计算机研究与发展
Journal of Computer Research and Development
https://crad.ict.ac.cn
62(5): 1248−1261, 2025

面向子空间聚类的多视图统一表示学习网络

林毓秀[1,2] 刘 慧[1,2] 于 晓[1,2] 张彩明[2,3]
[1](山东财经大学计算机与人工智能学院 济南 250014)
[2](山东省数字媒体技术重点实验室(山东财经大学) 济南 250014)
[3](山东大学软件学院 济南 250101)
(linyx0109@mail.sdufe.edu.cn)

A Multi-View Unified Representation Learning Network for Subspace Clustering

Lin Yuxiu[1,2], Liu Hui[1,2], Yu Xiao[1,2], and Zhang Caiming[2,3]
[1](*School of Computing and Artificial Intelligence*, *Shandong University of Finance and Economics*, *Jinan* 250014)
[2](*Digital Media Technology Key Lab of Shandong Province* (*Shandong University of Finance and Economics*), *Jinan* 250014)
[3](*School of Software*, *Shandong University*, *Jinan* 250101)

Abstract Multi-view subspace clustering aims to explore rich information across views to guide the clustering process. The key lies in effectively learning the unified representation and subspace representation between views. Recently, deep clustering methods have achieved promising effects due to the powerful representation capability of neural networks. However, the multi-source heterogeneity inherent in multi-view data allows existing methods to encode each view independently with a unimodal encoder, increasing the number of model parameters and limiting the model's generalization capability. Besides, low-rank subspace representation has been shown to facilitate clustering performance, while traditional nuclear norm regularization does not consider the difference between different singular values, leading to biased estimation. To tackle these two problems, we propose a novel multi-view unified representation learning network (namely, MURLN) for subspace clustering. Specifically, MURLN first uses the Transformer as the encoder architecture, which projects different views into the low-dimensional feature space with the same mapping rule by sharing parameters. In addition, a weighted fusion strategy for intra-view samples is conducted to learn a unified representation rationally. Finally, the weighted Schatten *p*-norm is introduced as the low-rank constraint of the subspace representation matrix. Extensive experiments on seven multi-view datasets verify the effectiveness and superiority of our proposed method.

Key words multi-view subspace clustering; Transformer; weighted fusion; low-rank representation; weighted Schatten *p*-norm

摘 要 多视图子空间聚类旨在挖掘多视图的丰富信息来指导高维数据聚类,其研究关键在于如何有效地学习多视图统一表示和子空间表示.近年来,深度聚类方法利用神经网络强大的表征能力取得了优异的性能.然而,多视图数据固有的多源异构性使得大多数现有方法以单模态编码器实现对各个视图的独立编码,不仅增加了模型参数量,同时限制了模型的泛化能力.另一方面,低秩子空间表示被证明能够提

收稿日期: 2024-06-03; **修回日期**: 2024-09-29
基金项目: 国家自然科学基金项目(62072274, U22A2033); 中央引导地方科技发展项目(YDZX2022009); 山东省泰山学者特聘专家基金项目(tstp20221137)
This work was supported by the National Natural Science Foundation of China (62072274, U22A2033), the Central Guidance on Local Science and Technology Development Project (YDZX2022009), and the Special Funds of Taishan Scholars Project of Shandong Province (tstp20221137).
通信作者: 刘慧(liuh_lh@sdufe.edu.cn)

升聚类性能，传统的核范数正则化优化没有考虑不同奇异值隐含的信息量差异，是矩阵秩的一个有偏估计. 为此，提出了一种面向子空间聚类的多视图统一表示学习网络. 首先，基于 Transformer 构建编码器，通过共享参数将异构视图以相同的映射规则投影到低维特征空间. 其次，针对每个样本在不同视图中可能具有不同的表现，采用视图内样本加权融合的方法学习多视图统一表示. 最后，引入加权 Schatten-p 范数对子空间表示矩阵施加低秩约束. 在 7 个多视图数据集上的广泛实验验证了所提方法的有效性和优越性.

关键词 多视图子空间聚类；Transformer；加权融合；低秩表示；加权 Schatten-p 范数

中图法分类号 TP181

DOI: 10.7544/issn1000-1239.202440431 **CSTR:** 32373.14.issn1000-1239.202440431

信息技术的普及和迅速发展，极大地丰富了数据在现实应用场景中的表现形式，使其呈现出高维、多源和异构属性[1]. 例如，在机器人领域，人机交互依赖视觉、听觉、触觉等方面的感官信息；在内容理解领域，多媒体片段包含图片、文本以及音频信号；在医疗诊断领域，脑磁共振成像根据扫描参数的不同分为 T1WI、T2WI、T2-FLAIR 和增强扫描；在人脸识别领域，提取图像的 LBP, HOG, Gabor 等特征以降低阴影及光照变化的影响. 广义地讲，从多角度、多渠道对同一观测对象进行的多样性描述被定义为多视图数据，每一种模态或特征都是描述对象的一个视图[2]. 对比单视图数据，多视图数据包含 2 个重要的信息分量，即一致性信息和互补性信息[3]，能够实现对数据样本的全面理解.

聚类作为一种数据预处理技术，旨在将无标签数据按内在相似性划分到不同类别，已被广泛应用于智能医学[4]、推荐系统[5]、模式识别[6]等任务场景. 因此，协同利用多视图数据蕴含的丰富信息来提升聚类效果具有重要的研究意义. 当前的多视图聚类方法可以粗略地划分为基于协同训练的方法[7]、基于多核学习的方法[8]、多视图图聚类[9]和多视图子空间聚类(multi-view subspace clustering, MvSC)[10-11]四类. MvSC 通过将高维多视图数据映射到低维特征空间，并从中获取统一的特征表示用于子空间自表达学习，是实现高维数据聚类的一类有效方法. 得益于深度神经网络在特征提取和表示方面的显著优势，深度多视图子空间聚类能够探索复杂场景中样本更深层的特征表达，成为近年来的研究热点[12-14].

自编码器(autoencoders, AEs)是一种典型的深度无监督学习模型，通过对称的网络结构将无标签数据样本抽象为特征向量[15]. 国内外学者将其与多视图子空间聚类结合，提出了一系列方法. Abavisani 等人[16]为每个视图分别训练单独的卷积自编码器，然后强制学到的潜在表示具有相同的子空间结构. 在此基础上，Zhu 等人[17]使用 2 组自编码器表征每个视图，分别构建一组特定于视图的子空间表示和一个共享的子空间表示. Zhang 等人[18]提出一种深度嵌套的自编码器模型 AE^2-Nets(autoencoder in autoencoder networks)来退化出视图间共享的紧致表示. 为了提高表示学习的质量，Li 等人[19]提出利用多尺度特征学习策略为每个视图提取不同尺度的低维表示. 总的来说，由于多视图数据来自不一致的数据空间，这些方法通常采用“1 对 1”的编码网络对各个视图进行独立编码. 特别是在处理跨模态的多视图数据时，需要根据具体数据类型选择不同的自编码器架构[20]，如图 1 所示.

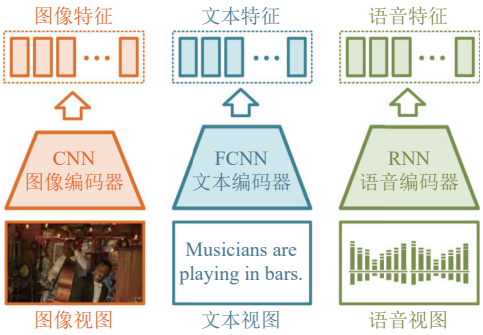


Fig. 1 Schematic diagram of multi-view encoding architecture

图 1 多视图编码架构示意图

近年来，Transformer[21]在自然语言处理和计算机视觉中展现出强大的建模能力以及良好的并行计算能力，逐渐成为处理文本和图像的通用编码架构[22-23]. 为此，本文提出了一种新的面向子空间聚类的多视图统一表示学习网络(multi-view unified representation learning network, MURLN). 在多视图特征提取方面，结合自注意力机制以及多视图学习的特点，提出基于 Transformer 的多视图共享编码框架学习每个视图的潜在表示，同时施加多样性约束以挖掘视图间的互补信息. 编码框架以共享参数的形式进行训练，提高了模型的泛化能力. 在多视图特征融合方面，考虑到同一个观测样本在不同视图中的不同聚类表现，

将多视图融合的目标粒度从整个视图细化到视图内样本, 提出一种样本加权的融合方法来从低维潜在空间中构建统一表示. 通过迭代优化模型参数, 自适应地学习视图中每个样本的合适权重. 最后, 为追求子空间表示的低秩性, 设计基于加权 Schatten-p 范数[24]的正则化约束. 网络结构如图 2 所示.

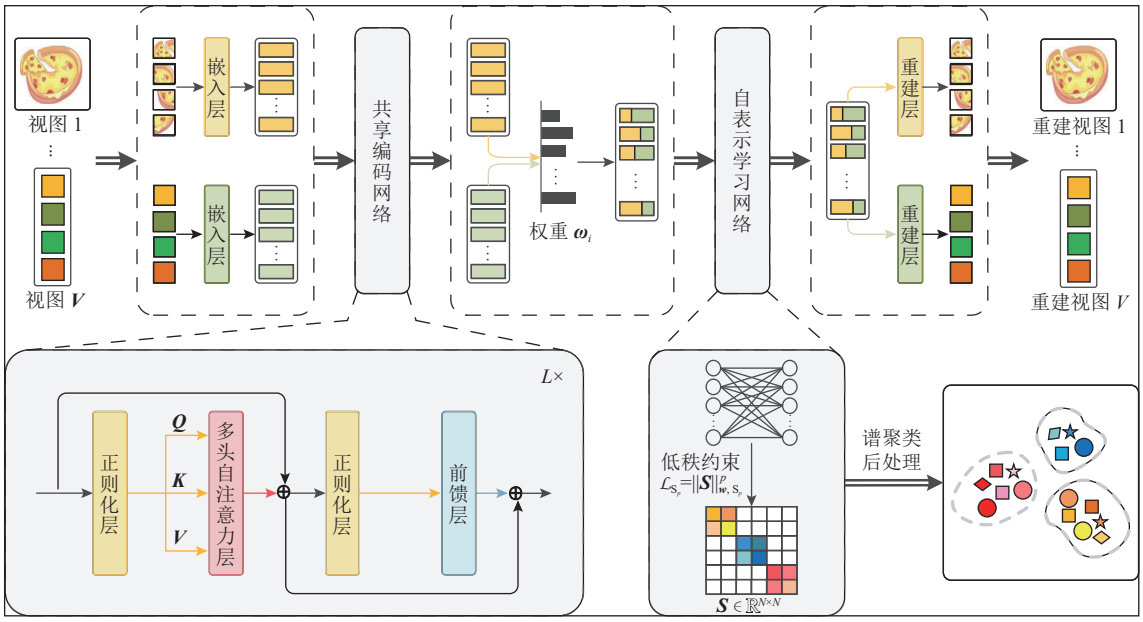


Fig. 2 Framework of MURLN

图 2 MURLN 架构

本文主要贡献包括 4 个方面:

1) 提出基于 Transformer 的多视图共享编码框架, 通过共享网络参数对齐不同视图的潜在表示, 消除多视图数据的异构性差异.

2) 设计了一种样本加权的多视图特征融合方法 MURLN, 降低视图内低质量样本对融合结果的影响, 加强可靠样本的权重.

3) 提出基于加权 Schatten-p 范数的低秩正则化约束, 与现有方法中常用的核范数相比, 由于其考虑了奇异值之间的显著差异, 生成的子空间表示能够更好地逼近原始秩函数.

4) 在多个类型的多视图数据集上进行了大量实验. 结果表明, 与其他方法相比, MURLN 实现了更优越的聚类性能.

1 相关工作

1.1 符号约定和问题定义

为方便描述, 表 1 列出文中常用的符号及定义.

给定具有 V 个视图的数据集 $\{\boldsymbol{X}^{(1)},\boldsymbol{X}^{(2)},\cdots,\boldsymbol{X}^{(V)}\}$, 其中, $\boldsymbol{X}^{(v)}=(\boldsymbol{x}_1^{(v)},\boldsymbol{x}_2^{(v)},\cdots,\boldsymbol{x}_N^{(v)})\in\mathbb{R}^{d_v\times N}$ 表示第 v 个视图维度为 d_v 的数据矩阵. 本文假设数据完整, 即所有视图的样本数量均为 N. 多视图子空间聚类的核心思想是通过整合多视图的一致性信息和互补性信息求解一个自表示系数矩阵 $\boldsymbol{S}\in\mathbb{R}^{N\times N}$, 并在此基础上构建亲和矩阵 $\boldsymbol{W}\in\mathbb{R}^{N\times N}$ 用于谱聚类.

Table 1 Some Main Notations

表 1 一些主要符号

符号	定义
$\boldsymbol{x}_i^{(v)}\in\mathbb{R}^{d_v}$	视图 v 中的第 i 个样本
$\boldsymbol{I}$	单位矩阵
$\lvert\boldsymbol{A}\rvert$	矩阵绝对值操作
$\sigma_i(\boldsymbol{A})$	矩阵 $\boldsymbol{A}$ 的第 i 个奇异值
$\mathrm{diag}(\boldsymbol{A})$	以向量的形式返回矩阵 $\boldsymbol{A}$ 的对角线值
$\mathrm{tr}(\cdot)$	迹函数

1.2 子空间聚类

基于谱聚类的子空间聚类依赖于数据的自表达特性[25-26], 它假设空间中每个样本都可以由同一子空间中其他样本的线性组合表示, 非同一子空间中的样本对应的表示系数几乎为 0. 经典的单视图子空间聚类可以形式化为如下优化问题:

$$\min_{\boldsymbol{S}}\|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{S}\|_{\mathrm{r}}+\lambda\|\boldsymbol{S}\|_{\Omega},\ \text{s.t. } \mathrm{diag}(\boldsymbol{S})=\boldsymbol{0}, \tag{1}$$

其中 $\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N)\in\mathbb{R}^{D\times N}$, D 表示视图维度. $\|\cdot\|_{\mathrm{r}}$ 表示重建损失, $\|\cdot\|_{\Omega}$ 表示正则化约束项. 可选约束项

$\mathrm{diag}(\boldsymbol{S})=\boldsymbol{0}$用来避免得到平凡解$\boldsymbol{S}=\boldsymbol{I}$. 通过优化目标函数式(1), 得到表示系数矩阵$\boldsymbol{S}$, 并计算亲和矩阵$\boldsymbol{W}=(|\boldsymbol{S}|+|\boldsymbol{S}^{\mathrm{T}}|)/2$. 最后, 执行谱聚类算法, 如 Ncut[27], 得到聚类结果.

然而, 利用单视图的有限信息进行聚类存在较多局限, 一旦数据受损可能产生较差的聚类效果. 另外, 由于多视图数据的异构性, 直接使用单视图聚类方法处理多视图数据也是不合理的.

1.3 多视图子空间聚类

近年来, 多视图子空间聚类方法有了很大发展[28-29]. Cao 等人[30]设计基于希尔伯特-施密特独立准则的正则化约束, 隐式地增强了多视图表示的多样性. 直接从原始数据空间进行自表达重构可能会受到噪声影响, Zhang 等人[10]提出将输入数据映射到一个隐空间, 并在隐空间中学习统一的低秩子空间表示 LMSC (latent multiview subspace clustering). 与浅层模型不同, 深度方法利用神经网络提取表达能力更强的数据表征, 受到了广泛关注[31-32]. 在 LMSC 的基础上, Zhang 等人[33]基于神经网络提出一个更泛化的模型 gLMSC. DMSC-UDL[12]通过卷积自编码器整合局部和全局结构信息, 并添加哈达玛积判别约束强化视图内样本的簇间差异. Wang 等人[34]构造一个深度结构化多路径网络(structured multi-pathway network, SMpNet)用于多视图子空间聚类, 结合编码器中间层产生的隐藏特征信息, 提高了自表示系数矩阵的质量.

深度 MvSC 常用的自编码器可分为卷积自编码器和全连接自编码器. 卷积自编码器[35]侧重于通过卷积层的级联保留 2 维图像数据的空间信息, 全连接自编码器则更适用于处理 1 维信号, 例如文本、特征描述子等. 由于视图间的异构性, 现有工作往往采用模态特定的编码器分别对不同视图独立编码, 有较高的训练成本或优化成本. 尽管可以依赖预训练好的模型(例如, Faster-RCNN[36]和 BERT[37])预先提取数据样本的抽象特征[38], 但模态间缺少交互性. Transformer 模型的出现在很大程度上为该问题提供了一种有前途的解决方法. 根据文献 [23], 依赖于自注意力机制的 Transformer 允许网络根据输入内容动态收集相关特征, 被证明能够作为通用的基础编码模型用于不同模态数据的特征学习.

为了区分不同视图对聚类性能提升的贡献程度, Wang 等人[39]使用一个视图级融合算子实现多个潜在表示的加权融合. Lu 等人[40]引入注意力机制获得每个视图的动态权重, 使融合过程更加关注重要的视图. 实际上, 质量较差的视图可能存在有价值的观测样本, 质量较好的视图同样可能存在不重要或者部分受损的样本. 因此, 将每个视图视为一个整体来分配权重的全局融合方法, 忽视了同一视图中不同样本的不同重要性, 从而对统一表示的学习造成不利影响.

1.4 低秩表示学习

基于谱聚类的子空间聚类算法性能高度依赖于自表示系数矩阵的质量. 大多数现有方法通过构建适当的正则化约束来保证系数矩阵满足某些特定属性, 常见的包括稀疏子空间聚类(sparse subspace clustering, SSC)[25]和低秩表示(low-rank representation, LRR)[26]. SSC 使用ℓ_0或ℓ_1范数正则化来约束子空间表示的稀疏性. LRR 将系数矩阵的秩函数凸松弛为核范数, 通过求解核范数最小化加强子空间表示的低秩性. 在多视图子空间聚类问题中, 低秩表示由于其良好的聚类性能和噪声鲁棒性而被广泛研究[41-43]. 根据标准核范数的定义, 其在最小化过程中将所有非零奇异值平均化, 是矩阵秩的一个有偏估计. 为此, 构建低秩正则化约束需要充分考虑奇异值之间的差异性.

2 本文方法

本文设计一个面向子空间聚类的多视图统一表示学习网络 MURLN. 一方面基于 Transformer 结构, 通过参数共享将每个视图以相同映射规则投影到低维潜在空间, 同时引入多样性约束来提高多视图的互补性. 另一方面通过向子空间表示学习引入基于加权 Schatten-p 范数的低秩约束正则化项来获得数据的最低秩表示. 网络结构如图 2 所示.

2.1 模型结构

2.1.1 特征嵌入模块

MURLN 首先提出一个特征嵌入模块, 将不同视图的输入转换为相同维度的特征嵌入. 更具体地说, 对于包含V个视图的数据集合, 特征嵌入模块由V个特定于视图的嵌入层组成. 以第v个视图输入$\boldsymbol{X}^{(v)}\in\mathbb{R}^{d_v\times N}$为例, 嵌入层通常包含 3 步操作: 1)将输入数据划分为M个不重叠的块(patch); 2)将块线性映射(通常由可学习的线性层实现)到k维的嵌入空间, 得到视图嵌入$\boldsymbol{X}_{\mathrm{emb}}^{(v)}\in\mathbb{R}^{M\times k\times N}$; 3)构造位置编码$\boldsymbol{X}_{\mathrm{pos}}^{(v)}\in\mathbb{R}^{M\times k\times N}$来保留数据内的相对关系. 嵌入层的输出$\boldsymbol{X}_{\mathrm{E}}^{(v)}\in\mathbb{R}^{M\times k\times N}$最终表示为带有位置编码的视图嵌入:

$$\boldsymbol{X}_{\mathrm{E}}^{(v)}=\boldsymbol{X}_{\mathrm{emb}}^{(v)}+\boldsymbol{X}_{\mathrm{pos}}^{(v)}. \tag{2}$$

由于视图间的异构性, 嵌入层中的块划分和位

置编码策略取决于输入视图的模态类型. 以长和宽分别为H和W的C通道2维图像视图$\boldsymbol{X}^{(v)}\in\mathbb{R}^{H\times W\times C\times N}$为例, 参考文献[44], 将其划分为$M=H\times W/P^2$个边长为$P$的图像块并拉伸成一维特征向量, 然后按照图像块的位次索引编码位置. 由于特征嵌入模块仅用于统一输入维度, 且不同模态的块划分和位置编码策略已经得到了广泛的研究, 因此其具体实现方式不作为本文研究的重点.

2.1.2 共享编码器模块

在统一视图嵌入维度的基础上, 进一步设计了一个基于Transformer结构的共享编码器模块来挖掘不同视图的高级语义信息. 该编码器通过参数共享的形式将各个视图嵌入以相同的映射规则投影到低维潜在空间, 有利于实现视图间的对齐, 同时避免维度灾难和可能存在的噪声干扰. 具体来说, 编码器由L个相同的网络层组堆叠而成, 每个层组包括一个多头自注意力层和一个带有GELU激活函数的全连接前馈层. 在每层之前执行归一化操作, 每层之间由残差结构连接. 给定第v个视图的嵌入$\boldsymbol{X}_{\mathrm{E}}^{(v)}$, 编码器中第$l$层组的计算过程为

$$\boldsymbol{X}_0^{(v)}=\boldsymbol{X}_{\mathrm{E}}^{(v)}, \tag{3}$$

$$\bar{\boldsymbol{X}}_l^{(v)}=MSA\left(LN\left(\boldsymbol{X}_{l-1}^{(v)}\right)\right)+\boldsymbol{X}_{l-1}^{(v)}, \tag{4}$$

$$\boldsymbol{X}_l^{(v)}=FFN\left(LN\left(\bar{\boldsymbol{X}}_l^{(v)}\right)\right)+\bar{\boldsymbol{X}}_l^{(v)}, \tag{5}$$

其中$\boldsymbol{X}_l^{(v)}(l=1,2,\cdots,L)$是编码模块中第$l$层的输出, $MSA(\cdot)$, $LN(\cdot)$, $FFN(\cdot)$分别表示多头自注意力层、归一化层和全连接前馈层. 经过嵌入模块和共享编码器模块, 原始高维空间中的数据$\boldsymbol{X}^{(v)}$被逐步映射到低维特征空间, 并最终得到潜在表示$\boldsymbol{F}^{(v)}=\boldsymbol{X}_L^{(v)}\in\mathbb{R}^{d\times N}$, 其中$d=M\times k$为潜在表示的维度.

2.1.3 样本加权融合模块

多视图特征表示的有效融合能够强化特征的表达能力, 实现信息的共享和互补, 提升聚类性能. 为此, MURLN提出一种样本加权的多视图特征融合方法, 将每个视图的潜在表示$\boldsymbol{F}^{(v)}$融合为统一特征表示$\boldsymbol{F}\in\mathbb{R}^{d\times N}$. 考虑到观测样本在不同视图中的质量差异, 简单地拼接特征容易忽略视图间以及视图内的复杂关系和不平衡性. 另一方面, 在没有先验知识的情况下, 人工设计权重具有盲目性. 本文将视图内样本间的权重关系融入目标函数, 通过网络迭代优化动态地进行更新. 加权融合的过程形式化为

$$\boldsymbol{F}=\sum_{v=1}^{V}\sum_{i=1}^{N}\omega_i^{(v)}\boldsymbol{f}_i^{(v)}, \tag{6}$$

其中$\boldsymbol{\omega}^{(v)}=(\omega_1^{(v)},\omega_2^{(v)},\cdots,\omega_N^{(v)})\in\mathbb{R}^N$表示第$v$个视图对应的权重向量.

2.1.4 自表示学习模块

自表示学习模块包含一个不带偏置和非线性激活函数的全连接层, 它能够自然地模拟数据的自表达特性, 从而描述成对样本之间的关系, 称为自表达层. 遵循多视图数据的一致性, 所有视图应具有相同的子空间结构. 因此, 将统一表示$\boldsymbol{F}$和特定视图的潜在表示$\boldsymbol{F}^{(v)}$共同作为自表达层的输入, 学习所有视图共享的一致子空间表示$\boldsymbol{S}\in\mathbb{R}^{N\times N}$. $\boldsymbol{S}=(s_{i,j})$, 其中$s_{i,j}$反映了样本i和样本j的相似性关系.

现有子空间聚类方法表明[33], 低秩表示可以保持数据的结构特征, 并且减少噪声等冗余信息的干扰, 有利于学习更理想的系数矩阵. 为此, 本文在自表示学习模块中定义一个基于加权Schatten-p范数的低秩约束来强化子空间表示的低秩性. 与传统的基于核范数的低秩约束相比, 加权Schatten-p范数不仅在形式上更加灵活, 同时还考虑了不同秩分量之间的显著差异(较大奇异值通常表示数据矩阵中嵌入的重要结构信息), 已在图像去噪、矩阵补全等任务中验证了有效性[45-46]. 定义为

$$\|\boldsymbol{S}\|_{\boldsymbol{w},S_p}=\left(\sum_{i=1}^{N}w_i\sigma_i(\boldsymbol{S})^p\right)^{1/p},\quad \text{s.t. } w_i=1/(\sigma_i(\boldsymbol{S})+\varepsilon), \tag{7}$$

其中$0<p\leqslant 1$. $\boldsymbol{w}=(w_1,w_2,\cdots,w_i,\cdots,w_N)$是非负权重向量, 通过不同的权重因子实现对非零奇异值不同程度的收缩. ε为一个极小值用于避免分母为0, 通常$\varepsilon=10^{-16}$. 为避免开根过程, 式(7)改写为:

$$\|\boldsymbol{S}\|_{\boldsymbol{w},S_p}^p=\sum_{i=1}^{N}w_i\sigma_i(\boldsymbol{S})^p=\mathrm{tr}(\boldsymbol{W}\boldsymbol{\Delta}^p), \tag{8}$$

其中$\boldsymbol{W}$和$\boldsymbol{\Delta}$为对角矩阵, 主对角线上的元素分别由w_i和σ_i组成. 当$p=1$时, 式(8)退化为加权核范数; 当$p\to 0$时, Schatten-0范数等价于求解系数矩阵的秩最小化问题. 因此, Schatten-p范数在形式上比核范数更接近于目标矩阵的秩函数, 对低秩问题的求解也更为灵活.

2.1.5 视图重建模块

对于无监督的多视图聚类任务, 验证编码器能否学到紧致的特征表示主要通过解码器重构输入数据$\hat{\boldsymbol{X}}^{(v)}\in\mathbb{R}^{d_v\times N}$, 并最小化重建误差来实现. MURLN使用单层全连接层作为解码器的具体实现.

2.2 模型优化

网络模型的训练过程可以形式化为对应目标函

数的优化问题. 首先, 通过最小化输入数据与重构数据之间的重建损失, 端到端地训练编码器-解码器. 重建损失$\mathcal{L}_{\text{re}}$定义为

$$\mathcal{L}_{\text{re}}=\sum_{v=1}^{V}\left\|\boldsymbol{X}^{(v)}-\hat{\boldsymbol{X}}^{(v)}\right\|_{\text{F}}^{2}. \tag{9}$$

将特定于视图的潜在表示输入样本加权融合模块, 构造多视图统一表示, 提高特征表达能力. 另外, 为了探索视图间的互补性信息, MURLN 利用哈达玛积设计正则化约束, 增强多视图表示之间的多样性. 融合损失定义为

$$\mathcal{L}_{\text{fusion}}=\sum_{v=1}^{V}\left\|\boldsymbol{F}-\boldsymbol{F}^{(v)}\right\|_{\text{F}}^{2}+\sum_{v\neq u}^{V}\left\|\boldsymbol{F}^{(v)}\odot\boldsymbol{F}^{(u)}\right\|_{1}, \tag{10}$$

其中⊙表示哈达玛积, 隐式地鼓励多视图统一表示编码互补信息.

自表示系数矩阵的学习受到统一表示和特定视图潜在表示 2 个方面的共同约束, 定义为

$$\mathcal{L}_{\text{self}}=\frac{1}{V}\sum_{v=1}^{V}\left\|\boldsymbol{F}^{(v)}-\boldsymbol{F}^{(v)}\boldsymbol{S}\right\|_{\text{F}}^{2}+\left\|\boldsymbol{F}-\boldsymbol{F}\boldsymbol{S}\right\|_{\text{F}}^{2}. \tag{11}$$

根据对各部分损失函数的分析, 模型的整体目标函数形式为:

$$\min\mathcal{L}=\min\mathcal{L}_{\text{re}}+\lambda_1\mathcal{L}_{\text{fusion}}+\lambda_2\mathcal{L}_{\text{self}}+\lambda_3\mathcal{L}_{\text{S}_p}, \tag{12}$$

其中$\lambda_1>0,\lambda_2>0,\lambda_3>0$为权衡参数. $\mathcal{L}_{\text{S}_p}=\|\boldsymbol{S}\|_{\boldsymbol{w},\text{S}_p}^{p}$是作用在子空间表示上的低秩约束正则化项.

2.3 训练细节

为了加速模型收敛, 本文使用“预训练+微调”的 2 阶段优化策略训练网络.

第 1 阶段预训练编码器和解码器. 首先从均值为零的高斯分布中随机采样初始化编码器和解码器参数, 并随机采样b个样本$\boldsymbol{x}_i^{(v)}(i=1,2,\cdots,b)$作为一个训练批次. 然后利用随机梯度下降算法最小化式(9)中的重建损失, 迭代更新编码器和解码器的参数.

基于上述初始值, 第 2 阶段对完整网络进行微调, 其中未预训练的参数被随机初始化. 由于自表达层学习所有样本的关联关系, 因此将全部数据作为微调阶段的输入. 利用梯度下降算法对式(12)迭代求解, 直至收敛. 一旦网络收敛, 学习到的自表示系数矩阵被用来计算亲和矩阵, 最后通过谱聚类后处理获得聚类结果. 具体的参数设置和实验结果在第 3 节中详细讨论. 算法 1 提供了本文模型的训练流程.

算法 1. MURLN 算法.

输入: 多视图数据$\boldsymbol{X}^{(v)}$, 样本数量N, 批大小b, 特征维度d, 次幂参数p, 权衡参数$\lambda_1,\lambda_2,\lambda_3$;

输出: 聚类结果.

初始化模型参数、超参数及系数矩阵$\boldsymbol{S}$;

/* 步骤 1. 预训练自编码器*/

① while 未收敛 do

② 采样一批数据: $\boldsymbol{x}_i^{(v)}$为$\boldsymbol{X}^{(v)}$的列向量, $i=1,2,\cdots,b$;

③ 最小化式(9), 更新特征嵌入模块、共享编码器模块和视图重建模块参数;

④ end while

/* 步骤 2. 微调完整网络*/

⑤ while 未收敛 do

⑥ 在数据集$\boldsymbol{X}^{(v)}$上最小化式(12), 更新所有可训练参数及系数矩阵$\boldsymbol{S}$;

⑦ end while

⑧ 计算亲和矩阵, 执行谱聚类;

⑨ 返回聚类结果.

3 实验结果与分析

3.1 数据集和实验设置

1)实验环境. 提出的方法在 Windows11 操作系统利用 PyTorch 深度学习库由 Python 编程语言实现. 实验环境的主要硬件参数为 NVIDIA GeForce RTX 3080 Ti GPU, 12 GB 显存, Intel® Core™ i9-10900F CPU, 32 GB 内存.

2)数据集. 实验在 7 个公开的多视图数据集上进行, 涉及新闻文档(BBCSport[33])、手写数字(MNIST-USPS[20])、人脸目标(Yale[30], Extended YaleB[19])和通用对象(RGB-D[17], Caltech101-20[31], MSRCV1[10])等多种场景. 表 2 列出了数据集的详细信息.

Table 2 Details of Datasets

表 2 数据集详细信息

场景类型	数据集	类别数	样本数	视图数	视图维度
新闻文档	BBCSport	5	544	2	[3 183, 3 203]
手写数字	MNIST-USPS	10	5 000	2	[28×28, 28×28]
人脸目标	Yale	15	165	3	[4 096, 3 304, 6 750]
	Extended YaleB	10	640	3	[2 500, 3 304, 6 750]
通用对象	MSRCV1	7	210	6	[1 302, 48, 512, 100, 256, 210]
	RGB-D	50	500	2	[64×64, 64×64]
	Caltech101-20	20	2 386	6	[48, 40, 254, 1 984, 512, 928]

3)评价指标. 本文使用 6 种通用的评价指标[41]综合评估方法的聚类性能, 包括准确性(accuracy, *ACC*)、标准化互信息(normalized mutual information,

NMI)、调整兰德指数(adjusted Rand index, *AR*)、*F* 分数(*F*-measure, *FM*)、精度(precision, *P*)和召回率(recall, *R*). 所有指标的值越高, 聚类质量越好.

4)对比方法. MURLN 将与 12 种先进的多视图聚类方法进行比较, 包括基于多样性约束的多视图子空间聚类(DiMSC[30])、基于低秩表示的多视图聚类(LMSC[10], RMSL[28], DSS-MSC[29], FCMSC[43], FLMSC[11], JSMC[47])、深度多视图子空间聚类(MSCNLG[31], DMSC-UDL[12], SDMSC[19], MSESC[48], D^{2}MVSC[32]). 所有对比方法遵循原文建议调整参数. 在处理多特征数据集时, DMSC-UDL 使用的卷积自编码器被替换为文献 [32] 中的全连接自编码器. 为了减少随机性的影响, 所有方法独立运行 20 次, 报告各指标在每个数据集中的平均值.

5)实验参数. 采用 AdamW 优化器对网络进行优化, 初始学习率设置为 5E−4, 动量因子和权值衰减分别设为 0.9 和 0.05. 优化器根据任务损失产生的梯度来更新模型参数. 在预训练阶段, 设置训练迭代次数为 500, 批大小为 256(数据量不足时, 批大小为样本数量). 在微调阶段, 设置训练迭代次数为 200, 批大小为输入数据的样本量, 以学习完整的表示. 用四元组(分块数 M, 潜在表示的特征维度 d, 网络层数 L, 注意力头个数)表示编码器的超参数. 对于 Yale 数据集, 编码器结构设置为(4, 256, 4, 4). 对于 MSRCV1 数据集, 编码器结构设置为(4, 512, 4, 8). 在 MNIST-USPS, BBCSport, Caltech101-20, Extended YaleB 数据集中, 编码器结构设置为(8, 512, 8, 16). 通过参数敏感性分析, 本文将模型在不同数据集上的超参数统一设置为 $\lambda_1 = 0.01$, $\lambda_2 = 0.5$, $\lambda_3 = 1$.

3.2 性能比较与分析

3.2.1 聚类性能比较

实验结果如表 3~5 所示. 从实验结果中可发现, 所提出的方法 MURLN 在所有数据集上都取得了有竞争力的聚类性能. 对结果的进一步分析表明, MURLN 的聚类有效性来自于 2 个方面.

一方面, MURLN 能够利用多视图中的互补性信息表示数据复杂的非线性相关性, 这对提升聚类结果至关重要. 观察表 3 在人脸数据集 Yale 上的聚类结果, 获得前 3 名的方法分别是: MURLN, D^{2}MVSC, MSCNLG, 反映了使用深度神经网络非线性地提取特征可以实现更好的聚类效果. Extended YaleB 数据集由于包含不同光线强度下的人脸图像, 样本关系与 Yale 相比更加复杂. 在 *ACC* 和 *NMI* 指标上, MURLN 比传统浅层方法中最优的 DSS-MSC 分别提高了 18.75 个百分点和 16.36 个百分点. 结合其他数据集中的实验结果, MURLN 无论是处理 2 维图像数据还是 1 维特征数据, 性能都较对比方法有明显提升, 验证了多视图统一表示学习网络在特征表示方面的出色能力.

另一方面, 模型使用 Schatten-p 范数正则化约束学习到了更低秩的自表示系数矩阵以及更清晰的簇结构. 在 Caltech101-20 数据集的 *ACC* 和 *NMI* 指标上, MURLN 比次优方法 D^{2}MVSC 分别提高了 7.46 个百

Table 3 Clustering Results in Yale and Extended YaleB Datasets (mean)

表 3 Yale 和 Extended YaleB 数据集上的聚类结果(均值) %

方法	Yale						Extended YaleB					
	ACC	*NMI*	*AR*	*FM*	*P*	*R*	*ACC*	*NMI*	*AR*	*FM*	*P*	*R*
DiMSC	70.90	72.70	53.50	56.40	54.30	58.60	61.50	63.50	45.30	50.40	48.10	53.40
LMSC	75.20	73.50	55.10	56.40	54.30	57.10	52.42	52.14	26.82	35.10	30.60	41.16
RMSL	74.70	75.50	55.64	58.55	54.51	63.30	53.92	52.75	28.93	36.81	32.72	42.08
DSS-MSC	78.20	77.90	60.10	61.30	59.20	62.20	78.65	76.86	64.81	68.47	65.48	71.85
FCMSC	76.36	78.90	63.67	65.97	63.52	68.61	53.20	53.24	27.51	35.72	31.09	41.98
FLMSC	72.72	74.02	55.67	58.47	56.58	60.49	57.85	55.86	39.49	47.90	46.35	49.64
JSMC	70.91	74.28	56.24	59.01	56.85	61.33	45.85	45.57	24.24	32.51	29.38	36.38
MSCNLG	91.67	90.30	82.10	83.21	81.76	84.70	51.23	48.88	25.45	33.75	29.91	38.72
DMSC-UDL	77.09	75.86	56.01	58.92	54.36	64.32	92.55	88.69	82.12	83.94	81.51	86.54
SDMSC	84.73	84.85	66.35	69.15	86.62	82.73	98.17	95.90	94.03	95.82	94.10	94.88
MSESC	81.21	82.19	71.29	73.22	71.71	74.53	94.15	88.39	87.09	88.37	87.93	88.81
D^{2}MVSC	94.50	**93.90**	88.90	90.40	90.50	**90.40**	**99.32**	**98.71**	**98.24**	**99.01**	**98.94**	**98.93**
MURLN(本文方法)	**94.95**	93.88	**89.37**	**90.59**	**90.68**	89.72	97.40	93.22	92.36	93.09	94.31	95.27

注: 最好的结果用黑体标注, 次优结果用下划线标注.

Table 4 Clustering Results in Caltech101-20 and MSRCV1 Datasets (mean)

表 4 Caltech101-20 和 MSRCV1 数据集上的聚类结果（均值） %

方法	Caltech101-20						MSRCV1					
	ACC	*NMI*	*AR*	*FM*	*P*	*R*	*ACC*	*NMI*	*AR*	*FM*	*P*	*R*
DiMSC	39.12	51.96	26.29	32.23	65.03	21.43	80.36	68.61	62.07	67.41	66.35	68.51
LMSC	55.78	65.12	42.95	48.88	76.61	35.89	80.60	65.30	59.90	65.20	61.20	66.30
RMSL	40.49	47.17	30.82	37.46	63.35	26.61	71.74	63.95	56.10	62.39	60.41	64.52
DSS-MSC	44.11	63.20	31.71	37.43	72.67	25.21	86.57	78.64	72.94	76.74	75.83	77.68
FCMSC	45.98	63.15	34.44	40.44	71.95	28.13	80.90	70.10	63.58	68.70	67.76	69.73
FLMSC	48.16	64.15	35.39	40.98	77.09	28.91	97.57	**96.12**	95.70	**96.16**	96.11	96.21
JSMC	65.26	68.11	53.90	59.21	83.62	45.83	62.38	52.98	42.09	50.32	49.14	51.56
MSCNLG	50.23	58.24	36.65	46.70	47.98	45.51	93.08	80.69	83.16	87.09	89.00	85.26
DMSC-UDL	56.76	64.22	48.20	55.86	62.55	51.12	78.36	66.97	59.85	65.46	65.46	65.92
SDMSC	62.00	59.60	48.56	54.04	61.90	62.00	97.06	90.37	92.06	93.99	**97.76**	96.59
MSESC	60.26	69.18	51.88	63.07	76.38	46.74	91.43	86.07	81.25	83.89	82.64	85.19
D^2MVSC	72.73	73.85	56.35	59.15	76.62	72.73	97.78	93.00	95.74	95.36	96.45	**97.03**
本文方法	**80.19**	**78.37**	**85.21**	**87.09**	**86.54**	**89.30**	**98.34**	95.99	**96.41**	96.07	96.82	96.87

注：最好的结果用黑体标注，次优结果用下划线标注.

Table 5 Clustering Results in BBCSport and MNIST-USPS Datasets (mean)

表 5 BBCSport 和 MNIST-USPS 数据集上的聚类结果（均值） %

方法	BBCSport						MNIST-USPS					
	ACC	*NMI*	*AR*	*FM*	*P*	*R*	*ACC*	*NMI*	*AR*	*FM*	*P*	*R*
DiMSC	96.00	88.50	92.00	92.90	91.40	93.00	63.72	59.94	50.08	55.14	54.32	55.99
LMSC	91.20	82.60	84.20	88.70	87.30	87.70	73.33	75.48	65.35	68.99	65.57	72.81
RMSL	97.61	91.81	92.90	95.34	95.23	94.47	44.64	33.19	32.88	24.71	30.25	36.02
DSS-MSC	96.60	88.40	89.80	92.30	92.70	91.80	82.93	87.54	80.99	83.00	78.40	88.17
FCMSC	96.51	89.04	91.06	93.17	94.21	92.15	76.42	77.88	72.54	68.97	69.32	76.51
FLMSC	92.72	85.56	84.57	89.65	89.05	88.53	93.27	86.29	86.93	88.23	88.15	88.32
JSMC	86.58	79.40	81.37	85.94	83.50	88.52	70.21	64.50	58.75	63.89	59.07	65.19
MSCNLG	88.79	72.50	73.39	79.91	77.73	82.23	67.60	77.95	62.66	67.04	57.12	81.12
DMSC-UDL	95.96	87.94	89.00	91.64	91.30	91.97	78.36	66.97	59.85	65.46	65.46	65.92
SDMSC	94.86	86.82	89.36	89.93	86.00	87.92	81.42	88.83	86.99	83.06	78.10	80.04
MSESC	94.85	86.89	85.70	89.05	91.39	86.78	83.76	89.45	82.49	84.34	79.37	89.98
D^2MVSC	98.10	93.80	94.90	96.40	96.40	96.41	93.86	90.72	90.60	90.36	89.01	89.32
本文方法	**98.70**	**95.07**	**94.96**	**96.41**	**96.93**	**96.82**	**95.63**	**93.11**	**92.87**	**93.49**	**92.55**	**93.94**

注：最好的结果用黑体标注，次优结果用下划线标注.

分点和 4.52 个百分点. 这是因为加权 Schatten-*p* 范数赋予不同奇异值不同的权重, 重视较大奇异值的作用, 保证了子空间表示的低秩性. 与 D^2MVSC 方法使用的核范数相比, 加权 Schatten-*p* 范数在求解低秩问题上具有更好的灵活性与有效性.

3.2.2 共享编码器效果分析

为深入探究基于 Transformer 的共享编码器在复杂多视图场景中的优势, 本文合成一个具有 2 种模态 4 个视图的人造数据集(MNIST-USPS-4V). 具体地, 在图像数据集 MNIST-USPS 上增加 2 个分别由 MNIST 和 USPS 数据集提取的 LBP 特征. 实验设计 3 种使用不同编码器的 MURLN 模型用于比较, 包括基于全连接编码器的 MURLN_FC、卷积自编码器的 MURLN_CNN 和本文所提方法 MURLN. 表 6 提供了这 3 种模型分别在 MNIST-USPS 和 MNIST-USPS-4V 数据集上的 *ACC* 和 *NMI* 结果.

Table 6 Comparison of Clustering Performance for Three Coding Networks (mean)

表 6 3 种编码网络的聚类性能比较（均值） %

数据集	视图（维数）	*ACC*			*NMI*		
		MURLN_CNN	MURLN_FC	MURLN（本文方法）	MURLN_CNN	MURLN_FC	MURLN（本文方法）
MNIST-USPS	视图-1 (28×28)	76.71	72.25	86.74	72.89	74.40	82.64
	视图-2 (28×28)	61.20	75.33	71.16	59.10	69.93	74.15
	MNIST-USPS	88.94	85.40	**95.63**	86.14	84.82	**93.11**
MNIST-USPS-4V	视图-1 (28×28)		41.31	85.67		19.56	78.40
	视图-2 (28×28)		50.90	72.78		42.23	67.44
	视图-3 (944)		52.56	62.81		44.08	59.66
	视图-4 (944)		35.04	67.34		19.75	62.19
	MNIST-USPS-4V		79.28	**93.27**		77.38	**91.78**

注：最好的结果用黑体标注.

通过观察发现：1)MURLN 具备处理不同模态数据的能力，在 2 种数据场景中均获得了最好的聚类效果，反映出通过 Transformer 进行的多模态特征编码对性能的提高至关重要. 例如，在 MNIST-USPS-4V 上比在 MURLN_FC 的 *ACC* 和 *NMI* 指标分别提高了 13.99 个百分点和 14.4 个百分点. 2)MURLN_CNN 通过卷积操作能够保留图像数据的空间信息，但其感受野有限，很难提取全局语义信息，不适用于处理 1 维特征. 3)MURLN_FC 可以通过将图像拉伸成 1 维向量来编码图像数据，但向量化会导致图像空间信息的丢失，进而降低聚类性能.

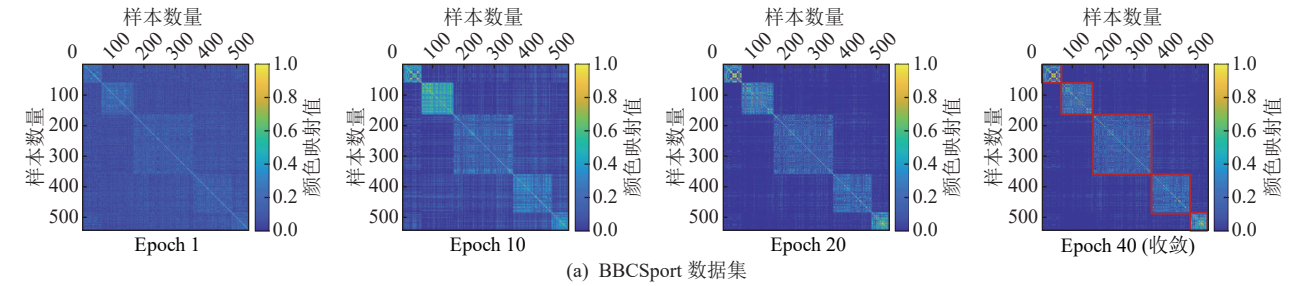
进一步，使用峰值信噪比(peak signal-to-noise ratio, PSNR)衡量 MURLN 与卷积自编码器对图像数据的表示能力. PSNR 值越高，表明重建结果越好. 选取 RGB-D 和 MNIST-USPS 数据集上的部分图像进行实验可知，MURLN_CNN 的平均 PSNR 值为 35.25 dB，MURLN 则达到了 54.36 dB. 显然，MURLN 能够更好地学习样本的有效抽象特征.

3.3 可视化分析

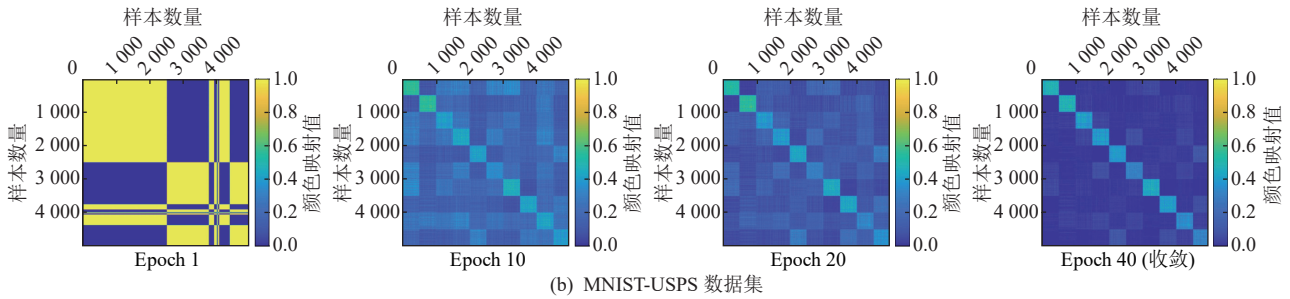
图 3 和图 4 给出了 MURLN 在新闻文档 BBCSport 数据集以及手写数字 MNIST-USPS 数据集上学习得到的亲和矩阵和 t-SNE 图. 显然，学习到的亲和矩阵具有理想的块对角结构和较少的噪声，能清楚地揭示样本之间的关系. 随着迭代次数的增加，聚类结构变得更加清晰，进一步验证了 MURLN 的有效性.

3.4 参数敏感性分析

MURLN 引入 4 个超参数 $\lambda_1,\lambda_2,\lambda_3,p$，并在各数据集上以网格搜索策略来确定最优参数组合.



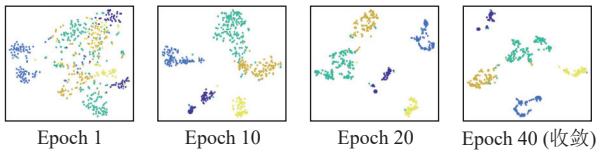
(a) BBCSport 数据集



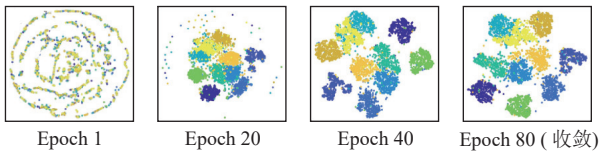
(b) MNIST-USPS 数据集

Fig. 3 Visualization of the affinity matrix

图 3 亲和矩阵可视化



(a) BBCSport 数据集



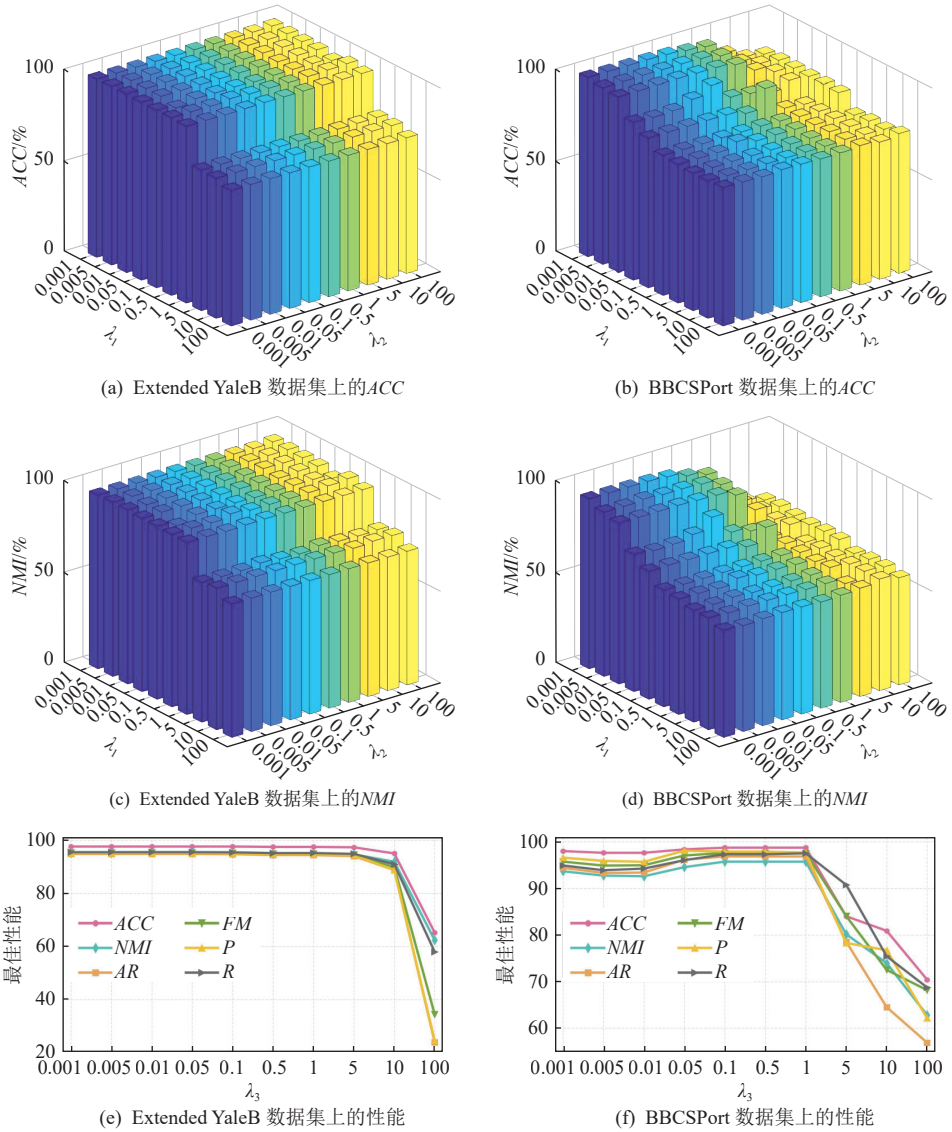
(b) MNIST-USPS 数据集

Fig. 4 Visualization of t-SNE clustering

图 4 t-SNE 聚类可视化

如图 5 所示，对于权衡超参数$\lambda_1,\lambda_2,\lambda_3$，结合控制变量策略，在 Extend YaleB, BBCSport 数据集中，使用不同超参数组合得到聚类结果. 以 BBCSport 为例，固定$\lambda_3=0.001$，$\lambda_1\in\{0.001, 0.005, 0.01\}$，$\lambda_2\in\{0.005, 0.01, 0.05, 0.1, 0.5\}$，可以实现较好的聚类效果. 综合其他数据集上的结果，设置$\lambda_1=0.01$，$\lambda_2=0.5$. 固定λ_1和λ_2，调整λ_3，如图 5 (e)(f)所示. 当$\lambda_3\in\{0.1, 0.5, 1.0\}$时，聚类性能相对稳定. 经过实验分析，本文设置$\lambda_3=1.0$.

图 6 提供了不同参数p在 4 个数据集中的 *ACC* 和 *NMI* 结果. p以 0.1 的步长在{0.1, 0.2, …, 1.0}范围内取值. 通过观察发现，参数p的选择受数据特性的影响. 例如，对于 Caltech101-20 和 BBCSport 数据集，p设置为 0.7 和 0.8 可以实现最好的聚类性能. 对于 Extended YaleB 和 MSRCV1 数据集，当$p=1$时可以获得最优结果. 显然，加权核范数作为秩函数的凸松弛并不总是近似求解秩最小化问题的最优选择，说明



(a) Extended YaleB 数据集上的*ACC*

(b) BBCSPort 数据集上的*ACC*

(c) Extended YaleB 数据集上的*NMI*

(d) BBCSPort 数据集上的*NMI*

(e) Extended YaleB 数据集上的性能

(f) BBCSPort 数据集上的性能

Fig. 5 Sensitivity analysis of parameter

图 5 参数敏感性分析

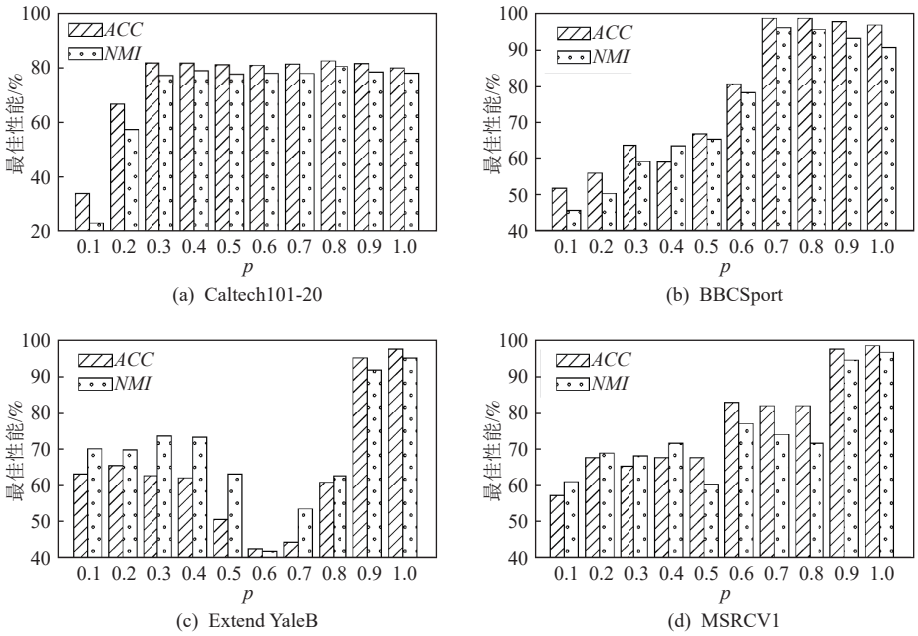


Fig. 6 Effect of parameter p on clustering performance in different datasets

图 6 不同数据集中参数 p 对聚类性能的影响

本文利用 Schatten-p 范数设计低秩约束的合理性.

3.5 消融研究

为了进一步分析不同模块对 MURLN 的贡献，全面说明 MURLN 设计的合理性和有效性，本节设计了 5 组消融实验方法. 在 Yale 和 MSRCV1 数据集上的实验结果如表 7 所示. 方法 *A* 和方法 *E* 分别代表了仅使用重建损失(即 $\lambda_1=\lambda_2=\lambda_3=0$)的基线模型和使用了全部损失的 MURLN 完整模型. 方法 *B*、方法 *C* 和方法 *D* 则是在基线模型的基础上，依次增加了融合损失、自表示损失和低秩损失.

实验结果验证了 5 个结论：1)本文使用的 3 类约束均提升了模型在聚类任务中的表现；2)融合损失能够约束样本级的融合权重，提高视图内高质量样本的影响，同时保证了视图间的多样性；3)自表示损失促进系数矩阵学习到数据样本对之间的相关性，带来了显著的性能提升；4)低秩损失强化了子空间表示的低秩性，使其充分保留多视图数据的结构特征. 5)整合所有模块的 MURLN 在 2 个数据集上均取得了最优性能，每个模块都各自发挥作用，去掉任何一个模块都会影响聚类效果.

3.6 收敛性分析

图 7 展示了 MURLN 在 2 个不同类型数据集上的收敛曲线. 可以看出，随着迭代次数增加，目标函数值逐渐下降并趋于稳定. 尽管在开始的几次迭代中发生振荡，但最终大约在 80 次迭代时收敛. 类似地，聚类性能(*ACC* 和 *NMI*)在前几次迭代过程迅速升高并逐渐收敛. 表明 MURLN 具有良好的收敛性，能够实现快速收敛. 在其他数据集上也得到了类似的收敛分析.

4 结 论

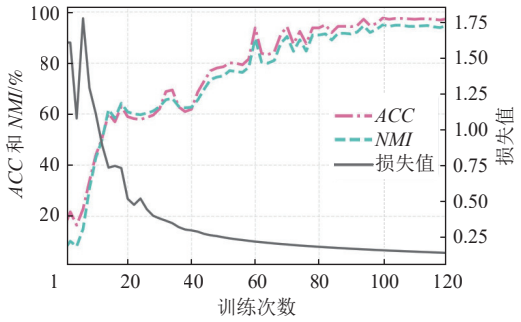
本文提出了一种新的多视图子空间聚类方法

Table 7 Ablation Study of Loss Functions on MSRCV1 and Yale Datasets

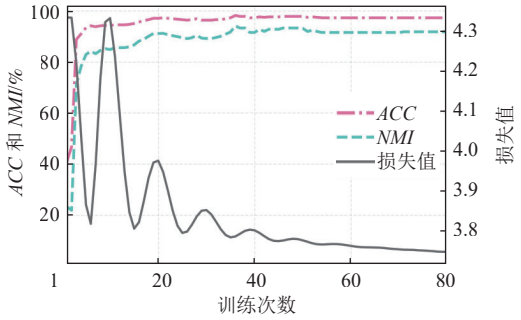
表 7 MSRCV1 和 Yale 数据集上损失函数的消融研究

方法	损失项				MSRCV1				Yale			
	$\mathcal{L}_{\text{re}}$	$\mathcal{L}_{\text{fusion}}$	$\mathcal{L}_{\text{self}}$	$\mathcal{L}_{S_p}$	*ACC*	*NMI*	*AR*	*FM*	*ACC*	*NMI*	*AR*	*FM*
A	√				87.13	78.59	76.43	75.86	86.78	83.61	80.29	80.05
B	√	√			92.86	91.37	87.21	88.03	93.16	91.21	85.79	84.73
C	√		√		95.51	89.97	91.76	90.44	92.37	92.01	86.19	88.32
D	√			√	94.77	90.40	88.84	89.16	91.53	89.86	84.15	88.35
E	√	√	√	√	**98.34**	**95.99**	**96.41**	**96.07**	**94.95**	**93.88**	**89.37**	**90.59**

注：最好的结果用黑体标注，“√”表示 MURLN 模型包含该模块.



(a) Extended YaleB 数据集



(b) BBCSPort 数据集

Fig. 7 Convergence experiments

图 7 收敛性实验

(MURLN). 与现有方法为不同视图使用单独的编码器不同, MURLN 基于 Transformer 结构设计一个多视图统一表示学习网络, 所有视图各自经过特定的嵌入层来统一输入形式和维度, 然后共享同一个编码模型. 此外, 提出基于加权 Schatten-p 范数的低秩正则化约束, 提高自表示系数矩阵学习的质量. 在多个公开数据集上的大量实验证明了所提方法的有效性和优越性.

利用全连接层实现数据自表达特性的方法普遍存在 full-batch training 的局限. 具体体现在随着样本数量的增多, 全连接层的参数量迅速提升, 需要较多的计算和存储资源. 后续工作将关注于模型参数与样本数量的解耦研究, 提高大规模多视图数据集聚类的效率和精度. 此外, 由于数据收集和传输的复杂性, 现实场景中可能存在部分视图缺失现象[49], 这也是未来的研究重点.

作者贡献声明: 林毓秀负责网络模型的设计和实现, 论文的撰写和修改; 刘慧提出研究方向, 把握论文的创新性, 并指导论文修改; 于晓负责实验的整理和可视化分析; 张彩明提出指导意见并修改论文.

参 考 文 献

[1] Liu Xinwang. Research on the fundamental theory and methods of multi-view learning[J]. China Basic Science, 2022, 24(3): 27−34 (in Chinese)
(刘新旺. 多视图学习的基础理论和方法研究[J]. 中国基础科学, 2022, 24(3): 27−34)

[2] Liang Jiye, Liu Xiaolin. Research progress and prospects of multi-view clustering[J]. Journal of Shanxi University: Natural Science Edition, 2022, 45(3): 612−621 (in Chinese)
(梁吉业, 刘晓琳. 多视图聚类研究进展与展望[J]. 山西大学学报: 自然科学版, 2022, 45(3): 612−621)

[3] Li Jinxing, Zhou Chuhao, Ji Xiaoqiang, et al. Multi-view instance attention fusion network for classification[J]. Information Fusion, 2024, 101: 101974

[4] Zhang Dongxu, Yan Yang , Huang Yulin, et al. Unsupervised Cryo-EM images denoising and clustering based on deep convolutional autoencoder and K-Means++[J]. IEEE Transactions on Medical Imaging, 2023, 42(5): 1509−1521

[5] Gong Weihua, Jin Rong, Pei Xiaobing, et al. Collaborative recommendation method based on community co-clustering in location based social networks[J]. Journal of Computer Research and Development, 2019, 56(11): 2506−2517 (in Chinese)
(龚卫华, 金蓉, 裴小兵, 等. LBSN 中基于社区联合聚类的协同推荐方法[J]. 计算机研究与发展, 2019, 56(11): 2506−2517)

[6] Cao Congqi, Zhang Xin, Zhang Shizhou, et al. Weakly supervised video anomaly detection based on cross-batch clustering guidance[C]//Proc of the 14th Int Conf on Multimedia and Expo. Piscataway, NJ: IEEE, 2023: 2723−2728

[7] Kumar A, Rai P, Daumé H. Co-regularized multi-view spectral clustering[C]//Proc of the 25th Annual Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2011: 1413−1421

[8] Xia Dongxue, Yang Yan, Wang Hao, et al. Late fusion multi-view clustering based on local multi-kernel learning[J]. Journal of Computer Research and Development, 2020, 57(8): 1627−1638 (in Chinese)
(夏冬雪, 杨燕, 王浩, 等. 基于邻域多核学习的后融合多视图聚类算法[J]. 计算机研究与发展, 2020, 57(8): 1627−1638)

[9] Tang Chang, Liu Xinwang, Zhu Xinzhong, et al. CGD: Multi-view clustering via cross-view graph diffusion[C]//Proc of the 34th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 5924−5931

[10] Zhang Changqing, Hu Qinghua, Fu Huazhu, et al. Latent multi-view subspace clustering[C]//Proc of the Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 4333−4341

[11] Zhang Guangyu, Huang Dong, Wang Changdong. Facilitated low-rank multi-view subspace clustering[J]. Knowledge-Based Systems, 2022, 260: 110141

[12] Wang Qianqian, Cheng Jiafeng, Gao Quanxue, et al. Deep multi-view subspace clustering with unified and discriminative learning[J]. IEEE Transactions on Multimedia, 2021, 23: 3483−3493

[13] Liu Zhaohu, Song Peng. Deep low-rank tensor embedding for multi-view subspace clustering[J]. Expert Systems with Applications, 2023, 237: 121518

[14] Wang Jing, Feng Songhe, Lyu Gengyu, et al. Triple-granularity contrastive learning for deep multi-view subspace clustering[C]//Proc of the 31st ACM Int Conf on Multimedia. New York: ACM, 2023: 2994−3002

[15] Rumelhart D, Hinton G, Williams R. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533−536

[16] Abavisani M, Patel V. Deep multimodal subspace clustering networks[J]. IEEE Journal of Selected Topics in Signal Processing, 2018, 12(6): 1601−1614

[17] Zhu Pengfei, Hui Binyuan, Zhang Changqing, et al. Multi-view deep subspace clustering networks[J]. arXiv preprint, arXiv: 1908.01978, 2019

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

[19] Li Kai, Liu Hongfu, Zhang Yulun, et al. Self-guided deep multiview subspace clustering via consensus affinity regularization[J]. IEEE Transactions on Cybernetics, 2022, 52(12): 12734−12744

[20] Cui Chenhang, Ren Yazhou, Pu Jingyu, et al. Deep multi-view subspace clustering with anchor graph[C]//Proc of the 32nd Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2023: 3577−3585

[21] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Proc of the 31st Annual Conf on Neural Information Processing Systems. Cambridge, MA: MIT Press, 2017: 5998−6008

[22] Xu Peng, Zhu Xiatian, Clifton D. Multimodal learning with Transformers: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(10): 12113−12132

[23] Han Xue, Wang Yitong, Feng Junlan, et al. A survey of transformer-based multimodal pre-trained modals[J]. Neurocomputing, 2023, 515: 89−106

[24] Xie Yuan, Gu Shuhang, Liu Yan, et al. Weighted schatten *p*-norm minimization for image denoising and background subtraction[J]. IEEE Transactions on Image Processing, 2016, 25(10): 4842−4857

[25] Elhamifar E, Vidal R. Sparse subspace clustering: Algorithm, theory, and applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2765−2781

[26] Liu Guangcan, Lin Zhouchen, Yan Shuicheng, 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

[27] Shi Jianbo, Malik J. Normalized cuts and image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888−905

[28] Li Ruihuang, Zhang Changqing, Fu Huazhu, et al. Reciprocal multi-layer subspace learning for multi-view clustering[C]//Proc of the IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 8171−8179

[29] Zhou Tao, Zhang Changqing, Peng Xi, et al. Dual shared-specific multiview subspace clustering[J]. IEEE Transactions on Cybernetics, 2020, 50(8): 3517−3530

[30] Cao Xiaochun, Zhang Changqing, Fu Huazhu, et al. Diversity-induced multi-view subspace clustering[C]//Proc of the Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 586−594

[31] Zheng Qinghai, Zhu Jihua, Ma Yuanyuan, et al. Multi-view subspace clustering networks with local and global graph information[J]. Neurocomputing, 2021, 449: 15−23

[32] Wang Jiao, Wu Bin, Ren Zhenwen, et al. Decomposed deep multi-view subspace clustering with self-labeling supervision[J]. Information Sciences, 2024, 653: 119798

[33] Zhang Changqing, Fu Huazhu, Hu Qinghua, et al. Generalized latent multi-view subspace clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(1): 86−99

[34] Wang Qianqian, Tao Zhiqiang, Gao Quanxue, et al. Multi-view subspace clustering via structured multi-pathway network[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(5): 7244−7250

[35] Masci J, Meier U, Ciresan D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]//Proc of the 21st Int Conf on Artificial Neural Networks. Berlin: Springer, 2011: 52−59

[36] Ren Shaoqing, He Kaiming, Girshick R, et al. Faster-RCNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39((6):): 1137−1149

[37] Devlin J, Chang Mingwei, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint, arXiv: 1810.04805, 2018

[38] Wei Xia, Wang Tianxiu, Gao Quanxue, et al. Graph embedding contrastive multi-modal representation learning for clustering[J]. IEEE Transactions on Image Processing, 2023, 32: 1170−1183

[39] Wang Qianqian, Lian Huanhuan, Sun Gan, et al. iCmSC: Incomplete cross-modal subspace clustering[J]. IEEE Transactions on Image Processing, 2020, 30: 305−317

[40] Lu Runkun, Liu Jianwei, Zuo Xin. Attentive multi-view deep subspace clustering net[J]. Neurocomputing, 2021, 435: 186−196

[41] Xie Yuan, Tao Dacheng, Zhang Wensheng, et al. On unifying multi-view self-representations for clustering by tensor multi-rank minimization[J]. International Journal of Computer Vision, 2018, 126(11): 1157−1179

[42] 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)

[43] Zheng Qinghai, Zhu Jihua, Li Zhongyu, et al. Feature concatenation multi-view subspace clustering[J]. Neurocomputing, 2019, 379: 89−102

[44] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint, arXiv: 2010.11929, 2020

[45] Xie Yuan, Qu Yanyun, Tao Dacheng, et al. Hyperspectral image

restoration via iteratively regularized weighted schatten *p*-norm minimization[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(8): 4642−4659

[46] Nie Feiping, Huang Heng, Ding C. Low-rank matrix recovery via efficient schatten *p*-norm minimization[C]//Proc of the 26th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2012: 655−661

[47] Cai Xiaosha, Huang Dong, Zhang Guangyu, et al. Seeking commonness and inconsistencies: A jointly smoothed approach to multi-view subspace clustering[J]. Information Fusion, 2023, 91: 364−375

[48] Cui Jinrong, Li Yuting, Fu Yulu, et al. Multi-view self-expressive subspace clustering network[C]//Proc of the 31st ACM Int Conf on Multimedia. New York: ACM, 2023: 417−425

[49] Zhao Boyu, Zhang Changqing, Chen Lei, et al. Generative model for partial multi-view clustering[J]. Acta Automatica Sinica, 2021, 47(8): 1867−1875 (in Chinese)
(赵博宇, 张长青, 陈蕾, 等. 生成式不完整多视图数据聚类[J]. 自动化学报, 2021, 47(8): 1867−1875)



Lin Yuxiu, born in 1996. PhD candidate. Her main research interests include data mining and visualization, and multi-view learning.

林毓秀, 1996 年生. 博士研究生. 主要研究方向为数据挖掘与可视化、多视图学习.



Liu Hui, born in 1978. PhD, professor, PhD supervisor. Member of CCF. Her main research interests include machine learning, data mining, and image processing.

刘　慧, 1978 年生. 博士, 教授, 博士生导师. CCF 会员. 主要研究方向为机器学习、数据挖掘、图像处理.



Yu Xiao, born in 1981. PhD, associate professor, master supervisor. Member of CCF. Her main research interests include statistical machine learning and data mining.

于　晓, 1981 年生. 博士, 副教授, 硕士生导师. CCF 会员. 主要研究方向为统计机器学习、数据挖掘.



Zhang Caiming, born in 1955. PhD, professor, PhD supervisor. Senior member of CCF. His main research interests include CAGD&CG, information visualization, and image processing.

张彩明, 1955 年生. 博士, 教授, 博士生导师. CCF 高级会员. 主要研究方向为 CAGD&CG、信息可视化、图像处理.