

1. 问题陈述

现有的不完整多视角聚类（IMVC）方法存在两个主要问题：一是对缺失数据的填补或填充不准确，可能会影响聚类性能；二是融合后的特征质量可能会受到低质量视图的干扰，尤其是不准确插补的视图。

2. 方法概述

为了解决这些问题，作者提出了采用了无填补（imputation-free）和无融合（fusion-free）的策略的深度不完整多视图聚类方法（DIMVC 框架）。

DIMVC 不需要对缺失数据进行填补，也不依赖于多个视角的特征融合，从而避免了因不准确填补而引入的噪声和低质量视角对聚类结果的干扰。

方法：

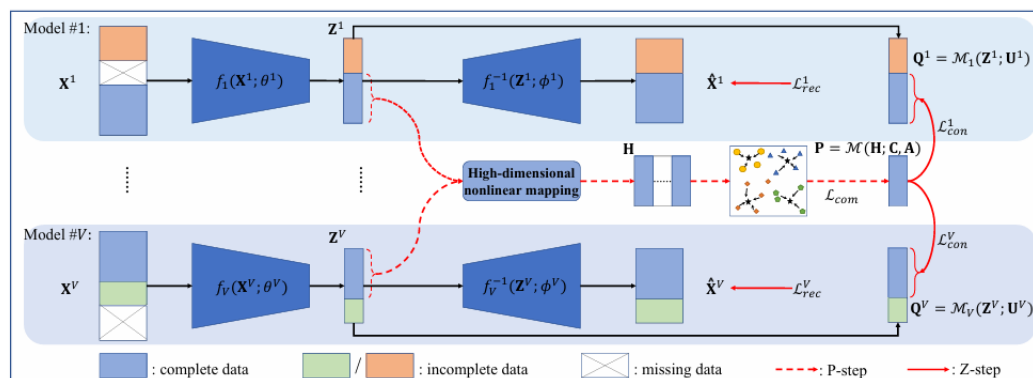


Figure 1: The framework of DIMVC. For the v -th view, the encoder f_v and decoder f_v^{-1} learn the embedding features Z^v , from which the mapping M_v predicts the cluster assignments Q^v . The high-dimensional nonlinear mapping is proposed to obtain the linearly separable features H . The mapping M aims to generate supervised information P with high confidence. For all views, P-step mines their complementary information by optimizing L_{com} . Z-step achieves the feature learning and clustering consistency across multiple views by optimizing $\{L_{rec}^v + L_{con}^v\}_{v=1}^V$.

作者设计了基于深度自编码器（autoencoder）的特征学习与聚类框架，用于处理每个视角的数据，采用自动编码器将异构多视图数据转换为便于聚类的嵌入特征。对于第 v 个视图，嵌入特征表示为 Z^v ，由编码器和解码器学习得到，然后在 Z^v 上通过一个 clustering head 来获得对每个样本的 clustering assignment。

所有视图的 X^v 和 \hat{X}^v 之间的重构损失为

$$\begin{aligned}
\mathcal{L}_{rec} &= \sum_{v=1}^V \mathcal{L}_{rec}^v = \sum_{v=1}^V \|\mathbf{X}^v - f_v^{-1}(\mathbf{Z}^v)\|_F^2 \\
&= \sum_{v=1}^V \sum_{i=1}^{N_v} \|\mathbf{x}_i^v - f_v^{-1}(f_v(\mathbf{x}_i^v))\|_2^2.
\end{aligned} \tag{1}$$

为了获得聚类预测，对于每个视图，利用参数化映射来获得软聚类分配 \mathbf{Q}^v

$$q_{ij}^v = \frac{(1 + \|\mathbf{z}_i^v - \mathbf{u}_j^v\|_2^2)^{-1}}{\sum_{j=1}^K (1 + \|\mathbf{z}_i^v - \mathbf{u}_j^v\|_2^2)^{-1}} \in \mathbf{Q}^v, \tag{2}$$

直到上一步，都只是在视图内的学习，视图间并未交互。

作者设计了一个加权的特征拼接映射 H ，将所有视角的嵌入特征映射到一个高维空间，其中，权重 w_v 通过视角的聚类分离度计算。

$$\mathbf{H} = \mathcal{H}(\{\mathbf{Z}^v\}_{v=1}^V) = (w_1 \mathbf{Z}^1, w_2 \mathbf{Z}^2, \dots, w_V \mathbf{Z}^V), \tag{3}$$

where $\mathbf{H} \in \mathbb{R}^{|\mathcal{X}| \times \sum_{v=1}^V d_v}$ denotes the obtained high-dimensional features. w_v is the weight calculated by

$$w_v = 1 + \log \left(1 + \frac{\sigma(\mathbf{U}^v)}{\sum_v \sigma(\mathbf{U}^v)} \right), \tag{4}$$

在高维空间 \mathbf{H} 中，对所有视角的聚类中心重新优化，使聚类结果更一致。作者认为：高度可分的视图意味着视图内类间可分离性强，相应地类中心有较大的方差。因此，通过加权的方式，让统一的表征 \mathbf{H} 更注重分离性强的视图，同时减弱分离性差的视图的影响。

$$\begin{aligned}
\mathcal{L}_{com} &= \min_{\{\mathbf{C}^v\}_{v=1}^V} \sum_{i \in \mathcal{X}} \sum_{j=1}^K \sum_{v=1}^V \|w_v \mathbf{z}_i^v - \mathbf{c}_j^v\|_2^2 \\
&= \min_{\mathbf{C}} \sum_{i \in \mathcal{X}} \sum_{j=1}^K \|\mathbf{h}_i - \mathbf{c}_j\|_2^2,
\end{aligned}$$

到这一步，作者在表征空间 \mathbf{H} 上对完整数据重新计算了类中心

在统一表征 \mathbf{H} 上进行 cluster assignments，类似于 softmax，并把在 \mathbf{H} 上获得的 cluster assignments 视为监督信息，用来监督 view-specific 的特征学习。通过这种方式，将挖掘出的互补聚类信息转换为具有高置信度的监督信息。此外，还对所有视图中 \mathbf{P} 和 \mathbf{Q}^v 之间的交叉熵损失进行优化：

$$\mathcal{L}_{con} = \sum_{v=1}^V H(\mathbf{P}, \mathbf{Q}^v) = - \sum_{v=1}^V \sum_{i \in \mathcal{X}} \mathbf{p}_i \log \mathbf{q}_i^v. \tag{9}$$

通过此方法可以实现视图间的聚类一致性。样本的聚类预测标签由所有视图的预测的和得到

$$y_i = \arg \max_j \sum_v q_{ij}^v. \quad (10)$$

在训练深度模型之后，所获得的所有视图的模型都是无需融合的，因为每个视图都有自己的模型来进行特征提取和聚类，这显著地不依赖于其他视图的特征融合。接下来是优化方法。

损失函数由三部分组成

$$\mathcal{L} = \mathcal{L}_{rec} + \mathcal{L}_{com} + \mathcal{L}_{con}$$

在初始化阶段，利用自动编码器提取嵌入特征，并使用 K-means 初始化聚类中心。之后交替更新 P-step 和 Z-step，P 步骤：优化 P, C, A 通过 K-means 更新 C 并通过匈牙利算法解决聚类匹配问题以更新 A 。Z 步骤：固定 P, C, A 利用梯度下降更新 Z^v 和聚类中心 U^v 。

Algorithm 1: Optimization of the proposed DIMVC

Input: dataset $\{\mathbf{X}^v \in \mathbb{R}^{N_v \times D_v}\}_{v=1}^V$, number of clusters K

Output: clustering predictions

- 1: **Initialization:** initialize $\{\mathbf{Z}^v\}_{v=1}^V$ by Eq. (1), initialize $\{\mathbf{U}^v\}_{v=1}^V$ by K-means, $\mathbf{A} = \mathbf{I}_K$
 - 2: **while** not reaching the maximal iterations **do**
 - 3: **P-step:** fix $\{\mathbf{Z}^v, \mathbf{U}^v\}_{v=1}^V$
 - 4: update \mathbf{C} by Eq. (5)
 - 5: update \mathbf{A} by Eq. (12)
 - 6: update \mathbf{P} by Eq. (6)
 - 7: **Z-step:** fix $\{\mathbf{P}, \mathbf{C}, \mathbf{A}\}$
 - 8: update $\{\mathbf{Z}^v, \mathbf{U}^v\}_{v=1}^V$ by Eq. (13)
 - 9: **end while**
 - 10: Calculate the clustering predictions by Eqs. (2 and 10)
-

本文提出了一种无输入、无融合的深度不完全多视图聚类框架。首先，提出了一种挖掘多视图间互补信息的新策略，即将所有视图的嵌入特征映射到连接加权特征空间 (CWFS) 中。因此，由于多视图集群的互补性，CWFS 中的新特性更有可能是线性可分的。在此基础上，采用一种替代 (EM-like) 优化策略进行特征学习和聚类，将互补信息转化为监督信息，实现多视图的一致性。

论文整体思路遵循了作者另一篇工作 SDMVC: 《Self-supervised Discriminative Feature Learning for Deep Multi-view Clustering》，并应用到了缺失多视图聚类上。相较于 SDMVC，此工作在融合 view-specific 表征时提出了加权融合策略。

3. 实验

作者使用了以下四个数据集：

BDGP: 包含视觉和文本视图数据, 共 2,500 个样本, 5 个类别。

Caltech: 包含 HOG 和 GIST 视图数据, 共 2,386 个样本, 20 个类别。

RGB-D: 包含视觉和文本视图数据, 共 1,449 个样本, 13 个类别。

Scene: 包含 GIST、PHOG 和 LBP 视图数据, 共 4,485 个样本, 15 个类别。

作者提供了代码: [SubmissionsIn/DIMVC: Deep Incomplete Multi-View Clustering via Mining Cluster Complementarity](#)

作者与以下 7 种多视角聚类方法进行了对比:

传统方法: SRLC (Zhuge et al., 2019)、APMC (Guo and Ye, 2019)、TMBSD (Li et al., 2021)、IMVTSC-MVI (Wen et al., 2021)

深度学习方法: DiMVMC (Wei et al., 2020)、CDIMC-net (Wen et al., 2020)、COMPLETER (Lin et al., 2021)

实验结果表明, DIMVC 在所有四个数据集上都达到了最优的聚类性能, 尤其是在高缺失率 (0.5 和 0.7) 的情况下, 与其他方法相比具有显著优势。以下是部分关键结果的简要描述:

1. 在 BDGP 数据集上, DIMVC 的 ACC、NMI 和 ARI 在不同缺失率下均为最佳。例如, 缺失率为 0.1 时, 分别达到了 0.964、0.892 和 0.912。
2. 在 Caltech 数据集上, DIMVC 同样表现最佳, 例如缺失率为 0.1 时, ACC 为 0.772, NMI 为 0.726, ARI 为 0.870。
3. 在 RGB-D 数据集上, DIMVC 的 ACC、NMI 和 ARI 总体优于其他方法, 尽管提升幅度较小。
4. 在 Scene 数据集上, DIMVC 的 ACC、NMI 和 ARI 在大多数情况下领先, 例如缺失率为 0.1 时, ACC 为 0.474, NMI 为 0.465, ARI 为 0.306。

作者的方法 (DIMVC) 在四个数据集上的效果显著优于传统和现有深度学习方法, 特别是在处理缺失率较高的数据时表现更加稳健。

4. 思考

论文提出了一个无需填充或融合的深度不完全多视角聚类框架, 通过交替优化策略 (P-step 和 Z-step) 挖掘视角间互补信息, 实现了优异的聚类性能。但存在计算复杂度较高、实验验证数据集多样性不足及超参数敏感性分析不全面等问题。未来可探索降低计算复杂度、改进交替优化机制、处理高冗余视角、以及增强模型可解释性和理论分析深度等方向。

提出了一个新颖的、无需数据填充或视角融合的多视角聚类框架, 避免了传统方法中对缺失视角进行补全或直接融合的局限性, 开发了将所有视角的嵌入特征映射到加权特征空间的策略 (CWFS)。设计了一种交替的 (类似 EM) 优化策略, 分为 P-step 和 Z-step, 其中将互补信息转化为监督信息以实现多个视图的一致性。方法能够有效地从完整数据中探索多视图的互补信息, 也具有处理不完整数据的泛化能力。同时通过实验验证达到了 state-of-the-art 的聚类性能。

尽管高维嵌入特征表现出稳定的性能, 但高维度可能带来计算复杂度和存储需求的显著增加, 这在大规模数据集上可能成为瓶颈。尽管文中描述了方法的关键部分, 但对 P-step 和 Z-step

具体交互机制的理论分析不够深入，特别是在不同视角数据分布较为不均衡的情况下，是否依然能够保持一致性，缺乏进一步讨论。缺乏对更加复杂、真实世界多视角数据的全面测试。方法的核心依赖于不同视角之间的互补信息。如果数据视角间的冗余较高或者互补性较差（如视角间高度相关），可能会限制方法的有效性。

未来可以探索低维投影或稀疏表示方法，以减少高维嵌入特征在实际应用中的计算开销，同时保证特征的表征能力。改进 **P-step** 和 **Z-step** 的交替优化机制，例如引入自适应的权重更新策略，动态平衡两者之间的贡献，从而提高优化效率。针对数据视角间冗余信息较多的场景，探索如何减少冗余视角对聚类效果的干扰，例如引入视角选择或去冗余技术。在特定应用场景中引入先验知识或监督信号（如弱监督、约束聚类），以进一步提高模型性能和适应性。

5. 其它（选填）

需要特别记录的其它笔记