

# 无监督多视图特征选择研究进展\*

吴建生<sup>1,4,5</sup>, 李艳兰<sup>2</sup>, 黄冲<sup>3</sup>, 闵卫东<sup>1,4,5</sup>



<sup>1</sup>(南昌大学 数学与计算机学院, 江西 南昌 330031)

<sup>2</sup>(南昌大学 软件学院, 江西 南昌 330047)

<sup>3</sup>(南昌大学 信息化办公室, 江西 南昌 330031)

<sup>4</sup>(南昌大学 元宇宙研究院, 江西 南昌 330031)

<sup>5</sup>(江西省智慧城市重点实验室 (南昌大学), 江西 南昌 330031)

通信作者: 闵卫东, E-mail: [minweidong@ncu.edu.cn](mailto:minweidong@ncu.edu.cn)

**摘要:** 多视图数据从不同角度描述数据对象, 数据在不同视图中的特征表示之间存在着相关性、互补性及多样性信息. 综合利用这些信息对多视图数据处理至关重要. 然而, 多视图数据通常具有高维度特点, 且常含有噪声特征, 这为多视图数据的处理与分析带来了许多困难. 无监督多视图特征选择无需样本标记信息, 从多个视图源的原始高维特征中学习更紧凑、更准确的特征表示以提高数据分析的效果, 在多视图数据处理领域起着重要作用. 根据已有的无监督多视图特征选择模型的工作机制的异同, 对这些模型进行归纳和总结, 分析其中存在的不足, 并指出未来研究的方向.

**关键词:** 多视图数据; 无监督特征选择; 聚类; 结构保持; 流数据

**中图法分类号:** TP391

中文引用格式: 吴建生, 李艳兰, 黄冲, 闵卫东. 无监督多视图特征选择研究进展. 软件学报. <http://www.jos.org.cn/1000-9825/7168.htm>

英文引用格式: Wu JS, Li YL, Huang C, Min WD. Recent Advances in Unsupervised Multi-view Feature Selection. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/7168.htm>

## Recent Advances in Unsupervised Multi-view Feature Selection

WU Jian-Sheng<sup>1,4,5</sup>, LI Yan-Lan<sup>2</sup>, HUANG Chong<sup>3</sup>, MIN Wei-Dong<sup>1,4,5</sup>

<sup>1</sup>(School of Mathematics and Computer Sciences, Nanchang University, Nanchang 330031, China)

<sup>2</sup>(School of Software, Nanchang University, Nanchang 330047, China)

<sup>3</sup>(Informatization Office, Nanchang University, Nanchang 330031, China)

<sup>4</sup>(Institute of Metaverse, Nanchang University, Nanchang 330031, China)

<sup>5</sup>(Jiangxi Key Laboratory of Smart City (Nanchang University), Nanchang 330031, China)

**Abstract:** Multi-view data depicts objects from different perspectives, with features in different views exhibiting correlations, complementary, and diverse information. Therefore, it is crucial to make full use of this information for the processing of multi-view data. However, the processing and analysis of multi-view data will be difficult due to the inherent challenges of dealing with a vast number of features and the presence of noise features in multi-view data. Unsupervised multi-view feature selection, emerging as a critical component in multi-view data learning, efficiently learns more accurate and compact representations from the original high-dimensional multi-view data without relying on label information to remarkably improve the performance of data analysis. This study reviews and categorizes these models based on the similarities and differences in the working mechanisms of existing unsupervised multi-view feature selection models, while also detailing their limitations. Furthermore, this study points out promising future research directions in the field of

\* 基金项目: 国家自然科学基金 (62066027, 62076117); 江西省自然科学基金 (20212BAB212011); 江西省智慧城市重点实验室 (20192BCD40002)

收稿时间: 2023-08-11; 修改时间: 2023-10-22; 采用时间: 2024-02-14; jos 在线出版时间: 2024-08-14

unsupervised feature selection.

**Key words:** multi-view data; unsupervised feature selection; clustering; structure preservation; stream data

近年来,随着计算机科学技术和网络技术的不断发展,各行各业产生了大量的多视图数据,比如社交网络数据、多媒体数据和物联网传感器数据等.这些数据通常来自多个异构的视图源,这些视图源中的特征表示分别描述了数据特定的物理特性和统计特性.例如,图像可以用 HOG<sup>[1]</sup>、GIST<sup>[2]</sup>、SIFT<sup>[3]</sup>和 LBP<sup>[4]</sup>等多种特征描述符表示,也可以通过标注文字来描述;微博等社交网络数据往往包含文本、图像、视频等多种类型的信息;一个患者的情况可由多种组学资料、临床记录、生理参数和其他结构化数据集来表示<sup>[5]</sup>.多视图数据表示已被广泛应用于社交网络<sup>[6,7]</sup>、机器视觉<sup>[8,9]</sup>、模式识别<sup>[8,10]</sup>和数据挖掘<sup>[11,12]</sup>等多个研究领域.

然而,在处理多视图数据时通常会遇到高维度难题<sup>[13]</sup>.比如,当使用 SIFT 特征描述图像时,词袋大小往往超过 1000,如果采用分层、分块的描述子 LBP、HOG,数据维度通常会更大;微博博文中往往含有大量的文字等.高维数据不仅会带来维度灾难问题<sup>[14,15]</sup>,还会降低多维数据处理与分析的效果<sup>[16,17]</sup>.这是因为:1) 高维数据的特征之间通常存在关联性,从而导致数据中存在冗余特征;2) 在获取数据时,数据中经常会引入噪声特征,这些噪声特征会极大地影响聚类算法的性能<sup>[18,19]</sup>.因此,数据降维成为提高多视图数据处理分析效果的重要一环.

特征选择是数据降维的有效方法之一,它从原始特征中选取重要的、具有代表性的特征来代表原始数据,是模式识别和机器学习中数据预处理的关键步骤<sup>[20-22]</sup>.一般来说,特征选择有 3 个优点:1) 选取的特征子集比原始特征集维数更低,可以避免维数灾难,并提高后续学习模型的计算性能;2) 选择最具有代表性的特征,丢弃不相关、冗余和噪声特征,提升后续学习任务的准确性及其模型的可解释性和泛化能力<sup>[21,23,24]</sup>;3) 相对于常用的数据降维方法,如 PCA 等,特征选择模型产生的特征子集具有更好的解释性.

根据数据标签是否可用,现有的特征选择方法可以分为 3 类,即有监督的特征选择方法、半监督的特征选择方法和无监督的特征选择方法<sup>[25]</sup>.有监督的特征选择方法<sup>[26,27]</sup>通常根据训练数据的外在信息——人工标定标签来选择具有判别性的特征.显然,这种方法挑选出的特征子集能最大程度保留数据的外在信息.半监督的特征选择方法<sup>[28,29]</sup>用来处理输入样本中只有少部分样本具有类标签的情况.它一般是通过有标签数据和无标签数据协同选择特征.但在实践中,获取标签是一项耗时费力的工作,未标记的数据比标记的数据更容易获取.无监督的特征选择<sup>[30,31]</sup>的目标就是在无标签的情况下挖掘数据间的关联性,从原始特征集中选取更准确、更相关和更具有代表性的特征子集.无监督特征选择有 3 个重要优势:1) 无监督特征选择是无偏的,即它们不依赖于数据的类别信息,仅根据数据本身的属性、特征之间的统计性质或数据分布来评估特征的重要性,在没有先验知识的情况下仍表现良好<sup>[30-32]</sup>;2) 与监督特征选择方法相比,它们不会对特定的类别或任务产生偏好,从而有助于降低数据过拟合的风险<sup>[30,31]</sup>;3) 监督特征选择方法可能无法处理有新类别的数据,而无监督特征选择的特点使其在处理不同类型的数据或解决多个领域的问题时更具灵活性和通用性,比如文本挖掘、生物信息学、图像检索等<sup>[30]</sup>.

根据特征选择方法的评价策略与学习算法的交互方式的异同,特征选择方法也可以分为过滤型(filter)、封装型(wrapper)和嵌入型(embedded)这 3 种<sup>[25,33]</sup>.过滤型特征选择方法<sup>[34,35]</sup>利用相关评分标准直接独立地计算每个特征的重要性评估指数,从中选择评分较大的特征.它们不需要训练模型,方法简单,但也忽略了不同特征之间的相关性,因而容易选出冗余特征<sup>[36]</sup>.封装型特征选择方法<sup>[35,37]</sup>一般针对特定的学习模型进行特征选择.它们首先从初始特征集中选择特征子集来构造样本的新特征表示,然后根据新的特征表示训练学习模型,并根据学习模型的性能来对特征子集进行评价,直到选择出最佳的特征子集.封装型特征选择方法在最终的模型性能方面要优于过滤型特征选择方法.但它在特征选择的过程中需要反复训练学习器,因而计算开销过高,并且可能导致学习模型出现训练过拟合的现象.嵌入型特征选择方法<sup>[6,8]</sup>将特征选择和数据间关联性信息的挖掘过程集成到统一的框架中,在挖掘数据间关联性信息的过程中进行特征选择.这种方法不仅考虑了特征之间的相关性,还避免了封装型交叉验证的巨大计算开销.现有的无监督多视图特征选择大多是嵌入型方法.

样本的特征表示在很大程度上决定了学习任务的性能.得益于数据在不同视图下的多样性和一致性信息,我们可以从多视图数据中获得更有用、更全面的信息,减少噪声特征和冗余特征的影响,从而进一步实现更好的数

据处理性能<sup>[38,39]</sup>. 无监督多视图特征选择 (UMVFS) 不需要标记信息, 通过挖掘数据内在结构信息, 从原始的多个视图源的高维特征中学习更紧凑、更准确的特征表示以提高数据分析的效果, 在多视图数据学习领域起着重要作用. 本文将针对近年来所提出的无监督多视图特征选择工作 (后统称无监督多视图特征选择) 进行系统性的介绍和总结. 根据无监督多视图特征选择的主要方法和特点, 将无监督多视图特征选择方法分为基于图的模型、基于伪标签的模型和基于数据流的模型, 并对同一类模型的使用方法、特点及复杂度进行了对比和总结.

近年来, 一些综述文献对现有特征选择算法的理论和应用进行了总结<sup>[20-25,28,29]</sup>, 它们主要聚焦于单视图特征选择模型. 然而, 至今仍缺乏一份全面整理无监督多视图特征选择的研究进展的综述. 虽然文献 [25] 提出了一个自适应学习视图权重的多视图特征选择框架, 但它关注的内容仍然是单视图数据特征选择选择和特征融合. 总的来说, 本文与文献 [25] 的主要区别如下: 首先, 本文对无监督多视图特征选择的研究进展进行了全面回顾, 而文献 [25] 主要针对单视图数据方法进行了综述, 并未重点关注无监督多视图数据特征选择. 其次, 鉴于目前的无监督多视图特征选择绝大部分是嵌入型方法, 本文根据无监督多视图特征选择模型的工作机制, 将其划分为基于图的模型、基于伪标签的模型和基于数据流的模型, 而文献 [25] 则从通用的分类角度——搜索策略 (过滤型、封装型和嵌入型) 方面来介绍特征选择研究的最新工作.

本文第 1 节中介绍无监督多视图特征选择的框架, 包括其结构和分类. 第 2 节介绍基于图的无监督多视图特征选择模型, 包括基于固定图的模型和基于自适应图的模型. 第 3 节介绍基于伪标签的无监督多视图特征选择模型, 包括直接学习数据共识类标的模型和基于视图特定的数据类标间接学习数据的共识类标的模型. 第 4 节介绍基于数据流的无监督多视图特征选择模型. 第 5 节描述常用的公开数据集、评价标准以及 9 个模型的模拟实验结果. 第 6 节介绍无监督多视图特征选择模型在实际中的一些应用. 最后, 第 7 节对本文进行总结, 并指出无监督多视图特征选择研究未来面临的挑战和发展方向.

## 1 无监督多视图特征选择框架

相较于无监督单视图特征选择模型, 无监督多视图特征选择模型 (UMVFS) 不仅需要解决数据结构学习和特征选择这两个本质问题<sup>[40,41]</sup>, 还需要充分挖掘视图间互补性和相关性信息. 因此, 多视图特征选择的关键是如何有效地利用多视图数据的多样性和一致性来协同衡量相关特征的重要性, 从而从原始特征中选取具有辨识性的特征.

### 1.1 无监督多视图特征选择结构

稀疏学习模型通过降低变量空间维数有效地解决高维问题<sup>[42,43]</sup>. 它们不仅可以过滤大量冗余特征与噪声特征, 只保留与样本最相关的特征, 简化数据特征表示, 降低计算量和传输存储, 而且具有很好的解释性, 便于数据可视化, 有效地解决高维数据集建模的诸多问题. 现在的绝大部分无监督多视图特征选择模型都是基于稀疏模型的方法.

一般而言, 无监督多视图特征选择模型的目标是通过低维嵌入, 挖掘数据间的几何结构, 并为每个视图找到一个稀疏的转换矩阵 (投影矩阵/特征选择矩阵)  $W^{(v)}$ , 将原始高维数据投影到低维嵌入所表示的低维空间中, 进而使投影后的数据可以保持数据间的几何结构. 无监督多视图稀疏特征选择方法的目标函数通常可以由公式 (1) 表示:

$$\min_{W^{(v)}} \sum_{v=1}^m \text{loss}(X^{(v)}, W^{(v)}) + \lambda \sum_{v=1}^m \|W^{(v)}\|_{2,p}^p \quad (1)$$

其中, 第 1 项为数据投影损失项 (经验误差), 第 2 项为惩罚项 (正则化项). 损失项主要用于挖掘数据间的内部结构性信息, 并结合特征重要性评价方法来评估投影后的数据在保持该内部结构方面的性能. 惩罚项是稀疏约束项, 能够防止模型过拟合现象, 提高了模型的泛化能力.  $\lambda$  为正则化参数, 是损失项和惩罚项之间的权衡系数. 不同损失项或惩罚项的稀疏学习模型表现出不同的性能. 为了使所选取的特征子集能够尽可能保持数据间的多种内部结构性信息, 一些稀疏学习模型甚至使用多个正则化参数来平衡惩罚因素<sup>[44,45]</sup>.  $p$  是  $l_{2,p}$  范数中的参数, 其取值范围为  $0 < p \leq 1$ , 大部分无监督多视图特征选择模型令  $p=1$ , 即使用  $l_{2,1}$  范数对特征选择矩阵施加行稀疏来达到特征选择的

目的<sup>[8,14]</sup>,也有个别方法将 $p$ 设为其他值来满足不同的稀疏性要求<sup>[46]</sup>.

基于稀疏学习的特征选择模型是一个优化问题,即最小化由正则化项惩罚的经验误差.求解该优化问题后,使用特征选择矩阵 $W^{(v)}$ 的行向量范数来分别计算每个视图中的每个特征的权重,权重越大,则该特征越重要.由于稀疏项的约束,矩阵 $W^{(v)}$ 的某些行将会收缩到零向量,这些零向量对应的特征即为冗余特征和噪声特征.基于稀疏学习模型的无监督多视图特征选择方法通常使用嵌入式方法,即在同一个框架中同步完成稀疏学习和特征重要性评价.基于稀疏模型的无监督多视图特征选择流程如图1所示.

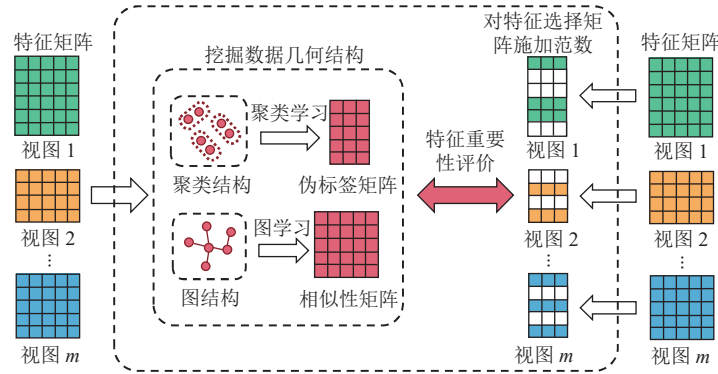


图1 基于稀疏模型的无监督多视图特征选择流程

## 1.2 无监督多视图特征选择分类

在无监督的学习模型中,由于缺少数据类标信息,数据间的几何结构常用于直接描述数据间的关联性信息,是用于特征选择和数据降维的一个重要指标和关键先验信息,具有出色的信息挖掘能力<sup>[47,48]</sup>.在无监督多视图特征选择研究中,数据间的相似性结构<sup>[28,49-52]</sup>和数据的聚类结构<sup>[12,53-55]</sup>常被用来表示数据间的几何结构.

本文根据无监督多视图特征选择模型的工作机制,将这些模型分为基于图的无监督多视图特征选择模型、基于伪标签的无监督多视图特征选择模型和基于数据流的无监督多视图特征选择模型.基于图的无监督多视图特征选择模型利用数据的相似性图刻画数据的关联关系,并借助图理论分析工具<sup>[50,51,56-58]</sup>来分析特征对图几何结构的保持能力.基于伪标签的无监督多视图特征选择模型通常利用非负矩阵分解<sup>[59]</sup>和谱聚类<sup>[60]</sup>两种聚类方法来获取数据的伪类标信息.基于数据流的无监督多视图特征选择模型考虑了现实应用中的大规模数据和数据流问题,利用在线学习方式和缓存技术对大规模数据和数据流进行处理.此外,无监督多视图特征选择模型为了尽可能地挖掘视图信息,可能会同时将不同的方法组合到一个框架中,一个算法从不同角度可归为不同类,因此该分类并非绝对分类.

## 1.3 符号表示

本文使用大写字母和小写粗体字母,比如 $X$ 和 $\mathbf{x}$ ,分别表示矩阵和列向量.对于一个任意给定的矩阵 $X \in \mathbb{R}^{n_1 \times n_2}$ ,我们用 $X_{ij}$ 和 $\mathbf{x}_k$ 分别表示它的第 $i$ 行第 $j$ 列的元素和第 $k$ 列列向量.那么,它的Frobenius范数和 $l_{p,q}$ 范数定义如公式(2)和公式(3)所示:

$$\|X\|_F = \sqrt{\sum_{i=1}^{n_1} \sum_{j=1}^{n_2} |X_{ij}|^2} \quad (2)$$

$$\|X\|_{p,q} = \left( \sum_{i=1}^{n_1} \left( \sum_{j=1}^{n_2} |X_{ij}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \quad (3)$$

如果 $X$ 是方阵,那么 $X$ 的迹可以用 $\text{Tr}(X)$ 表示.

对于三阶张量,本文使用手写风格的大写字母表示,例如 $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ .此外,张量 $\mathcal{X}$ 基于张量奇异值分解的张

量核范数定义如公式 (4) 所示:

$$\|\mathcal{X}\|_{\otimes} = \sum_{i=1}^{n_3} \|\mathcal{X}_f^{(i)}\|_* = \sum_{i=1}^{n_3} \sum_{j=1}^{\min(n_1, n_2)} |\mathcal{S}_f^{(i)}(j, j)| \quad (4)$$

其中,  $\mathcal{X}_f$  表示沿第 3 个维度对该张量进行快速傅立叶变换获得的数据表示,  $\mathcal{S}_f^{(i)}$  是对  $\mathcal{X}_f$  的第  $i$  个前切片作奇异值分解得到的奇异值构成的对角矩阵.

此外, 后文中多次提及希尔伯特-施密特独立性准则 (HSIC)<sup>[61,62]</sup>, 它常用于度量两个随机变量之间相互关系, 捕捉它们之间的非线性依赖, 旨在检测这两个随机变量之间的独立性. 给定两个随机变量  $X$  和  $Y$ , 它们的 HSIC 定义如公式 (5) 所示:

$$HSIC(X, Y) = \frac{1}{(n-1)^2} \text{Tr}(HK_x HK_y) \quad (5)$$

其中,  $K_x$  和  $K_y$  分别是与  $X$  和  $Y$  相关的核矩阵,  $H$  是中心化矩阵. HSIC 的值越大, 表示  $X$  和  $Y$  之间的关联性越强, 较小则表明它们之间更可能是独立的. HSIC 提供了一种不依赖于具体分布的方法, 适用于各种数据类型, 被广泛用于特征选择等任务<sup>[61]</sup>.

本文中常用的符号如表 1 所示.

表 1 符号说明

符号	含义	符号	含义
$n$	样本数量	$I$	单位矩阵
$m$	视图数量	$c$	数据的聚类簇数
$d_v$	第 $v$ 个视图的维度	$H$	中心化矩阵
$d$	所有视图的维度之和	$s$	投影后的数据维度
$X$	全体样本的特征矩阵	$\bar{d}$	数据的平均维度
$X^{(v)}$	第 $v$ 个视图的特征矩阵	$\mathbf{1}$	元素全为 1 的列向量
$x_i$	第 $i$ 个对象的多视图特征向量	$s_v$	第 $v$ 个视图投影的维度
$x_i^{(v)}$	第 $i$ 个对象在第 $v$ 个视图内的特征向量	$k$	$k$ -近邻方法的最近邻数

## 2 基于图的无监督多视图特征选择模型

无监督特征选择模型的一个重要准则是局部流形结构的保持: 高维数据中包含着低维流形结构<sup>[49]</sup>. 局部结构的保持有利于学习和获得紧凑的低维表示, 而数据间的局部结构常使用图 (相似性矩阵) 进行刻画<sup>[28,49-52]</sup>. 基于图的无监督多视图特征选择模型通常从多视图数据学习获得一个共识图 (或多个视图特定的图), 并通过挖掘数据特征与该图结构的关联关系来衡量数据特征的重要性, 以便从特征集中选取能够保持数据间图结构的特征子集作为数据的新表达. 这类方法一般需要借助谱图分析工具<sup>[56,63]</sup>将数据在低维子空间中展开, 获得数据在低维子空间中的表示. 根据图结构的定义方式不同, 这类模型又可分为基于固定图的特征选择模型和基于自适应图的特征选择模型.

### 2.1 基于固定图的模型

基于固定图的无监督多视图特征选择模型为每个视图内的数据预先定义一个相似性矩阵用于描述数据在该视图内的结构关系. 在模型训练过程中, 这些相似性矩阵保持不变, 即图的构造过程独立于特征选择的过程. 基于固定图的无监督多视图特征选择流程如图 2 所示.

为了提高人体运动数据检索的准确性, Wang 等人<sup>[9]</sup>提出了自适应多视图特征选择模型 (AMFS). 具体而言, AMFS 首先使用多种视觉特征描述人体运动数据, 构成人体运动多视图表示. 基于此, AMFS 为每个视图计算基于预定义图的拉普拉斯矩阵  $L^{(v)}$ , 将其作为运动数据在每个视图中的局部结构描述, 并通过自适应视图权重  $\alpha$ , 将这些矩阵线性组合来达到利用多视图数据互补性的目的. 其次, Wang 等人要求投影后的数据仍能够保持运动数据特

征之间的局部结构信息和全局结构信息,因此采用了传统 Fisher Score 特征选择方法中的迹比准则将结构保持和特征选择统一到了一个框架中. AMFS 目标函数如公式 (6) 所示:

$$\min_{W, \alpha} \frac{\text{Tr}\left(W^T X \left(\sum_{v=1}^m \alpha_v L^{(v)}\right) X^T W\right)}{\text{Tr}(W^T X H X^T W)}, \text{ s.t. } \sum_{v=1}^m \alpha_v = 1, \alpha_v \geq 0, W \in \{0, 1\}^{d \times s} \quad (6)$$

其中,  $H = I - \frac{1}{n} \mathbf{1}\mathbf{1}^T$  表示中心化矩阵,  $r$  是一个参数, 用于避免只有一个视图被赋予权重的平凡解.

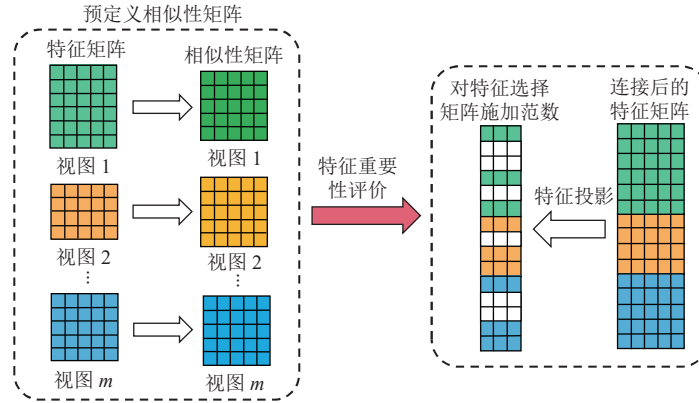


图2 基于固定图的无监督多视图特征选择流程

AMFS<sup>[9]</sup>显式地为每个视图分配了不同的权重, 而 Wangila 等人<sup>[45]</sup>则引入了无参学习策略<sup>[64]</sup>, 提出了一种混合稀疏正则化的无监督多视图特征选择模型 (MSMFS), 可以隐式地为每个视图学习不同的权重. 同时, MSMFS 对特征选择矩阵  $W$  施加了组稀疏正则化来消除异常特征的负面影响. MSMFS 的目标函数如公式 (7) 所示:

$$\min_W \sum_{v=1}^m \sqrt{\text{Tr}(W^T X L^{(v)} X^T W)} + \lambda_1 \|W\|_{2,1} + \lambda_2 \sum_{v=1}^m \|W^{(v)}\|_F, \text{ s.t. } W^T W = I \quad (7)$$

其中,  $\lambda_1$  和  $\lambda_2$  是用于平衡不同项的超参数, MSMFS 的视图权重隐式定义如公式 (8) 所示:

$$\alpha_v = 1 / \left( 2 \sqrt{\text{Tr}(W^T X L^{(v)} X^T W)} \right), v = 1, \dots, m \quad (8)$$

核方法是解决非线性模式分析问题的一种有效途径, 近年来在多视图数据学习领域被广泛研究<sup>[15,55,65]</sup>. Wei 等人<sup>[15]</sup>基于核方法, 提出了基于交叉扩散和矩阵对齐的无监督多视图特征选择模型 (CDMA-FS). CDMA-FS 为每个视图构建一个核矩阵  $K^{(v)}$ , 并通过交叉扩散将这些核融合成一个鲁棒交叉扩散矩阵  $G$ . 在这个交叉扩散矩阵的指导下, 该模型利用希尔伯特-施密特独立性准则 (HSIC)<sup>[61]</sup>度量所选取特征与数据结构之间的关联性, 并针对每个视图分别进行特征选择. 与 AMFS 和 MSMFS 明显不同的是, CDMA-FS 通过矩阵对齐学习了一个公共的共识图  $G$  用于表示数据跨视图的结构信息, 并且它可以评估特征间的非线性关系. CDMA-FS 的目标函数如公式 (9) 所示:

$$\min_{s^{(v)}} -\text{Tr}(HGHK^{(v)}), \text{ s.t. } \sum_{p=1}^{d_v} s_p^{(v)} = c^{(v)}, s_p^{(v)} \in \{0, 1\}, p = 1, \dots, d_v \quad (9)$$

其中,  $s^{(v)}$  是每个视图的特征选择指示向量,  $c^{(v)}$  是第  $v$  个视图选择的特征数.

综上所述, 基于固定图的无监督多视图特征选择模型的特点是: 图的构造过程独立于特征选择的过程, 且图结构在特征选择的过程中保持不变. 然而, 对于基于图的特征选择模型而言, 特征选择的性能在很大程度上依赖于图的质量. 虽然基于固定图的模型已被证明在一些数据集上是有效的, 但由于原始的高维数据中通常包含着噪声特征和异常值, 在原始特征空间中学习到的相似性图无法准确地刻画多视图数据间的真实关系, 从而不能指导特征选择得到最优结果<sup>[65-69]</sup>. 鉴于这些因素, 当前主流的无监督多视图特征选择研究趋向于摒弃采用固定图的策略. 除

了以上共同点, 这几个模型在多视图数据信息挖掘方面和噪声特征处理方面存在一些差别.

1) 在多视图数据信息挖掘方面, AMFS 和 MSMFS 将视图特定的拉普拉斯矩阵作为多视图数据的多样性表示, 但忽略了多视图数据之间的潜在一致性. 此外, AMFS 和 MSMFS 将所有视图中的特征进行连接合成一个较大的数据表示, 并在该数据表示的基础上衡量数据特征与拉普拉斯表示之间的关联性. 然而, 不同视图的特征特性不同, 通过这种方法所选择的特征通常难以进行解释和进一步分析, 弱化了多个视图之间的多样性, 违背了多视图特征选择的初衷. 相比之下, CDMA-FS 则采用视图特定的核矩阵和融合的交叉扩散矩阵分别表示多视图数据的多样性和一致性, 并利用这两种表示指导特征选择.

2) 在噪声特征处理方面, AMFS 和 MSMFS 分别利用线性判别分析和组稀疏去除冗余特征和无关特征; 而 CDMA-FS 为每个视图分别进行特征选择, 因此可能会选择出表示相同信息的冗余特征. 此外, AMFS 和 CDMA-FS 没有引入额外的手动设定参数, 具有模型简单的优点; CDMA-FS 还使用 *HSIC* 衡量特征间的非线性关系.

表 2 对以上模型进行了总结.

表 2 基于固定图的无监督多视图特征选择模型

算法简称	特点	优缺点	时间复杂度
AMFS <sup>[9]</sup>	人体运动检索应用、迹比准则去除冗余特征、通过样本的合成特征数据矩阵学习	无参方法, 模型简单; 忽略了多视图数据一致性	$O(T(dn^2 + nd^2 + d^3))$
MSMFS <sup>[45]</sup>	混合组稀疏正则化、无参方法学习视图权值、通过样本的合成特征数据矩阵学习	无参权重学习, 减少模型参数; 忽略了多视图数据一致性	$O(T(d^3 + dn^2 + nd^2))$
CDMA-FS <sup>[15]</sup>	核学习方式、交叉扩散矩阵指导特征选择	无参方法, 模型简单; 可以评估特征的非线性可用性; 可能会选择出冗余特征	$O(T(mdn^2 + md^2))$

## 2.2 基于自适应图的模型

在没有数据实例标签的情况下, 样本的局部几何结构通常作为先验知识来引导特征选择过程, 以选取能够保持数据间几何结构的特征子集来构造数据的新表达. 因此, 相似性矩阵质量的好坏在很大程度上影响着模型的最终性能<sup>[70]</sup>. 为了有效地解决基于固定图的无监督多视图特征选择模型存在的问题, 研究者们提出了将数据间结构学习和特征选择结合的思想, 即基于自适应图的无监督多视图特征选择模型. 具体来讲, 这种方法将图的构造和特征的选择组合到一个框架中, 交替进行相似性矩阵的优化学习和特征的重要性评估. 这种自适应学习机制使模型在很大程度上避免了噪声特征或不相关特征对相似性图学习的影响, 进而保证模型能够选择出更具有辨别性的特征. 基于自适应图的无监督多视图特征选择流程如图 3 所示.

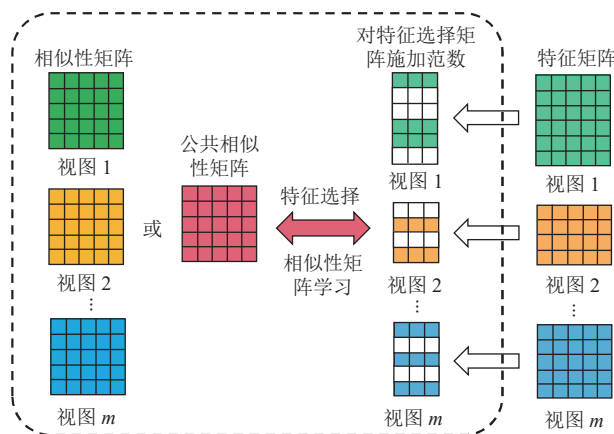


图 3 基于自适应图的无监督多视图特征选择流程

Dong 等人<sup>[71]</sup>提出的自适应协同相似性学习的无监督多视图特征选择模型 (ACSL) 基于每个视图的预定义相似性矩阵  $S^{(v)}$  自适应地学习协同相似性矩阵  $S$ . 为了使最终的协同矩阵达到理想状态, 即图的连通分量数目正好等于聚类数目  $c$ <sup>[72]</sup>, ACSL 对拉普拉斯矩阵  $L_S$  施加了秩约束  $rank(L_S) = n - c$ , 并通过  $\text{Tr}(F^T L_S F)$  将离散优化问题转成为迹形式的连续优化问题. ACSL 的目标函数如公式 (10) 所示:

$$\begin{cases} \min_{S, F, W, U} \sum_{j=1}^n \left\| s_j - \sum_{v=1}^m U_{vj} s_j^{(v)} \right\|_F^2 + \alpha \text{Tr}(F^T L_S F) + \beta (\|X^T W - F\|_F^2 + \gamma \|W\|_{2,1}) \\ \text{s.t. } \forall j, \mathbf{1}^T s_j = 1, s_j \geq 0, \mathbf{u}_j^T \mathbf{1} = 1, F \in R^{n \times c}, F^T F = I \end{cases} \quad (10)$$

其中,  $U$  是相似性矩阵的权重矩阵,  $L_S = D_S - \frac{S^T + S}{2}$  为拉普拉斯矩阵,  $D_S = \text{diag}(S\mathbf{1})$  是度矩阵. 继 ACSL 之后, 刘欣宇等人<sup>[73]</sup>在 ACSL 的基础上使用稀疏性更强的  $l_{1,2}$  范式替代了公式 (10) 中的  $l_{2,1}$  范式, 以更好地处理噪声特征.

Du 等人<sup>[74]</sup>提出的基于迹比准则的加权共识图无监督多视图特征选择模型 (CGWFS) 与 ACSL 类似, 它同样为每个视图预定义一个相似性矩阵  $S^{(v)}$  (对应视图级图), 用于表示特定视图的局部结构信息. 不同的是, CGWFS 还通过特征级的相似性矩阵  $G$  (对应特征级图) 来保持数据基于单个特征表示的跨视图数据局部结构. 为了融合这两种图的信息, CGWFS 学习了一个最优的共识图  $S$ , 使不同视图的特征和单个特征之间达成一致, 同时保持数据的局部结构. 为此, CGWFS 通过两级加权线性组合来实现共识图和特征级图之间的近似; 通过样本相似性矩阵对齐来保证共识图和视图级图之间的一致性. CGWFS 的目标函数如公式 (11) 所示:

$$\min_{S, \mathbf{w}, \mathbf{u}^{(v)}} \frac{\left\| S - \sum_{v=1}^m w_v \left( \sum_{i=1}^{d_v} u_i^{(v)} G^{(v)} \right) \right\|_F^2}{\text{Tr} \left( S \left( \sum_{v=1}^m w_v S^{(v)} \right) \right)}, \text{ s.t. } \mathbf{w} \geq 0, \sum_{v=1}^m w_v = 1, \mathbf{u}^{(v)} \geq 0, \sum_{i=1}^{d_v} u_i^{(v)} = 1, S \geq 0, \sum_{i=1}^n S_{vi} \geq 0 \quad (11)$$

其中,  $G^{(v)}$  为数据基于第  $v$  个视图内的第  $i$  个特征的相似性矩阵,  $\mathbf{w}$  和  $\mathbf{u}^{(v)}$  分别是视图级图和特征级图的权重向量.

ACSL 和 CGWFS 都需要在预先的相似性矩阵基础上学习一个协同相似性矩阵来指导特征选择. 与基于固定图的模型类似, 预先构造的相似性矩阵会受到噪声特征影响, 可能导致协同相似性矩阵不够准确. 针对这个问题, Hou 等人<sup>[14]</sup>受局部保持投影 (LPP)<sup>[57]</sup>和局部线性嵌入 (LLE)<sup>[49]</sup>的启发, 提出了一个局部相似性和视图权重自适应学习的无监督多视图特征选择模型 (ASVW). 该模型综合数据在所有视图中的分布, 同步学习投影矩阵和相似性矩阵以获得数据跨视图的全局结构  $S$ , 刻画数据跨视图的共性信息, 避免了预定义相似性矩阵带来的偏差. ASVW 的目标函数如公式 (12) 所示:

$$\begin{cases} \min_{W^{(v)}, S, \alpha} \sum_{v=1}^m \sum_{i,j=1}^n \alpha_v^n \left\| (W^{(v)})^T \mathbf{x}_i^{(v)} - (W^{(v)})^T \mathbf{x}_j^{(v)} \right\|^2 (S_{ij})^2 + \lambda \sum_{v=1}^m \|W^{(v)}\|_{2,p}^p \\ \text{s.t. } (W^{(v)})^T W^{(v)} = I, \sum_{v=1}^m \alpha_v = 1, \alpha_v \geq 0, \sum_{j=1}^n S_{ij} = 1, S_{ij} \geq 0, \|s_i\|_0 = k \end{cases} \quad (12)$$

其中,  $S$  表示多视图数据的公共局部几何结构, 约束项  $\|s_i\|_0 = k$  用于保证每个样本拥有  $k$  个最优近邻. ASVW 为相似性矩阵的构造提供了一个新的思路, 即这个相似性矩阵不是预先定义的, 而是需要根据数据在各视图内的分布情况进行自适应地学习, 因此被后续的很多模型学习并扩展.

在 MSMFS<sup>[45]</sup>和 ASVW<sup>[14]</sup>的基础上, Wang 等人<sup>[38]</sup>提出了基于图优化的权重自学习的多视图特征选择模型 (MFSGL). MFSGL 将 MSMFS 中的视图权重无参学习模型扩展到更一般的情形, 它能根据数据在各视图内的分布情况, 自动为视图分配不同的权重. 此外, 该模型也将离散的秩约束  $rank(L_S) = n - c$  转换为了连续的迹形式的优化问题. MFSGL 的目标函数如公式 (13) 所示:

$$\begin{cases} \min_{W^{(v)}, S, F} \sum_{v=1}^m \left( \alpha_v \text{Tr} \left( (W^{(v)})^T (X^{(v)})^T L_S X^{(v)} W^{(v)} \right) + \gamma \|W^{(v)}\|_{2,1} \right) + \mu \sum_{i,j} S_{ij}^2 + 2\lambda \text{Tr}(F^T L_S F) \\ \text{s.t. } s_i^T \mathbf{1} = 1, 0 \leq S_{ij} \leq 1, F^T F = I, F \in R^{n \times c}, (W^{(v)})^T W^{(v)} = I \end{cases} \quad (13)$$



其中, 第  $v$  个视图的权重  $\alpha_v$  定义如公式 (14) 所示:

$$\alpha_v = \frac{p}{2\text{Tr}\left((W^{(v)})^T(X^{(v)})^T L_S X^{(v)} W^{(v)}\right)^{\frac{2-p}{2}}} \quad (14)$$

传统的基于最小回归的特征选择算法通过两个步骤获得特征权重: 先学习投影矩阵  $W$ , 然后通过  $l_{2,1}$  范数计算得到特征权重, 但这种方法可能无法找到投影矩阵的全局稀疏解. 此外, 在实际应用中, 我们往往很难确定样本的邻域大小, 并且, 不同视图中的样本甚至同一视图中的不同样本都很可能具有不同大小的邻域, 而 ASVW 中限制样本邻域大小为  $k$  的约束过于理想化, 不符合现实场景中真实的节点分布情况. 针对以上问题, Zhang 等人<sup>[10]</sup>提出了动态学习样本结构的无监督多视图特征选择模型 (DSS-UFS). 它使用自表达的方法, 根据数据的实际分布来自动学习样本间的近邻及它们之间的关联关系, 并引入权重因子  $\theta$  直接度量特征的重要性, 能够避免两步法造成的非最优解情况. DSS-UFS 的目标函数如公式 (15) 所示:

$$\min_{W^{(v)}, S} \sum_{v=1}^m \left\| (\hat{W}^{(v)})^T X^{(v)} (I - S) \right\|_F^2 + \gamma \left\| [W^{(1)}; \dots; W^{(m)}] \right\|_{2,1}^2, \text{ s.t. } (W^{(v)})^T W^{(v)} = I, s_i^T \mathbf{1} = 1, S_{ii} = 0, S_{ij} \geq 0 \quad (15)$$

其中,  $\hat{W}^{(v)} = \theta^{(v)} W^{(v)}$ ,  $\theta^{(v)} = \text{diag}\left(\sqrt{\theta_1^{(v)}}, \dots, \sqrt{\theta_{p^{(v)}}^{(v)}}\right)$  是缩放特征权重矩阵.

局部结构和全局结构是两种刻画原始数据分布的重要工具, 它们从不同方面描述了数据分布情况: 局部结构描述了数据及其近邻间的关系, 全局结构则反映了数据在整个数据空间中的分布情况. 综合考虑数据的局部和全局结构可以更全面地表示数据间的相关性信息. 为了同时学习全局结构和局部结构信息, Sun 等人<sup>[75]</sup>和 Wan 等人<sup>[68]</sup>提出了自适应全局结构和局部结构学习的特征选择模型 (分别对应 ME-ASOS 和 ASE-UMFS). 这两种方法的主要区别在于全局结构信息的刻画方式不同. ME-ASOS<sup>[75]</sup>利用面向视图的多输出正则化投影<sup>[76]</sup>将不同视图的高维数据映射到一个共享的低维嵌入子空间  $U$  中, 如公式 (16) 所示:

$$\min_{W^{(v)}, U} \sum_{v=1}^m \xi_v^r \left\| X^{(v)} - U(W^{(v)})^T X^{(v)} \right\|_F^2 + \mu \|U\|_1, \text{ s.t. } U^T U = I \quad (16)$$

其中,  $U$  反映了数据间的全局流形结构,  $\xi$  用于平衡重构误差. ASE-UMFS<sup>[68]</sup>则通过自表达的形式<sup>[77]</sup>学习数据间的全局关联信息  $Q$ , 如公式 (17) 所示:

$$\min_{W^{(v)}, Q} \sum_{v=1}^m \left\| (W^{(v)})^T X^{(v)} - (W^{(v)})^T X^{(v)} Q \right\|_F^2 + \mu \|Q\|_1, \text{ s.t. } Q_{ii} = 0 \quad (17)$$

其中,  $Q$  反映了数据间的全局关联关系.

目前大多数方法对每个视图的投影矩阵  $W^{(v)}$  进行优化, 并根据这些投影对特征进行联合评估. 这些方法往往需要设定每个投影空间的维度, 但不同视图中的特征通常具有不同的解释和数目, 因此, 如何确定投影子空间的维度仍是一个难题. 此外, 当不同视图的特征评估矩阵的维度差异较大时, 对异构特征进行联合评估是不合理的, 这会导致选择结果存在偏差. 为了解决以上问题, Zhang 等人<sup>[78]</sup>提出的基于多级投影的近邻图自适应学习的多视图特征选择模型 (MAMFS) 将投影矩阵和特征选择矩阵角色分离, 即每个视图的投影矩阵  $U^{(v)}$  和一个统一的特征选择矩阵  $W$ . 通过这个多级投影, MAMFS 从每个视图的特征  $X^{(v)}$  和多视图数据连接后的特征  $X$  中学习了异质特征的全局结构和局部结构, 充分利用了视图内和视图间的信息. 其次, 为了消除冗余特征, MAMFS 对特征选择矩阵施加了不相关约束  $W^T S_i W = I$  ( $S_i = X H X^T$  为总散列矩阵). MAMFS 的目标函数如公式 (18) 所示:

$$\begin{cases} \min_{W, S, \alpha, U^{(v)}} \sum_{v=1}^m \alpha_v^\gamma \sum_{i,j} \left\| (U^{(v)})^T \mathbf{x}_i^{(v)} - (U^{(v)})^T \mathbf{x}_j^{(v)} \right\|_2^2 S_{ij} + \lambda \|S\|_F^2 + \beta \text{Tr}(W^T X L_S X^T W) + \eta \|W\|_{2,1} \\ \text{s.t. } W^T S_i W = I, (U^{(v)})^T U^{(v)} = I, \alpha^T \mathbf{1} = 1, \alpha \geq 0, S \mathbf{1} = \mathbf{1}, S \geq 0, \|s_i\|_0 = k \end{cases} \quad (18)$$

低秩张量学习方法能够有效地捕获多视图数据的一致性信息, 近年来得到了广泛的关注<sup>[79,80]</sup>. Yuan 等人<sup>[81]</sup>提出的基于张量低秩的无监督多视图特征选择模型 (TLR-MVFS) 在 ASVW<sup>[14]</sup>的基础上, 学习每个视图的相似性矩阵  $S^{(v)}$  以保留数据的特定视图信息并构造张量  $\mathcal{Z}$ ; 然后使用张量低秩约束来捕获多视图数据的一致性信息. TLR-

MVFS 的目标函数如公式 (19) 所示:

$$\begin{cases} \min_{W^{(v)}, S^{(v)}} \sum_{v=1}^m \sum_{i=1}^n \sum_{j=1}^n \left\| (W^{(v)})^T \mathbf{x}_i^{(v)} - (W^{(v)})^T \mathbf{x}_j^{(v)} \right\|^2 S_{ij}^{(v)} + \lambda_1 \sum_{v=1}^m \|W^{(v)}\|_{2,1} + \lambda_2 \|Z\|_{\otimes} \\ \text{s.t. } (W^{(v)})^T W^{(v)} = I, \sum_{j=1}^n S_{ij}^{(v)} = 1, S_{ij}^{(v)} \geq 0, S_{ii}^{(v)} = 0, Z = \Phi(S^{(1)}, S^{(2)}, \dots, S^{(m)}) \end{cases} \quad (19)$$

其中,  $\|Z\|_{\otimes}$  表示张量  $Z$  的秩,  $\Phi(S^{(1)}, S^{(2)}, \dots, S^{(m)})$  表示关于  $\{S^{(v)}\}_{v=1}^m$  的  $n \times m \times n$  的三阶张量.

基于自适应图的方法通过学习机制自适应地学习数据间的结构, 降低了冗余特征和噪声特征对数据间几何结构学习的影响. 根据几何结构学习策略不同, 以上基于自适应图的无监督多视图特征选择模型可分为两类.

1) 基于固定图的自适应学习方法, 包括 ACSL 和 CGWFS. 它们通过融合每个视图预定义的相似性图的方式来生成公共的相似性图, 因此这种方式得到的公共相似性矩阵很可能无法准确地反映数据的真实分布情况. 除此之外, ACSL 和 CGWFS 还存在以下差异: a) ACSL 对相似性矩阵施加了秩约束从而使其达到理想聚类结构, 促使后续的特征选择能够选取保持数据聚类结构的特征, 增加特征子集的辨识性. 但它简单地将多视图数据合并成单视图数据进行学习, 弱化了多视图数据的多样性信息. b) CGWFS 通过使用视图级图和特征级图, 同时保持了数据在每个视图内的局部几何结构和数据跨视图的全局几何结构信息.

2) 根据数据分布自适应学习, 如 ASVW 和 MFSGL 等. 此类方法根据数据分布情况, 在低维投影空间中自适应地学习数据间相似性矩阵, 尽可能地降低噪声特征对相似性矩阵构造的影响, 从而有效地进行特征选择. 除了上述联系之外, 这些方法之间还有许多差异: a) MFSGL、ME-ASOS、ASE-UMFS 和 MAMFS 均沿用了 ASVW 的学习方式, 自适应学习一个公共相似性矩阵以表示多视图数据的一致性几何结构信息. 它们主要区别在于: MFSGL 对相似性矩阵施加秩约束使其达到理想聚类结构, 以获得更准确的相似性图; 但它过于强调一致性, 忽略了不同视图之间的多样性和相关性. 而除了保持多视图数据的局部几何结构外, ME-ASOS、ASE-UMFS 还分别利用多输出正则化投影和自表达机制学习了多视图数据的全局结构, 能够同时利用多视图数据的全局和局部几何结构信息指导特征选择. MAMFS 则基于子空间学习, 通过多级投影同时学习视图内和视图间的信息, 有效地提高了特征选择的性能. b) TLR-MVFS 虽然也采纳了 ASVW 方法, 但它并非学习统一的相似性矩阵, 而是为每个视图都学习一个相似性矩阵以捕捉视图特定的信息, 并基于张量低秩约束挖掘视图之间的潜在一致性信息. c) 与以上基于 ASVW 的模型不同, DSS-UFS 利用自表达策略提出了一种新的自适应相似性矩阵学习方式, 能够自动地为样本学习不同数目的近邻. 但 DSS-UFS 只考虑了多视图数据一致性, 忽略了多视图数据间的多样性.

表 3 对基于自适应图的无监督多视图特征选择模型进行了总结.

表 3 基于自适应图的无监督多视图特征选择模型

算法简称	特点	优缺点	时间复杂度
ACSL <sup>[71]</sup>	对样本的合成特征数据矩阵学习、从预先计算的图中学习共识图、施加秩约束以达到理想聚类结构	忽视了视图多样性; 预定义的相似性矩阵存在噪声	$O(T(dn^2 + n^2d + n^3m))$
CGWFS <sup>[74]</sup>	两级加权线性组合学习共识图、从预先计算的图中学习共识图	无参方法, 模型简单; 学习了局部和跨视图信息; 预定义的相似性矩阵存在噪声	—
ASVW <sup>[14]</sup>	完全自适应学习公共相似性图	易于扩展到其他多视图学习中; 忽略了多视图数据的多样性	$O\left(T \sum_{v=1}^m d_v(d_v^2 + nk_s v)\right)$
MFSGL <sup>[38]</sup>	无参策略学习视图权重、施加秩约束以达到理想聚类结构	无参数赋权方式减少了模型的参数; 忽视了视图多样性	$O\left(T \left(\sum_{v=1}^m d_v^3 + n^3\right)\right)$
DSS-UFS <sup>[10]</sup>	显式定义方法度量特征重要性、基于自表达策略自适应学习相似性矩阵	可以直接获得各特征的重要性; 忽视了视图多样性	$O\left(T \sum_{v=1}^m (d_v^2 n + n^2 d_v + d_v^3)\right)$

表3 基于自适应图的无监督多视图特征选择模型(续)

算法简称	特点	优缺点	时间复杂度
ASE-UMFS <sup>[68]</sup>	全局、局部结构学习、自表达的形式学习数据间的全局信息	自适应学习了局部结构和全局结构	$O\left(T \sum_{v=1}^m d_v(d_v^2 + nk_{s_v})\right)$
ME-ASOS <sup>[75]</sup>	全局、局部结构学习、多输出正则化投影学习共享子空间	自适应学习了局部结构和全局结构	$O\left(T \sum_{v=1}^m d_v(d_v^2 + nk_{s_v})\right)$
MAMFS <sup>[78]</sup>	多级投影子空间、视图内特征和连接特征联合学习、施加不相关约束以消除冗余特征	自适应学习了视图内和跨视图信息、局部结构和全局结构	$O(Tnd^2s + d^3 + n^2 \log(n))$
TLR-MVFS <sup>[81]</sup>	学习视图特定相似性矩阵、施加张量低秩正则化约束	学习了多视图数据一致性与多样性	$O(T(n^2 \log(n) + \sum_{v=1}^m d_v^3 s))$

### 2.3 小结

基于图的无监督多视图特征选择模型旨在挖掘数据特征与数据间图结构的关联性, 并以此来衡量数据特征的重要性, 从而选取能够保持数据间图结构的特征子集作为数据的新表达. 在这个框架下, 基于固定图模型和基于自适应图模型呈现不同的工作方式和特点.

基于固定图的模型为每个视图内的数据预先定义一个相似性矩阵, 用于描述数据在该视图内的结构关系. 在模型训练过程中, 这些相似性矩阵保持不变, 即图的构造过程独立于特征选择的过程. 这类模型的优势在于其模型简单、计算开销小、参数较少. 但由于需要提前构建图, 并且通常使用诸如  $k$ -近邻等方法直接在原始数据集上构建图, 因此这类模型容易受到噪声特征和异常值的影响. 在主流研究中, 基于固定图的策略逐渐被摒弃. 若能够有鲁棒且有效的构造图的方法, 这类模型的性能将有望得到进一步提升.

基于自适应图的模型则是在特征选择的过程中动态地构造和更新相似性矩阵, 即图的构造过程与特征选择的过程相辅相成. 这使得图的质量可以随着特征选择的过程不断提高, 更准确地反映数据间的真实关系以及更好地适应多视图数据的特点, 相比基于固定图的模型取得了更好的性能. 然而, 这也导致了模型的复杂性增加、计算开销加大、参数数量增多等问题.

## 3 基于伪标签的无监督多视图特征选择模型

聚类是探索数据底层(簇)结构的重要方法之一<sup>[82]</sup>. 其基本思想是根据某种准则对一组数据进行划分, 使相似的数据可以归入同一个簇, 不相似的数据划分到不同的类簇<sup>[83]</sup>. 根据多视图数据聚类的观点, 不同特征视图的样本标签应该是相同的, 且不依赖于不同的特征分布<sup>[54,84]</sup>. 因此, 基于伪标签的模型通常采用矩阵分解或谱分析来学习一个数据共识类标矩阵, 并利用线性回归或判别分析使投影后的数据趋近该类标矩阵, 以此来指导模型选择出能够保持数据聚类结构信息的特征子集. 根据学习伪标签的方式不同, 该类模型又可以分为: 直接学习数据共识类标的模型和基于视图特定的数据类标间接学习数据共识类标的模型. 基于伪标签的无监督多视图特征选择流程如图4所示.

### 3.1 直接学习数据共识类标

直接学习数据共识类标矩阵的模型遵循多视图聚类一致性的观点, 直接通过聚类算法为所有视图的数据学习一个共识类标矩阵指导特征选择.

为了提升视觉概念识别的准确性、处理社交网络媒体中多视图数据的高维度问题, Feng 等人<sup>[8]</sup>和 Tang 等人<sup>[6]</sup>率先提出了基于类标学习的无监督多视图数据特征选择模型 AUMFS 和 MVFS. 这两个模型先通过谱聚类获得多视图数据的共识伪标签  $F$ , 然后在伪标签空间中通过稀疏回归进行特征选择. 不同点在于: 1) AUMFS 将所有视图中的特征进行拼接形成数据的高维表示, 并通过该表示衡量数据与伪标签之间的关联性, MVFS 则是独立地在每个视图内进行特征重要性度量; 2) AUMFS 可以自适应学习视图的权重, 而 MVFS 需要预先设定视图权重, 并

在模型优化过程中保持该权重固定.

AUMFS 的目标函数如公式 (20) 所示:

$$\min_{W, F, \alpha} \text{Tr} \left( F^T \left( \sum_{v=1}^m \alpha_v L^{(v)} \right) F \right) + \lambda \|X^T W - F\|_{2,1} + \beta \|W\|_{2,1}, \text{ s.t. } \sum_{v=1}^m \alpha_v = 1, \alpha_v \geq 0, F^T F = I, F \geq 0 \quad (20)$$

其中,  $F \geq 0$  表示  $F$  的所有元素都是非负的,  $\alpha$  是视图权重向量.

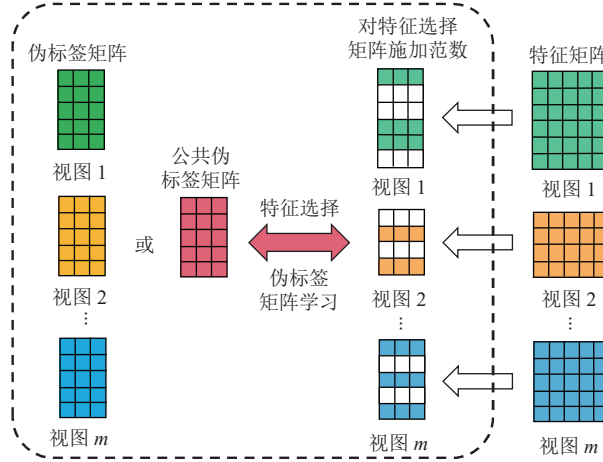


图 4 基于伪标签的无监督多视图特征选择流程

MVFS 的目标函数则如公式 (21) 所示:

$$\min_{W^{(v)}, F, \alpha} \lambda \sum_{v=1}^m \left( \left\| (X^{(v)})^T W^{(v)} - F \right\|_F^2 + \beta \|W^{(v)}\|_{2,1} \right) + \sum_{v=1}^m \alpha_v \text{Tr} (F^T L^{(v)} F), \text{ s.t. } F^T F = I, F \geq 0 \quad (21)$$

Shi 等人<sup>[44]</sup>基于迹比准则和谱聚类,提出了聚类结构和数据特征全局相关性保持的多视图特征选择模型 (CSP-UFS). 该模型利用判别分析,最小化投影数据的类内散列矩阵和总散列矩阵的迹的比值,使得同一类内的投影数据间的相似性更大,进而得到能够保留聚类结构的判别性特征子集. CSP-UFS 的目标函数如公式 (22) 所示:

$$\min_{W, F, \alpha} \frac{\text{Tr} (W^T \tilde{X} (I_n - FF^T) \tilde{X}^T W)}{\text{Tr} (W^T \tilde{X} \tilde{X}^T W)} + \lambda_1 \sum_{v=1}^m \alpha_v \text{Tr} (F^T L^{(v)} F) + \lambda_2 \|W\|_{2,1}, \text{ s.t. } F^T F = I, F \geq 0, \sum_{v=1}^m \alpha_v = 1, \alpha_v \geq 0 \quad (22)$$

其中,  $\tilde{X} = XH$  表示去中心的数据矩阵. Liu 等人<sup>[85]</sup>提出的判别多视图特征选择及融合模型与 CSP-UFS 相近.

基于伪标签的无监督多视图特征选择模型根据伪标签进行特征选择,因此伪标签的质量在很大程度上决定了所选特征的质量. Liu 等人<sup>[53]</sup>基于鲁棒的多视图 K-means 聚类算法<sup>[86]</sup>,提出了鲁棒的无监督多视图特征选择模型 (RMFS),高效地生成高质量的伪标签. RMFS 的目标函数如公式 (23) 所示:

$$\min_{W^{(v)}, F, C^{(v)}, G^{(v)}} \sum_{v=1}^m \alpha_v \left\| (X^{(v)})^T - FG^{(v)} \right\|_{2,1} + \sum_{v=1}^m \left\| (X^{(v)})^T W^{(v)} - FC^{(v)} \right\|_F^2 + \beta \sum_{v=1}^m \|W^{(v)}\|_{2,1}, \text{ s.t. } F \in \{0, 1\}^{n \times c} \quad (23)$$

其中,  $C^{(v)}$  表示  $X^{(v)} W^{(v)}$  和  $F$  之间的对齐矩阵,  $G^{(v)}$  是第  $v$  个视图数据  $X^{(v)}$  的质心矩阵.

在基于伪标签的模型中,常用的聚类方法之一——谱聚类通常需要预定义相似性矩阵,而原始多视图数据中普遍存在噪声特征和异常特征,这将导致预定义相似性矩阵不够准确. Bai 等人<sup>[87]</sup>提出非负结构化图学习的多视图特征选择模型 (NSGL),将公共局部几何结构学习和非负伪标签学习融合在一个框架中,有效地避免了原始特征空间中噪声特征的影响,提高了聚类的质量,从而提升了特征选择的效果. NSGL 的目标函数如公式 (24) 所示:

$$\begin{cases} \min_{W, F, S, p} \sum_{v=1}^m \left( p_v \sum_{i,j} \|x_i^{(v)} - x_j^{(v)}\|^2 S_{ij} + \alpha S_{ij}^2 \right) + 2\lambda \text{Tr} (F^T L_S F) + \beta \left( \|X^T W - F\|_F^2 + \|W\|_{2,1} \right) \\ \text{s.t. } \forall i, S_i^T \mathbf{1} = 1, 0 \leq S_{ij} \leq 1, \sum_{v=1}^m p_v = 1, 0 \leq p_v \leq 1, F^T F = I, F \geq 0 \end{cases} \quad (24)$$

其中,  $p_v$  是第  $v$  个视图的权重系数.

为了挖掘多视图数据的全阶相关性, TUFs<sup>[88]</sup> 借助张量表示来探索数据特征跨视图的高阶和低阶相关性信息. 与其他基于张量的方法不同的是, TUFs 无需显式构造张量, 从而更加高效和灵活. TUFs 的目标函数如公式 (25) 所示:

$$\min_{W^{(v)}, F, \alpha} \left\| \left( \prod_{v=1}^m * \left( Z^{(v)} \right)^T W^{(v)} \right) \right\|_{F} (W^{(m+1)})^T - F \left\|_{F}^2 + \gamma \sum_{v=1}^{m+1} \|W^{(v)}\|_{2,1} + \lambda \sum_{v=1}^m \alpha_v \text{Tr}(F^T L^{(v)} F), \text{ s.t. } F^T F = I, F \geq 0 \quad (25)$$

其中, 运算符  $*$  表示矩阵点乘,  $Z^{(v)} = [X^{(v)}; \mathbf{1}^T]$ .

Liang 等人<sup>[89]</sup> 同样运用张量方法, 提出了基于张量鲁棒主成分分析和共识图学习的无监督多视图特征选择模型 (TRCA-CGL). 首先, Liang 等人定义了每个视图的相似性矩阵  $P^{(v)}$ . 然而, 考虑到原始数据中可能存在噪声信息, TRCA-CGL 运用张量鲁棒主成分分析的方法将  $P^{(v)}$  分解为无噪声的相似性矩阵  $Z^{(v)}$  和显式定义的噪声表示矩阵  $E^{(v)}$ . 基于这一分解, 该模型借助张量低秩约束, 挖掘了多个视图之间的高阶关联性. 为了融合这些相似性矩阵的信息, TRCA-CGL 继而学习了一致性相似性矩阵  $S$ , 用于表示所有视图公共的局部几何结构. TRCA-CGL 的目标函数如公式 (26) 所示:

$$\begin{cases} \min_{Z, W, S, E, \alpha} \sum_{v=1}^m \alpha_v \|S - Z^{(v)}\|_F^2 + \|Z\|_{\otimes} + \gamma \|E\|_{2,1} + 2\lambda \text{Tr}(F^T L_S F) + \beta (\|XW - F\|_{2,1} + \delta \|W\|_{2,1}) \\ \text{s.t. } s_i^T \mathbf{1} = 1, 0 \leq S_{ij} \leq 1, F^T F = I, F \geq 0, \mathcal{P} = Z + E \end{cases} \quad (26)$$

其中,  $\mathcal{P}$ 、 $Z$  和  $E$  分别是由  $P^{(v)}$ 、 $Z^{(v)}$  和  $E^{(v)}$  堆叠而成的张量数据.

上述多视图特征选择研究假设所有视图都是完整的, 即每个样本在所有视图中都存在. 然而, 在现实应用中, 这种假设并非总是成立<sup>[90]</sup>. 现实场景中往往因为数据收集和整理的困难, 会出现数据不完整的情况, 导致一些视图中缺失若干实例. 例如, 在摄像头网络中, 摄像头会出现暂时失效或被某些物体遮挡等情况, 导致某些摄像头拍摄的实例缺失. 这种不完整视图会导致多视图数据分布扭曲和各视图之间数据的不一致性<sup>[65,90]</sup>. 针对不完全多视图数据, Huang 等人<sup>[46]</sup> 提出了基于互补和共识学习的不完整多视图无监督特征选择模型 ( $C^2$ IMUFS). 它首先对数据进行预处理, 用存在样本的均值对缺失样本进行填充, 再利用加权非负矩阵分解中的权重矩阵来区分存在样本和缺失样本. 同时,  $C^2$ IMUFS 从数据局部结构保持方面挖掘多视图数据信息, 采用了一种互补的自适应相似性学习方式, 通过线性组合其他视图的相似性矩阵来学习每个视图的相似性矩阵, 从而能够在不完整视图场景中获得完整的、互补的相似性图.  $C^2$ IMUFS 的目标函数如公式 (27) 所示:

$$\begin{cases} \min_{W^{(v)}, F, S^{(v)}, \alpha, U} \sum_{v=1}^m (\alpha^{(v)})^\gamma \{ \| (X^{(v)} - W^{(v)} F^T) \Omega^{(v)} \|_F^2 + \beta (\text{Tr}(F^T L_S^{(v)} F) + \|S^{(v)} - \sum_{i=1, i \neq v}^m S^{(i)} U_{i v}\|_F^2 + \|U\|_F^2 + \lambda \|W^{(v)}\|_{2,p}^2) \} \\ \text{s.t. } F \geq 0, F^T F = I, W^{(v)} \geq 0, 0 \leq \alpha^{(v)} \leq 1, \sum_{v=1}^m \alpha^{(v)} = 1, 0 \leq S^{(v)} \leq 1, (S^{(v)})^T \mathbf{1} = \mathbf{1}, S_{ii}^{(v)} = 0, 0 \leq U_{i v} \leq 1, \sum_{i=1, i \neq v}^m U_{i v} = 1, U_{v v} = 0 \end{cases} \quad (27)$$

其中,  $U$  是相似性矩阵的权重矩阵.

综上所述, 直接学习数据共识类标的模型的共同点是他们都基于多视图数据之间的聚类一致性, 为多个视图的数据学习一个统一的共识类标矩阵, 然后在该类标矩阵的指导下进行特征选择. 不同点如下.

1) 在聚类方面, AUMFS、MVFS、CSP-UFS、NSGL、TUFs 和 TRCA-CGL 均通过谱聚类学习公共类标矩阵. 但除了 NSGL 和 TRCA-CGL, 其余模型都需要预先构造相似性矩阵, 因而可能会受到噪声特征的影响进而得到不准确的结果; NSGL 和 TRCA-CGL 则将相似性矩阵学习和聚类结合到一个框架中, 尽可能地避免以上问题. RMFS 则采用鲁棒的多视图 K-means 聚类算法来生成高质量的数据类标矩阵, 具有计算效率高、时间复杂度低的优点.

2) 在多视图数据信息挖掘方面, a) 在多样性方面: AUMFS、MVFS、CSP-UFS 和 TUFs 预先构造视图特定的拉普拉斯矩阵来挖掘数据的多样性信息; RMFS 分别对每个视图的数据进行聚类和特征选择以利用多视图数据的

异构信息; TRCA-CGL 则自适应学习视图特定的相似性矩阵来保持准确的几何结构信息. 而 NSGL 仅专注于多视图数据的一致性, 从多个视图中学习公共的相似性矩阵, 忽略了不同视图的差异, 因此可能会导致次优效果. b) 在相关性方面, AUMFS、CSP-UFS 和 NSGL 将所有视图中的特征拼接成一个长向量来进行特征选择, 弱化了多视图数据的差异性, 可能无法得到具有判别性的特征子集. 虽然 MVFS 和 RMFS 在每个视图内衡量特征与公共类标矩阵的联系, 但他们仅在每个视图内衡量特征的重要性, 即不同视图的特征选择矩阵的优化是相互独立的, 忽略了多视图数据之间的相关性, 可能会错过一些重要特征而保留部分冗余特征. TUFS 和 TRCA-CGL 则利用张量方法挖掘了视图之间的高阶相关性, 有利于提高特征选择的性能.

3)  $C^2IMUFS$  考虑了现实中数据不完整的情况, 通过加权矩阵分解和均值填充来解决缺失样本的不平衡问题, 是无监督多视图特征选择应对非完整数据集挑战的尝试. 然而, 它采取的均值填充策略无法有效地恢复能够反映数据内在结构的缺失视图信息, 因此可能无法提升模型的性能, 这也是目前已有的非完整数据集无监督多视图特征选择模型存在的问题. 尽管如此,  $C^2IMUFS$  仍有值得借鉴之处: 它提出了一种新颖的互补学习策略, 能够自适应地从其他视图学习互补信息来估计每个视图的完整相似性图, 从而提高了聚类结果的准确性; 但需要注意的是, 它要求两个实例之间的相似关系至少存在于一个视图中, 因此在样本缺失率较高的场景下, 该模型可能无法有效地工作. 因此, 在处理非完整数据集方面, 未来的研究仍需要寻找缺失数据恢复或适用于高缺失率数据等特殊情况的模型或策略, 以克服数据不完整性带来的挑战.

表 4 对直接学习数据共识类标的无监督多视图特征选择模型进行了总结.

表 4 直接学习数据共识类标的无监督多视图特征选择模型

算法简称	特点	优缺点	时间复杂度
AUMFS <sup>[8]</sup>	视觉概念识别应用	对全体样本的特征数据矩阵学习, 弱化了视图多样性	$O(Tn(n+s)(n+d))$
MVFS <sup>[6]</sup>	社交媒体数据应用	需要预先设定视图权重	$O\left(T \sum_{v=1}^m \min(d_v^2 n, d_v n^2)\right)$
CSP-UFS <sup>[45,86]</sup>	采用迹比准则保持聚类结构	采用迹比准则保持聚类结构	$O\left(T \left(\sum_{v=1}^m n d_v^3 + n^2 d + n d^2 + d^3\right)\right)$
RMFS <sup>[53]</sup>	鲁棒多视图K-means学习伪标签	鲁棒K-means生成高质量伪标签	$O\left(T n c m(d + c m) + T \sum_{v=1}^m (d_v^3)\right)$
NSGL <sup>[87]</sup>	图学习与标签学习结合、施加秩约束以达到理想聚类结构	从原始数据空间中学习图, 无法准确反映数据之间的关系; 对全体样本的特征数据矩阵学习, 忽视了视图多样性	$O(T(d n^2 + \min(d n^2, n d^2)))$
TUFS <sup>[88]</sup>	引入跨视图信息度量、使用张量探索多视图数据的全阶交互	探索了多视图数据的全阶交互	$O\left(T \left(\sum_{v=1}^m s^3 d_v^2 n + s^3 d n^2 + n^2 c m^2\right)\right)$
TRCA-CGL <sup>[89]</sup>	图学习与标签学习结合、张量鲁棒主成分分析学习无噪声图、使用张量探索多视图数据的高阶关联性	探索了多视图数据的高阶关联性、多样性和一致性	$O(T(n^2 m^2 + d^3 + n^2 m \log(n)))$
$C^2IMUFS$ <sup>[46]</sup>	不完整视图处理、自适应相似性互补学习	挖掘了数据的一致性和多样性; 不适用于缺失率较高的情况	$O(T n \max(d_v, n) c)$

### 3.2 基于视图特定的数据类标间接学习数据的共识类标

基于视图特定的数据类标间接学习数据的共识类标的模型除了学习统一的共识类标矩阵的模型外, 还为每个视图学习了不同的数据类标矩阵. 在无监督多视图特征选择模型中, 一个关键的问题是如何综合利用数据在不同视图中的分布信息, 挖掘数据间潜在公共结构. 这类模型的通常做法是, 首先为每个视图学习特定的数据类标矩阵, 然后, 依据多视图数据聚类一致性的观点, 通过集成学习或协同正则化方式, 让每个视图的数据类标矩阵趋近于统一的共识类标矩阵. 它们从聚类结构方面学习多视图数据的多样性和一致性, 以更全面地挖掘不同视图中的

信息及公共结构.

Tang 等人<sup>[54]</sup>将嵌入的无监督特征选择模型 EUFS<sup>[91]</sup>扩展到多视图数据, 提出了集成指导的无监督多视图特征选择模型 (CGMV-UFS). 该模型对每个视图的数据进行非负矩阵分解和非负图分解, 得到每个视图的聚类指标矩阵  $F^{(v)}$ , 并让这些聚类指标矩阵趋近共识类标矩阵  $F_*$ . CGMV-UFS 的目标函数如公式 (28) 所示:

$$\begin{cases} \min_{W^{(v)}, F^{(v)}, F_*, \alpha, \beta, \gamma} \sum_{v=1}^m \left\| X^{(v)} - F^{(v)} (W^{(v)})^T \right\|_F^2 + \alpha_v \left\| F^{(v)} - F_* \right\|_F^2 + \gamma_v^2 \text{Tr} \left( (F^{(v)})^T L^{(v)} F^{(v)} \right) + \beta_v^2 \left\| W^{(v)} \right\|_{2,1} \\ \text{s.t. } (F^{(v)})^T F^{(v)} = I, F_*^T F_* = I, F_v \geq 0, F_* \geq 0, \alpha_v \geq 0, \beta_v \geq 0, \gamma_v \geq 0, \sum_{v=1}^m \alpha_v = 1, \sum_{v=1}^m \beta_v = 1, \sum_{v=1}^m \gamma_v = 1 \end{cases} \quad (28)$$

Hsieh 等人<sup>[12]</sup>提出的自适应协同正则化的无监督多视图特征选择模型 (ASCRA) 通过协同正则化显式度量特定视图聚类类标矩阵  $F^{(v)}$  和共识类标矩阵  $F_*$  之间的结构化差异, 从而可以处理视图之间更复杂的、非线性的关系. ASCRA 的目标函数如公式 (29) 所示:

$$\begin{cases} \sum_{v=1}^m \text{Tr} \left( (F^{(v)})^T L^{(v)} F^{(v)} \right) + \alpha \sum_{v=1}^m \left\| (X^{(v)})^T W^{(v)} - F_* \right\|_F^2 + \sum_{v=1}^m \frac{2}{p_v} \left( 1 - \frac{\text{Tr} \left( F^{(v)} (F^{(v)})^T F_* F_*^T \right)}{\sqrt{sc}} \right) + \beta \sum_{v=1}^m \left\| W^{(v)} \right\|_{2,1} \\ \text{s.t. } (F^{(v)})^T F^{(v)} = I, F_* \in \text{Ind}, p_v \geq 0, \sum_{v=1}^m p_v = 1 \end{cases} \quad (29)$$

Liu 等人<sup>[55]</sup>提出的通用无监督多视图特征选择 (gMUFS) 采用希尔伯特-施密特独立性准则  $HSIC$ <sup>[70]</sup> 最大化投影后的数据  $X^T W$  和集成伪标签矩阵  $F_*$  之间的依赖性, 因此能够挖掘聚类结构与数据分布之间比线性关系更复杂的相关性. gMUFS 的目标函数如公式 (30) 所示:

$$\begin{cases} \min_{W, F^{(v)}, F_*} \sum_{v=1}^m \text{Tr} \left( (F^{(v)})^T L^{(v)} F^{(v)} \right) - \beta \text{Tr} (K_F H K_L H) - \alpha \sum_{v=1}^m \text{Tr} \left( F^{(v)} (F^{(v)})^T F_* F_*^T \right) + \gamma \left\| W \right\|_{2,1} \\ \text{s.t. } (F^{(v)})^T F^{(v)} = I, F_*^T F_* = I, F^{(v)} \geq 0, F_* \geq 0 \end{cases} \quad (30)$$

其中,  $K_F, K_L$  分别为公共伪标签和特征选择后的数据对应的核矩阵.

Cao 等人<sup>[92]</sup>基于子空间学习, 提出了共识聚类结构指导的无监督多视图特征选择模型 (CCSFS). 与以往基于伪标签方法不同的是, 该模型在子空间学习框架中进行图分解来生成每个视图的数据类标矩阵  $F^{(v)}$ . 通过图分解,  $F^{(v)} (F^{(v)})^T$  可以表示每个视图内数据间的相似性, 且具有低秩的特性. CCSFS 的目标函数如公式 (31) 所示:

$$\begin{cases} \min_{F^{(v)}, F_*, W^{(v)}, \alpha} \sum_{v=1}^m \left\| X^{(v)} - X^{(v)} F^{(v)} (F^{(v)})^T \right\|_F^2 + \lambda \sum_{v=1}^m \alpha_v \left\| F^{(v)} (F^{(v)})^T - F_* F_*^T \right\|_F^2 + \gamma \sum_{v=1}^m \left\| W^{(v)} \right\|_{2,1} + \sum_{v=1}^m \beta \left\| (X^{(v)})^T W^{(v)} - F_* \right\|_F^2 \\ \text{s.t. } F^{(v)} \geq 0, F^{(v)} (F^{(v)})^T \mathbf{1} = \mathbf{1}, F_* \geq 0 \end{cases} \quad (31)$$

为了充分利用多视图数据的互补信息, 无监督多视图特征选择需要探索不同视图之间的相关性和交互作用. 现有的方法通常仅利用了每个视图内的信息, 而忽略了多个视图特征之间的相关性. 为此, CR-UMVFS<sup>[5]</sup>、CRV-DCL<sup>[93]</sup>和 CvLP-DCL<sup>[69]</sup>在无监督多视图特征选择模型中引入了多视图数据跨视图信息度量, 目的是挖掘多视图数据间的跨视图信息, 降低冗余特征的影响, 进而提高所选特征的判别性.

具体地, CR-UMVFS<sup>[5]</sup>采用谱聚类为每个视图内的数据学习视图特定的聚类表示, 并要求聚类表示相同的数据, 其投影向量也相同. CR-UMVFS 的目标函数如公式 (32) 所示:

$$\begin{cases} \min_{W^{(v)}, W^{(u)}, F^{(v)}} \alpha \left( \left\| (W^{(v)})^T X^{(v)} \left( (W^{(u)})^T X^{(u)} \right)^T - F^{(v)} (F^{(u)})^T \right\|_2^2 \right) + \alpha \left( \left\| (W^{(v)})^T X^{(v)} - F^{(v)} \right\|_2^2 + \lambda \left\| W^{(v)} \right\|_{2,1} \right) + \text{Tr} \left( F^{(v)} L^{(v)} (F^{(v)})^T \right) \\ \text{s.t. } F^{(v)} (F^{(v)})^T = I, F^{(u)} (F^{(u)})^T = I \end{cases} \quad (32)$$

相较于 CR-UMVFS, CRV-DCL<sup>[93]</sup>则采用了视图投影协同优化, 要求同一数据在不同视图内的投影向量尽量一致, 同一视图内相似数据的投影向量尽可能接近. 其次, 与之前将视图特定的类标矩阵与共识类标矩阵逼

近的方法不同, CRV-CDL 将每个视图的数据投影到一个共识的语义标签空间中, 并将该空间分为一致性部分  $F_*$  和多样性部分  $F^{(v)}$ , 以此挖掘不同视图的共性信息和区分不同视图的特性信息. CRV-DCL 的目标函数如公式 (33) 所示:

$$\begin{cases} \min_{W^{(v)}, F^{(v)}, F_*} \sum_{v=1}^m \left\| (X^{(v)})^T W^{(v)} - (F_* + F^{(v)}) \right\|_F^2 + \beta \sum_{v=1}^m \sum_{u=1}^m \text{Tr} \left( (W^{(v)})^T X^{(v)} L (X^{(u)})^T W^{(u)} \right) + \alpha \sum_{v=1}^m \left( \|W^{(v)}\|_{2,1} + \|F^{(v)}\|_1 \right) \\ \text{s.t. } F_*^T F_* = I, F_* \geq 0 \end{cases} \quad (33)$$

其中, 对多样性部分  $F_*$  施加  $l_1$  范数, 可以最小化不同视图的噪声信息.  $L$  是由同一视图内的和不同视图之间的拉普拉斯矩阵组合成的大矩阵, 定义如公式 (34) 所示:

$$L = \begin{bmatrix} L^{11} & L^{12} & \dots & L^{1m} \\ L^{21} & L^{22} & \dots & L^{2m} \\ \vdots & \vdots & \ddots & \vdots \\ L^{m1} & L^{m2} & \dots & L^{mm} \end{bmatrix} \in \mathbf{R}^{m \times mn} \quad (34)$$

因此, 通过公式 (33) 中的第 2 项, CRV-DCL 可以挖掘并保持同一视图内的和跨视图的局部几何结构信息.

但是, CRV-DCL 将所有视图视为同等重要. 这虽然避免了权值的计算, 但削弱了模型对异构特征的区分能力. 在 CRV-DCL 基础之上, CvLP-DCL<sup>[69]</sup> 从不同视图的相似性图中学习一个协作相似性图, 保留了多视图数据的一致性几何结构信息. 此外, 它用数据跨视图的关联关系来自适应学习不同视图的权重, 以避免两个高度相似的视图同时被赋予较大的权值, 从而降低所选特征的冗余性. CvLP-DCL 的目标函数如公式 (35) 所示:

$$\begin{cases} \min_{W^{(v)}, F^{(v)}, F_*, S, \gamma} \sum_{v=1}^m \left\| (X^{(v)})^T W^{(v)} - (F_* + F^{(v)}) \right\|_F^2 + \lambda \text{Tr} (F_*^T L_S F_*) + \alpha \left\| S - \sum_{v=1}^m \gamma^{(v)} S^{(v)} \right\|_F^2 \\ + \frac{\beta}{2} \gamma^T M \gamma + \sum_{v=1}^m \left( \gamma^{(v)} \|F^{(v)}\|_1 + (1 - \gamma^{(v)}) \|W^{(v)}\|_{2,1} \right) \\ \text{s.t. } F_*^T F_* = I, F_* \geq 0, M_{pq} = \text{Tr} (S^p)^T S^q, \gamma^T \mathbf{1} = \mathbf{1}, \gamma^{(v)} > 0, \bar{S} \mathbf{1} = \mathbf{1}, \bar{S}_{ij} \geq 0 \end{cases} \quad (35)$$

其中,  $\gamma$  是视图权重向量,  $M$  是用于衡量不同视图间相关性的矩阵.

基于视图特定的数据类标间接学习数据共识类标的模型需要学习各视图的类标矩阵以及多视图数据的共识类标矩阵, 从数据的聚类表示方面同时挖掘多视图数据间的多样性和一致性, 选择具有判别性的、低冗余的特征子集. 此外, 这些模型之间还存在不同之处.

1) CGMV-UFS、ASCRA、gMUFS 和 CCSFS 通过不同的方法来度量视图特定的类标矩阵与共识类标矩阵之间的差异, 表现出了不同的优势: CGMV-UFS 能够自动学习模型中用于平衡不同项的参数, 减少了模型训练的成本; ASCRA 采用的协同正则化可以处理视图之间更复杂的、非线性的关系; gMUFS 中运用的希尔伯特-施密特独立性准则能够探索共识聚类结构与所选择的特征之间的非线性相关性; CCSFS 提出了一种新颖的数据类标学习方式, 可由数据类标矩阵直接得到低秩图, 便于进行后续基于图的学习任务. 但 CGMV-UFS、ASCRA 和 CCSFS 独立衡量每个视图内的投影后的特征与公共数据类标矩阵的关联, 忽略了多个视图的特征之间的相关性. gMUFS 通过公共的特征选择矩阵直接从所有视图连接后的数据中选择特征, 同样无法挖掘多视图数据中存在的相关性, 而且它将所有视图同等对待, 忽略了不同视图的重要性, 降低了模型性能.

2) 为了充分利用多视图数据的互补性、相关性和交互作用, CR-UMVFS、CRV-DCL 和 CvLP-DCL 在模型中引入了多视图数据的跨视图信息度量. 其中, a) CR-UMVFS 通过伪标签使得视图之间的投影向量相互接近, 但它没有学习共识类标矩阵, 也没有使用其他的学习方式来提取多视图数据之间的一致性信息, 而由于异质的多视图数据在不同视图内特征的差异性, 这种跨视图信息度量方式亦无法准确地表示多视图数据之间的一致性信息. 同时, CR-UMVFS 也没有区分不同视图重要性的能力, 因此可能无法选择出具有判别性的特征. b) CRV-DCL 和 CvLP-DCL 与以往的方法不同, 它们直接将投影后的数据划分为一致性表示和多样性表示, 以此挖掘不同视图的共性信息和区分不同视图的特性信息. 两者相比, CRV-DCL 将所有的视图特定的相似性图组合成一个较大的矩阵



进行聚类, 不仅加重了计算负担, 还增加了模型对预定义相似性矩阵的依赖; 与此同时, 它将所有视图一视同仁, 削弱了模型对异构特征的区分能力. CvLP-DCL 则在 CRV-DCL 的基础上用数据跨视图的关联关系来自适应学习不同视图的权重, 并为多视图数据学习了共识相似性矩阵, 保持了多视图数据的一致性几何结构信息.

表 5 总结了基于视图特定的数据类标间接学习数据共识类标的无监督多视图特征选择模型进行.

表 5 基于视图特定的数据类标间接学习数据共识类标的无监督多视图特征选择模型

算法简称	特点	优缺点	时间复杂度
CGMV-UFS <sup>[54]</sup>	自动学习平衡参数	自动学习参数减少模型训练的成本	$O\left(T\left(\sum_{v=1}^m ncd_v + n^2c\right)\right)$
ASCRA <sup>[12]</sup>	协同正则化学习	能够处理视图之间更复杂的、非线性的关系	$O\left(T\left(n^2s + \sum_{v=1}^m d_v^3\right)\right)$
gMUFS <sup>[55]</sup>	利用独立性准则挖掘非线性关系	挖掘了聚类结构与数据分布之间的复杂相关性; 无法区分不同视图的重要性	$O\left(T\left(cn^2 + dnc + cd^2 + d^3\right)\right)$
CCSFS <sup>[92]</sup>	子空间学习、图分解学习数据类标矩阵、学习特定视图的和公共的类标矩阵	挖掘了数据的一致性和多样性; 由数据类标可生成低秩图	$O\left(T\sum_{v=1}^m d_v^3\right)$
CR-UMVFS <sup>[5]</sup>	跨视图信息度量、肿瘤亚型识别应用	挖掘了数据的多样性; 忽略了数据的一致性; 无法区分不同视图的重要性	$O\left(T\left(\sum_{v=1}^m (nd_v^2 + d_vn^2) + n^3\right)\right)$
CRV-DCL <sup>[93]</sup>	跨视图信息度量	挖掘了数据的一致性和多样性; 无法区分不同视图的重要性; 合成的相似性矩阵加重计算负担	$O\left(T\sum_{v=1}^m (n\max(d_v, n)n + d_v^3)\right)$
CvLP-DCL <sup>[69]</sup>	跨视图信息度量, 使用数据关系自适应学习视图权重	挖掘了数据的一致性和多样性; 使用数据关系自适应学习视图权重	$O\left(T\sum_{v=1}^m (n\max(d_v, n)n + d_v^3)\right)$

### 3.3 小 结

基于伪标签的无监督多视图特征选择模型通过非负矩阵分解或谱聚类等聚类方法来获取数据的伪类标信息, 并将其作为数据的内在表示. 根据学习伪标签方式的不同, 可以将这些模型分为直接学习数据共识类标的模型和基于视图特定的数据类标间接学习数据的共识类标的模型. 这两类模型存在以下差异.

直接学习数据共识类标的模型遵循多视图聚类一致性的观点, 直接从多视图数据中学习一个共同的类标矩阵, 并将其作为多视图数据的一致性表示. 这类模型专注于多视图数据的一致性, 有助于减少视图间的差异性, 从而筛选出一致性较强的特征. 在学习数据的多样性和相关性方面, 部分模型如 TRCA-CGL 和 TUFs, 通过视图特定的图表示或张量方法挖掘信息, 而一些模型如 AUMFS 和 NSGL, 则忽略或弱化了视图间的多样性, 可能导致信息的丢失或冗余.

基于视图特定的数据类标间接学习数据的共识类标的模型从聚类结构方面学习多视图数据的多样性和一致性. 具体而言, 这类模型通常先在每个视图的特征空间中对数据进行聚类, 得到视图特定的数据类标, 然后将这些视图特定的类标融合, 形成数据的共识类标. 他们的优势在于可以充分利用视图间的多样性, 保留更多的视图内信息. 值得注意的是, 一些模型如 CvLP-DCL, 通过跨视图度量挖掘了多视图数据的互补性、相关性和交互作用, 展现了强大的学习能力. 然而, 个别模型如 gMUFS, 则在区分不同视图的重要性方面存在不足, 相对于其他模型还有待改进.

此外, 我们可以观察到: 大多数基于伪标签的无监督多视图特征选择模型采用了谱聚类方法或结合了图学习来指导特征选择. 尽管这些方法在标签学习的基础上引入了图方法来挖掘数据的潜在流形结构, 但也涉及了相似性矩阵的学习或预处理过程, 增加了模型对相似性矩阵的依赖性, 存在和基于固定图的方法类似的问题: 1) 预先构造的相似性矩阵会受到噪声特征的影响; 2) 通过手动定义参数构造的相似性矩阵可能不够准确, 从而导致模型性能不佳.

#### 4 基于数据流的无监督多视图特征选择模型

在传感器网络异常检测、信用卡欺诈行为监测、天气预报和互联网用户日志分析等众多实际应用问题中,数据大都以流的形式不断产生.这些数据往往是不不断动态变化的,并且随着时间的推移不断增长,其体量往往无限,具有这些特征的数据被称为流式数据或数据流<sup>[94,95]</sup>.传统的数据流挖掘与分析方法一般假设数据独立同分布.事实上,在真实的数据流环境中,数据分布往往会随着时间而发生改变,例如,天气预报所依据的规律可能会随着季节变化而发生改变.一般地,把这种数据流中的数据分布随着时间以某种方式发生不可预知变化的现象称为概念漂移<sup>[94,95]</sup>.近年来,针对数据流的挖掘研究引起了越来越多研究者的关注,人们在这方面也取得了长足的进步<sup>[96,97]</sup>,其中不乏基于数据流的无监督多视图特征选择方法.

为了应对大规模数据、概念漂移及在数据流情况下利用多视图数据的一致性和互补性信息这3个难题,基于数据流的无监督多视图特征选择模型在以往方法的基础上,利用在线学习方式和缓存技术逐块处理多视图数据.这种方法可以在没有明显的性能牺牲的前提下,大大降低存储成本,加快数据处理速度并将无监督多视图特征选择模型扩展到大数据应用<sup>[65,98]</sup>.

Shao 等人<sup>[98]</sup>首次提出了在线无监督多视图特征选择模型 (OMVFS).该模型在嵌入式无监督特征选择模型 EUFS<sup>[91]</sup>的基础上,将无监督特征选择嵌入到基于非负矩阵分解及谱聚类的联合聚类框架中.OMVFS 的目标函数如公式 (36) 所示:

$$\min_{W^{(v)}, F_{[t]}} \sum_{v=1}^m \|X_{[t]}^{(v)} - F_{[t]}(W^{(v)})^T\|_F^2 + \alpha_v \text{Tr}(F_{[t]}^T L_{[t]}^{(v)} F_{[t]}) + \beta_v \|W^{(v)}\|_{2,1}, \text{ s.t. } F_{[t]}^T F_{[t]} = I, F_{[t]} \geq 0, W^{(v)} \geq 0 \quad (36)$$

其中,  $X_{[t]}^{(v)}$  表示截至时间  $t$  时接收到的第  $v$  个视图的数据,  $F_{[t]}$  是截至时间  $t$  时收到的所有实例的集成聚类指标矩阵,  $L_{[t]}^{(v)}$  是由  $X_{[t]}^{(v)}$  构造的拉普拉斯矩阵.公式 (36) 在更新  $F_{[t]}$  和  $W^{(v)}$  的时候,需要使用到所有的数据,这有悖于在线学习模型的特点.由于流数据中新的数据块与邻近的数据块的关联性较它与更早的数据块之间的关联性更紧密,所以在更新  $F$  的时候,OMVFS 仅考虑临近的  $s$  个时间片段内的数据块及它们的伪标签.而在更新  $W^{(v)}$  的时候,OMVFS 通过等价变换,将  $W^{(v)}$  的更新转换为仅关于  $t-1$  和  $t$  这两个时间片段的数据及伪类标的形式.

针对不完整多视图数据流, Xu 等人<sup>[65]</sup>提出了不完整多视图数据流特征选择模型 (CVFS).CVFS 是第 1 个可以处理非完整数据集的无监督多视图特征选择模型.它在 OMVFS<sup>[98]</sup>的基础上,使用存在数据的均值作为缺失数据的估计值,并利用加权非负矩阵分解来衡量缺失样本和非缺失样本之间的差异.同时, CVFS 通过希尔伯特-施密特独立性准则衡量每个视图特征选择的独立性,以选择出能够充分保留异构特征多样性的、低冗余的特征子集.CVFS 的目标函数如公式 (37) 所示:

$$\begin{cases} \min_{W^{(v)}, F_{[t]}} \sum_{v=1}^m \left\| \Omega^{(v)} \left( \left( X_{[t]}^{(v)} \right)^T - F_{[t]} \left( W^{(v)} \right)^T \right) \right\|_F^2 + \beta \sum_{v=1}^m \text{Tr} \left( F_{[t]}^T L_{[t]}^{(v)} F_{[t]} \right) + \alpha \sum_{v=1}^m \|W^{(v)}\|_{2,1} + \gamma \sum_{v=1}^m \sum_{z=1, z \neq v}^m \text{Tr} \left( W^{(v)} H K^{(z)} H \left( W^{(z)} \right)^T \right) \\ \text{s.t. } F_{[t]}^T F_{[t]} = I, F_{[t]} \geq 0, W^{(v)} \geq 0 \end{cases} \quad (37)$$

其中,  $K^{(v)} = (W^{(v)})^T W^{(v)}$ , 对角阵  $\Omega^{(v)}$  是数据权重矩阵.

Huang 等人<sup>[99]</sup>在 CVFS 的基础上,提出了针对不完全多视图数据的动态增量无监督特征选择 ( $I^2$ MUFS).类似地,这个模型将特征选择嵌入到加权非负矩阵分解模型和谱聚类的联合聚类框架中.不同的是,  $I^2$ MUFS 从数据嵌入表示方面考虑多视图数据的多样性和一致性,为每个视图学习一个数据类标矩阵,然后使他们尽可能接近共识聚类指标矩阵.  $I^2$ MUFS 的目标函数如公式 (38) 所示:

$$\begin{cases} \min_{W^{(v)}, F_{[t]}^{(v)}, F_{s[t]}^{(v)}, \alpha} \sum_{v=1}^m \left( \left( \alpha^{(v)} \right)^\lambda \left\| \left( X_{[t]}^{(v)} - W^{(v)} \left( F_{[t]}^{(v)} \right)^T \right) \Omega^{(v)} \right\|_F^2 + \eta^{(v)} \|W^{(v)}\|_{2,1} + \beta^{(v)} \left\| \Omega^{(v)} \left( F_{[t]}^{(v)} - F_{s[t]}^{(v)} \right) \right\|_F^2 + \theta^{(v)} \text{Tr} \left( \left( F_{[t]}^{(v)} \right)^T L_{[t]}^{(v)} F_{[t]}^{(v)} \right) \right) \\ \text{s.t. } F_{[t]}^{(v)} \geq 0, W^{(v)} \geq 0, F_{s[t]}^{(v)} \geq 0, \left( F_{[t]}^{(v)} \right)^T F_{[t]}^{(v)} = I, \alpha^{(v)} \geq 0, \sum_{v=1}^m \alpha^{(v)} = 1 \end{cases} \quad (38)$$

其中,  $\theta^{(v)}$ 、 $\beta^{(v)}$  和  $\eta^{(v)}$  是权衡参数.

基于数据流的模型将无监督多视图特征选择应用到大规模数据和数据流上.目前,基于数据流的或针对非完

整数据集的模型还较少, 已有的几个模型采用的方法和基于伪标签的方法类似, 利用非负矩阵分解和谱聚类对数据进行分析 and 选择. 这几种模型之间还存在以下差别.

1) OMVFS 首次提出了利用在线方式进行无监督多视图特征选择, 加快了大规模数据的计算处理速度, 降低了存储成本. 但在多视图信息学习方面, OMVFS 仅学习了一个公共的数据类标矩阵, 然后在此基础上令每个视图的特征选择矩阵独立优化, 从而无法利用多视图数据的多样性和相关性. 此外, 它采用了预定义视图权重并使其在模型优化过程中保持固定, 因此可能无法合理地评估不同视图的重要性.

2) CVFS 和  $I^2$ MUFS 与  $C^2$ IMUFS 类似, 都是通过加权矩阵分解和均值填充来解决缺失样本的不平衡问题. 同样地, 它们并没有针对缺失视图信息的恢复提供有效的解决方案, 从而可能限制了模型性能的提升. 除了以上共同点外, CVFS 和  $I^2$ MUFS 利用不同的方法对 OMVFS 进行了改进, 并在一些数据集上证明了其有效性: CVFS 通过希尔伯特-施密特独立性准则衡量每个视图的特征选择矩阵的独立性, 降低了所选择的特征的冗余性; 但它直接丢弃了视图权重, 忽略了视图间信息不平衡的负面因素, 导致它同样无法挖掘多视图数据的多样性.  $I^2$ MUFS 同时学习了每个视图的数据类标矩阵和共识类标矩阵以从聚类角度来保持数据的多样性和一致性, 并通过自适应学习视图权重来评估不同视图的重要程度; 但它也没有考虑到多视图数据之间的相关性.

表 6 对基于数据流的无监督多视图特征选择模型进行了总结.

表 6 基于数据流的无监督多视图特征选择模型

算法简称	特点	优缺点	时间复杂度
OMVFS <sup>[98]</sup>	大规模和流数据处理	降低了计算和存储成本, 能够扩展到大数据应用	$O(Tmndc)$
CVFS <sup>[65]</sup>	大规模和流数据处理	挖掘了数据的一致性和多样性; 降低了计算和存储成本, 能够扩展到大数据应用	$O(Tmndc)$
$I^2$ MUFS <sup>[99]</sup>	大规模和流数据处理, 不完整视图处理	挖掘了数据的一致性和多样性; 降低了计算和存储成本, 能够扩展到大数据应用	$O\left(T\left(\sum_{v=1}^m(d_v c + n^2 c + d_v n c) + n c m\right)\right)$

## 5 实验分析

本节介绍了无监督多视图特征选择研究中常用的公开数据集和评价标准. 为了更深入地评估无监督多视图特征选择模型的性能, 我们从已介绍的模型中挑选了 9 个, 在这些数据集上进行了特征选择和聚类分析.

### 5.1 公开数据集

无监督多视图特征选择研究的常用的公开数据集包括但不限于 Mfeat、MSRC-v1、Caltech101-7、Outdoor Scene、WebKB、BBCSport 和 Yale. 这些数据集的详细信息如表 7.

表 7 基准数据集的详细信息 (特征类型 (维度))

数据集	视图1	视图2	视图3	视图4	视图5	视图6	样本数	聚类数
Mfeat	FAC (216)	FOU (76)	KAR (64)	MOR (6)	PIX (240)	ZER (47)	2 000	10
MSRC-v1	CM (24)	HOG (576)	GIST (512)	LBP (256)	CENTRIST (1 302)	—	210	7
Caltech101-7	CMT (48)	HOG (100)	GIST (512)	LBP (256)	CENTRIST (1 302)	DoG-SIFT (441)	441	7
Outdoor Scene	CM (432)	HOG (256)	GIST (512)	LBP (48)	—	—	2 688	8
WebKB	Text (1 703)	Link (230)	Text (230)	—	—	—	203	4
BBCSport	Text (3 183)	Text (3 203)	—	—	—	—	544	5
3Sources	Text (3 560)	Text (3 631)	Text (3 068)	—	—	—	169	6
Yale	Intensity (4 096)	LBP (3 304)	Gabor (6 750)	—	—	—	165	15

(1) Mfeat 数据集<sup>[71]</sup>包含 2 000 个手写数字 (0–9). 每个数字有 200 种模式. 这些数字由 6 种的特征表示, 包括 FOU、FAC、KAR、PIX、ZER 和 MOR.

(2) MSRC-v1 数据集<sup>[71]</sup>包含 7 个类别的 210 张图像. 每张图像由 5 视觉描述符表示, 分别是 CM、HOG、

GIST、LBP 和 CENTRIST.

(3) Caltech101-7 数据集<sup>[14]</sup>包含 7 个类别的 441 张图像. 每张图像由 6 种视觉描述符表示, 包括 CMT、HOG、GIST、LBP、CENTRIST 和 DoG-SIFT.

(4) Outdoor Scene 数据集<sup>[71]</sup>包含 8 个主题的 2 688 张图像. 每张图像由 4 种视觉描述符表示, 即 CM、LBP、HOG 和 GIST.

(5) WebKB 数据集<sup>[100]</sup>由 4 个类别的 203 个页面组成. 每个页面分为 3 个视图, 即页面文本内容、超链接文本和标题文本. 这个数据集存在视图不充分问题, 因为数据在超链接和标题这两个视图的特征表示不足以描述数据的分布特点.

(6) BBCSport 数据集<sup>[54]</sup>包含 544 篇体育文章. 这些文章涵盖了 5 个不同主题, 即田径、板球、足球、橄榄球和网球. 每篇文章分为两段, 分别有 3 183 个字和 3 203 个字.

(7) 3Sources 数据集<sup>[81]</sup>包含从 BBC、路透社和卫报收集来的 169 条新闻. 每条新闻属于以下 6 个主题之一: 商业、娱乐、健康、政治、体育和技术.

(8) Yale 数据集<sup>[101]</sup>是一个流行的人脸图像数据集. 它由 165 张灰度图像组成, 涵盖了 15 个人物, 每个人物各有 11 张图像. 每张图像由 3 种描述符表示: Intensity、LBP 和 Gabor.

## 5.2 评价标准

大多数无监督多视图特征选择研究通常将全体样本的特征矩阵  $X$  (即将所有视图中的数据特征连接形成的矩阵) 作为基准 (Baseline), 并通过对  $X$  和模型选择出的特征子集进行 K-means 聚类来评估模型的性能<sup>[65,68,89]</sup>. 聚类准确性 (clustering accuracy,  $Acc$ ) 和归一化互信息 (normalized mutual information,  $NMI$ ) 是两种常用于评估聚类性能的指标, 这两个指标的值越大, 表明聚类性能越好, 即所选择的特征越具有判别性<sup>[78,81,92]</sup>.

聚类准确性 ( $Acc$ ) 表示数据中正确聚类的实例的比例, 其计算公式如公式 (39) 所示:

$$Acc = \frac{1}{n} \sum_{i=1}^n \delta(r_i, \text{map}(p_i)) \quad (39)$$

其中,  $r_i$  和  $p_i$  分别表示第  $i$  个实例的真实类别和预测聚类标签,  $\text{map}(\cdot)$  是将每个聚类标签映射到真实类别的置换函数,  $\delta(x, y)$  是符号函数. 当  $x=y$  时,  $\delta(x, y)$  等于 1, 否则等于 0.

归一化互信息 ( $NMI$ ) 衡量一组真实标签与一组预测聚类标签之间的相似性. 令  $C$  表示真实标签集,  $\mathcal{L}$  表示聚类标签集,  $NMI$  的计算公式如公式 (40) 所示:

$$NMI(C, \mathcal{L}) = \frac{MI(C, \mathcal{L})}{0.5(H(C) + H(\mathcal{L}))} \quad (40)$$

其中,  $H(C)$  和  $H(\mathcal{L})$  分别表示  $C$  和  $\mathcal{L}$  的熵,  $MI(C, \mathcal{L})$  是互信息函数, 用于衡量两个分布  $C$  和  $\mathcal{L}$  之间的相关性.

## 5.3 实验设置与结果分析

我们从已介绍的模型中挑选了 9 个模型, 分别是 ASVW、ACSL、CGMVFS、ASE-UMFS、TUFS、CvLP-DCL、TLR-MVFS、CCSFS 和 TRCA-CGL, 并根据这些模型在第 5.1 节中介绍的数据集上的特征选择结果进行评估.

在实验中, 我们按照每个模型的作者的建议进行参数调整. 其次, 鉴于如何确定模型所选特征的数量一直是一个难以解决的难题<sup>[14]</sup>, 我们在所有数据集上都选取 100、200、300、400 和 500 个特征, 以观察在不同特征数量下模型的性能变化. 同时, 我们采用不同的随机初始化重复实验 20 次, 并取其聚类结果的平均值来评价模型性能, 以降低随机初始化的影响. 实验结果如图 5 和图 6 所示, 它们分别展示了在不同特征选择数量情况下模型的  $Acc$  和  $NMI$  结果. 通过对实验结果深入分析, 我们可以得到以下结论.

(1) 总体性能方面: 在综合考虑多个数据集时, 很难确定一个绝对的最佳模型. 然而, 通过仔细观察实验结果, 我们发现 ACSL、CCSFS 和 TRCA-CGL 在几乎所有数据集上选出的特征子集都能显著地提升 K-means 的聚类性能, 体现了这些模型对不同类型数据的鲁棒性. 相较之下, CvLP-DCL 在 Yale 数据集上选出的特征子集对聚类性能提升相对不足, 但在其余数据集上选出的子集均呈现出出色的可辨识度. 尤其在 BBCSport 数据集上的选出 300 个特征时, CvLP-DCL 选出的特征子集在聚类后的  $Acc$  和  $NMI$  指标值分别接近 0.8 和 0.7, 达到了最高值.

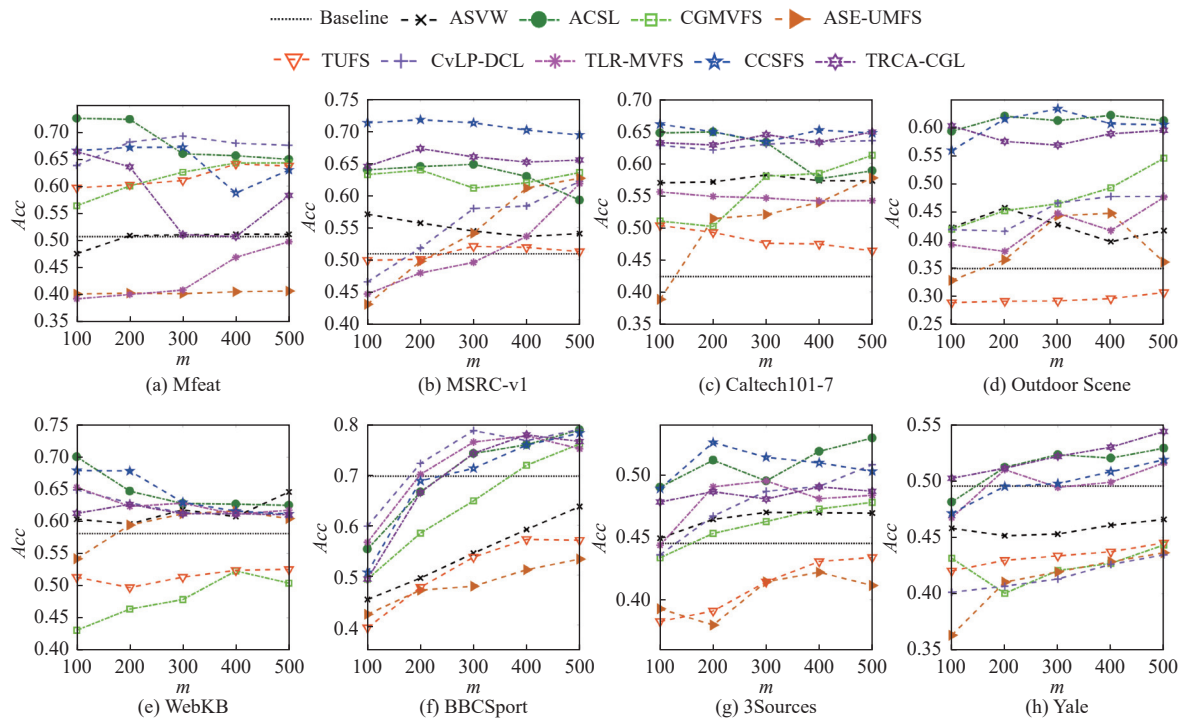


图5 特征选择模型在不同特征选择数量下的  $Acc$  值

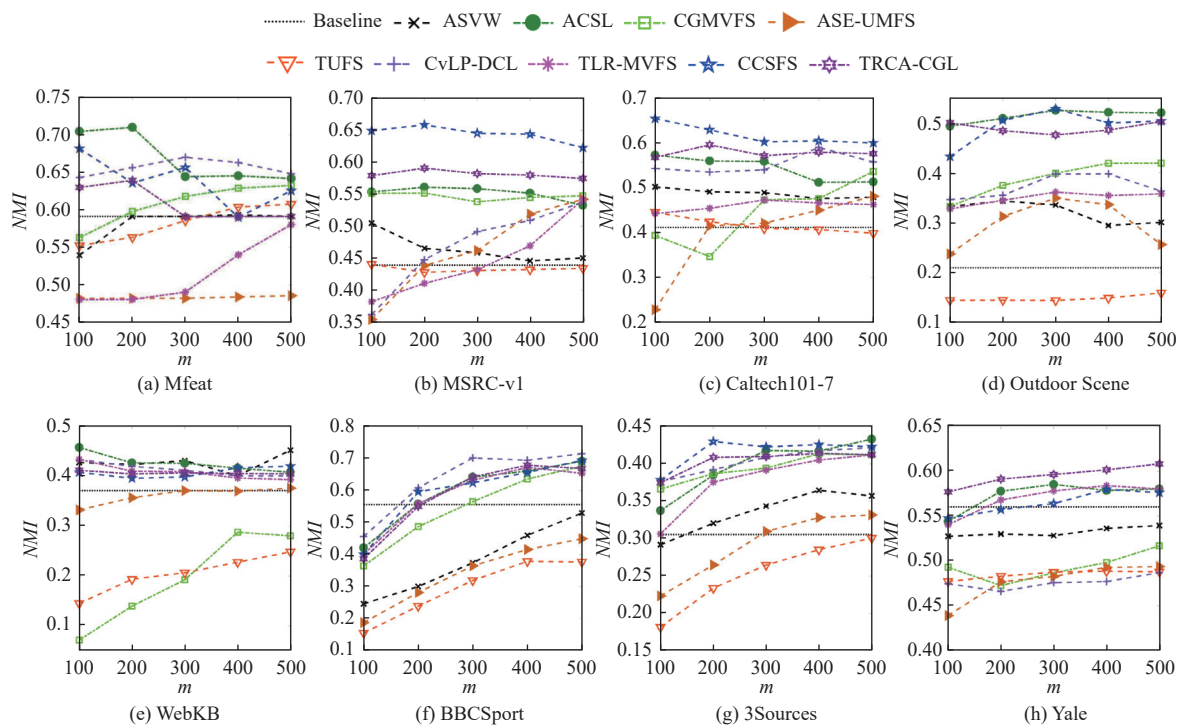


图6 特征选择模型在不同特征选择数量下的  $NMI$  值

对于其他模型,包括 ASVW、CGMVFS、ASE-UMFS、TUFS 和 TLR-MVFS,它们在不同数据集上选出的特征子集的辨识性存在明显差异. 它们在某些数据集上表现良好,但在个别数据集上相对不足. 具体而言,ASVW 和 CGMVFS 在 Yale 数据集上选出的特征子集的聚类 *Acc* 和 *NMI* 指标值明显低于基准水平. 然而,除了 CGMVFS 在 WebKB 数据集上和 ASVW 在 BBCSport 数据集上的表现之外,ASVW 和 CGMVFS 在其余数据集上选择的特征子集仍提升了聚类性能. TLR-MVFS 的情况类似,除 Mfeat 数据集外,在其余数据集上的表现较好. ASE-UMFS 则相对较弱,在 4 个数据集 (MFeat、BBCSport、3Sources 和 Yale) 上选出的特征子集性能较差,但在其余数据集上性能仍然可观. 然而, TUFS 表现最差,在 5 个数据集上,包括 Outdoor Scene、WebKB、BBCSport、3Sources 和 Yale,选出的特征子集的聚类性能均未达到基准水平.

(2) 与所选特征数量的关系方面: 我们可以观察到,在 Mfeat、MSRC-v1、Caltech101-7、Outdoor Scene 和 WebKB 数据集上,当所选特征的数量不超过 300 时,ASVW、ACSL、CvLP-DCL、CCSFS 和 TRCA-CGL 选出的特征子集能够获得出色的聚类表现. 这表明无监督多视图特征选择模型能够从数据中选择出具有判别性和代表性的特征,有效地降低数据的维数,提高聚类任务的性能. 在剩余的数据集上 (BBCSport、3Sources 和 Yale),尽管 ASVW、ACSL、CvLP-DCL、CCSFS 和 TRCA-CGL 选出的特征子集的聚类性能总体随着特征数量的增加而提高,但仍能在选择的特征数量为 500 时表现出优秀的效果.

此外,值得注意的是,在视图不充分的数据集 WebKB 上,ASVW、ACSL、CvLP-DCL、TLR-MVFS、CCSFS 和 TRCA-CGL 这几个模型选择出的特征子集仍改善了聚类效果,体现了无监督多视图特征选择模型的强大信息综合及数据挖掘能力.

综上所述,选择特征选择模型时,应考虑数据的特性以及所需的特征数量等因素. 此外,在大数据时代,除了聚类性能之外,特征选择的其他特性,如可扩展性、稳定性、安全性等,也同样值得关注.

## 6 应用动态

无监督多视图数据特征选择是一种在多视图数据分析中非常重要的预处理方法或知识发现工具,它可以有效利用不同视图数据之间的互补性和一致性,从高维数据集中选择出最具代表性的特征子集,以提高后续数据分析和学习的效果<sup>[78,88]</sup>. 本节对无监督多视图特征选择的热点应用进行了整理.

(1) 多媒体检索: 网络技术的迅速发展产生了越来越多的高维多视图数据集. 由此引发的维度灾难和多视图数据分析问题给现有的多媒体检索与理解技术带来了挑战<sup>[5,7]</sup>. 为了解决这一问题,无监督多视图特征选择成为一项关键技术,它可以从文本、图像、音频、视频等多媒体高维数据中选择出能够描述其内容、风格、情感等方面的代表性特征,从而支持相似性度量、分类、聚类等任务<sup>[7,8]</sup>. 例如,Gui 等人<sup>[102]</sup>首次提出了一种专注于异构人脸识别问题的无监督多视图特征选择算法. 该算法充分利用多视图人脸数据,显著提高了人脸识别任务的准确率. Yang 等人<sup>[103]</sup>提出了一种无监督多任务多视图的特征选择和融合方法用于多媒体分析,在多个动作识别数据集和图像分类数据集上实现了良好的效果. 在极化合成孔径雷达图像的分类任务中,Huang 等人<sup>[104]</sup>基于广义典型相关分析方法,结合稀疏正则化和流形正则化,提出了一种无监督多视图特征选择方法,有效地改善了极化合成孔径雷达图像的分类性能.

(2) 生物信息应用: 多年来,生物信息学中数据的高维性催生了大量特征选择技术. 这些技术在序列分析、微阵列分析和质谱分析、单核苷酸多态性分析等研究领域得到了广泛应用,从而为疾病诊断、药物发现等应用提供了有力的支持<sup>[20,28]</sup>. 在生物信息学中,生物数据往往具有多个异构的视图源. 例如,特定基因组的基因特征可以通过基因表达数据、基因本体、蛋白质相互作用网络等数据进行描述<sup>[105]</sup>. 在这一背景下,Acharya 等人<sup>[105]</sup>提出了一种基于共识多视图多目标聚类的基因选择方法,显著地提高了基因表达数据的分类效果. 不久后,Acharya 等人<sup>[106]</sup>又提出了一种无监督多视图特征选择方法,用于从多视图高维基因表达数据中选择出相关的、非冗余的基因标记. 实验证明该方法能够有效地减少基因空间的维度,检测出具有生物学意义的基因标记. Wang 等人<sup>[107]</sup>提出了一种基于共识图学习和低秩张量学习的无监督多视图特征选择模型,在单细胞多组学数据集上进行了广泛的对比实验,证明了该模型在选择非冗余和具有代表性的特征方面的有效性.

(3) 辅助诊断: 尽管深度学习在数据分类和聚类任务中取得了显著进展, 但其生成的数据内在表示通常缺乏可解释性, 难以揭示某个特征对于目标的影响。相比之下, 特征选择是一种更为直观的方法, 通过从数据特征中挑选具有明确意义和辨识性的特征, 有助于更深入地理解数据并提升模型的可解释性。以计算机辅助诊断为例, 识别相关特征是该领域的核心问题<sup>[28]</sup>。在这方面, 无监督多视图特征选择成为从多视图的高维医学数据中选择出代表性特征的有效手段, 从而提高疾病诊断的准确性。例如, EL-Manzalawy 等人<sup>[108]</sup>设计了一种无监督多视图特征选择算法, 并将其应用于多组学数据的整合分析任务中。他们在卵巢癌多组学数据集上进行了生存预测实验, 结果表明该算法选择出的特征能够提高预测性能, 且具有潜在的生物标志物意义。为了评估胰腺切除术后新发糖尿病的风险, Hu 等人<sup>[109]</sup>采用了一种基于子空间聚类的无监督多视图特征选择模型, 从患者的多视图临床特征数据中进行特征选择。该模型提高了后续预测任务的准确性, 在临床实践中具有潜在的应用价值。另外, Li 等人<sup>[110]</sup>提出了一种基于 Exclusive Lasso 学习和稀疏学习的无监督多视图特征选择模型, 用于研究阿尔茨海默病的诊断。该模型提高了阿尔茨海默病的诊断效率和准确性, 改善了辅助诊断结果。

## 7 总结与展望

本文将现有的无监督多视图特征选择模型分为了基于图的无监督多视图特征选择模型、基于伪标签的无监督多视图特征选择模型和基于数据流的无监督多视图特征选择模型, 并对这些模型进行了介绍和总结。无监督多视图特征选择方法从多视图特征集中选取具有高辨识性的特征子集, 过滤数据中的冗余特征和噪声特征, 提高数据的可分性。它们一方面可以挖掘多视图数据样本的结构和分布信息, 另一方面可以起到降低数据维度的作用。

尽管现有的无监督多视图特征选择算法已经取得了丰富的成果, 但仍存在一些亟待解决的问题。

(1) 大规模数据应用。大规模数据应用已成为当今各个领域的主要趋势。然而, 现代大数据应用往往需要综合分析多个视角的数据。目前, 许多无监督多视图特征选择方法只适用于小规模数据的有效学习, 当面对大规模数据时, 需要耗费大量的计算和时间。以 MAMFS<sup>[78]</sup>为例, 其图优化的时间和空间复杂度随样本数量呈二次方增长, 导致算法在处理大规模样本时效率低下。因此, 未来的发展可以考虑优化算法的时间复杂度, 或者引入诸如分布式计算、采样技术<sup>[111]</sup>等大数据处理技术, 以提高处理效率和克服大规模数据在存储等方面所面临的挑战, 从而实现高效的特征选择。

(2) 提高数据间结构刻画的质量。目前, 图方法在无监督多视图特征选择模型中占主导地位, 图的质量很大程度上影响了这些模型性能。而大部分模型在构造图的过程中忽略了噪声特征可能带来的影响, 因此无法保证所得到的图能够准确反映样本之间的真实关系。其次, 当前模型中的图主要是基于成对相似性测量形成的相似性矩阵, 而在现实应用中, 对象及其特征往往呈现出复杂的多层关系而非简单的成对关系, 简单的成对相似性测量无法很好地反映出高阶数据之间的关系<sup>[112]</sup>。为此, 可以考虑使用更复杂的图结构, 例如超图<sup>[113]</sup>或高阶图<sup>[114]</sup>, 来更全面地反映数据之间的复杂关系。此外, 结合深度学习中的图卷积网络 (GCN)<sup>[115]</sup>等嵌入技术, 有助于捕捉节点之间的关系, 从而进一步提升无监督多视图特征选择模型的性能。

(3) 高质量的数据内在表示学习。数据的内在表示在很大程度上决定了学习任务的性能, 提高数据内在表示的质量对特征选择具有重要意义。然而, 当前的无监督多视图特征选择研究主要依赖于线性降维方法, 如矩阵分解方法或者谱分析方法来获得的数据内在表示, 并据此衡量数据特征的重要性。与此同时, 深度学习以其卓越的特征学习能力和出色的复杂数据关系建模能力而备受推崇。因此, 未来的研究可以考虑将深度学习技术引入无监督多视图特征选择方法中, 尤其是深度神经网络<sup>[116]</sup>和注意力机制<sup>[117]</sup>。这样的整合有望充分利用深度学习在数据内在表示学习方面优势, 指导模型从数据中选择更具辨识性的特征, 在一定程度上利用特征选择的特点来弥补深度学习模型在可解释性方面的不足。例如, 可以通过无监督的自编码器来学习更好的特征表示, 以支持特征选择过程。然而, 与深度学习结合也伴随着一些挑战, 如模型对大规模数据和计算资源的需求等。因此, 未来的研究还需要全面考虑如何平衡深度学习的优势与挑战, 推动无监督多视图特征选择方法的进一步发展。这可能包括模型的简化、新的网络结构设计以及对抗性训练等方面的探索, 以更好地适应多视图数据的特点。

(4) 处理数据的质量和动态性问题。在现实应用中, 高维多视图数据除了存在特征维度较高和噪声数据的问题之外, 还涉及一些常常被忽略的实际因素, 如样本的不完整 (样本缺失及样本特征缺失)、数据流的存在和数据分布的

漂移等. 而以往的无监督多视图特征选择算法侧重于提高分类或聚类性能, 对上述实际因素的关注相对较少, 模型的发展空间较大. 因此, 未来研究可以聚焦于缺失样本处理和流式学习等方法, 以更有效地处理数据的质量和动态性问题, 改善模型在具体应用中的表现. 具体而言, 针对样本不完整的问题, 可以考虑采用基于模型的填充方法<sup>[118]</sup>, 或者利用其他视图中的信息进行填充<sup>[119]</sup>, 以最大限度地保留数据的完整性. 对于处理动态性问题, 可以运用时间窗口、在线学习和增量技术等方法<sup>[120,121]</sup>, 与无监督多视图特征选择相结合, 以捕捉数据不断变化的趋势和模式.

## References:

- [1] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proc. of the 2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. San Diego: IEEE, 2005. 886–893. [doi: [10.1109/CVPR.2005.177](https://doi.org/10.1109/CVPR.2005.177)]
- [2] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int'l Journal of Computer Vision*, 2001, 42(3): 145–175. [doi: [10.1023/A:1011139631724](https://doi.org/10.1023/A:1011139631724)]
- [3] Lowe DG. Object recognition from local scale-invariant features. In: Proc. of the 1999 IEEE Int'l Conf. on Computer Vision. Kerkyra: IEEE, 1999. 1150–1157. [doi: [10.1109/ICCV.1999.790410](https://doi.org/10.1109/ICCV.1999.790410)]
- [4] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 971–987. [doi: [10.1109/TPAMI.2002.1017623](https://doi.org/10.1109/TPAMI.2002.1017623)]
- [5] Imangaliyev S, Levin E. Unsupervised multi-view feature selection for tumor subtype identification. In: Proc. of the 8th ACM Int'l Conf. on Bioinformatics, Computational Biology, and Health Informatics. 2017. 491–499. [doi: [10.1145/3107411.3107413](https://doi.org/10.1145/3107411.3107413)]
- [6] Tang JL, Hu X, Gao HJ, Liu H. Unsupervised feature selection for multi-view data in social media. In: Proc. of the 13th SIAM Int'l Conf. on Data Mining. Austin: SIAM, 2013. 270–278. [doi: [10.1137/1.9781611972832.3](https://doi.org/10.1137/1.9781611972832.3)]
- [7] Li YX, Shi X, Du CL, Liu Y, Wen YG. Manifold regularized multi-view feature selection for social image annotation. *Neurocomputing*, 2016, 204: 135–141. [doi: [10.1016/j.neucom.2015.07.151](https://doi.org/10.1016/j.neucom.2015.07.151)]
- [8] Feng YF, Xiao J, Zhuang YT, Liu XM. Adaptive unsupervised multi-view feature selection for visual concept recognition. In: Proc. of the 11th Asian Conf. on Computer Vision. Daejeon: Springer, 2013. 343–357. [doi: [10.1007/978-3-642-37331-2\\_26](https://doi.org/10.1007/978-3-642-37331-2_26)]
- [9] Wang Z, Feng YF, Qi T, Yang XS, Zhang JJ. Adaptive multi-view feature selection for human motion retrieval. *Signal Processing*, 2016, 120: 691–701. [doi: [10.1016/j.sigpro.2014.11.015](https://doi.org/10.1016/j.sigpro.2014.11.015)]
- [10] Zhang LY, Liu ML, Wang RF, Du TT, Li JY. Multi-view unsupervised feature selection with dynamic sample space structure. In: Proc. of the 2019 IEEE Symp. Series on Computational Intelligence. Xiamen: IEEE, 2019. 2641–2648. [doi: [10.1109/SSCI44817.2019.9003084](https://doi.org/10.1109/SSCI44817.2019.9003084)]
- [11] Cao RW, Zhu JH, Hao WY, Zhang CQ, Zhang ZH, Li ZY. Dual weighted multi-view subspace clustering. *Ruan Jian Xue Bao/Journal of Software*, 2022, 33(2): 585–597 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6148.htm> [doi: [10.13328/j.cnki.jos.006148](https://doi.org/10.13328/j.cnki.jos.006148)]
- [12] Hsieh TY, Sun YW, Wang SH, Honavar V. Adaptive structural co-regularization for unsupervised multi-view feature selection. In: Proc. of the 2019 IEEE Int'l Conf. on Big Knowledge. Beijing: IEEE, 2019. 87–96. [doi: [10.1109/ICBK.2019.00020](https://doi.org/10.1109/ICBK.2019.00020)]
- [13] Emami CK, Cullot N, Nicolle C. Understandable big data: A survey. *Computer Science Review*, 2015, 17: 70–81. [doi: [10.1016/j.cosrev.2015.05.002](https://doi.org/10.1016/j.cosrev.2015.05.002)]
- [14] Hou CP, Nie FP, Tao H, Yi DY. Multi-view unsupervised feature selection with adaptive similarity and view weight. *IEEE Trans. on Knowledge and Data Engineering*, 2017, 29(9): 1998–2011. [doi: [10.1109/TKDE.2017.2681670](https://doi.org/10.1109/TKDE.2017.2681670)]
- [15] Wei XK, Cao BK, Philip YS. Multi-view unsupervised feature selection by cross-diffused matrix alignment. In: Proc. of the 2017 Int'l Joint Conf. on Neural Networks. Anchorage: IEEE, 2017. 494–501. [doi: [10.1109/IJCNN.2017.7965894](https://doi.org/10.1109/IJCNN.2017.7965894)]
- [16] Lu CY, Min H, Gui J, Zhu L, Lei YK. Face recognition via weighted sparse representation. *Journal of Visual Communication and Image Representation*, 2013, 24(2): 111–116. [doi: [10.1016/j.jvcir.2012.05.003](https://doi.org/10.1016/j.jvcir.2012.05.003)]
- [17] Mi JX, Lei DJ, Gui J. A novel method for recognizing face with partial occlusion via sparse representation. *Optik*, 2013, 124(24): 6786–6789. [doi: [10.1016/j.jleo.2013.05.099](https://doi.org/10.1016/j.jleo.2013.05.099)]
- [18] Song LL. Research overview of class label noise. *Modern Computer*, 2016, (3): 20–23 (in Chinese with English abstract). [doi: [10.3969/j.issn.1007-1423.2016.03.005](https://doi.org/10.3969/j.issn.1007-1423.2016.03.005)]
- [19] Dave RN. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 1991, 12(11): 657–664. [doi: [10.1016/0167-8655\(91\)90002-4](https://doi.org/10.1016/0167-8655(91)90002-4)]
- [20] Li Y, Li T, Liu H. Recent advances in feature selection and its applications. *Knowledge and Information Systems*, 2017, 53(3): 551–557. [doi: [10.1007/s10115-017-1059-8](https://doi.org/10.1007/s10115-017-1059-8)]



- [21] Li JD, Liu H. Challenges of feature selection for big data analytics. *IEEE Intelligent Systems*, 2017, 32(2): 9–15. [doi: [10.1109/MIS.2017.38](https://doi.org/10.1109/MIS.2017.38)]
- [22] Liu Y, Cao JJ, Diao XC, Zhou X. Survey on stability of feature selection. *Ruan Jian Xue Bao/Journal of Software*, 2018, 29(9): 2559–2579 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5394.htm> [doi: [10.13328/j.cnki.jos.005394](https://doi.org/10.13328/j.cnki.jos.005394)]
- [23] Fakhraei S, Soltanian-Zadeh H, Fotouhi F. Bias and stability of single variable classifiers for feature ranking and selection. *Expert Systems with Applications*, 2014, 41(15): 6945–6958. [doi: [10.1016/j.eswa.2014.05.007](https://doi.org/10.1016/j.eswa.2014.05.007)]
- [24] Li ZQ, Du JQ, Nie B, Xiong WP, Huang CY, Li H. Summary of feature selection methods. *Computer Engineering and Applications*, 2019, 55(24): 10–19 (in Chinese with English abstract). [doi: [10.3778/j.issn.1002-8331.1909-0066](https://doi.org/10.3778/j.issn.1002-8331.1909-0066)]
- [25] Zhang R, Nie FP, Li XL, Wei X. Feature selection with multi-view data: A survey. *Information Fusion*, 2019, 50: 158–167. [doi: [10.1016/j.inffus.2018.11.019](https://doi.org/10.1016/j.inffus.2018.11.019)]
- [26] Fang Z, Zhang ZF. Discriminative feature selection for multi-view cross-domain learning. In: *Proc. of the 22nd ACM Int'l Conf. on Information & Knowledge Management*. San Francisco: ACM, 2013. 1321–1330. [doi: [10.1145/2505515.2505532](https://doi.org/10.1145/2505515.2505532)]
- [27] Zhang YS, Wu J, Cai ZH, Yu PS. Multi-view multi-label learning with sparse feature selection for image annotation. *IEEE Trans. on Multimedia*, 2020, 22(11): 2844–2857. [doi: [10.1109/TMM.2020.2966887](https://doi.org/10.1109/TMM.2020.2966887)]
- [28] Zhang DF, Chen HY, Wang JD. Survey of semi-supervised feature selection methods. *Application Research of Computers*, 2021, 38(2): 321–329 (in Chinese with English abstract). [doi: [10.19734/j.issn.1001-3695.2020.01.0001](https://doi.org/10.19734/j.issn.1001-3695.2020.01.0001)]
- [29] Sheikhpour R, Sarram MA, Gharaghani S, Zare Chahooki MA. A survey on semi-supervised feature selection methods. *Pattern Recognition*, 2017, 64: 141–158. [doi: [10.1016/j.patcog.2016.11.003](https://doi.org/10.1016/j.patcog.2016.11.003)]
- [30] Solorio-Fernández S, Carrasco-Ochoa JA, Martínez-Trinidad JF. A review of unsupervised feature selection methods. *Artificial Intelligence Review*, 2020, 53(2): 907–948. [doi: [10.1007/s10462-019-09682-y](https://doi.org/10.1007/s10462-019-09682-y)]
- [31] Xie JY, Ding LJ, Wang MZ. Spectral clustering based unsupervised feature selection algorithms. *Ruan Jian Xue Bao/Journal of Software*, 2020, 31(4): 1009–1024 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5927.htm> [doi: [10.13328/j.cnki.jos.005927](https://doi.org/10.13328/j.cnki.jos.005927)]
- [32] Guyon I, Elisseeff A. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 2003, 3(3): 1157–1182. [doi: [10.5555/944919.944968](https://doi.org/10.5555/944919.944968)]
- [33] Wu XY, Jiang BB, Lü SF, Wang XY, Chen QJ, Chen HH. A survey on causal feature selection based on Markov boundary discovery. *Pattern Recognition and Artificial Intelligence*, 2022, 35(5): 422–438 (in Chinese with English abstract). [doi: [10.16451/j.cnki.issn1003-6059.202205004](https://doi.org/10.16451/j.cnki.issn1003-6059.202205004)]
- [34] Alibeigi M, Hashemi S, Hamzeh A. Unsupervised feature selection based on the distribution of features attributed to imbalanced data sets. *Int'l Journal of Artificial Intelligence and Expert Systems*, 2011, 2(1): 14–22.
- [35] Zhou PY, Chan KCC. An unsupervised attribute clustering algorithm for unsupervised feature selection. In: *Proc. of the 2015 IEEE Int'l Conf. on Data Science and Advanced Analytics*. Paris: IEEE, 2015. 1–7. [doi: [10.1109/DSAA.2015.7344857](https://doi.org/10.1109/DSAA.2015.7344857)]
- [36] Hou CP, Nie FP, Yi DY, Wu Y. Feature selection via joint embedding learning and sparse regression. In: *Proc. of the 22nd Int'l Joint Conf. on Artificial Intelligence*. Barcelona: ACM, 2011. 1324–1329. [doi: [10.5555/2283516.2283617](https://doi.org/10.5555/2283516.2283617)]
- [37] Deepthi PS, Thampi SM. Unsupervised gene selection using particle swarm optimization and K-means. In: *Proc. of the 2nd ACM IKDD Conf. on Data Sciences*. Bangalore: ACM, 2015. 134–135. [doi: [10.1145/2732587.2732615](https://doi.org/10.1145/2732587.2732615)]
- [38] Wang Q, Jiang X, Chen ML, Li XL. Autoweighted multiview feature selection with graph optimization. *IEEE Trans. on Cybernetics*, 2022, 52(12): 12966–12977. [doi: [10.1109/TCYB.2021.3094843](https://doi.org/10.1109/TCYB.2021.3094843)]
- [39] Li YM, Yang M, Zhang ZF. A survey of multi-view representation learning. *IEEE Trans. on Knowledge and Data Engineering*, 2019, 31(10): 1863–1883. [doi: [10.1109/TKDE.2018.2872063](https://doi.org/10.1109/TKDE.2018.2872063)]
- [40] Nie FP, Zhu W, Li XL. Unsupervised feature selection with structured graph optimization. In: *Proc. of the 2016 AAAI Conf. on Artificial Intelligence*. Phoenix: AAAI, 2016. 1302–1308. [doi: [10.1609/aaai.v30i1.10168](https://doi.org/10.1609/aaai.v30i1.10168)]
- [41] Li ZC, Yang Y, Liu J, Zhou XF, Lu HQ. Unsupervised feature selection using nonnegative spectral analysis. In: *Proc. of the 2021 AAAI Conf. on Artificial Intelligence*. Toronto: AAAI, 2021. 1026–1032. [doi: [10.1609/aaai.v26i1.8289](https://doi.org/10.1609/aaai.v26i1.8289)]
- [42] Liu JW, Cui LP, Liu ZY, Luo XL. Survey on the regularized sparse models. *Chinese Journal of Computers*, 2015, 38(7): 1307–1325 (in Chinese with English abstract). [doi: [10.11897/SP.J.1016.2015.01307](https://doi.org/10.11897/SP.J.1016.2015.01307)]
- [43] Li XP, Wang YD, Ruiz R. A survey on sparse learning models for feature selection. *IEEE Trans. on Cybernetics*, 2022, 52(3): 1642–1660. [doi: [10.1109/TCYB.2020.2982445](https://doi.org/10.1109/TCYB.2020.2982445)]
- [44] Shi H, Li Y, Han YH, Hu QH. Cluster structure preserving unsupervised feature selection for multi-view tasks. *Neurocomputing*, 2016, 175: 686–697. [doi: [10.1016/j.neucom.2015.11.001](https://doi.org/10.1016/j.neucom.2015.11.001)]
- [45] Wangila KW, Gao K, Zhu PF, Hu QH, Zhang CQ. Mixed sparsity regularized multi-view unsupervised feature selection. In: *Proc. of the*

- 2017 IEEE Int'l Conf. on Image Processing. Beijing: IEEE, 2017. 1930–1934. [doi: [10.1109/ICIP.2017.8296618](https://doi.org/10.1109/ICIP.2017.8296618)]
- [46] Huang YY, Shen ZX, Cai YX, Yi XW, Wang DJ, Lv FM, Li TR. C<sup>2</sup>IMUFS: Complementary and consensus learning-based incomplete multi-view unsupervised feature selection. *IEEE Trans. on Knowledge and Data Engineering*, 2023, 35(10): 10681–10694. [doi: [10.1109/TKDE.2023.3266595](https://doi.org/10.1109/TKDE.2023.3266595)]
- [47] Liu XW, Wang L, Zhang J, Yin JP, Liu H. Global and local structure preservation for feature selection. *IEEE Trans. on Neural Networks and Learning Systems*, 2014, 25(6): 1083–1095. [doi: [10.1109/TNNLS.2013.2287275](https://doi.org/10.1109/TNNLS.2013.2287275)]
- [48] Zhou N, Xu YY, Cheng H, Fang J, Pedrycz W. Global and local structure preserving sparse subspace learning: An iterative approach to unsupervised feature selection. *Pattern Recognition*, 2016, 53: 87–101. [doi: [10.1016/j.patcog.2015.12.008](https://doi.org/10.1016/j.patcog.2015.12.008)]
- [49] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, 290(5500): 2323–2326. [doi: [10.1126/science.290.5500.2323](https://doi.org/10.1126/science.290.5500.2323)]
- [50] Saul KL, Roweis ST. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research*, 2003, 4: 119–155. [doi: [10.1162/153244304322972667](https://doi.org/10.1162/153244304322972667)]
- [51] He XF, Cai D, Yan SC, Zhang HJ. Neighborhood preserving embedding. In: *Proc. of the 2005 IEEE Int'l Conf. on Computer Vision*. Beijing: IEEE, 2005. 1208–1213. [doi: [10.1109/ICCV.2005.167](https://doi.org/10.1109/ICCV.2005.167)]
- [52] Cai D, Zhang CY, He XF. Unsupervised feature selection for multi-cluster data. In: *Proc. of the 16th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. Washington: ACM, 2010. 333–342. [doi: [10.1145/1835804.1835848](https://doi.org/10.1145/1835804.1835848)]
- [53] Liu HF, Mao HY, Fu Y. Robust multi-view feature selection. In: *Proc. of the 16th IEEE Int'l Conf. on Data Mining*. Barcelona: IEEE, 2016. 281–290. [doi: [10.1109/ICDM.2016.0039](https://doi.org/10.1109/ICDM.2016.0039)]
- [54] Tang C, Chen JJ, Liu XW, Li MM, Wang PC, Wang MH, Lu P. Consensus learning guided multi-view unsupervised feature selection. *Knowledge-based Systems*, 2018, 160: 49–60. [doi: [10.1016/j.knosys.2018.06.016](https://doi.org/10.1016/j.knosys.2018.06.016)]
- [55] Liu Y, Zhang CQ, Zhu PF, Hu QH. Generalized multi-view unsupervised feature selection. In: *Proc. of the 27th Int'l Conf. on Artificial Neural Networks and Machine Learning*. Rhodes: Springer, 2018. 469–478. [doi: [10.1007/978-3-030-01421-6\\_45](https://doi.org/10.1007/978-3-030-01421-6_45)]
- [56] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. In: *Proc. of the 14th Int'l Conf. on Neural Information Processing Systems: Natural and Synthetic*. Vancouver: ACM, 2001. 585–591. [doi: [10.5555/2980539.2980616](https://doi.org/10.5555/2980539.2980616)]
- [57] He XF, Niyogi P. Locality preserving projections. In: *Proc. of the 16th Int'l Conf. on Neural Information Processing Systems*. Whistler: ACM, 2004. 153–160. [doi: [10.5555/2981345.2981365](https://doi.org/10.5555/2981345.2981365)]
- [58] Wang WW, Li XP, Feng XC, Wang SQ. A survey on sparse subspace clustering. *Acta Automatica Sinica*, 2015, 41(8): 1373–1384 (in Chinese with English abstract). [doi: [10.16383/j.aas.2015.c140891](https://doi.org/10.16383/j.aas.2015.c140891)]
- [59] Lee DD, Seung HS. Algorithms for non-negative matrix factorization. In: *Proc. of the 13th Int'l Conf. on Neural Information Processing Systems*. Denver: ACM, 2000. 535–541. [doi: [10.5555/3008751.3008829](https://doi.org/10.5555/3008751.3008829)]
- [60] Von Luxburg U. A tutorial on spectral clustering. *Statistics and Computing*, 2007, 17(4): 395–416. [doi: [10.1007/s11222-007-9033-z](https://doi.org/10.1007/s11222-007-9033-z)]
- [61] Gretton A, Bousquet O, Smola A, Schölkopf B. Measuring statistical dependence with Hilbert-Schmidt norms. In: *Proc. of the 16th Int'l Conf. on Algorithmic Learning Theory*. Singapore: Springer, 2005. 63–77. [doi: [10.1007/11564089\\_7](https://doi.org/10.1007/11564089_7)]
- [62] Wang ZF, Salehi B, Gritsenko A, Chowdhury K, Ioannidis S, Dy J. Open-world class discovery with kernel networks. In: *Proc. of the 2020 Int'l Conf. on Data Mining*. Sorrento: IEEE, 2020. 631–640. [doi: [10.1109/ICDM50108.2020.00072](https://doi.org/10.1109/ICDM50108.2020.00072)]
- [63] Bach FR, Jordan MI. Learning spectral clustering. In: *Proc. of the 16th Int'l Conf. on Neural Information Processing Systems*. Whistler: ACM, 2003. 305–312. [doi: [10.5555/2981345.2981384](https://doi.org/10.5555/2981345.2981384)]
- [64] Nie FP, Li J, Li XL. Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification. In: *Proc. of the 25th Int'l Joint Conf. on Artificial Intelligence*. New York: ACM, 2016. 1881–1887. [doi: [10.5555/3060832.3060884](https://doi.org/10.5555/3060832.3060884)]
- [65] Xu YY, Yin Y, Wang J, Wei JM, Liu J, Yao LN, Zhang WJ. Unsupervised cross-view feature selection on incomplete data. *Knowledge-based Systems*, 2021, 234: 107595. [doi: [10.1016/j.knosys.2021.107595](https://doi.org/10.1016/j.knosys.2021.107595)]
- [66] Nie FP, Wang XQ, Jordan M, Huang H. The constrained laplacian rank algorithm for graph-based clustering. In: *Proc. of the 2016 AAAI Conf. on Artificial Intelligence*. Phoenix: AAAI, 2016. 1969–1976. [doi: [10.1609/aaai.v30i1.10302](https://doi.org/10.1609/aaai.v30i1.10302)]
- [67] Pang YW, Zhou B, Nie FP. Simultaneously learning neighborhood and projection matrix for supervised dimensionality reduction. *IEEE Trans. on Neural Networks and Learning Systems*, 2019, 30(9): 2779–2793. [doi: [10.1109/TNNLS.2018.2886317](https://doi.org/10.1109/TNNLS.2018.2886317)]
- [68] Wan Y, Sun SZ, Zeng C. Adaptive similarity embedding for unsupervised multi-view feature selection. *IEEE Trans. on Knowledge and Data Engineering*, 2021, 33(10): 3338–3350. [doi: [10.1109/TKDE.2020.2969860](https://doi.org/10.1109/TKDE.2020.2969860)]
- [69] Tang C, Zheng X, Liu XW, Zhang W, Zhang J, Xiong J, Wang LZ. Cross-view locality preserved diversity and consensus learning for multi-view unsupervised feature selection. *IEEE Trans. on Knowledge and Data Engineering*, 2022, 34(10): 4705–4716. [doi: [10.1109/TKDE.2020.3048678](https://doi.org/10.1109/TKDE.2020.3048678)]

- [70] Nie FP, Cai GH, Li XL. Multi-view clustering and semi-supervised classification with adaptive neighbours. In: Proc. of the 2017 AAAI Conf. on Artificial Intelligence. San Francisco: AAAI, 2017. 2408–2414. [doi: [10.1609/aaai.v31i1.10909](https://doi.org/10.1609/aaai.v31i1.10909)]
- [71] Dong X, Zhu L, Song XM, Li JJ, Cheng ZY. Adaptive collaborative similarity learning for unsupervised multi-view feature selection. In: Proc. of the 27th Int'l Joint Conf. on Artificial Intelligence. Stockholm: ACM, 2018. 2064–2070. [doi: [10.5555/3304889.3304946](https://doi.org/10.5555/3304889.3304946)]
- [72] Nie FP, Wang XQ, Huang H. Clustering and projected clustering with adaptive neighbors. In: Proc. of the 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM, 2014. 977–986. [doi: [10.1145/2623330.2623726](https://doi.org/10.1145/2623330.2623726)]
- [73] Liu XY, Han XH, Song K. Unsupervised multi-view feature selection based on adaptive similarity. Computer Engineering and Design, 2021, 42(11): 3158–3163 (in Chinese with English abstract). [doi: [10.16208/j.issn1000-7024.2021.11.021](https://doi.org/10.16208/j.issn1000-7024.2021.11.021)]
- [74] Du L, Lv XL. Consensus graph weighting via trace ratio criterion for multi-view unsupervised feature selection. In: Proc. of the 2019 Int'l Conf. on Data Mining Workshops. Beijing: IEEE, 2019. 615–619. [doi: [10.1109/ICDMW.2019.00093](https://doi.org/10.1109/ICDMW.2019.00093)]
- [75] Sun SZ, Wan Y, Zeng C. Multi-view embedding with adaptive shared output and similarity for unsupervised feature selection. Knowledge-based Systems, 2019, 165: 40–52. [doi: [10.1016/j.knosys.2018.11.017](https://doi.org/10.1016/j.knosys.2018.11.017)]
- [76] Yu SP, Yu K, Tresp V, Kriegel HP. Multi-output regularized feature projection. IEEE Trans. on Knowledge and Data Engineering, 2006, 18(12): 1600–1613. [doi: [10.1109/TKDE.2006.194](https://doi.org/10.1109/TKDE.2006.194)]
- [77] Zhou T, Zhang CQ, Peng X, Bhaskar H, Yang J. Dual shared-specific multiview subspace clustering. IEEE Trans. on Cybernetics, 2020, 50(8): 3517–3530. [doi: [10.1109/TCYB.2019.2918495](https://doi.org/10.1109/TCYB.2019.2918495)]
- [78] Zhang H, Wu DY, Nie FP, Wang R, Li XL. Multilevel projections with adaptive neighbor graph for unsupervised multi-view feature selection. Information Fusion, 2021, 70: 129–140. [doi: [10.1016/j.inffus.2020.12.007](https://doi.org/10.1016/j.inffus.2020.12.007)]
- [79] Wang MH, Wang Q, Hong DF, Roy SK, Chanussot J. Learning tensor low-rank representation for hyperspectral anomaly detection. IEEE Trans. on Cybernetics, 2023, 53(1): 679–691. [doi: [10.1109/TCYB.2022.3175771](https://doi.org/10.1109/TCYB.2022.3175771)]
- [80] Xia W, Gao QX, Wang QQ, Gao XB. Tensor completion-based incomplete multiview clustering. IEEE Trans. on Cybernetics, 2022, 52(12): 13635–13644. [doi: [10.1109/TCYB.2021.3140068](https://doi.org/10.1109/TCYB.2021.3140068)]
- [81] Yuan HL, Li JY, Liang Y, Tang YY. Multi-view unsupervised feature selection with tensor low-rank minimization. Neurocomputing, 2022, 487: 75–85. [doi: [10.1016/j.neucom.2022.02.005](https://doi.org/10.1016/j.neucom.2022.02.005)]
- [82] Xu R, Wunsch D. Survey of clustering algorithms. IEEE Trans. on Neural Networks, 2005, 16(3): 645–678. [doi: [10.1109/TNN.2005.845141](https://doi.org/10.1109/TNN.2005.845141)]
- [83] Xu YM, Wang CD, Lai JH. Weighted multi-view clustering with feature selection. Pattern Recognition, 2016, 53: 25–35. [doi: [10.1016/j.patcog.2015.12.007](https://doi.org/10.1016/j.patcog.2015.12.007)]
- [84] Wen J, Zhang Z, Fei LK, Zhang B, Xu Y, Zhang Z, Li JX. A survey on incomplete multiview clustering. IEEE Trans. on Systems, Man, and Cybernetics: Systems, 2023, 53(2): 1136–1149. [doi: [10.1109/TSMC.2022.3192635](https://doi.org/10.1109/TSMC.2022.3192635)]
- [85] Liu YB, Liao BB, Han YH. Discriminative multi-view feature selection and fusion. In: Proc. of the 2015 IEEE Int'l Conf. on Multimedia and Expo. Turin: IEEE, 2015. 1–6. [doi: [10.1109/ICME.2015.7177432](https://doi.org/10.1109/ICME.2015.7177432)]
- [86] Cai X, Nie FP, Huang H. Multi-view K-means clustering on big data. In: Proc. of the 23rd Int'l Joint Conf. on Artificial Intelligence. Beijing: ACM, 2013. 2598–2604. [doi: [10.5555/2540128.2540503](https://doi.org/10.5555/2540128.2540503)]
- [87] Bai XP, Zhu L, Liang C, Li JJ, Nie XS, Chang XI. Multi-view feature selection via nonnegative structured graph learning. Neurocomputing, 2020, 387: 110–122. [doi: [10.1016/j.neucom.2020.01.044](https://doi.org/10.1016/j.neucom.2020.01.044)]
- [88] Zhang YS, Wang XX, Cai ZH, Zhou YC, Philip SY. Tensor-based unsupervised multi-view feature selection for image recognition. In: Proc. of the 2021 IEEE Int'l Conf. on Multimedia and Expo. Shenzhen: IEEE, 2021. 1–6. [doi: [10.1109/ICME51207.2021.9428428](https://doi.org/10.1109/ICME51207.2021.9428428)]
- [89] Liang C, Wang LZ, Liu L, Zhang HX, Guo F. Multi-view unsupervised feature selection with tensor robust principal component analysis and consensus graph learning. Pattern Recognition, 2023, 141: 109632. [doi: [10.1016/j.patcog.2023.109632](https://doi.org/10.1016/j.patcog.2023.109632)]
- [90] Hu ML, Chen SC. One-pass incomplete multi-view clustering. In: Proc. of the 2019 AAAI Conf. on Artificial Intelligence. Honolulu: AAAI, 2019. 3838–3845. [doi: [10.1609/aaai.v33i01.33013838](https://doi.org/10.1609/aaai.v33i01.33013838)]
- [91] Wang SH, Tang JL, Liu H. Embedded unsupervised feature selection. In: Proc. of the 2015 AAAI Conf. on Artificial Intelligence. Austin: AAAI, 2015. 470–476. [doi: [10.1609/aaai.v29i1.9211](https://doi.org/10.1609/aaai.v29i1.9211)]
- [92] Cao ZW, Xie XJ, Sun FX, Qian JB. Consensus cluster structure guided multi-view unsupervised feature selection. Knowledge-based Systems, 2023, 271: 110578. [doi: [10.1016/j.knosys.2023.110578](https://doi.org/10.1016/j.knosys.2023.110578)]
- [93] Tang C, Zhu XZ, Liu XW, Wang LZ. Cross-view local structure preserved diversity and consensus learning for multi-view unsupervised feature selection. In: Proc. of the 2019 AAAI Conf. on Artificial Intelligence. Honolulu: AAAI, 2019. 5101–5108. [doi: [10.1609/aaai.v33i01.33015101](https://doi.org/10.1609/aaai.v33i01.33015101)]
- [94] Sun YG. Research on classification for data streams with concept drift [Ph.D. Thesis]. Beijing: Beijing Jiaotong University, 2019 (in Chinese with English abstract). [doi: [10.26944/d.cnki.gbjfu.2019.000738](https://doi.org/10.26944/d.cnki.gbjfu.2019.000738)]

- [95] Li PP. Concept drifting detection and classification on data streams [Ph.D. Thesis]. Hefei: Hefei University of Technology, 2012 (in Chinese with English abstract). [doi: 10.7666/d.Y2299134]
- [96] Zhai TT, Gao Y, Zhu JW. Survey of online learning algorithms for streaming data classification. *Ruan Jian Xue Bao/Journal of Software*, 2020, 31(4): 912–931 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5916.htm> [doi: 10.13328/j.cnki.jos.005916]
- [97] Ahmad S, Lavin A, Purdy S, Agha Z. Unsupervised real-time anomaly detection for streaming data. *Neurocomputing*, 2017, 262: 134–147. [doi: 10.1016/j.neucom.2017.04.070]
- [98] Shao WX, He LF, Lu CT, Wei XK, Philip SY. Online unsupervised multi-view feature selection. In: Proc. of the 16th IEEE Int'l Conf. on Data Mining. Barcelona: IEEE, 2016. 1203–1208. [doi: 10.1109/ICDM.2016.0160]
- [99] Huang YY, Guo KJ, Yi XW, Li Z, Li TR. Incremental unsupervised feature selection for dynamic incomplete multi-view data. *Information Fusion*, 2023, 96: 312–327. [doi: 10.1016/j.inffus.2023.03.018]
- [100] Wang H, Yang Y, Liu B. GMC: Graph-based multi-view clustering. *IEEE Trans. on Knowledge and Data Engineering*, 2020, 32(6): 1116–1129. [doi: 10.1109/TKDE.2019.2903810]
- [101] Chen MS, Huang L, Wang CD, Huang D. Multi-view clustering in latent embedding space. In: Proc. of the 2020 AAAI Conf. on Artificial Intelligence. New York: AAAI, 2020. 3513–3520. [doi: 10.1609/aaai.v34i04.5756]
- [102] Gui J, Li P. Multi-view feature selection for heterogeneous face recognition. In: Proc. of the 2018 IEEE Int'l Conf. on Data Mining. Singapore: IEEE, 2018. 983–988. [doi: 10.1109/ICDM.2018.00122]
- [103] Yang ZW, Wang HY, Han YH, Zhu XL. Discriminative multi-task multi-view feature selection and fusion for multimedia analysis. *Multimedia Tools and Applications*, 2018, 77(3): 3431–3453. [doi: 10.1007/s11042-017-5165-0]
- [104] Huang XY, Nie XL. Multi-view feature selection for PolSAR image classification via  $l_{2,1}$  sparsity regularization and manifold regularization. *IEEE Trans. on Image Processing*, 2021, 30: 8607–8618. [doi: 10.1109/TIP.2021.3118976]
- [105] Acharya S, Cui LZ, Pan Y. A consensus multi-view multi-objective gene selection approach for improved sample classification. *BMC Bioinformatics*, 2020, 21(13): 386. [doi: 10.1186/s12859-020-03681-5]
- [106] Acharya S, Cui LZ, Pan Y. Multi-view feature selection for identifying gene markers: A diversified biological data driven approach. *BMC Bioinformatics*, 2020, 21(18): 483. [doi: 10.1186/s12859-020-03810-0]
- [107] Wang LZ, Liang C, Dong WJ, Chen WL. Multi-view unsupervised feature selection via consensus guided low-rank tensor learning. In: Proc. of the 2022 IEEE Int'l Conf. on Bioinformatics and Biomedicine. Las Vegas: IEEE, 2022. 575–580. [doi: 10.1109/BIBM55620.2022.9995699]
- [108] El-Manzalawy Y, Hsieh TY, Shivakumar M, Kim D, Honavar V. Min-redundancy and max-relevance multi-view feature selection for predicting ovarian cancer survival using multi-omics data. *BMC Medical Genomics*, 2018, 11(3): 71. [doi: 10.1186/s12920-018-0388-0]
- [109] Hu PJ, Li X, Lu N, Dong KQ, Bai XL, Liang TB, Li JS. Prediction of new-onset diabetes after pancreatectomy with subspace clustering based multi-view feature selection. *IEEE Journal of Biomedical and Health Informatics*, 2023, 27(3): 1588–1599. [doi: 10.1109/JBHI.2022.3233402]
- [110] Li JY, Wu L, Wen GQ, Li Z. Exclusive feature selection and multi-view learning for Alzheimer's disease. *Journal of Visual Communication and Image Representation*, 2019, 64: 102605. [doi: 10.1016/j.jvcir.2019.102605]
- [111] Feng WM, Yin YT. Survey on theory of distributed sampling. *Ruan Jian Xue Bao/Journal of Software*, 2022, 33(10): 3673–3699 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6372.htm> [doi: 10.13328/j.cnki.jos.006372]
- [112] Zhang WJ, Wang BZ, Zhang ZH. A survey of feature selection algorithms based on graph. *Journal of Anhui University (Natural Sciences)*, 2017, 41(1): 10–20 (in Chinese with English abstract). [doi: 10.3969/j.issn.1000-2162.2017.01.003]
- [113] Huang HN, Zhou GX, Liang NY, Zhao QB, Xie SL. Diverse deep matrix factorization with hypergraph regularization for multi-view data representation. *IEEE/CAA Journal of Automatica Sinica*, 2023, 10(11): 2154–2167. [doi: 10.1109/JAS.2022.105980]
- [114] Wang R, Wang PL, Wu DY, Sun ZS, Nie FP, Li XL. Multi-view and multi-order structured graph learning. *IEEE Trans. on Neural Networks and Learning Systems*, 2023: 1–12. [doi: 10.1109/TNNLS.2023.3279133]
- [115] Shao ZZ, Xu YJ, Wei W, Wang F, Zhang Z, Zhu FD. Heterogeneous graph neural network with multi-view representation learning. *IEEE Trans. on Knowledge and Data Engineering*, 2023, 35(11): 11476–11488. [doi: 10.1109/TKDE.2022.3224193]
- [116] Huang ZY, Zhou JT, Zhu HY, Zhang CQ, Lv JC, Peng X. Deep spectral representation learning from multi-view data. *IEEE Trans. on Image Processing*, 2021, 30: 5352–5362. [doi: 10.1109/TIP.2021.3083072]
- [117] Lu RK, Liu JW, Zuo X. Attentive multi-view deep subspace clustering net. *Neurocomputing*, 2021, 435: 186–196. [doi: 10.1016/j.neucom.2021.01.011]
- [118] Wen J, Zhang Z, Zhang Z, Zhu L, Fei LK, Zhang B, Xu Y. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. In: Proc. of the 2021 AAAI Conf. on Artificial Intelligence. AAAI, 2021. 10273–10281. [doi: 10.1609/aaai.v35i11.

- 17231]
- [119] Yang JH, Fu LL, Chen C, Dai HN, Zheng ZB. Cross-view graph matching for incomplete multi-view clustering. *Neurocomputing*, 2023, 515: 79–88. [doi: 10.1016/j.neucom.2022.10.007]
- [120] Zubaroglu A, Atalay V. Data stream clustering: A review. *Artificial Intelligence Review*, 2021, 54(2): 1201–1236. [doi: 10.1007/s10462-020-09874-x]
- [121] Chen J, Yang SX, Wang Z. Multi-view representation learning for data stream clustering. *Information Sciences*, 2022, 613: 731–746. [doi: 10.1016/j.ins.2022.09.045]

#### 附中文参考文献:

- [11] 曹容玮, 祝继华, 郝问裕, 张长青, 张苗涵, 李钟毓. 双加权多视角子空间聚类算法. *软件学报*, 2022, 33(2): 585–597. <http://www.jos.org.cn/1000-9825/6148.htm> [doi: 10.13328/j.cnki.jos.006148]
- [18] 宋磊磊. 类标噪声研究综述. *现代计算机*, 2016, (3): 20–23. [doi: 10.3969/j.issn.1007-1423.2016.03.005]
- [22] 刘艺, 曹建军, 刁兴春, 周星. 特征选择稳定性研究综述. *软件学报*, 2018, 29(9): 2559–2579. <http://www.jos.org.cn/1000-9825/5394.htm> [doi: 10.13328/j.cnki.jos.005394]
- [24] 李郅琴, 杜建强, 聂斌, 熊旺平, 黄灿奕, 李欢. 特征选择方法综述. *计算机工程与应用*, 2019, 55(24): 10–19. [doi: 10.3778/j.issn.1002-8331.1909-0066]
- [28] 张东方, 陈海燕, 王建东. 半监督特征选择综述. *计算机应用研究*, 2021, 38(2): 321–329. [doi: 10.19734/j.issn.1001-3695.2020.01.0001]
- [31] 谢娟英, 丁丽娟, 王明钊. 基于谱聚类的无监督特征选择算法. *软件学报*, 2020, 31(4): 1009–1024. <http://www.jos.org.cn/1000-9825/5927.htm> [doi: 10.13328/j.cnki.jos.005927]
- [33] 吴兴宇, 江兵兵, 吕胜飞, 王翔宇, 陈秋菊, 陈欢欢. 基于马尔科夫边界发现的因果特征选择算法综述. *模式识别与人工智能*, 2022, 35(5): 422–438. [doi: 10.16451/j.cnki.issn1003-6059.202205004]
- [42] 刘建伟, 崔立鹏, 刘泽宇, 罗雄麟. 正则化稀疏模型. *计算机学报*, 2015, 38(7): 1307–1325. [doi: 10.11897/SP.J.1016.2015.01307]
- [58] 王卫卫, 李小平, 冯象初, 王斯琪. 稀疏子空间聚类综述. *自动化学报*, 2015, 41(8): 1373–1384. [doi: 10.16383/j.aas.2015.c140891]
- [73] 刘欣宇, 韩晓红, 宋可. 基于自适应相似性的无监督多视图特征选择. *计算机工程与设计*, 2021, 42(11): 3158–3163. [doi: 10.16208/j.issn1000-7024.2021.11.021]
- [94] 孙艳歌. 概念漂移数据流分类算法研究 [博士学位论文]. 北京: 北京交通大学, 2019. [doi: 10.26944/d.cnki.gbfju.2019.000738]
- [95] 李培培. 数据流中概念漂移检测与分类方法研究 [博士学位论文]. 合肥: 合肥工业大学, 2012. [doi: 10.7666/d.Y2299134]
- [96] 翟婷婷, 高阳, 朱俊武. 面向流数据分类的在线学习综述. *软件学报*, 2020, 31(4): 912–931. <http://www.jos.org.cn/1000-9825/5916.htm> [doi: 10.13328/j.cnki.jos.005916]
- [111] 凤维明, 尹一通. 分布式采样理论综述. *软件学报*, 2022, 33(10): 3673–3699. <http://www.jos.org.cn/1000-9825/6372.htm> [doi: 10.13328/j.cnki.jos.006372]
- [112] 张文静, 王备战, 张志宏. 基于图的特征选择算法综述. *安徽大学学报(自然科学版)*, 2017, 41(1): 10–20. [doi: 10.3969/j.issn.1000-2162.2017.01.003]



吴建生(1986—), 男, 博士, 讲师, CCF 专业会员, 主要研究领域为机器学习, 特征选择, 数据挖掘。



黄冲(1983—), 男, 工程师, 主要研究领域为数据挖掘, 大数据分析。



李艳兰(1998—), 女, 硕士生, CCF 学生会会员, 主要研究领域为机器学习, 数据挖掘。



闵卫东(1966—), 男, 博士, 教授, 博士生导师, CCF 杰出会员, 主要研究领域为图像和视频处理, 人工智能, 大数据, 分布式系统, 智慧城市信息技术。