



通信学报
Journal on Communications
ISSN 1000-436X, CN 11-2102/TN

《通信学报》网络首发论文

题目：面向视角非对齐数据的多视角聚类方法
作者：李鹭，冯聪，牛宇童，徐士彪，张英涛，孙广路
收稿日期：2022-03-21
网络首发日期：2022-07-21
引用格式：李鹭，冯聪，牛宇童，徐士彪，张英涛，孙广路. 面向视角非对齐数据的多视角聚类方法[J/OL]. 通信学报.
<https://kns.cnki.net/kcms/detail/11.2102.TN.20220720.1354.018.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

面向视角非对齐数据的多视角聚类方法

李鹭¹, 冯聪¹, 牛宇童¹, 徐士彪², 张英涛³, 孙广路¹

(1. 哈尔滨理工大学计算机科学与技术学院, 黑龙江 哈尔滨 150080;

2. 北京邮电大学人工智能学院, 北京 100876;

3. 哈尔滨工业大学计算机科学与技术学院, 黑龙江 哈尔滨 150080)

摘 要: 如何在视角对齐关系错位时有效进行非对齐多视角学习是一类新的挑战性问题。针对这一问题, 提出面向视角非对齐数据的多视角学习方法。一方面, 为了捕获多视角异构特征的跨视角相似度信息, 基于多视角非负矩阵分解进行表示学习, 将原始特征嵌入一个可度量的低维同构空间。另一方面, 在低维同构空间中, 以二部图最优匹配模型建模视角对齐关系, 并提出参考视角概念将模型推广至多视角情形。将表示学习和视角对齐关系学习整合到统一的 Bi-level 优化框架, 使其在迭代中相互促进, 进一步提高模型对视角非对齐数据的学习能力。在视角非对齐数据聚类应用上的大量实验结果表明, 相比于 8 种先进的多视角聚类方法, 所提方法在 3 个数据集上的多项性能指标均取得较优的性能。

关键词: 聚类分析; 多视角学习; 视角非对齐数据; 非负矩阵分解

中图分类号: TP391

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2022134

Multiview clustering method for view-unaligned data

LI Ao¹, FENG Cong¹, NIU Yutong¹, XU Shibiao², ZHANG Yingtao³, SUN Guanglu¹

1. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China

2. Artificial Intelligence School, Beijing University of Posts and Telecommunications, Beijing 100876, China

3. School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150080, China

Abstract: A new challenge for multi-view learning was posed by corrupted view-correspondences. To address this issue, an effective multi-view learning method for view-unaligned data was proposed. First, to capture cross-view latent affinity in multi-view heterogeneous feature spaces, representation learning was employed based on multi-view non-negative matrix factorization to embed original features into a measurable low-dimensional subspace. Second, view-alignment relationships were modeled as optimal matching of a bipartite graph, which could be generalized to multiple-views situations via the proposed concept reference view. Representation learning and data alignment were further integrated into a unified Bi-level optimization framework to mutually boost the two learning processes, effectively enhancing the ability to learn from view-unaligned data. Extensive experimental results of view-unaligned clustering on three public datasets demonstrate that the proposed method outperforms eight advanced multiview clustering methods on multiple evaluation metrics.

Keywords: clustering analysis, multiview learning, view-unaligned data, non-negative matrix factorization

收稿日期: 2022-03-21; 修回日期: 2022-06-21

基金项目: 国家自然科学基金资助项目 (No.62071157); 黑龙江省青年创新人才计划 (No.UNPYSCT-2018203); 黑龙江省自然科学基金优秀青年基金 (No.YQ2019F011); 黑龙江省高等学校基本科研业务专项资金资助项目 (No.LGYC2018JQ013); 黑龙江省博士后启动基金资助项目 (No.LBH-Q19112)

Foundation Items: The National Natural Science Foundation of China (No.62071157), The University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (No.UNPYSCT-2018203), The Natural Science Foundation of Heilongjiang Province (No.YQ2019F011), The Fundamental Research Foundation for University of Heilongjiang Province (No.LGYC2018JQ013), Postdoctoral Foundation of Heilongjiang Province (No.LBH-Q19112)

0 引言

近年来,多视角学习在机器学习领域引起了广泛的研究兴趣^[1-2]。一方面,在众多研究领域中,数据自然地呈现出多种表达方式^[3-4]。另一方面,数据的各种表达方式之间存在着隐式的一致性和互补性^[5-6]。因此,融合数据的多视角特征,能更全面地揭示数据的分布规律,提高多种下游任务的性能。

多视角聚类是多视角学习的典型应用,它旨在融合多个视角的线索,更好地挖掘数据中的本质簇结构^[7-8]。考虑基于多视角数据的相似性进行融合, Nie 等^[9]提出参数无关的自加权多视角聚类方法,该方法重新表述标准的谱聚类算法,为每一视角分配最适权重。文献^[10]寻求一个低秩相似性矩阵作为各视角图结构的质心,从质心中直接获取统一的类簇指示矩阵。Liang 等^[11]在统一的框架中建模多视角数据的一致性和不一致性,进一步提高聚类性能。为了克服静态相似性矩阵的缺陷, Wang 等^[12]提出基于动态相似性融合的多视角聚类方法,实现视角专属相似性矩阵和公共相似性矩阵的相互促进。降维技术是应对多视角数据冗余性的有效方法。Liu 等^[13]提出基于联合非负矩阵分解的多视角聚类方法,对各视角系数矩阵施加一致性约束,使其融合为一致的类簇指示矩阵。张祎等^[14]在矩阵分解的基础上加入流形约束,有效保持了样本空间的流形结构。张量低秩约束是一种利用多视角数据高阶相关性的有效方法。Zhang 等^[15]提出基于低秩张量约束的多视角子空间聚类方法,将各视角的自表达矩阵堆叠成张量,再以张量低秩约束降低其冗余性。Wu 等^[16]提出基于关键张量学习的谱聚类方法,将各视角基于马尔可夫链的概率转移矩阵堆叠成张量并施加张量低秩约束,捕获了多视角数据的高阶相关性。

尽管上述多视角聚类方法取得了良好的性能,但在真实场景下的应用仍有一定局限性。由于信道时延、传输异步、网络拥塞等多种客观因素,获取的多视角数据会发生视角的非对齐现象,影响现有多视角聚类方法的性能。这类问题是多视角学习中的一类初探新兴问题,以三视角数据为例进行说明。如图 1(a)所示,给定一个对齐的多视角数据集 $\mathbf{X} = \{\mathbf{X}^v\}_{v=1}^3$, 其所有视角的特征矩阵 \mathbf{X}^v 的第 i 列 \mathbf{X}_i^v 表示同一样本 i 在不同视角 v 下的特征,即同一样本在各视角的特征矩阵中是沿列对齐的。然而,

由于信道时延、传输异步等因素,如图 1(b)所示,数据的原有视角对齐关系可能会发生错位,即各视角特征矩阵 \mathbf{X}^v 的同一列可能不再来自同一样本。由于现有多视角聚类方法大多要求不同视角之间存在严格的对齐关系以实现特征融合^[2], 视角非对齐现象将误导现有方法将不同类别样本的特征加以融合,降低了学习表示的鉴别力,从而限制了现有多视角学习方法的应用。

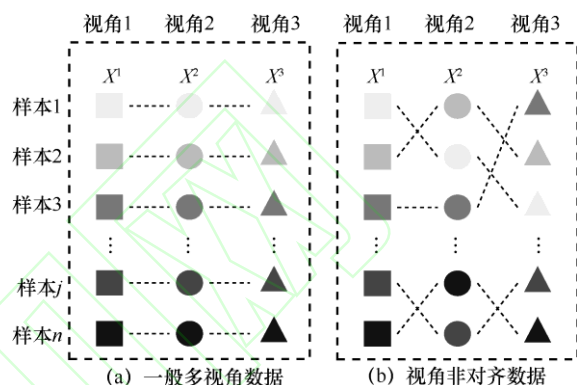


图 1 一般多视角数据与视角非对齐多视角数据对比

面向这一新兴问题的研究并不多见,尽本文的调研所知,目前仅包括文献^[17-18]这 2 个前人工作。文献^[17]提出部分视角对齐聚类 (PVC, partially view-aligned clustering) 方法,设计了一个可微的对齐模块,并将其集成于深度神经网络以重建视角对齐关系。文献^[18]提出一种多视角噪声稳健的对比学习 (MvCLN, multi-view contrastive learning with noise-robust loss) 模型,使对齐样本在表示空间中相互靠近,而非对齐样本则互相远离。然而,上述前人工作只适用于两视角数据。

前人工作揭示了一种处理视角非对齐数据的可行思路:先基于特征的相似度信息学习视角对齐关系,再重建视角之间的对应关系以缓解视角非对齐现象对多视角学习的影响。在此基础上,本文提出了一种新的、适用于多视角的视角非对齐数据聚类方法。首先,本文方法基于多视角非负矩阵分解进行表示学习,将异构的多视角特征嵌入可度量的低维同构空间。其次,在低维同构空间中,基于二部图最优匹配学习视角对齐关系,并提出参考视角的概念将对齐模型推广至具有 2 个以上视角的非对齐数据。最后,将上述表示学习和数据对齐模块放入统一的 Bi-level 优化框架,通过交替求解上层问题 (表示学习) 和下层问题 (数据对齐),使二者在迭代中相互促进,以提升模型的总体性能。本文

的贡献如下。

1) 提出一种新颖的面向视角非对齐数据的多视角学习方法, 将表示学习和数据对齐置于同一目标函数中交替优化, 学习到具有 2 个以上视角的复杂非对齐数据的紧凑表示。

2) 提出了一种基于 Bi-level 优化的目标函数, 将表示学习和数据对齐抽象为上下层问题。设计了一种基于交替优化的数值求解算法, 该算法具有较好的参数稳定性和高效的收敛性能。

3) 将本文方法应用于视角非对齐数据的多视角聚类任务, 在 3 个公开数据集、3 种视角非对齐率和 3 种评价指标上, 本文方法均优于 8 种先进的多视角聚类方法。本文方法的总体结构如图 2 所示。

1 相关工作

1.1 多视角非负矩阵分解

多视角非负矩阵分解^[13-14] (MultiNMF, multiview non-negative matrix factorization) 是一种有效的多视角表示学习方法, 能够在降低原始特征维度的同时融合各视角的特征, 学习到紧凑的多视角公共表示。假设 $\mathbf{X} = \{\mathbf{X}^i\}_{i=1}^{N_v}$, $\mathbf{X}^v \geq 0$ 是具有 N_v 个视角的多视角数据集。对各视角的特征矩阵 \mathbf{X}^v 进行非负矩阵分解, 使 $\mathbf{X}^v \approx \mathbf{U}^v (\mathbf{V}^v)^T$, 得到基矩阵 \mathbf{U}^v 和系数矩阵 \mathbf{V}^v , 则 \mathbf{V}^v 可作为该视角的低维表示, 再对 \mathbf{V}^v 施加一致性约束, 可得到公共表示矩阵 \mathbf{V}^* 。该方法的目标函数为

$$\min_{\mathbf{U}^v, \mathbf{V}^v, \mathbf{V}^*} \sum_{v=1}^{N_v} \|\mathbf{X}^v - \mathbf{U}^v (\mathbf{V}^v)^T\|_F^2 + \sum_{v=1}^{N_v} \lambda_v \|\mathbf{V}^v - \mathbf{V}^*\|_F^2 \quad (1)$$

$$\text{s.t. } \forall 1 \leq k \leq K, \|\mathbf{U}_k^v\|_1 = 1, \mathbf{U}^v, \mathbf{V}^v, \mathbf{V}^* \geq 0$$

1.2 Kuhn-Munkres 算法

二部图是一种有着广泛应用的图论模型。对于二部图 G 和它的 2 个大小相等的顶点集 V_1, V_2 , G 上的最优匹配是指一个 V_1, V_2 的顶点之间的一一对应, 使所有相互匹配的顶点对之间的边权总和最小。假设 \mathbf{W} 是图 G 的边权矩阵, $\mathbf{W}_{i,j}$ 表示顶点 i 与顶点 j 之间的边权; 定义指示变量 $\mathbf{P}_{i,j}$, 当顶点 i 与顶点 j 互相匹配时, $\mathbf{P}_{i,j} = 1$, 否则 $\mathbf{P}_{i,j} = 0$, 二部图最优匹配问题的目标函数为

$$\min_{\mathbf{P}_{i,j}} \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}_{i,j} \mathbf{W}_{i,j} \quad (2)$$

$$\text{s.t. } \sum_{j=1}^n \mathbf{P}_{i,j} = 1, \sum_{i=1}^n \mathbf{P}_{i,j} = 1, \mathbf{P}_{i,j} \in \{0,1\}, 1 \leq i, j \leq n$$

式(2)又称为线性指派问题, 是一类特殊的整数线性规划问题, 分别由 Kuhn 和 Munkres 在 20 世纪 50 年代独立解决, 一般把该算法称为 Kuhn-Munkres 算法。该算法能在多项式时间内精确求解二部图的最优匹配问题^[19]。

2 本文方法描述及其数值算法

2.1 数据降维的必要性分析

多视角数据通常具有较高的特征维度, 含有较多

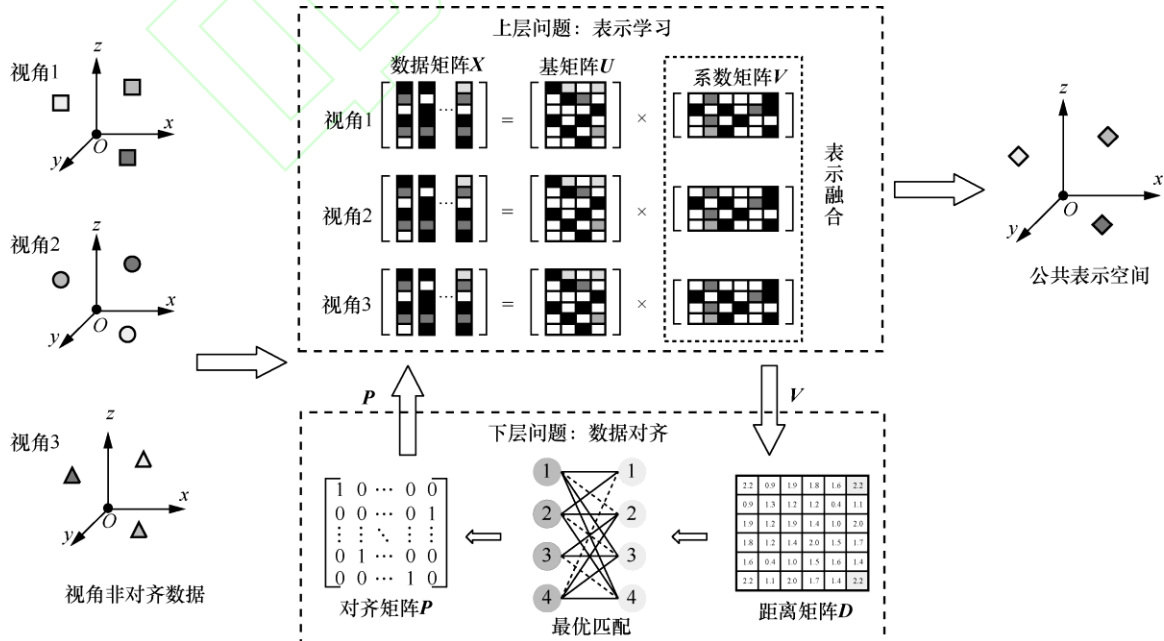


图2 本文视角非对齐数据表示学习方法的总体结构

冗余信息, 判别性较低, 以人脸数据集 ORL 的视角 1 为例加以说明。该数据集的判别性信息随奇异值的分布规律如图 3 所示。如图 3(a)所示, 尽管特征总数多达 400 个, 但其前 50 个特征值已包含了该数据集中超过 60% 的主要判别信息。如图 3(b)和图 3(c)所示, 与原始特征矩阵相比, 后 30% 小奇异值重建数据矩阵的取值较为均匀, 说明其中的判别性信息较少。因此, 对原始高维冗余数据进行降维, 保留其主要判别分量, 可以有效减少数据的冗余信息, 提高特征的判别性。



图 3 ORL 数据集视角 1 判别性信息分析

2.2 面向视角非对齐数据的多视角聚类方法

面向视角非对齐数据聚类, 前人的思路可以概

括为: 从多视角数据中学习潜在的视角对齐关系, 以指导模型从视角非对齐数据中学习有效的表示。在此基础上, 本文提出将学习到的表示转化为一种跨视角相似度信息并反馈到视角对齐模型中, 通过二者的相互促进, 以学习到的对齐关系指导表示学习模型尽可能地融合来自同类样本的视角信息, 以消除视角不对齐因素的影响, 改善表示特征的判别性, 进而提高聚类性能。

本文的思路来自两方面的观察。一方面, 对于数据对齐来说, 保持跨视角一致性的表示学习是必要的。在缺少监督信号的情况下, 表示特征之间的相似度信息是对齐关系学习的重要依据。但由于多视角特征的异构性, 需要先将原始多视角特征嵌入一个低维同构的表示空间, 才能有效地度量跨视角相似度。另一方面, 对于视角非对齐数据的表示学习来说, 对齐关系学习也是不可或缺的。由于视角非对齐现象的影响, 模型无法直接对多视角数据进行特征融合, 不能很好地利用多视角潜在一致性。因此, 需要为每个特征找到与之对齐的其他视角特征, 使它们对应相同的类簇, 才能确保特征融合的有效性。受此启发, 本文提出在 2 个模块之间建立双向数据通道, 以提高联合学习性能。

为了实现这一目的, 本文将上述 2 个模块嵌入基于文献[20]的 Bi-level 优化框架中进行交替优化, 使它们在迭代中相互促进, 以提升模型的总体性能。本文面向视角非对齐数据的统一表示学习和数据对齐模型为

$$\begin{aligned}
 \min_{\mathbf{U}^v, \mathbf{V}^v, \mathbf{P}^v, \mathbf{V}^*, \mathbf{V}^v} \quad & \sum_{v=1}^{N_v} \left\| \mathbf{X}^v \mathbf{P}^v - \mathbf{U}^v (\mathbf{V}^v)^T \right\|_F^2 + \sum_{i=1}^{N_v} \lambda_v \left\| \mathbf{V}^v - \mathbf{V}^* \right\|_F^2 \\
 \text{上层问题: 表示学习} \\
 \text{s.t.} \quad & \mathbf{P}^v = \arg \min_{\mathbf{P}} \text{Tr}(\mathbf{D}^v \mathbf{P}^T), \forall v \neq u, \mathbf{P}^u = \mathbf{I} \\
 & \mathbf{D}_{i,j}^v = d(\mathbf{V}_i^v, \mathbf{V}_j^u), \mathbf{P}^v \in \{0,1\}^{N \times N}, \mathbf{P}^v \mathbf{1} = \mathbf{1}, (\mathbf{P}^v)^T \mathbf{1} = \mathbf{1} \\
 \text{下层问题: 数据对齐}
 \end{aligned} \quad (3)$$

其中, $\mathbf{X} = \{\mathbf{X}^i\}_{i=1}^{N_v}$ 是具有 N_v 个视角的非对齐多视角数据集, λ_v 是权衡参数。 $\mathbf{U}^v, \mathbf{V}^v$ 分别是各视角的基矩阵和系数矩阵, \mathbf{V}^* 是公共表示矩阵。视角 u 是参考视角 (参考视角的概念将在下文说明), $\mathbf{P}^v, \mathbf{D}^v, v \neq u$ 分别是视角 v 相对于参考视角的对齐矩阵和距离矩阵。 \mathbf{V}_j^v 是原始特征 \mathbf{X}_j^v 的低维同构表示。 $d: \mathbb{R}^K \times \mathbb{R}^K \rightarrow \mathbb{R}^+$ 是一种距离函数, K 是低维表示的维度, \mathbb{R}^+ 是非负实数集。

2.3 上层问题：表示学习

如上文所述，非负矩阵分解是一种有效的降维技术，能滤除数据中的冗余信息，提高特征的判别性。此外，该技术学到的特征具有较好的跨视角一致性，在跨模态检索领域有着广泛应用^[21-22]。受此启发，本文基于该技术设计表示学习模块，以促进视角对齐关系学习并获得多视角公共表示。

在多视角非负矩阵分解中，基矩阵 U^v 捕获了原始特征空间中的潜在概念，而 V^v 则是这些潜在概念的叠加系数，从而把原始特征 X^v 嵌入一个以潜在概念为基底构成的坐标空间。由于不同视角所描述的样本是一致的，各视角的叠加系数 V^v 具有很高的相关性，对 V^v 施加一致性约束后，可视为原始特征的一种同构表示，具有跨视角一致性，即属于相同类簇的样本的不同视角的特征，在表示空间中也具有较高的相似性。

然而，传统的多视角非负矩阵分解只适用于视角对齐数据。对于视角非对齐数据，引入基于参考视角的对齐矩阵 $P^v, 1 \leq v \leq N_v, v \neq u$ (视角 u 是参考视角, $P^u = I$, I 是单位矩阵)。将 P^v 右乘于原始特征矩阵 X^v ，可调整 X^v 的视角对齐关系，再对对齐后的特征矩阵 $X^v P^v, 1 \leq v \leq N_v$ 进行多视角非负矩阵分解，得出表示学习的目标函数为

$$\min_{U^v, V^v, V^*} \sum_{v=1}^{N_v} \|X^v P^v - U^v (V^v)^T\|_F^2 + \sum_{i=1}^{N_v} \lambda_v \|V^v - V^*\|_F^2 \quad (4)$$

$$\text{s.t. } U^v, V^v, V^* \geq 0, \forall 1 \leq k \leq K, \|U_k^v\|_1 = 1$$

其中， λ_v 是权衡参数， P^v 是由数据对齐模块学习到的视角对齐矩阵。

2.4 下层问题：数据对齐

数据对齐模块的功能是从视角非对齐数据中学习潜在的视角对齐关系，为表示学习模块提供对齐矩阵 P^v 。本文提出一种基于参考视角的对齐方法，即选定一个参考视角，以之为基准逐个对齐其他视角，使所有视角都与参考视角保持对齐。该方法将一个 N_v 视角的对齐问题转化为 $N_v - 1$ 个两视角对齐的子问题。因此，可以先针对两视角对齐问题设计模型，再由参考视角扩展到多视角的情况。

由于不同视角特征的跨视角相似度越大，它们属于相同类簇的可能性也就越大，应该令所有相互匹配的特征对 (X_i^v, X_j^u) 之间的跨视角相似度总和最大。然而，由于原始特征的异构性， (X_i^v, X_j^u) 之

间的相似度无法度量，只能用相应的低维同构表示 (V_i^v, V_j^u) 之间的距离 $d(V_i^v, V_j^u)$ 来近似估计，由此得到两视角对齐模型的目标函数为

$$\min_C \sum_{(i,j) \in C} d(V_i^v, V_j^u) \quad (5)$$

$$\text{s.t. } C = \{(i, j) | X_i^v \text{ matches } X_j^u\}$$

其中， C 表示一种可能的特征匹配方案。注意到式(5)等价于一个二部图最优匹配模型。将 X^v, X^u 分别作为二部图的 2 个顶点集合 V_1, V_2 ，将距离 $d(V_i^v, V_j^u)$ 作为 2 个顶点之间的边权 $W_{i,j}$ ，将指示变量 $P_{i,j} \in \{0,1\}$ 表示特征对 (X_i^v, X_j^u) 是否匹配，可得到如式(2)所示的二部图最优匹配模型。

基于参考视角，可将式(5)推广至多视角情况。设 u 为参考视角， v 为任意视角，距离矩阵 D^v 为视角 v, u 特征之间的成对距离， $D_{i,j}^v = d(V_i^v, V_j^u)$ ，则求解视角 v 关于参考视角 u 的对齐矩阵 P^v 的目标函数为

$$P^v = \arg \min_P \text{Tr}(D^v P^T), \forall v \neq u \quad (6)$$

$$\text{s.t. } P \in \{0,1\}^{N \times N}, P \mathbf{1} = \mathbf{1}, P^T \mathbf{1} = \mathbf{1}$$

2.5 目标函数求解

目标函数式(3)是 Bi-level 优化问题，上层问题是非凸的，下层问题是整数规划问题，难以同时求解。本文采用交替优化策略，通过交替求解上下层问题来获得近似解。

2.5.1 上层问题求解

上层问题的目标函数如式(4)所示，该问题可通过交替优化变量 V^v, U^v, V^* 求解。为了消除对基向量 U_k^v 的 l_1 范数约束，引入辅助变量 Q^v

$$Q^v = \text{diag} \left(\sum_{i=1}^{M_v} U_{i,1}^v, \sum_{i=1}^{M_v} U_{i,2}^v, \dots, \sum_{i=1}^{M_v} U_{i,K}^v \right) \quad (7)$$

其中， $Q^v \in \mathbb{R}^{K \times K}$ 为对角矩阵， M_v 为视角 v 的特征维数， K 为低维表示的维数。借助辅助变量 Q^v ，把式(4)等价变形为

$$\min_{U^v, V^v, V^*} \sum_{v=1}^{N_v} \|X^v P^v - U^v (V^v)^T\|_F^2 + \sum_{i=1}^{N_v} \lambda_v \|V^v Q^v - V^*\|_F^2 \quad (8)$$

$$\text{s.t. } U^v, V^v, V^* \geq 0$$

当 V^* 固定时，对于任意的 2 个视角 v_1, v_2 ($v_1 \neq v_2$)， U^{v_1}, V^{v_1} 的求解不依赖于 U^{v_2}, V^{v_2} ，

因此对固定的 v 求解 U^v -子问题或者 V^v -子问题时, 式(8)可化简为 (为简便起见, 省略上标和下标 v)

$$\min_{U,V} \|XP - UV^T\|_F^2 + \lambda \|VQ - V^*\|_F^2 \quad \text{s.t. } U, V \geq 0 \quad (9)$$

从式(9)可将目标函数拆成如下的子问题。

U-子问题。 将式(9)用矩阵的迹展开并引入拉格朗日乘子 Φ , 得到无约束拉格朗日函数

$$L_1 = \text{Tr}(UV^T VU^T - 2XPVU^T) + \lambda R_1 + \text{Tr}(\Phi^T U) \quad (10)$$

$$R_1 = \text{Tr}(VQQ^T P^T - 2VQ(V^*)^T)$$

对 R_1 求导得

$$C_{i,k} = \frac{\partial R_1}{\partial U_{i,k}} = 2 \left(\sum_{l=1}^M U_{l,k} \sum_{j=1}^N V_{j,k}^2 - \sum_{j=1}^N V_{j,k} V_{j,k}^* \right) \quad (11)$$

对 L_1 求导并令其导数等于 0, 结合 KKT (Karush-Kuhn-Tucker) 条件得

$$\begin{cases} \frac{\partial L_1}{\partial U} = -2XPV + 2UV^T V + \lambda C + \Phi = 0 \\ \Phi_{i,k} U_{i,k} = 0, \forall 1 \leq i \leq d, 1 \leq k \leq K \end{cases} \quad (12)$$

从式(12)中解得 U 的更新规则为

$$U_{i,k} \leftarrow U_{i,k} \frac{(XPV)_{i,k} + \lambda \sum_{j=1}^N V_{j,k} V_{j,k}^*}{(UV^T V)_{i,k} + \lambda \sum_{l=1}^M U_{l,k} \sum_{j=1}^N V_{j,k}^2} \quad (13)$$

V-子问题。 由于式(9)中 Q 的作用是对 U 进行 l_1 范数归一化, 而 U 在 V -子问题中是常量, 因此可通过对 U, V 进行显式的归一化来消除辅助变量 Q

$$U \leftarrow UQ^{-1}, V \leftarrow VQ \quad (14)$$

由于式(14)的归一化不改变式(9)中的目标函数值, 因此在显式归一化 U, V 后, V -子问题可化简为

$$\min_{V,V^*} \|XP - UV^T\|_F^2 + \lambda \|V - V^*\|_F^2 \quad \text{s.t. } V \geq 0 \quad (15)$$

将式(15)用矩阵的迹展开并引入拉格朗日乘子 Ψ , 得到无约束拉格朗日函数

$$L_2 = \text{Tr}(UV^T VU^T - 2XPVU^T) + \lambda R_2 + \text{Tr}(\Psi^T V) \quad (16)$$

$$R_2 = \text{Tr}(VV^T - 2V(V^*)^T)$$

对 L_2 求导并令其导数等于 0, 结合 KKT 条件得

$$\begin{cases} \frac{\partial L_2}{\partial V} = 2VU^T U - 2P^T X^T U + 2\lambda(V - V^*) + \Psi = 0 \\ \Psi_{j,k} V_{j,k} = 0, \forall 1 \leq j \leq N, 1 \leq k \leq K \end{cases} \quad (17)$$

从式(17)解得 V 的更新规则为

$$V_{j,k} \leftarrow V_{j,k} \frac{(P^T X^T U)_{j,k} + \lambda V_{j,k}^*}{(VU^T U)_{j,k} + \lambda V_{j,k}} \quad (18)$$

V^* -子问题。 对于式(9), 只考虑与 V^* 有关的部分, 可得到目标函数为

$$\min_{V^*} \sum_{i=1}^M \lambda_i \|V^v Q^v - V^*\|_F^2 \quad \text{s.t. } V^* \geq 0 \quad (19)$$

式(19)是一个最小二乘问题, 求导并令导数等于 0, 可解得 V^* 的闭式解为

$$V^* = \frac{\sum_{v=1}^{N_v} \lambda_v V^v Q^v}{\sum_{v=1}^{N_v} \lambda_v} \geq 0 \quad (20)$$

2.5.2 下层问题求解

式(6)是一类特殊的整数线性规划问题, 可采用文献[19]中多项式时间的精确算法对其进行求解。本文的数值求解算法如算法 1 所示。

算法 1 目标函数式(3)的数值求解算法

输入 视角非对齐数据 $\{X^v\}_{v=1}^M$, 低维公共表示的维度 K 和权衡参数 $\{\lambda_v\}_{v=1}^M$

输出 多视角公共表示矩阵 V^* , 聚类结果 Y

- 1) 归一化: $\|X^v\|_1 = 1, 1 \leq v \leq N_v$;
- 2) 初始化: $V^v, U^v, P^v, V^*, 1 \leq v \leq N_v, P^u = I$;
- 3) while
- 4) while // 表示学习
- 5) while
- 6) for $v=1: N_v$
- 7) 根据式(13)更新 U^v ;
- 8) 根据式(14)归一化 U^v, V^v ;
- 9) 根据式(18)更新 V^v ;
- 10) end for
- 11) end while
- 12) 根据式(20)更新 V^* ;
- 13) end while
- 14) for $v=2: N_v$ // 数据对齐
- 15) 根据式(6)更新 P^v ;
- 16) end for
- 17) end while
- 18) 对 V^* 实施 K-means 算法获得聚类结果 Y

3 实验结果及分析

3.1 数据集描述

本节在 3 个公开数据集上进行多视角聚类实

验, 数据集的具体情况阐述如下。

MSRC-v1 数据集。该数据集是一个物体图像数据集, 选择其中 7 个类别的样本, 共 210 张图像。提取 5 种不同的特征作为 5 个视角, 分别为色矩 (CM, color moment) 特征、局部二进制模式 (LBP, local binary pattern) 特征、CENTRIST 特征、GIST 特征以及 SIFT 特征。

ORL 数据集。该数据集是一个人脸图像数据集, 由 400 张人脸图像组成。提取 4 种不同的特征作为 4 个视角, 分别为强度特征 (Intensity)、LBP 特征、Gabor 特征以及灰度值特征 (Gray)。

Digits 数据集。该数据集包含 10 种手写数字的图像, 样本总数为 2 000。提取 3 种不同的特征作为 3 个视角, 分别为剖面相关特征、傅里叶系数特征以及形态学特征。

3.2 对比实验设置及实验结果分析

由于文献[17]中的部分视角对齐聚类方法 (PVC) 和文献[18]中的多视角稳健对比学习方法

(MvCLN) 仅能处理具有 2 个视角的数据聚类问题, 因此本文分别在两视角和 2 个以上视角这 2 种情况下进行对比实验。由于视角非对齐现象可以看作一类针对多视角数据的复杂噪声干扰, 为了尽可能的公平, 在包含 2 个以上视角的聚类实验中, 本文选取了若干对一般性噪声稳健的先进传统多视角聚类方法进行比较, 使对比实验能够在一定程度上客观地说明本文方法的优势。2 个以上视角情况下的对比方法包括: 文献[12]中的基于图的多视角聚类方法 (GMC)、文献[11]中的非一致图融合方法 (DGF)、文献[10]中的自加权多视角聚类方法 (SwMC)、文献[16]中的基于关键张量学习的多视角谱聚类方法 (ETLMSC)、文献[9]中的无参多图学习方法 (AMGL) 和文献[13]中的多视角非负矩阵分解方法 (MultiNMF)。

对于上述 6 种对比方法, 首先在常用的三视角设定下进行实验, 分别采用 MSRC-v1 和 ORL 的前 3 个视角及 Digits 的全部视角, 实验结果如图 4(a)~图 4(c)所示。其次, 与 PVC^[17]和 MvCLN^[18]在上述

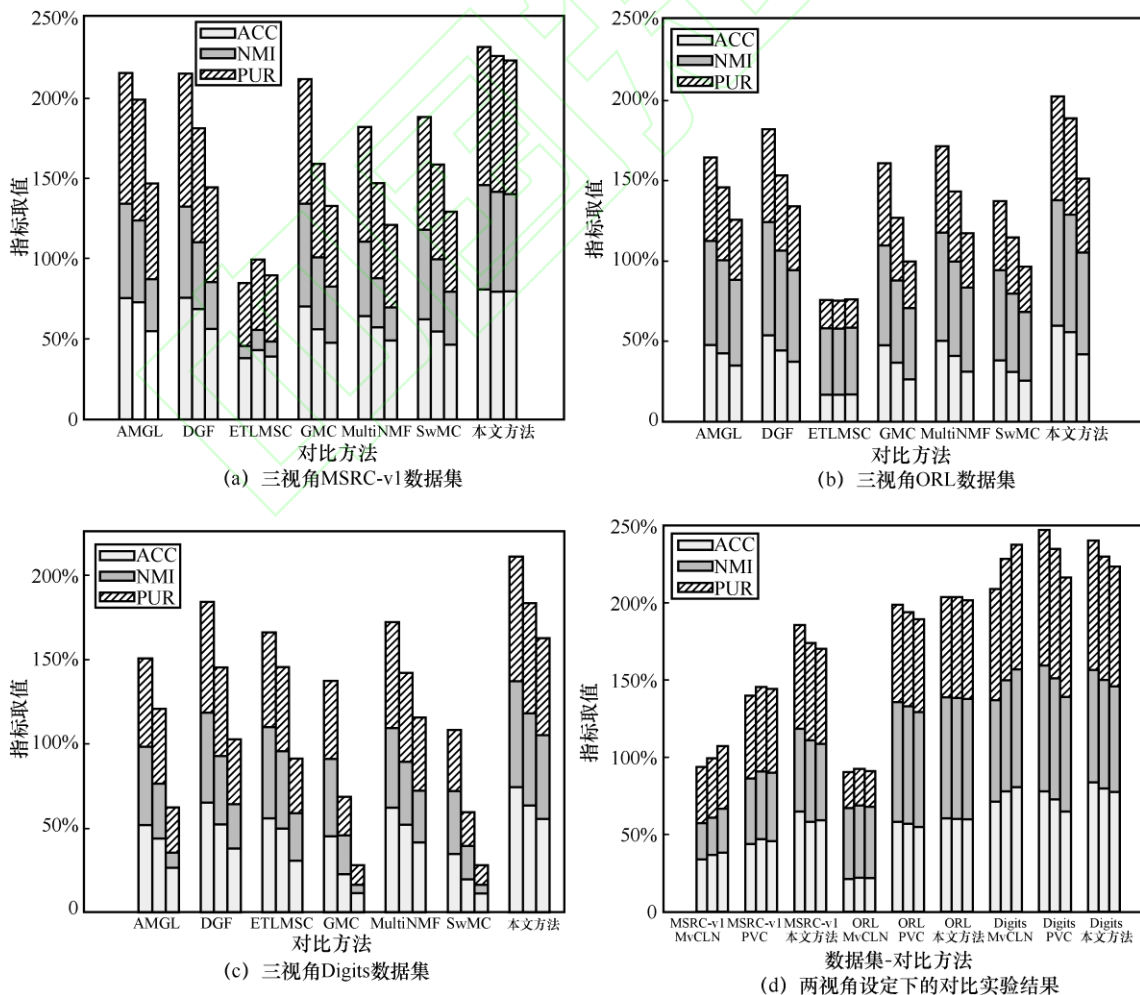


图4 本文方法在三视角和两视角设定下的对比实验结果

数据集的两视角子集上进行对比实验, 分别采用 MSRC-v1 的 CM 和 CENTRIST 视角, ORL 的 LBP 和 Gabor 视角以及 Digits 的傅里叶系数和剖面特征视角, 实验结果如图 4(d)所示。最后, 为了证明本文方法对一般多视角数据集的有效性, 分别在 MSRC-v1 和 ORL 的全部视角上与上述 6 种传统多视角聚类方法进行对比实验, 实验结果如图 4(d)所示。实验中设置了 3 种视角不对齐率 (0.3、0.5 和 0.7), 采用准确率 (ACC, accuracy)、归一化互信息 (NMI, normalized mutual information) 和纯度 (PUR, purity) 作为评价指标, 所有实验均重复 5 次, 取其均值作为最终结果。对本文方法的参数设定如下: 参考文献[13,14]中的做法, 将表示学习中各视角的权衡参数 λ_v 设置为相同值 λ ; 将参考视角 u 设置为视角 1; 取高斯核函数作为距离函数, 即 $d(\mathbf{x}, \mathbf{y}) = -\exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right)$, 其中 σ 为高斯核的尺度参数, 实验中取 $\lambda = 0.01, \sigma = 1$ 。

如图 4(a)~图 4(c)所示, 在绝大多数情况下, 本文方法的各项性能指标都超过了其他对比方法。以不同不对齐率下的指标均值来考量对比方法的平均性能, 本文方法在 MSRC-v1、ORL、Digits 这 3 个数据集上, ACC 指标依次超过次优方法 7%、7% 和 13%; NMI 指标依次超过次优方法 8%、8% 和 14%; PUR 指标依次超过次优方法 7%、8% 和 13%。当不对齐率逐渐增大时, 其他对比方法的性能下降趋势明显, 而本文方法的性能变化趋势相对平缓。在 MSRC-v1 数据集上, 当不对齐率逐渐增大时, 本文方法的各项指标几乎保持不变。在 Digits 数据集上, 本文方法的性能下降趋势更平缓, 例如当不对齐率从 0.5 上升至 0.7 时, 对比方法的 NMI 指标平均下降了 16.4%, 而本文方法的 NMI 指标只下降了 5.1%。

本文方法与 PVC、MvCLN 方法的比较结果如图 4(d)所示。可以看到, 本文方法在 MSRC-v1 和 ORL 数据集的评价指标上的性能均优于 PVC 和 MvCLN。在 Digits 数据集上, ACC 指标取得了几乎最优的性能, NMI 和 PUR 这 2 个指标在部分实验条件下取得了次优的性能, 说明本文方法有效改进了该领域的前人工作。本文方法在四、五视角设定下的实验结果如图 5 所示。可以看到, 各项性能指标在四视角 ORL 数据集和五视角 MSRC-v1 数据集上的变化趋势与在相应的三视角子集上基本一致, 说明本文方法对一般的多视角数据集同样有效。

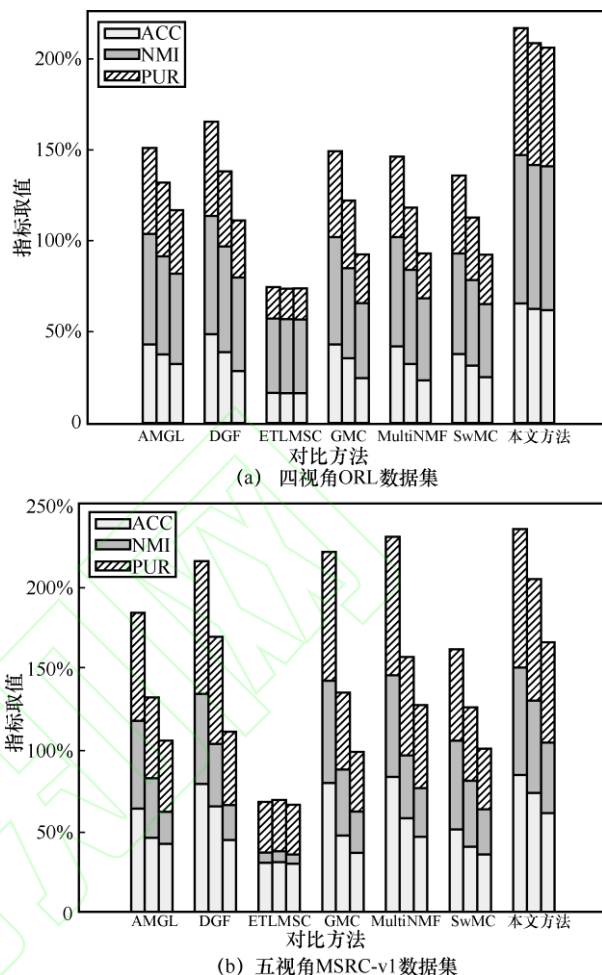


图 5 本文方法在四、五视角设定下的对比实验结果

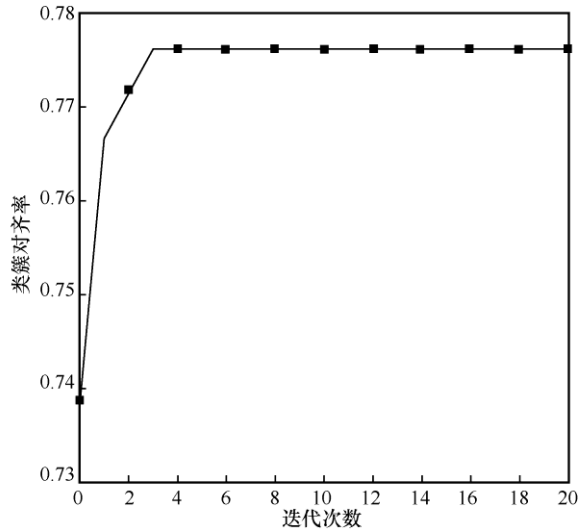
综上所述, 本文方法在多个数据集、多种视角数量和多种不对齐率下的实验中都取得了最优性能。上述实验结果证明了本文方法框架对视角非对齐数据表示学习及其聚类应用的有效性和稳健性。

3.3 数据对齐模块的有效性验证

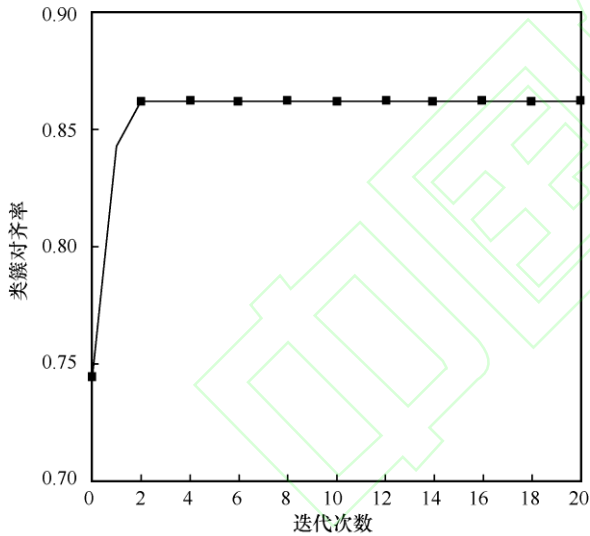
为了验证数据对齐模块的有效性, 选取类簇对齐率作为视角对齐程度的评价指标, 并绘制不同视角的类簇对齐率随迭代次数变化的曲线, 这里展示在不对齐率为 0.3 时 MSRC-v1 数据集上的实验结果。如图 6 所示, 在初始状态下, 视角 2 和视角 3 的类簇对齐率处于较低水平 (仅略高于数据本身的样本对齐率 0.7)。随着迭代数的增加, 对齐率显著提高, 最后稳定在高于初始状态 4.5%~12% 的水平。

注意到图 6 的曲线变化趋势与图 7(b) 的收敛曲线的变化趋势基本一致: 经过约 3 轮迭代后, 数值算法开始收敛, 数据集对齐程度趋于饱和。这里提出一种提高该饱和程度的可能策略: 在迭代过程中动态地调整一致性约束项的权衡系数 λ 。理由如

下：在算法迭代初期，数据不对齐程度较高，不应强制 \mathbf{V}^v 趋于一致，而应设置较小的 λ 值；在迭代后期，数据对齐程度有所提高，可加快 \mathbf{V}^v 融合以提高其跨视角一致性，应适当增大 λ 值。



(a) 视角2



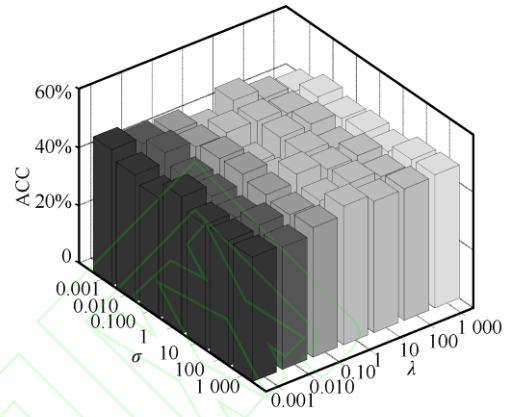
(b) 视角3

图6 视角2和视角3在MSRC-v1数据集上的类簇对齐率变化曲线

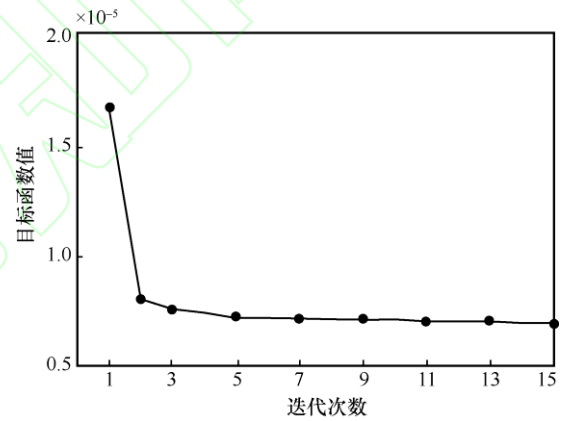
3.4 参数稳定性与收敛性分析

本文方法共包括2个参数：表示学习中的权衡参数 λ 和数据对齐中的高斯核尺度参数 σ 。本文按照 $\{0.001, 0.01, 0.1, 1, 10, 100, 1000\}$ 的变化范围进行网格搜索，分析聚类性能对不同参数组合的敏感程度。图7(a)是在不对齐率为0.7的MSRC-v1数据集上，ACC指标关于参数 λ 和 σ 的变化分布。可以看出，在较大的参数范围内，性能指标的变化幅度相对较小，说明本文方法具有较好的参数稳定性，从而具有较好的易用性。

为了证明本文数值算法的收敛性，以迭代次数和目标函数值分别为横、纵坐标绘制收敛曲线。在不对齐率为0.7的MSRC-v1数据集上的实验结果如图7(b)所示。从图7中可以看出，本文数值算法具有较好的收敛性能。



(a) ACC-参数分布情况



(b) 数值算法的收敛曲线

图7 参数稳定性与收敛性分析

4 结束语

面向视角非对齐数据的多视角学习问题，本文提出了一种基于多视角非负矩阵分解的学习方法。在3个公开数据集、多种不同的视角数量和3种不对齐率上，本文方法的聚类性能均优于8种先进的多视角聚类方法。在此基础上，后续研究可考虑引入核技巧扩展本文的表示学习模型，增强其非线性适应能力。此外，在视角非对齐设定下，表示学习的稳定性及相关参数的影响分析也是值得探索的后续研究方向。

参考文献：

- [1] ZHAO J, XIE X J, XU X, et al. Multi-view learning overview: recent progress and new challenges[J]. Information Fusion, 2017, 38: 43-54.
- [2] CHAO G Q, SUN S L, BI J B. A survey on multi-view clustering[J].

- IEEE Transactions on Artificial Intelligence, 2021, 2(2): 146-168.
- [3] KANG Z, LIN Z, ZHU X, et al. Structured graph learning for scalable subspace clustering: from single view to multiview[J]. IEEE Transactions on Cybernetics, 2021: doi.org/ 10.1109/TCYB.2021.3061660.
- [4] HUANG A P, CHEN W L, ZHAO T S, et al. Joint learning of latent similarity and local embedding for multi-view clustering[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2021, 30: 6772-6784.
- [5] HUANG Z M, REN Y Z, PU X R, et al. Dual self-paced multi-view clustering[J]. Neural Networks, 2021, 140: 184-192.
- [6] TAO Z Q, LIU H F, LI S, et al. Marginalized multiview ensemble clustering[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(2): 600-611.
- [7] GAO Q X, XIA W, WAN Z Z, et al. Tensor-SVD based graph learning for multi-view subspace clustering[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 3930-3937.
- [8] 李鹭, 王卓, 于晓洋, 等. 多核低冗余表示学习的稳健多视图子空间聚类方法[J]. 通信学报, 2021, 42(11): 193-204.
- LI A, WANG Z, YU X Y, et al. Robust multiview subspace clustering method based on multi-kernel low-redundancy representation learning[J]. Journal on Communications, 2021, 42(11): 193-204.
- [9] NIE F, LI J, LI X. Parameter-free auto-weighted multiple graph learning: a framework for multiview clustering and semi-supervised classification[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. [S.l.:s.n.], 2016: 1881-1887.
- [10] NIE F P, LI J, LI X L. Self-weighted multiview clustering with multiple graphs[C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. [S.l.:s.n.], 2017: 2564-2570.
- [11] LIANG Y W, HUANG D, WANG C D. Consistency meets inconsistency: a unified graph learning framework for multi-view clustering[C]//Proceedings of 2019 IEEE International Conference on Data Mining. Piscataway: IEEE Press, 2019: 1204-1209.
- [12] WANG H, YANG Y, LIU B. GMC: graph-based multi-view clustering[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 32(6): 1116-1129.
- [13] LIU J L, WANG C, GAO J, et al. Multi-view clustering via joint nonnegative matrix factorization[C]//Proceedings of the 2013 SIAM International Conference on Data Mining. Philadelphia: Society for Industrial and Applied Mathematics, 2013: 252-260.
- [14] 张伟, 孔祥维, 王振帆, 等. 基于多视图矩阵分解的聚类分析[J]. 自动化学报, 2018, 44(12): 2160-2169.
- ZHANG Y, KONG X W, WANG Z F, et al. Matrix factorization for multi-view clustering[J]. Acta Automatica Sinica, 2018, 44(12): 2160-2169.
- [15] ZHANG C Q, FU H Z, LIU S, et al. Low-rank tensor constrained multiview subspace clustering[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2015: 1582-1590.
- [16] WU J L, LIN Z C, ZHA H B. Essential tensor learning for multi-view spectral clustering[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2019, 28(12): 5910-5922.
- [17] HUANG Z Y, HU P, ZHOU J T, et al. Partially view-aligned clustering[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: Curran Associates Inc., 2020: 2892-2902.
- [18] YANG M X, LI Y F, HUANG Z Y, et al. Partially view-aligned representation learning with noise-robust contrastive loss[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 1134-1143.
- [19] CUI H, ZHANG J J, CUI C F, et al. Solving large-scale assignment problems by Kuhn-Munkres algorithm[C]//Proceedings of the 2nd International Conference on Advances in Mechanical Engineering and Industrial Informatics (AMEII 2016). Paris: Atlantis Press, 2016: 822-827.
- [20] COLSON B, MARCOTTE P, SAVARD G. An overview of bilevel optimization[J]. Annals of Operations Research, 2007, 153(1): 235-256.
- [21] KANG P P, LIN Z H, YANG Z G, et al. Intra-class low-rank regularization for supervised and semi-supervised cross-modal retrieval[J]. Applied Intelligence, 2022, 52(1): 33-54.
- [22] ZHANG D L, WU X J. Robust and discrete matrix factorization hashing for cross-modal retrieval[J]. Pattern Recognition, 2022, 122: 108343.

作者简介



李鹭(1986—), 男, 黑龙江哈尔滨人, 博士, 哈尔滨理工大学副教授、博士生导师, 主要研究方向为模式识别、机器学习等。

冯聪(1997—), 男, 广东佛山人, 哈尔滨理工大学硕士生, 主要研究方向为机器学习、数据挖掘等。

牛宇童(1997—), 男, 黑龙江哈尔滨人, 哈尔滨理工大学硕士生, 主要研究方向为机器学习、视觉目标检测等。

徐士彪(1985—), 男, 吉林舒兰人, 博士, 北京邮电大学教授, 主要研究方向为在三维计算机视觉、模式识别、深度学习、遥感图像处理及虚拟现实等。

张英涛(1979—), 女, 黑龙江哈尔滨人, 博士, 哈尔滨工业大学副教授, 主要研究方向为人工智能与信息处理等。

孙广路(1979—), 男, 黑龙江哈尔滨人, 博士, 哈尔滨理工大学教授、博士生导师, 主要研究方向为机器学习、网络安全等。