \text{其他}, \end{cases} \tag{7}$$

其中伪标记为−1 则意味着该样本当前置信度低, 不参与聚类层面的对比学习.

学得伪标记后, 本文将同类样本在视图内拉近, 不同类样本在视图内和视图间推远. 以样本 i 在视图 t 和视图 k 上的表示为例, 2 种操作可分别形式化为

$$\ell_{\text{pos}}^{(tk)}(i) = \sum_{j=1, j\neq i}^{n}\mathbb{I}(P_i = P_j)\,\mathrm{e}^{sim(\boldsymbol{B}_i^{(t)}, \boldsymbol{B}_j^{(t)})/\tau}, \tag{8}$$

$$\ell_{\text{neg}}^{(tk)}(i) = \sum_{j=1, j\neq i}^{n}\mathbb{I}(P_i \neq P_j)\sum_{v=t,k}\mathrm{e}^{sim(\boldsymbol{B}_i^{(t)}, \boldsymbol{B}_j^{(v)})/\tau}, \tag{9}$$

其中 sim 表示余弦相似度函数, τ 表示温度系数, 实验中将 τ 固定为 0.5. 由式(8)(9), 本文设计了 2 个视图上的聚类层面对比损失函数

$$\ell_{\text{clu}}^{(tk)} = -\sum_{i=1}^{n}\ln\frac{\ell_{\text{pos}}^{(tk)}(i)}{\ell_{\text{neg}}^{(tk)}(i)+\ell_{\text{pos}}^{(tk)}(i)}, \tag{10}$$

进一步将其推广到多视图场景, 得到相应聚类层面对比损失为

$$\mathcal{L}_{\text{ccl}} = \frac{1}{2N}\sum_{t=1}^{V}\sum_{k\neq t}\ell_{\text{clu}}^{(tk)}. \tag{11}$$

该损失能够增强所学表示的判别性, 从而提升算法性能.

2.4 整体目标函数和优化方法

综合上述内容, MCECC 的整体目标函数为

$$\mathcal{L} = \mathcal{L}_{\text{icl}} + \lambda_1\mathcal{L}_{\text{ort}} + \lambda_2\mathcal{L}_{\text{rec}} + \lambda_3\mathcal{L}_{\text{ccl}}, \tag{12}$$

其中 λ_1, λ_2, λ_3 是调节各项损失的超参数. 式(12)等号右侧的前 2 项对应图 2 中的表示解耦模块, 第 3 项对应视图重建模块, 最后 1 项对应聚类层面对比学习模块.

实验过程中, 先在前 100 轮使用式(12)等号右侧的前 3 项进行初始化训练, 目的是使模型能够生成稳定的表示, 提高伪标记的可靠性. 然后使用整体目标函数进行后续训练, 增强表示的判别性. 本文采用小批量梯度下降法(mini-batch gradient descent, MBGD)训练模型, 学习率设为 0.000 1, 样本批次大小设为 256.

MCECC 算法的流程如算法 1 所示.

算法 1. 消除互补性争议的多视图聚类算法.

输入: 多视图数据 $\{\boldsymbol{X}^{(v)}\}_{v=1}^{V}$, 迭代次数 E, 聚类个数 K, 平衡参数 α 和 β, 温度系数 τ, 超参数 λ_1, λ_2 和 λ_3, 阈值 γ;

输出: 多视图样本表示矩阵 $\boldsymbol{Z}$, 聚类结果.

① 利用式(2)~(4)初始化 $\{f_{\text{con}}^{(v)}, f_{\text{com}}^{(v)}, g^{(v)}\}_{v=1}^{V}$;

② for *epoch* = 1 to E do

③ 根据式(2)计算样本层面对比损失 $\mathcal{L}_{\text{icl}}$;

④ 根据式(3)计算正交损失 $\mathcal{L}_{\text{ort}}$;

⑤ 根据式(4)计算视图重建误差损失 $\mathcal{L}_{\text{rec}}$;

⑥ 利用式(5)(6)更新软标记 $\boldsymbol{Q}$;

⑦ 利用式(7)更新伪标记 $\boldsymbol{P}$;

⑧ 根据式(11)计算聚类层面对比损失$\mathcal{L}_{\mathrm{ccl}}$;

⑨ 利用式(12)更新$\{f_{\mathrm{con}}^{(v)}, f_{\mathrm{com}}^{(v)}, g^{(v)}\}_{v=1}^{V}$;

⑩ end for

⑪ 通过$\{f_{\mathrm{con}}^{(v)}, f_{\mathrm{com}}^{(v)}\}_{v=1}^{V}$得到$\{\boldsymbol{H}^{(v)}, \boldsymbol{S}^{(v)}\}_{v=1}^{V}$;

⑫ 通过视图融合得到多视图样本表示矩阵$\boldsymbol{Z}$;

⑬ 利用k-means对$\boldsymbol{Z}$聚类得到聚类结果.

3 实验与分析

3.1 实验方案

本文选取多视图聚类领域广泛使用的4个标准数据集进行实验[6,11,26-27], 分别是Caltech101-20[38], LandUse-21[39], Scene-15[40], Scene-15-3V[40]. 详细信息如表1所示.

Table 1 Detailed Information of Datasets

表1 数据集详细信息

数据集	聚类数	样本数	视图数	视图维度
Caltech101-20	20	2 386	2	1 984，512
LandUse-21	21	2 100	2	59，40
Scene-15	15	4 485	2	59，20
Scene-15-3V	15	4 485	3	59，20，40

3种具有代表性的聚类指标被用于衡量算法效果, 分别是准确率(accuracy, ACC)、标准化互信息(normalized mutual information, NMI)和调整兰德指数(adjusted rand index, ARI). 这3种指标的值越高意味着聚类效果越好.

为全面评估MCECC的性能, 本文选取了10个现有多视图聚类算法进行对比, 分别是深度典型相关分析(deep canonical correlation analysis, DCCA)[5]、深度典型相关自动编码器(deep canonically correlated autoencoders, DCCAE)[8]、双对齐不完整多视图聚类(doubly aligned incomplete multi-view clustering, DAIMC)[17]、二元多视图聚类(binary multi-view clustering, BMVC)[41]、双层嵌套编码器网络(autoencoder in autoencoder networks, AE^2-Nets)[10]、对比多视图聚类(contrastive multi-view clustering, CoMVC)[42]、基于多层次特征学习的对比多视图聚类(multi-level feature learning for contrastive multi-view clustering, MFLVC)[13]、共识谱旋转融合(consensus spectral rotation fusion, CSRF)[22]、跨视图对比学习(cross-view contrastive learning, CVCL)[27]和对偶对比预测(dual contrastive prediction, DCP)[26]. 其中DAIMC, BMVC, CSRF属于传统方法, 其余算法属于基于深度学习的方法. 所有算法均重复5次独立实验后取平均值作为最终结果.

3.2 实验结果分析

表2给出了各个多视图聚类算法在不同数据集上的性能评估结果.

分析表2可见, 所提出的MCECC在所有数据集上的聚类性能均显著优于其他算法. 例如, MCECC在3个数据集上的ACC相较于排名第2的算法分别提升了3.64个百分点、2.79个百分点、1.52个百分点. 在Caltech101-20数据集上的ARI相较DCP显著提升了5.68个百分点. 此外: 1)相较于只强调一致性而丢弃互补信息的方法MFLVC, MCECC在LandUse-

Table 2 Clustering Performance Comparison on Different Datasets

表2 不同数据集上的聚类性能比较 %

算法	Caltech101-20			LandUse-21			Scene-15		
	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
DCCA	41.89	59.14	33.39	15.51	23.15	4.43	36.18	38.92	20.87
DCCAE	44.05	59.12	34.56	15.62	24.41	4.42	36.44	39.78	21.47
DAIMC	45.48	61.79	32.40	24.35	29.35	10.26	32.09	33.55	17.42
BMVC	42.55	63.63	32.33	25.34	28.56	11.39	40.50	41.20	24.11
AE^2-Nets	49.10	65.38	35.66	24.79	30.36	10.35	36.10	40.39	22.08
CoMVC	38.67	61.48	31.38	25.58	31.92	13.00	30.64	30.31	13.62
MFLVC	45.05	49.59	40.20	21.73	24.90	8.38	39.97	42.52	24.38
CSRF	60.39	66.22	47.99	25.95	31.86	11.67	41.23	42.03	24.13
CVCL	40.46	61.22	34.00	25.99	30.05	12.09	40.62	42.92	25.19
DCP	70.18	68.06	76.88	26.23	30.65	13.70	41.07	45.11	24.78
MCECC（本文）	**73.82**	**71.71**	**82.56**	**29.02**	**33.73**	**15.44**	**42.75**	**45.58**	**26.66**

注：每列中最好的结果以黑体标注，次优的结果以下划线标注.

21 数据集上的 3 项指标分别提升了 7.29 个百分点、8.83 个百分点、7.06 个百分点,说明 MCECC 能避免信息浪费并充分利用互补信息. 2)相较于将一致性和互补性约束至同一特征空间而造成二者冲突的方法 DCP, MCECC 对应的 3 项指标在 LandUse-21 数据集上分别实现了 2.79 个百分点、3.08 个百分点、1.74 个百分点的提升,说明 MCECC 通过表示解耦,能够避免一致性和互补性的冲突. 由对比实验可知 MCECC 具有较好的聚类性能,进一步验证了本文动机的合理性.

进一步, Scene-15 数据集上随视图数的增加, DCP 和 MCECC 这 2 种算法聚类性能的变化及对比如图 3 所示. 图例中 2V-和 3V-分别表示使用表 1 中介绍的 Scene-15 数据集(2 个视图)和 Scene-15-3V 数据集(3 个视图),二者区别在于后者增加了 LBP 视图. 通过观察可以发现: 1)随着视图数的增加, 2 种算法的聚类性能均有提升,但 MCECC 的提升更为显著,例如 NMI 指标上 DCP 仅提升 0.12 个百分点而 MCECC 提升 1.3 个百分点. 2)仅使用 2 个视图数据的 MCECC 比使用 3 个视图数据的 DCP 具有更好的聚类性能. 3)在均使用 3 个视图数据时, MCECC 的各项指标值相较于 DCP 分别提升了 2.06 个百分点、1.65 个百分点、1.74 个百分点. 实验结果说明 MCECC 能够更好地兼顾一致性和互补性,既能保留新增视图中的一致信息,又能挖掘其中的互补信息,体现了多视图学习的优势.

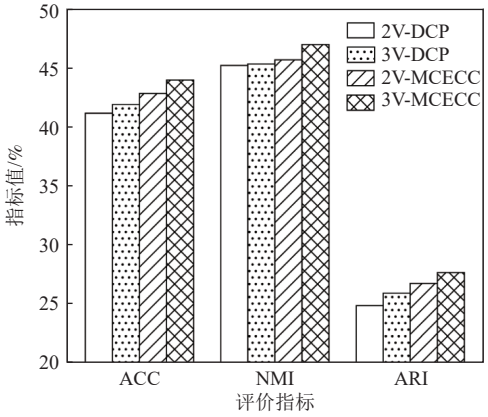
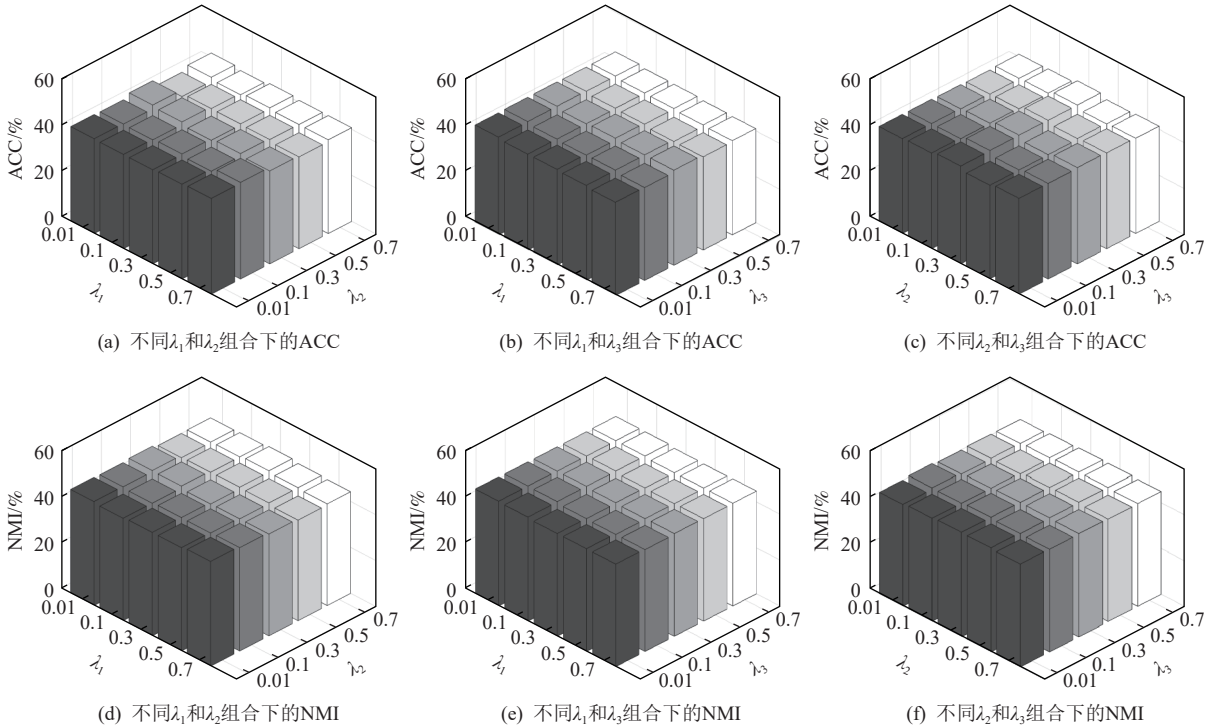


Fig. 3 Clustering performance of two algorithms with different number of views on Scene-15 dataset

图 3 Scene-15 数据集上 2 种算法在不同视图数时的聚类性能

3.3 参数敏感性分析

由式(12)可知, MCECC 包含 3 个超参数 λ_1, λ_2, λ_3, 本文使用网格搜索法来确定三者的最优参数组合,其中 λ_1, λ_2, λ_3 的取值均在 {0.01, 0.1, 0.3, 0.5, 0.7} 范围内进行调整. 本节以 Scene-15 数据集为例,将 3 个超参数两两组合,展示不同组合对应的 ACC 和 NMI 变化情况,如图 4 所示. 由图 4 可见 MCECC 在取值范围内对超参数不敏感,说明 MCECC 的性能较为稳定.



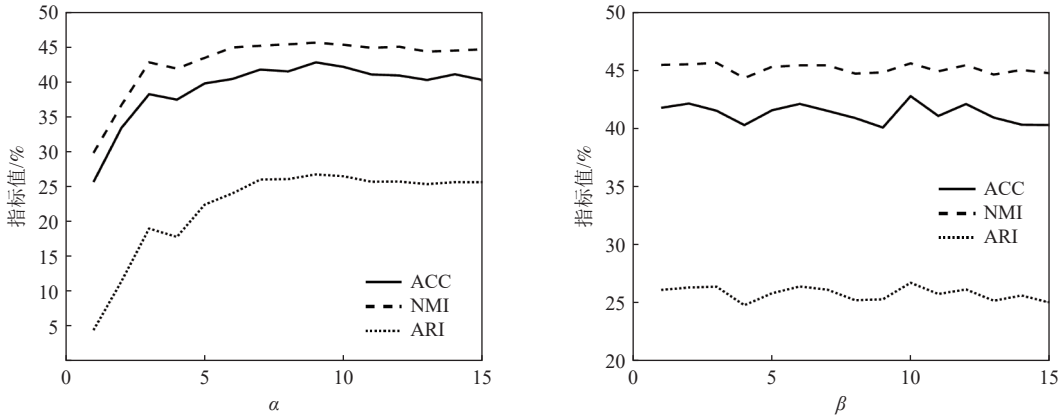
(a) 不同λ_1和λ_2组合下的ACC (b) 不同λ_1和λ_3组合下的ACC (c) 不同λ_2和λ_3组合下的ACC

(d) 不同λ_1和λ_2组合下的NMI (e) 不同λ_1和λ_3组合下的NMI (f) 不同λ_2和λ_3组合下的NMI

Fig. 4 Parameter sensitivity analysis on Scene-15 dataset

图 4 Scene-15 数据集上的参数敏感性分析

此外，本节分别针对式(1)中的平衡参数α和式(3)中的平衡参数β进行敏感性分析，展示不同取值时聚类性能的变化，如图 5 所示. 其中α和β的取值均在[1, 15]范围内进行调整，取值间隔为 1. 由图 5(a)可见，随着α值逐渐增大，算法的聚类性能先是明显提升，而后趋于稳定. 由图 5(b)可见，随着β值的变化，聚类性能虽出现波动，但整体趋势稳定. 由此可见，MCECC 对α的取值较为敏感，而对β的取值不敏感.



(a) 3个指标随平衡参数α的变化 (b) 3个指标随平衡参数β的变化

Fig. 5 Balance parameter sensitivity analysis on Scene-15 dataset

图 5 Scene-15 数据集上的平衡参数敏感性分析

3.4 消融实验

本节在 Scene-15 数据集上进行损失函数消融实验，实验结果如表 3 所示. 由于表示解耦必须同时依赖样本层面对比损失和正交损失，因此将二者视为一个组合进行分析. 由表 3 可知，该组合对于聚类性能的提升最为重要，其原因是样本层面对比损失可以保留视图间的一致信息，正交损失则可以在保留视图内互补信息的同时避免一致性和互补性的冲突. 此外，视图重建误差损失的增加使得性能有所提升，说明其能够保证模型学习到充分的视图信息. 进一步，聚类层面对比损失也有利于提升聚类性能，其原因可能是该损失能够增强所学表示的判别性. 综上，MCECC 的各项损失均有意义.

Table 3 Ablation Study on Scene-15 Dataset

表 3 Scene-15 数据集上的消融实验 %

$\mathcal{L}_{icl}+\mathcal{L}_{ort}$	$\mathcal{L}_{rec}$	$\mathcal{L}_{ccl}$	ACC	NMI	ARI
√			39.98	44.76	25.08
	√		26.40	30.03	10.52
√	√		41.37	44.92	26.06
√		√	40.49	44.63	25.96
	√	√	25.36	25.59	8.97
√	√	√	42.75	45.58	26.66

除了对损失函数进行消融实验，本节还进一步模拟现有工作中将一致信息和互补信息约束至同一特征空间的情况，通过去除模型中的互补编码器和正交损失函数，从而不对表示进行解耦，如图 6 所示. 此时的 3 项指标相较于进行解耦的 MCECC 均有所下降，说明针对一致性和互补性的表示解耦的确有利于多视图聚类任务.

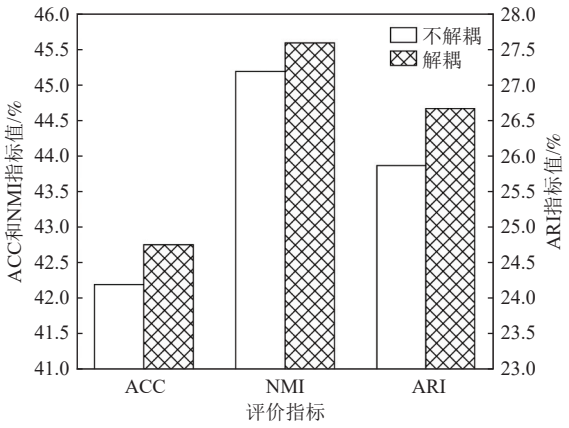


Fig. 6 Effect of representation disentanglement on Scene-15 dataset

图 6 在 Scene-15 数据集上表示解耦的影响

4 结 论

本文首先回顾了现有多视图聚类方法在互补性研究方面存在的主要争议，发现争议焦点并未触及互补性的本质，即一致性和互补性并非独立而是相互耦合，结果造成互补性研究的困境. 不同于现有方法，本文从互补性本质出发，提出了消除互补性争议

的多视图聚类算法(MCECC). 该算法利用互信息约束和正交约束来实现一致信息和互补信息的解耦, 具体使它们位于不同的特征子空间而非同一特征空间, 从而有效提取出互补信息并避免一致性与互补性的冲突. 此外, 本文进一步从聚类层面设计对比损失函数, 利用伪标记将同类样本在视图内拉近, 不同类样本在视图内和视图间同时推远, 从而增强所学表示的判别性. 在标准数据集上的对比实验验证了MCECC 的有效性. 未来工作将继续探索互补性在多视图聚类中的作用, 并针对其他形式的解耦和视图融合方法展开进一步研究, 给出理论证明.

作者贡献声明: 赵玉涵负责提出算法模型, 制定实验方案并完成实验, 撰写和修改论文; 陈松灿提出研究方向和提出指导意见, 并修改论文.

参 考 文 献

[1] Sun Shiliang. A survey of multi-view machine learning[J]. Neural Computing and Applications, 2013, 23: 2031−2038

[2] Chen Mansheng, Lin Jiaqi, Li Xianglong, et al. Representation learning in multi-view clustering: A literature review[J]. Data Science and Engineering, 2022, 7(3): 225−241

[3] Liu Jing, Jiang Yu, Li Zechao, et al. Partially shared latent factor learning with multiview data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(6): 1233−1246

[4] Xu Chang, Tao Dacheng, Xu Chao. A survey on multi-view learning[J]. arXiv preprint, arXiv: 1304.5634, 2013

[5] Andrew G, Arora R, Bilmes J, et al. Deep canonical correlation analysis[C]//Proc of the 30th Int Conf on Machine Learning. New York: ACM, 2013: 1247−1255

[6] Peng Xi, Huang Zhenyu, Lv Jianchen, et al. COMIC: Multi-view clustering without parameter selection[C]//Proc of the 36th Int Conf on Machine Learning. New York: ACM, 2019: 5092−5101

[7] Yu Xiao, Liu Hui, Lin Yuxiu, et al. Consensus guided auto-weighted multi-view clustering[J]. Journal of Computer Research and Development, 2022, 59(7): 1496−1508 (in Chinese)
(于晓, 刘慧, 林毓秀, 等. 一致性引导的自适应加权多视图聚类[J]. 计算机研究与发展, 2022, 59(7): 1496−1508)

[8] Wang Weiran, Arora R, Livescu K, et al. On deep multi-view representation learning[C]//Proc of the 32nd Int Conf on Machine Learning. New York: ACM, 2015: 1083−1092

[9] Yang Hao, Mao Hua, Woo W L, et al. Consistency enhancement-based deep multiview clustering via contrastive learning[J]. arXiv preprint, arXiv: 2401.12648, 2024

[10] Zhang Changqing, Liu Yeqing, Fu Huazhu. AE2-Nets: Autoencoder in autoencoder networks[C]//Proc of the 37th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2019: 2577−2585

[11] Lin Yijie, Gou Yuanbiao, Liu Zitao, et al. Completer: Incomplete multi-view clustering via contrastive prediction[C]//Proc of the 39th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 11174−11183

[12] Geng Chuanxing, Han Aiyang, Chen Songcan. View-labels are indispensable: A multifacet complementarity study of multi-view clustering[J]. arXiv preprint, arXiv: 2205.02507, 2022

[13] Xu Jie, Tang Huayi, Ren Yazhou, et al. Multi-level feature learning for contrastive multi-view clustering[C]//Proc of the 40th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 16051−16060

[14] Liu Suyuan, Liao Qing, Wang Siwei, et al. Robust and consistent anchor graph learning for multi-view clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(8): 4207−4219

[15] Jia Xiaodong, Jing Xiaoyuan, Zhu Xiaoke, et al. Semi-supervised multi-view deep discriminant representation learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(7): 2496−2509

[16] Li Jing, Gao Quanxue, Wang Qianqian, et al. Orthogonal non-negative tensor factorization based multi-view clustering[C]//Proc of the 37th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2023: 18186−18202

[17] Hu Menglei, Chen Songcan. Doubly aligned incomplete multi-view clustering[C]//Proc of the 27th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2018: 2262−2268

[18] Chen Zhe, Wu Xiaojun, Xu Tianyang, et al. Fast self-guided multi-view subspace clustering[J]. IEEE Transactions on Image Processing, 2023, 32: 6514−6525

[19] Luo Shirui, Zhang Changqing, Zhang Wei, et al. Consistent and specific multi-view subspace clustering[C]//Proc of the 32nd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2018: 3730−3737

[20] Liu Jiyuan, Liu Xinwang, Yang Yuexiang, et al. Contrastive multi-view kernel learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(8): 9552−9566

[21] Liu Xinwang. SimpleMKKM: Simple multiple kernel k-means[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 5174−5186

[22] Chen Jie, Mao Hua, Peng Dezhong, et al. Multiview clustering by consensus spectral rotation fusion[J]. IEEE Transactions on Image Processing, 2023, 32: 5153−5166

[23] Lin Zhiping, Kang Zhao. Graph filter-based multi-view attributed graph clustering[C]//Proc of the 30th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2021: 2723−2729

[24] Li Haobin, Li Yunfan, Yang Mouxing, et al. Incomplete multi-view clustering via prototype-based imputation[C]//Proc of the 32nd Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2023: 3911−3919

[25] Tang Huayi, Liu Yong. Deep safe multi-view clustering: Reducing the risk of clustering performance degradation caused by view increase[C]//Proc of the 40th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 202−211

[26] Lin Yijie, Gou Yuanbiao, Liu Xiaotian, et al. Dual contrastive prediction for incomplete multi-view representation learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(4): 4447−4461

[27] Chen Jie, Mao Hua, Woo W L, et al. Deep multiview clustering by contrasting cluster assignments[C]//Proc of the 19th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2023: 16752−16761

[28] Oord A, Li Yazhe, Vinyals O. Representation learning with contrastive predictive coding[J]. arXiv preprint, arXiv: 1807.03748, 2018

[29] Chen Ting, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//Proc of the 37th Int Conf on Machine Learning. New York: ACM, 2020: 1597−1607

[30] He Kaiming, Fan Haoqi, Wu Yuxin, et al. Momentum contrast for unsupervised visual representation learning[C]//Proc of the 38th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 9729−9738

[31] Tian Yonglong, Krishnan D, Isola P. Contrastive multiview coding[C]//Proc of the 16th European Conf on Computer Vision. Berlin: Springer, 2020: 776−794

[32] Tsai Y H, Wu Yue, Salakhutdinov R, et al. Self-supervised learning from a multi-view perspective[C/OL]//Proc of the 9th Int Conf on Learning Representations. 2021[2024-06-01]. https://openreview.net/forum?id=-bdp_8Itjwp

[33] Tang Huayi, Liu Yong. Deep safe incomplete multi-view clustering: Theorem and algorithm[C]//Proc of the 39th Int Conf on Machine Learning. New York: ACM, 2022: 21090−21110

[34] Ji Xu, Henriques J F, Vedaldi A. Invariant information clustering for unsupervised image classification and segmentation[C]//Proc of the 17th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 9865−9874

[35] Van der Maaten L, Hinton G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(86): 2579−2605

[36] Guo Xifeng, Gao Long, Liu Xinwang, et al. Improved deep embedded clustering with local structure preservation[C]//Proc of the 26th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2017: 1753−1759

[37] Sohn K, Berthelot D, Li C L, et al. FixMatch: Simplifying semi-supervised learning with consistency and confidence[C]//Proc of the 34th Int Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2020: 596−608

[38] Li Yeqing, Nie Feiping, Huang Heng, et al. Large-scale multi-view spectral clustering via bipartite graph[C]//Proc of the 29th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2015: 2750−2756

[39] Yang Yi, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification[C]//Proc of the 18th SIGSPATIAL Int Conf on Advances in Geographic Information systems. New York: ACM, 2010: 270−279

[40] Li Feifei, Perona P. A Bayesian hierarchical model for learning natural scene categories[C]//Proc of the 2005 IEEE Computer Society Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2005: 524−531

[41] Zhang Zheng, Liu Li, Shen Fumin, et al. Binary multi-view clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(7): 1774−1782

[42] Trosten D J, Lokse S, Jenssen R, et al. Reconsidering representation alignment for multi-view clustering[C]//Proc of the 39th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 1255−1265



Zhao Yuhan, born in 2000. Master. Student member of CCF. Her main research interests include machine learning and pattern recognition.

赵玉涵, 2000 年生. 硕士. CCF 学生会员. 主要研究方向为机器学习、模式识别.



Chen Songcan, born in 1962. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include pattern recognition, machine learning, and neural computing.

陈松灿, 1962 年生. 博士, 教授, 博士生导师. CCF 高级会员. 主要研究方向为模式识别、机器学习、神经计算.