

基于邻域多核学习的后融合多视图聚类算法

夏冬雪^{1,2} 杨燕¹ 王浩¹ 阳树洪²

¹(西南交通大学信息科学与技术学院 成都 611756)

²(广西科技大学计算机科学与通信工程学院 广西柳州 545006)

(vdx_swjtu@126.com)

Late Fusion Multi-View Clustering Based on Local Multi-Kernel Learning

Xia Dongxue^{1,2}, Yang Yan¹, Wang Hao¹, and Yang Shuhong²

¹(School of Information Science and Technology, Southwest Jiaotong University, Chengdu 611756)

²(School of Computer Science and Communication Engineering, Guangxi University of Science and Technology, Liuzhou, Guangxi 545006)

Abstract Graph-based multi-view clustering is one of the representative methods in that field. However, existing models still have problems as following. First, most of them do not consider the difference of clustering capacity among different views and force all views to share a common similarity graph. Next, some models construct the similarity graph and conduct clustering in separated steps, resulting in the constructed similarity graph is not optimal for the following clustering tasks. Finally, although there are many models using kernel learning to deal with the nonlinear relationship between data points, most of them calculate the self-expressive relationship in kernel space based on global models. Such global schemes are not conducive to fully explore local nonlinear relationship, and easy to bring about heavy computing load. Therefore, this paper proposes a late fusion multi-view clustering model based on local multi-kernel learning. We implement information fusion at the level of class partition space rather than similarity graph, and adopt local multi-kernel learning scheme to fully preserve the local nonlinear relationship as well as reduce the computational load. We also propose an alternative optimization scheme to solve the construction of similarity graph, combination of multi-kernel and generation of class indicator matrix in a unified framework. Experiments on multiple datasets show that the proposed method has good multi-view clustering effect.

Key words local structure; multi-kernel learning; spectral rotation; late fusion; multi-view clustering

摘要 基于图谱理论的多视图聚类是该领域的代表性方法之一。然而,现有模型尚存在3个问题。1)这类方法大多没有考虑不同视图之间的聚类性能差异,强制要求所有视图共享一个公共相似图;2)部分模型将相似图构建和聚类分步进行,导致所构建的相似图对于聚类任务并非最优;3)虽已有若干模型采用核学习处理数据间的非线性关系,但大多基于全局模型计算数据在核空间中的自表达关系,不利于充分挖掘局部非线性信息,且易带来沉重的计算负荷。为了应对以上问题,提出一种基于邻域多核学习的后融合多视图聚类算法,在类划分空间而不是数据相似图的层次进行信息融合,采用邻域多核学习方案在

收稿日期:2020-03-20;修回日期:2020-05-15

基金项目:国家自然科学基金项目(61976247,61572407)

This work was supported by the National Natural Science Foundation of China (61976247, 61572407).

通信作者:杨燕(yyang@swjtu.edu.cn)

充分保留局部非线性关系的同时减轻计算负荷,并提出一种交替优化方案将相似图构建、多核组合、类指示矩阵生成等子任务在统一的框架下进行协同优化.多个数据集上的实验表明:该算法具有良好的多视图聚类效果.

关键词 邻域结构;多核学习;谱旋转;后融合;多视图聚类

中图法分类号 TP301

机器学习的许多实际问题中,研究对象常由不同来源的数据或不同特征进行刻画.例如不同机构对同一新闻事件的各自报道,同一图像的不同特征表示,同一故事的视频或文字描述.类似场景中,从不同视角对同一对象集合进行描述的数据,称之为多视图数据^[1].多视图学习利用多视图数据来获得对象的全面理解,可以克服单视图数据可能导致的偏差或不足,因而成为近年的研究热点^[2].同时,在当前大数据时代,数据产生和收集的速度迅猛增长,人工标注变得非常昂贵和不切实际,聚类分析作为一种重要的无监督学习形式,日益受到研究者的广泛关注.本文主要关注多视图聚类.已有研究表明:多视图聚类能够充分利用视图之间的互补性和一致性,从而产生比单视图聚类更为精确和健壮的数据划分^[3-4].

现有的多视图聚类算法大致包括典型相关分析、多视图矩阵分解、图谱聚类等类别^[1].近年来,许多基于图谱理论的算法被陆续提出并取得良好效果^[3,5-7].这些方法通常以分步进行的方式完成多视图聚类,即首先构造每个视图的相似图 S (similarity graph),然后学习所有视图的公共相似图,最后对得到的公共相似图进行聚类分析^[3,6].

上述基于图谱理论的多视图聚类算法在许多场景中被证明是有效的,但仍存在 3 个缺陷.1) 信息融合可在原始数据级、特征级和决策级进行,构建公共相似图的算法本质上是特征级的信息融合^[3,7].然而,从相似图的性质来看,一者不同视图描述对象的视角、聚类能力存在较大差异^[8];二来由于实际应用中广泛存在的信息缺失、噪声等因素容易造成某个视图的相似图失真,强制所有视图共享一个公共相似图可能导致最终聚类结果不理想^[6].2) 目前基于图谱理论的多视图聚类多以分步形式完成:即先构造公共相似图,然后据此进行聚类.而不是在一个统一的过程中优化这 2 个任务,从而可能导致额外的 PAC(probably approximately correct) 边界^[3,9].3) 对于数据中广泛存在的非线性关系,尽管已有若干工

作采用核方法进行有效处理,但大多基于全局自表达学习的框架进行核变换^[10],基于局部结构进行多核学习的方法仍鲜见报导.然而,局部结构对于聚类分析和非线性关系建模有重要意义^[11-12].具体表现包括:在足够小的邻域范围内,样本间非线性关系通常能用局部线性嵌入一阶逼近,样本的聚类标签可用近邻回归进行预测^[12-13];在稍大的局部范围内,虽然低维输入空间中数据样本分布呈现出非线性流形关系,但若将其嵌入到高维特征空间中,则数据样本在高维空间中的分布仍与欧氏空间局部同胚,从而在局部范围内仍可借鉴欧氏空间的相关理论方法进行建模与分析^[13].

因此,本文提出一种基于邻域多核学习的后融合多视图聚类算法(local multi-kernel learning based late fusion multi-view clustering, LMLFMC).其主要贡献包括 3 个方面:

- 1) 与基于全局自表达的核方法不同,本文仅考虑近邻数据,无需学习整个数据集的自表达关系,从而在保持局部非线性结构的同时减轻计算负荷.
- 2) 尽管各视图的相似图可能存在差异,但不同视图的数据是从不同角度描述同一个样本集合,其簇类划分结构是跨越不同视图的全局结构,本文模型将信息融合推迟到数据划分空间进行,因而比基于相似图融合的多视图聚类模型更加鲁棒.
- 3) 本文模型在一个统一的框架下,对多核组合方式、各个视图的相似图构造及最终的簇类结构划分进行协同优化,从而使得上述子任务能以相互促进的方式迭代提升模型整体性能.

1 相关工作

由于多视图数据的广泛应用,近年来,许多将基于图谱理论的聚类模型扩展到多视图的算法被不断提出^[7,14-17].Saha 等人^[14]首先构造视图的子空间表示,然后在公共子空间实现多视图聚类,但没有考虑不同视图的权重.针对此问题,Nie 等人^[7]首先生成

每个视图的相似图,然后以不同权重将其整合得到一个公共相似图,再在公共相似图上进行聚类分析.后续若干算法对此进行持续研究,提出一系列更为有效的方法^[15-16].这些方法并不需要附加的 K -means 步骤,但仍以分步进行的方式完成信息融合:即首先生成各视图的相似图,保持不变,然后再进行信息融合.此外,上述基于图谱理论的多视图聚类方法强制所有视图统一到一个公共相似图,而该公共相似图并没有以最终的簇类划分为目标进行优化,从而导致对后者而言,该公共相似图并非最优.为了减轻相似度矩阵的计算负担,Wang 等人^[17]采用数据簇相似度矩阵计算数据点和簇中心之间、而不是对数据点之间的相似度,尽管该方法可以提高计算效率,但仍需一个附加的聚类步骤来获得最终的簇类划分.

本文模型采用谱旋转实现多视图信息的后融合.Yu 等人^[18]首次采用谱旋转实现多类别划分.Zelnikmanor 等人^[19]在此基础上提出自适应谱聚类.Huang 等人^[20]综合比较了谱聚类框架下的谱旋转和 K -means 算法,提出了一种基于谱旋转的谱聚类算法.Nie 等人^[6]用谱旋转代替 K -means 聚类步骤,将文献^[20]中的算法推广到多视图聚类,但仍以分步方式完成多视图聚类,即先构造各视图的相似图,在后续的聚类过程中这些相似图保持固定,这可能导致所构造的相似图对于后续的聚类任务并非最优.

核方法由于能有效建模数据点之间的非线性关系而被广泛应用到聚类分析中.Scholkopf 等人^[21]提出了核 K -means 聚类算法,Zhang 等人^[22]将其扩展为基于核 PCA 的通用核学习框架.Langone 等人^[23]将核方法引入到谱聚类.上述基于单个核函数的模型的性能严重依赖于核函数的选择.然而,从预定义的函数库中选择最佳核函数非常耗时甚至不切实际.为解决这一问题,Kang 等人^[24]采用多核学习对多个核函数进行优化加权,从而无需预先选择最佳核函数.此后,Huang 等人^[10]提出一种基于多核学习的多视图聚类算法.已有研究表明:局部结构对于聚类分析和非线性关系学习有重要意义^[11-13].但上述基于核方法的聚类模型,均基于全局自表达学习方案,即在核空间中学习所有样本对于特定样本的全局线性表达,该方法不利于充分挖掘数据间的局部结构,且对于在高维空间进行表达学习的核方法而言,由于需要计算所有数据点的表达系数,从而带来巨大的计算负荷.

2 本文模型

本节介绍本文提出的 LMLFMC 模型.后文采用斜粗体大/小写分别代表矩阵和向量,如 \mathbf{X}, \mathbf{x} . \mathbf{X} 的第 i 行和第 j 列分别用 $\mathbf{X}_{i\cdot}$ 和 $\mathbf{X}_{\cdot j}$ 表示.矩阵 \mathbf{X} 的元素用相应的斜体小写字母加下标表示如 x_{ij} . $\|\mathbf{X}\|_F$ 和 $\|\mathbf{x}\|$ 分别代表 \mathbf{X} 的 F 范数和 \mathbf{x} 的 l_2 范数.黑体 $\mathbf{1}$ 代表元素全为 1 的向量.用 $\mathbf{Ind} \triangleq \{\mathbf{Y} \in \{0, 1\}^{n \times c} \mid \mathbf{Y} \times \mathbf{1} = \mathbf{1}\}$ 代表类别指示矩阵.多视图数据的第 v 个视图,视上下文方便用上标 \mathbf{X}^v 或下标 \mathbf{X}_v 进行标注.

2.1 核空间中的邻接矩阵构造

对于给定的数据集 $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \in \mathbb{R}^{m \times n}$,相似图(或称之为邻接矩阵) \mathbf{S} 最常见的构建方法是 k 近邻图,即将 \mathbf{x}_i 与距其最近的 k 个数据点进行连接,两点之间边的权重由高斯核函数 $s_{ij} = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$ 确定.该方法的问题在于,实际应用中受噪声和离群点等因素影响,超参 σ 难以确定.Wright 等人^[25]指出,稀疏表达对噪声和离群点鲁棒.通常假定,两点之间边的权重应和距离成反比,故本文采用稀疏表达构建数据矩阵的相似图 $\mathbf{S}^{[3]}$,即:

$$\begin{aligned} \min_{\mathbf{S}} \sum_{i,j=1}^n \|\mathbf{x}_i - \mathbf{x}_j\|^2 s_{ij} + \zeta \|\mathbf{S}_i\|_1, \quad (1) \\ \text{s.t. } s_{ij} \geq 0. \end{aligned}$$

对邻接矩阵 \mathbf{S} 进行归一化,使得 $\mathbf{S}^T \mathbf{1} = \mathbf{1}$,则式(1)的第 2 项成为常数项,即此处的归一化操作等价于对 \mathbf{S} 进行稀疏化约束.故问题式(1)可改写为

$$\begin{aligned} \min_{\mathbf{S}} \sum_{i,j=1}^n \|\mathbf{x}_i - \mathbf{x}_j\|^2 s_{ij}, \quad (2) \\ \text{s.t. } s_{ij} \geq 0, \mathbf{S}^T \mathbf{1} = \mathbf{1}. \end{aligned}$$

但问题式(2)有平凡解:即对 $t = \arg \min_j \|\mathbf{x}_i - \mathbf{x}_j\|^2$ 的点 $\mathbf{x}_t, s_{it} = 1$;其他所有 $j \neq t$ 的点 \mathbf{x}_j 对应的 s_{ij} 都为零.因此,本文对问题式(2)加入一个先验约束^[3,24]:

$$\begin{aligned} \min_{\mathbf{S}} \sum_{i,j=1}^n \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 s_{ij} + \xi \|\mathbf{S}\|_F^2, \quad (3) \\ \text{s.t. } s_{ij} \geq 0, \mathbf{S}^T \mathbf{1} = \mathbf{1}. \end{aligned}$$

上述邻接矩阵 \mathbf{S} 的构建基于原始输入空间的欧氏距离.采用该距离测度进行聚类,通常要求数据集本身线性可分^[26],而该条件在真实数据上往往难以满足.为克服该局限性,本文用核技巧将输入

空间的数据映射到高维特征空间,以提高数据的可分离性。

具体地,令 $\Phi: \mathbb{R}^m \rightarrow \mathbf{H}$ 为从原始输入空间 \mathbb{R}^m 到核空间 \mathbf{H} 的非线性映射,则包含 n 个样本的原始输入数据集 $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \in \mathbb{R}^{m \times n}$ 可以映射到 $\Phi(\mathbf{X}) = (\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_n))$. 定义半正定核函数矩阵 $\mathbf{K} = \Phi(\mathbf{X})^T \Phi(\mathbf{X})$, 则数据样本在核空间中的内积可以用 $k_{ij} = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$ 描述. 在上述核函数矩阵 \mathbf{K} 的定义中,不需要显示定义非线性映射 $\Phi(\mathbf{X})$, 从而显著减少核函数矩阵的计算量. 利用 k_{ij} 将问题式(3)映射到核空间,得到核空间的相似图 \mathbf{S} 构造问题:

$$\begin{aligned} \min_{\mathbf{S}} \sum_{i,j=1}^n (k_{ii} - 2k_{ij} + k_{jj})s_{ij} + \xi \|\mathbf{S}\|_{\mathbb{F}}^2, \\ \text{s.t. } s_{ij} \geq 0, \mathbf{S}^T \mathbf{1} = \mathbf{1}. \end{aligned} \quad (4)$$

可验证,式(3)是式(4)采用线性核时的一个特例。

2.2 低秩约束的单视图聚类

通过求解问题式(4),可得到每个视图的相似图 \mathbf{S}^v , 继而通过谱聚类进行聚类分析. 然而,传统的谱聚类需要一个附加的 K -means 步骤来获得最终的聚类结果,而这个附加的聚类步骤可能会带来额外的 PAC 边界^[9]. 故此,本文提出一种在统一框架下对相似图构造和聚类分析进行协同求解的方法,使得所构造的相似图的连通分量数目正好等于聚类数目,从而省除附加的 K -means 步骤。

根据图谱理论^[27],令 $\mathbf{S} \in \mathbb{R}^{n \times n}$ 为数据集 \mathbf{X} 的邻接矩阵,定义其度矩阵 (degree matrix) 为 $\mathbf{D}_s = \text{diag}(\mathbf{S}^T \mathbf{1})$, 其中 $\text{diag}(\mathbf{y})$ 代表以其参数向量 \mathbf{y} 为对角元素的矩阵. \mathbf{S} 的拉普拉斯矩阵可以定义为 $\mathbf{L}_s = \mathbf{D}_s - \mathbf{S}$. 拉普拉斯矩阵 \mathbf{L}_s 具有如下重要性质^[27]:

定理 1. 拉普拉斯矩阵 \mathbf{L}_s 的特征值 0 的重数 c 等于 \mathbf{S} 所对应的图中连通分量的个数。

定理 1 在聚类分析中的意义在于,如果拉普拉斯矩阵的秩 $\text{rank}(\mathbf{L}_s) = n - c$, 则对应的图正好包含 c 个全连通分量,因此不需要附加的 K -means 步骤就可以得到最终的聚类结果。

受定理 1 的启发,本文对问题式(4)添加一个低秩约束,使得所构造的相似图的连通分量数目正好等于聚类数目,即:

$$\begin{aligned} \min_{\mathbf{S}} \sum_{i,j=1}^n (k_{ii} - 2k_{ij} + k_{jj})s_{ij} + \xi \|\mathbf{S}\|_{\mathbb{F}}^2, \\ \text{s.t. } s_{ij} \geq 0, \mathbf{S}^T \mathbf{1} = \mathbf{1}, \text{rank}(\mathbf{L}_s) = n - c. \end{aligned} \quad (5)$$

由于低秩约束不易处理,且 \mathbf{L}_s 依赖于优化变量 \mathbf{S} , 直接求解问题式(5)非常困难. 然而,由于 \mathbf{L}_s 是半正定矩阵^[27], 问题式(5)可转化为易于求解的等价形式。

因为半正定矩阵的所有特征值均大于等于零,令 $\sigma_i(\mathbf{L}_s)$ 表示 \mathbf{L}_s 的第 i 个最小特征值,约束 $\text{rank}(\mathbf{L}_s) = n - c$ 等价于 $\sum_{i=1}^c \sigma_i(\mathbf{L}_s) = 0$. 据 KyFan 定理^[28],最小的 c 个特征值之和可求解为

$$\sum_{i=1}^c \sigma_i(\mathbf{L}_s) = \min_{\mathbf{F}^T \mathbf{F} = \mathbf{I}} \text{Tr}(\mathbf{F}^T \mathbf{L}_s \mathbf{F}). \quad (6)$$

问题式(6)的解 $\mathbf{F}^* \in \mathbb{R}^{n \times c}$ 由 \mathbf{L}_s 的 c 个最小特征值对应的特征向量构成. 从谱聚类的求解过程可知^[27], \mathbf{F}^* 恰好是样本矩阵的谱嵌入,谱聚类正是通过对 \mathbf{F}^* 进行 K -means 聚类而得到最终的聚类结果。

将式(5)中的约束 $\text{rank}(\mathbf{L}_s) = n - c$ 用式(6)进行替换,问题式(5)转换为

$$\begin{aligned} \min_{\mathbf{S}, \mathbf{F}} \sum_{i,j=1}^n (k_{ii} - 2k_{ij} + k_{jj})s_{ij} + \\ \xi \|\mathbf{S}\|_{\mathbb{F}}^2 + \lambda \text{Tr}(\mathbf{F}^T \mathbf{L}_s \mathbf{F}), \\ \text{s.t. } s_{ij} \geq 0, \mathbf{S}^T \mathbf{1} = \mathbf{1}, \mathbf{F}^T \mathbf{F} = \mathbf{I}. \end{aligned} \quad (7)$$

当问题式(7)中的参数 λ 设置为一个足够大的值时,其解 \mathbf{F}^* 可使问题式(5)中约束 $\text{rank}(\mathbf{L}_s) = n - c$ 得到满足。

2.3 基于谱旋转的后融合多视图聚类

现有基于图谱理论的多视图聚类大多先利用各视图的邻接矩阵 \mathbf{S}^v 学习一个公共邻接矩阵 \mathbf{U} , 然后基于 \mathbf{U} 进行聚类分析^[3,7]. 但从邻接矩阵的角度来看,因不同视图描述对象的角度、可分类性存在差异^[8]; 且信息缺失、噪声等现象易造成单个视图的邻接矩阵 \mathbf{S}^v 失真. 迫使所有视图共享一个公共邻接矩阵可能会导致最终聚类结果不理想^[6]. 另一方面,从类别指示矩阵 \mathbf{Y} 来看,由于不同视图的数据都是从不同角度描述同一个样本集合,该集合的类别指示矩阵 \mathbf{Y} 是一个跨越不同视图的全局结构,且 \mathbf{Y} 并不由各视图的邻接矩阵 \mathbf{S}^v 直接求解得到,因此,对 \mathbf{S}^v 的聚类性能差异和信息失真, \mathbf{Y} 比公共邻接矩阵 \mathbf{U} 更加鲁棒. 因此,当将本文描述的聚类方案扩展到多视图时,本文先对每个视图求解问题式(7)得到邻接矩阵 \mathbf{S}_v^* 和对应的谱嵌入 \mathbf{F}_v^* , 然后再求解统一的分类指示矩阵 \mathbf{Y} .

具体而言,对于单个视图,考察式(7)中 \mathbf{F} 的相关目标函数项 $\text{Tr}(\mathbf{F}^T \mathbf{L}_s \mathbf{F})$ 和约束 $\mathbf{F}^T \mathbf{F} = \mathbf{I}$ 可知, \mathbf{F} 的最优解 \mathbf{F}^* 有旋转不变性,即:对任意正交矩阵

$\mathbf{R} \in \mathbb{R}^{c \times c}$, 有 $Tr(\mathbf{F}^T \mathbf{L}_s \mathbf{F}) = Tr((\mathbf{F}\mathbf{R})^T \mathbf{L}_s \mathbf{F}\mathbf{R})$ 和 $(\mathbf{F}\mathbf{R})^T \mathbf{F}\mathbf{R} = \mathbf{I}$, 因此, 若 $\{\mathbf{S}^*, \mathbf{F}^*\}$ 是问题式(7)的解, 则 $\{\mathbf{S}^*, \mathbf{F}^* \mathbf{R}\}$ 也是。

对某视图 \mathbf{X}^v 而言, 其谱嵌入 \mathbf{F}^v 是其分类指示矩阵 \mathbf{Y}^v 的一组正交基^[18], 同时, 因不同视图有公共的类别指示矩阵 \mathbf{Y} . 因此, 虽各视图的谱嵌入 \mathbf{F}^v 存在差异, 但因式(7)关于 \mathbf{F}^v 的解 \mathbf{F}^v_* 具有旋转不变性, 皆为 \mathbf{Y} 的正交基, 故可通过求解 \mathbf{F}^v 的最优旋转实现谱嵌入矩阵 \mathbf{F}^v 的信息融合, 得到最优公共分类指示矩阵 \mathbf{Y}^* . 因此, 本文构建基于谱旋转的后融合模型:

$$\begin{aligned} \min_{\mathbf{S}^v, \mathbf{F}^v, \mathbf{R}^v, \mathbf{Y}} \sum_{v=1}^V \sum_{i,j=1}^n d_{ij}^v s_{ij}^v + \xi \|\mathbf{S}^v\|_{\mathbb{F}}^2 + \\ \lambda Tr((\mathbf{F}^v)^T \mathbf{L}_s \mathbf{F}) + \gamma \|\mathbf{Y} - \mathbf{F}^v \mathbf{R}^v\|_{\mathbb{F}}^2, \quad (8) \\ \text{s.t. } s_{ij}^v \geq 0, (\mathbf{S}^v)^T \mathbf{1} = \mathbf{1}, (\mathbf{F}^v)^T \mathbf{F}^v = \mathbf{I}, \\ \mathbf{Y} \in \text{Ind}, (\mathbf{R}^v)^T \mathbf{R}^v = \mathbf{I}, \end{aligned}$$

其中, $d_{ij}^v = k_{ii}^v - 2k_{ij}^v + k_{jj}^v$, V 为视图个数. 式(8)中目标函数的最后一项 $\|\mathbf{Y} - \mathbf{F}^v \mathbf{R}^v\|_{\mathbb{F}}^2$ 可视为各个视图的谱嵌入矩阵的最佳旋转与统一类别指示矩阵 \mathbf{Y} 之间的差异的正则化项. 这一项的值越大, 表明该视图的类簇结构与统一类别指示矩阵的差异越大, 在信息融合中应该取更小的权重. 显然式(8)没有考虑该权重, 因此, 考虑不同视图在聚类融合中的贡献差异, 引入一个权重向量 \mathbf{w} , 用其分量 w_v 描述不同视图的权重差异, 将式(8)改写为

$$\begin{aligned} \min_{\mathbf{S}^v, \mathbf{F}^v, \mathbf{R}^v, \mathbf{Y}, \mathbf{w}} \left\{ \sum_{v=1}^V \sum_{i,j=1}^n d_{ij}^v s_{ij}^v + \xi \|\mathbf{S}^v\|_{\mathbb{F}}^2 + \right. \\ \left. \lambda Tr((\mathbf{F}^v)^T \mathbf{L}_s \mathbf{F}) + \frac{\gamma}{w_v} \|\mathbf{Y} - \mathbf{F}^v \mathbf{R}^v\|_{\mathbb{F}}^2 \right\}, \\ \text{s.t. } d_{ij}^v = k_{ii}^v - 2k_{ij}^v + k_{jj}^v, s_{ij}^v \geq 0, (\mathbf{S}^v)^T \mathbf{1} = \mathbf{1}, \\ \mathbf{F}_v^T \mathbf{F}_v = \mathbf{I}, \mathbf{Y} \in \text{Ind}, \mathbf{R}_v^T \mathbf{R}_v = \mathbf{I}, \mathbf{1}^T \mathbf{w} = 1. \quad (9) \end{aligned}$$

从上述分析可知, 本文模型是对谱嵌入矩阵而不是样本邻接矩阵进行信息融合。

2.4 基于多核学习的后融合多视图聚类

虽然式(9)描述的模型可在核空间中以统一框架学习邻接矩阵和聚类指示矩阵, 但仍面临核函数的选择问题: 核方法的性能与核函数的选择高度相关. 但针对特定问题预先选择一个最佳核函数非常困难. 在多视图学习中, 此问题更加棘手, 因为各视图的数据之间的非线性关系可能互不相同, 对于某个视图而言最佳的核函数却不一定适用于另一个视图. 为了克服单个核函数的局限性, 本节将式(9)描述的单核模型扩展到多核, 用多个预先定义的核函数自动学习最佳核函数组合, 从而避免核函数选择的问题。

若共有 m 个候选核函数 $\{\mathbf{K}_t\}_{t=1}^m$, 则相应地有 m 个不同的核空间 $\{\mathbf{H}_t\}_{t=1}^m$. 多核学习需利用候选核函数生成一个新核函数, 并保证所构建的新核函数仍满足 Mercer 条件. 本文使用映射 $\hat{\Phi}(\mathbf{x}) = [\sqrt{p_1} \Phi_1(\mathbf{x}), \sqrt{p_2} \Phi_2(\mathbf{x}), \dots, \sqrt{p_m} \Phi_m(\mathbf{x})]$ 连接所有核空间^[29], 生成核函数:

$$\tilde{\mathbf{K}}(\mathbf{x}, \mathbf{y}) = \sum_{t=1}^m p_t \mathbf{K}_t(\mathbf{x}, \mathbf{y}), \sum_{t=1}^m \sqrt{p_t} = 1, \quad (10)$$

其中, 向量 \mathbf{p} 为核函数的权重向量, $p_t \geq 0 (\forall t)$. 由式(10)所生成的核函数矩阵 $\tilde{\mathbf{K}}$ 是多个半正定矩阵的凸组合, 从而可以保证 $\tilde{\mathbf{K}}$ 是半正定的, 满足 Mercer 条件^[29].

将式(10)代入式(9), 并调整参数位置, 最终得到本文基于多核学习的后融合多视图聚类模型:

$$\begin{aligned} \min_{\mathbf{S}^v, \mathbf{F}^v, \mathbf{R}^v, \mathbf{Y}, \mathbf{w}, \mathbf{p}} \left\{ \sum_{v=1}^V \sum_{i,j=1}^n \mu d_{ij}^v s_{ij}^v + \alpha \|\mathbf{S}^v\|_{\mathbb{F}}^2 + \right. \\ \left. \lambda Tr((\mathbf{F}^v)^T \mathbf{L}_s \mathbf{F}) + \frac{1}{w_v} \|\mathbf{Y} - \mathbf{F}^v \mathbf{R}^v\|_{\mathbb{F}}^2 \right\}, \\ \text{s.t. } d_{ij}^v = \tilde{k}_{ii}^v - 2\tilde{k}_{ij}^v + \tilde{k}_{jj}^v, \tilde{\mathbf{K}} = \sum_{t=1}^m p_t \mathbf{K}_t(\mathbf{x}, \mathbf{y}), \\ p_t \geq 0 (\forall t), \sum_{t=1}^m p_t = 1, s_{ij}^v \geq 0, (\mathbf{S}^v)^T \mathbf{1} = \mathbf{1}, \\ \mathbf{F}_v^T \mathbf{F}_v = \mathbf{I}, \mathbf{Y} \in \text{Ind}, \mathbf{R}_v^T \mathbf{R}_v = \mathbf{I}, \mathbf{1}^T \mathbf{w} = 1. \quad (11) \end{aligned}$$

3 模型求解

因目标函数式(11)中的优化变量相互耦合、包含离散变量 \mathbf{Y} , 且约束亦不平滑, 故求解式(11)并不容易. 本文提出一个交替迭代的优化方案求解式(11). 下文按优化变量具体介绍其各自更新规则。

3.1 固定其他变量, 对每个视图更新 \mathbf{S}^v

当其他与 \mathbf{S}^v 无关的变量保持不变时, 问题式(11)的目标函数关于 \mathbf{S}^v 可加, 且关于 \mathbf{S}^v 的约束也可按视图分离. 因此, 可独立对每个视图依次更新 \mathbf{S}^v . 省略视图上下标, 则式(11)可改写为

$$\begin{aligned} \min_{\mathbf{S}} \sum_{i,j=1}^n \mu d_{ij} s_{ij} + \alpha \|\mathbf{S}\|_{\mathbb{F}}^2 + \lambda Tr(\mathbf{F}^T \mathbf{L}_s \mathbf{F}), \\ \text{s.t. } s_{ij} \geq 0, \mathbf{S}^T \mathbf{1} = \mathbf{1}. \quad (12) \end{aligned}$$

通常仅保留数据点 x_i 与其最近邻的若干个数据点之间的边, 是以在保持邻接矩阵 \mathbf{S} 的稀疏性的同时降低计算复杂度^[3,30]. 若为数据相似图中的每个节点保留 k 个邻居, 即邻接矩阵 \mathbf{S} 每行仅保留 k 个非零元素, 则 \mathbf{S} 的更新方案为

$$s_{ij}^+ = \begin{cases} \frac{g_{i(k+1)} - g_{ij}}{\left(kg_{i(k+1)} - \sum_{j=1}^k g_{ik}\right)}, & j \leq k; \\ 0, & j > k; \end{cases} \quad (13)$$

其中, $g_{ij} = \mu d_{ij} + \lambda \|F_{i\cdot} - F_{j\cdot}\|^2$, 详细求解过程见附录 A.

3.2 固定其他变量, 对每个视图更新 F^v

与 S^v 类似, 当其他变量保持不变时, F^v 亦可依视图进行独立更新, 则式(11)可改写为

$$\min_{F^T F = I} \lambda Tr(F^T L_s F) + \frac{1}{\omega_v} \|Y - FR\|_F^2, \quad (14)$$

因为此处 L_s, Y, R 皆为常量, 式(14)可变换为

$$\min_{F^T F = I} Tr(F^T L_s F) - \frac{1}{\lambda \omega_v} Tr(F^T Y R^T). \quad (15)$$

问题式(15)是一个典型的 Stiefel 流形上的二次优化问题 (quadratic problem on the Stiefel manifold, QPSM)^[31]. 由于式(15)的目标函数的 Hessian 矩阵 L_s 为半正定矩阵, 所以式(15)是一个凸问题, 其全局最优解 F^* 必然存在. 若干学者对 QPSM 进行了深入研究, 提出了若干高效稳定的迭代算法, 可用于求解式(15)^[31-32].

3.3 固定其他变量, 对每个视图更新 R^v

类似地, R^v 亦可依视图独立更新. 当其他变量保持不变时, 式(11)可变换为

$$\min_{R^T R = I} \|Y - FR\|_F^2 \Leftrightarrow \max_{R^T R = I} Tr(F^T Y R). \quad (16)$$

该问题的解析解为^[31]

$$R^+ = UV^T, \quad (17)$$

其中, U 和 V 分别为 $F^T Y$ 的 SVD 分解的左右奇异向量.

3.4 固定其他变量更新核函数权重向量 p^v

类似地, 核函数权重向量 p^v 也可依视图独立更新, 此时式(11)等价于:

$$\min_p \sum_{t=1}^m p_t \sum_{i,j=1}^n ((K_t)_{ii} - 2(K_t)_{ij} + (K_t)_{jj}) s_{ij},$$

$$\text{s.t.} \quad \sum_{t=1}^m \sqrt{p_t} = 1. \quad (18)$$

定义向量 z , 其第 t 个分量为

$$z_t = \sum_{i,j=1}^n ((K_t)_{ii} - 2(K_t)_{ij} + (K_t)_{jj}) s_{ij},$$

则式(18)可等价变换为

$$\min_p \{p^T z\},$$

$$\text{s.t.} \quad \sum_{t=1}^m \sqrt{p_t} = 1. \quad (19)$$

问题式(19)的拉格朗日函数为

$$L(p, \lambda) = p^T z - \lambda \left(1 - \sum_{t=1}^m \sqrt{p_t}\right). \quad (20)$$

据 KKT 条件可知, 令 $L(p, \lambda)$ 关于 p_j 的一阶导数为零, 并结合约束 $\sum_{j=1}^m \sqrt{p_j} = 1$, 可得 p_j 的解析解为

$$p_j^+ = \left(z_j \sum_{j=1}^m z_j^{-1}\right)^{-2}. \quad (21)$$

3.5 固定其他变量更新聚类指示矩阵 Y

固定其他所有变量, 对式(11)求解 Y 等价于:

$$\min_{Y \in \text{Ind}} \left\{ \frac{1}{\omega_v} \|Y - F^v R^v\|_F^2 \right\}. \quad (22)$$

令 $Q \triangleq \sum_v F^v R^v / \omega_v$, 则式(22)等价于:

$$\max_{Y \in \text{Ind}_{i,j=1}^n} \sum y_{ij} q_{ij}. \quad (23)$$

因 Y 为类别指示矩阵, 显然式(23)有解析解为

$$y_{ij}^+ = \begin{cases} 1, & j = \arg \max_k (q_{ik}); \\ 0, & \text{otherwise.} \end{cases} \quad (24)$$

3.6 固定其他变量更新视图权重向量 w

固定其他所有变量, 对式(11)求解 w 等价于:

$$\min_w \sum_{v=1}^V \frac{1}{\omega_v} \|Y - F^v R^v\|_F^2,$$

$$\text{s.t.} \quad w^T \mathbf{1} = 1, \omega_v \geq 0 (\forall v). \quad (25)$$

定义辅助向量 y 和 x , 其元素分别为 $y_v = \sqrt{\omega_v}$

和 $x_v = \|Y - F^v R^v\|_F / \sqrt{\omega_v}$. 则式(25)等价于:

$$\min_w \|x\|^2,$$

$$\text{s.t.} \quad w^T \mathbf{1} = 1, \omega_v \geq 0 (\forall v). \quad (26)$$

问题式(26)可求解如下. 据 Cauchy-Schwarz 不等式有:

$$\|x\|_2^2 \stackrel{(a)}{=} \|x\|_2 \|y\|_2 \stackrel{(b)}{\geq} (x^T y)^2, \quad (27)$$

其中, (a) 成立基于 $\|y\|^2 = w^T \mathbf{1} = 1$. 当 $x \propto y$ 时, 不等式(b)的等号成立, 即式(26)的目标函数最小值为

$$(x^T y)^2 = \|Y - F^v R^v\|_F^2, \quad (28)$$

将式(28)的结论回代到问题式(25), 简单计算可得:

$$\omega_v = \|Y - F^v R^v\|_F / \sum_{v=1}^V \|Y - F^v R^v\|_F. \quad (29)$$

总结本文模型 (LMLFMC) 的求解算法.

算法 1. LMLFMC.

输入: 多视图数据 X^v 、邻居个数 k 、类别数 c 、参数 μ 和 λ 、最大迭代次数 t_{\max} ;

输出: 类别指示矩阵 Y .

① 初始化: $t=0$, 生成核矩阵, 初始化 w, p .

② 对每个视图 X^v :

按式(13)更新 S^v ;
 求解子问题式(15)更新 F^v ;
 按式(17)更新 R^v ;
 按式(21)更新核矩阵权重向量 p^v .

- ③ 按式(24)更新 Y ;
- ④ 按式(29)更新视图权重向量 w ;
- ⑤ 重复②~④直到最大迭代次数.

4 实验与结果分析

4.1 实验方案

实验环境为 Intel Xeon Gold 6254@3.10 GHz (X2)CPU,768GB RAM,2×Tesla V100 GPU.操作系统为 Windows Server 2019,编程语言为 Matlab.

本文选取 6 个多视图聚类中广泛采用的公开数据集进行实验:文本数据集 3source 和 BBC,图像数据集 Caltech101_7,100leaves 和 ALOI,以及网络/文本数据集 WebKB.详情如表 1 所示:

Table 1 Detailed Information of Benchmark Datasets
表 1 基准数据集详细信息

Dataset	N	V	C	d_1	d_2	d_3	d_4	d_5	d_6
3source	169	3	6	3560	3631	3068			
BBC	685	4	5	4659	4633	4665	4684		
100leaves	1600	3	100	64	64	64			
Caltech101_07	1474	6	7	48	40	254	1984	512	928
ALOI	11025	4	100	77	13	64	64		
WebKB	203	3	4	1703	230	230			

表 1 中 N 、 V 、 C 分别代表数据集的样本个数,视图个数和类别数, d_i 代表第 i 个视图的特征维数.

实验对原始数据进行归一化处理,采用 12 个核函数进行多核学习^[24],包括:线性核 $K(x, y) = x^T y$;4 个多项式核 $K(x, y) = (a + x^T y)^b$,其中 $a \in \{0, 1\}$, $b \in \{2, 4\}$;7 个高斯核 $K(x, y) = \exp(-\|x - y\|^2 / \xi d_{\max}^2)$,其中 d_{\max} 为样本间的最大距离, $\xi \in \{0.01, 0.05, 0.1, 1, 10, 50, 100\}$.

本文选取了多个代表性多视图聚类算法进行对比分析.包括:MKC(multi-view K -means clustering)^[33],MVCNMF(multi-view clustering via non-negative matrix factorization)^[34],CRSC(co-regularized spectral clustering)^[35],RMSC(robust multi-view spectral clustering)^[36],ASMV(adaptive structure-based multi-view clustering)^[16],AMG(auto-weighted multiple graph-learning)^[7],

MKMVC(multi-kernel multi-view clustering)^[10],MGFSC(multi-graph fusion spectral clustering)^[8].同时,也将标准谱聚类 SC(spectral clustering)^[37]应用到每一个视图,并给出 SC 在各单视图的最佳和最差结果.所有算法运行 10 次,给出结果的均值和方差.

4.2 实验结果分析

本文采用 ACC, MNI, F -measure 和 ARI 这 4 个指标进行聚类效果评价.实验结果如表 2~7 所示.其中 $SC_w(V_i)$ 和 $SC_b(V_j)$ 分别表示将 SC 应用到相应数据集的各个单视图,分别在第 i 个和第 j 个视图上得到的最差(worst)和最佳结果(best).最好的 2 个结果用黑体显示.

Table 2 Cluster Performance on 3sources Dataset

表 2 3sources 数据集上的聚类性能

Method	ACC	NMI	F -measure	ARI
SC_w(V3)	33.55(0.29)	6.08(0.47)	36.44(0.21)	-1.05(0.34)
SC_b(V1)	34.91(0.00)	5.86(0.00)	36.96(0.00)	-0.36(0.00)
MVCNMF	51.18(1.73)	47.04(1.89)	45.64(1.40)	29.48(1.87)
MKC	47.99(3.35)	36.30(6.50)	38.25(3.98)	20.79(4.02)
CRSC	55.43(4.08)	51.57(2.62)	48.83(3.07)	34.72(4.20)
RMSC	48.17(1.85)	40.71(2.17)	42.34(2.13)	27.97(2.82)
ASMV	33.73(0.00)	8.96(0.00)	35.28(0.00)	-2.11(0.00)
AMG	67.51(3.12)	57.88(4.73)	60.10(4.78)	44.88(7.73)
MKMVC	72.17(0.00)	56.82(0.00)	59.63(0.00)	46.39(0.00)
GFSC	70.26(2.21)	59.21(2.42)	62.66(2.08)	47.20(3.96)
LMLFMC	78.11(0.00)	66.98(0.00)	70.87(0.00)	61.14(0.00)

Note: The top-2 results of every column are highlighted in bold.

Table 3 Cluster Performance on BBC Dataset

表 3 BBC 数据集上的聚类性能

Method	ACC	NMI	F -measure	ARI
SC_w(V1)	32.99(0.00)	2.09(0.00)	37.74(0.00)	0.13(0.00)
SC_b(V3)	33.87(0.00)	2.62(0.00)	38.04(0.00)	0.34(0.00)
MVCNMF	46.22(3.08)	32.58(2.93)	38.09(1.02)	11.81(2.24)
MKC	60.18(2.95)	48.22(4.72)	50.30(4.06)	34.16(6.38)
CRSC	46.66(0.59)	25.66(3.25)	46.52(2.31)	21.62(5.47)
RMSC	68.35(7.29)	53.45(3.07)	57.63(3.10)	44.93(3.55)
ASMV	33.72(0.00)	3.48(0.00)	37.81(0.00)	0.18(0.00)
AMG	50.32(10.16)	32.19(17.40)	51.10(7.80)	26.67(15.25)
MKMVC	71.87(0.00)	52.65(0.00)	69.42(0.00)	57.31(0.00)
GFSC	44.45(2.40)	29.32(1.14)	37.05(7.87)	7.79(2.21)
LMLFMC	87.88(0.00)	73.33(0.00)	81.39(0.00)	75.64(0.00)

Note: The top-2 results of every column are highlighted in bold.

Table 4 Cluster Performance on 100leaves Dataset

表 4 100leaves 数据集上的聚类性能

Method	ACC	NMI	F-measure	ARI
SC_w(V3)	33.77(0.49)	63.58(0.25)	20.34(0.38)	19.43(0.39)
SC_b(V1)	60.50(0.56)	81.93(0.20)	48.38(0.77)	47.81(0.79)
MVCNMF	67.65(1.67)	86.53(0.74)	59.36(2.05)	58.93(2.08)
MKC	1.00(0.00)	0.00(0.00)	1.86(0.00)	0.00(0.00)
CRSC	78.07(2.41)	74.80(2.26)	72.94(1.65)	74.45(1.94)
RMSC	75.18(2.39)	90.72(0.88)	69.82(2.81)	69.51(2.85)
ASMV	79.06(0.00)	90.09(0.00)	61.48(0.00)	61.04(0.00)
AMG	70.58(2.75)	87.41(0.77)	38.48(4.74)	37.60(4.85)
MKMVC	74.67(0.00)	91.52(0.00)	71.49(0.00)	69.37(0.00)
GFSC	54.37(2.17)	49.33(1.94)	39.73(1.84)	29.19(1.47)
LMLFMC	80.44(0.00)	92.07(0.00)	75.72(0.00)	75.47(0.00)

Note: The top-2 results of every column are highlighted in bold.

Table 5 Cluster Performance on Caltech101-7 Dataset

表 5 Caltech101-7 数据集上的聚类性能

Method	ACC	NMI	F-measure	ARI
SC_w(V1)	35.30(0.06)	22.97(0.13)	34.27(0.07)	15.34(0.09)
SC_b(V3)	55.09(0.00)	10.76(0.00)	57.10(0.00)	5.72(0.00)
MVCNMF	42.69(5.21)	42.06(2.50)	45.50(3.42)	29.39(3.73)
MKC	51.72(7.45)	46.58(5.37)	53.60(6.51)	38.18(7.06)
CRSC	40.91(2.38)	46.81(2.38)	44.52(2.06)	29.20(2.40)
RMSC	44.07(1.88)	45.47(1.63)	45.92(1.59)	30.57(1.88)
ASMV	54.75(0.00)	2.42(0.00)	55.81(0.00)	0.95(0.00)
AMG	65.41(8.92)	49.36(13.09)	61.60(6.79)	35.91(16.58)
MKMVC	63.57(0.00)	58.46(0.00)	59.31(0.00)	47.64(0.00)
GFSC	55.18(0.39)	5.36(0.07)	55.61(0.02)	2.08(0.34)
LMLFMC	64.32(0.00)	63.13(0.00)	64.93(0.00)	51.36(0.00)

Note: The top-2 results of every column are highlighted in bold.

Table 6 Cluster Performance on ALOI Dataset

表 6 ALOI 数据集上的聚类性能

Method	ACC	NMI	F-measure	ARI
SC_w(V1)	5.72(0.15)	20.41(0.19)	1.21(0.06)	0.20(0.06)
SC_b(V3)	70.53(1.24)	81.87(0.23)	56.11(0.84)	55.62(0.85)
MVCNMF	1.05(0.00)	4.07(0.00)	1.98(0.00)	0.03(0.00)
MKC	1.02(0.00)	0.00(0.00)	1.95(0.00)	0.00(0.00)
CRSC	53.41(2.13)	67.23(1.32)	51.91(1.18)	50.59(1.89)
RMSC	54.85(2.00)	73.70(0.51)	42.33(1.45)	41.71(1.48)
ASMV	49.37(0.00)	70.32(0.00)	10.05(0.00)	8.40(0.00)
AMG	51.30(1.51)	71.62(1.04)	20.50(1.72)	18.88(1.78)
MKMVC	72.31(0.00)	70.56(0.00)	55.47(0.00)	56.38(0.00)
GFSC	21.80(1.95)	61.94(1.49)	18.30(1.48)	16.87(1.52)
LMLFMC	76.51(0.00)	88.54(0.00)	70.47(0.00)	70.15(0.00)

Note: The top-2 results of every column are highlighted in bold.

Table 7 Cluster Performance on WebKB Dataset

表 7 WebKB 数据集上的聚类性能

Method	ACC	NMI	F-measure	ARI
SC_w(V2)	41.82(0.16)	3.76(0.51)	45.27(0.49)	-1.29(0.25)
SC_b(V3)	49.11(1.78)	4.92(1.00)	46.14(1.82)	-0.28(1.19)
MVCNMF	53.69(0.00)	3.83(0.00)	56.60(0.00)	1.43(0.00)
MKC	64.38(3.45)	27.05(1.66)	56.67(2.60)	30.19(2.87)
CRSC	61.48(3.01)	31.10(2.47)	55.69(2.50)	32.57(3.18)
RMSC	46.85(3.97)	22.52(2.57)	45.79(2.64)	21.30(3.56)
ASMV	68.97(8.31)	25.45(11.33)	62.14(4.37)	31.99(16.42)
AMG	66.06(10.97)	26.56(19.30)	62.57(5.82)	22.07(18.34)
MKMVC	68.78(0.00)	37.74(0.00)	61.51(0.00)	35.29(0.00)
GFSC	62.46(2.13)	33.23(3.81)	57.59(3.61)	34.77(4.89)
LMLFMC	69.46(0.00)	35.36(0.00)	62.87(4.37)	36.33(0.00)

Note: The top-2 results of every column are highlighted in bold.

1) 聚类效果分析

① 除了 Caltech101-7 数据集上的 ACC 指标和 WebKB 数据集的 NMI 指标结果稍次, 本文模型其他所有数据集的所有指标都比对比方法有显著提升.

② 从 SC 的结果可知, 各视图的聚类性能差异显著. 以 ALOI 为例, SC 在 V3 视图上有较好效果, 但在 V1 上的结果基本不可用. 说明不同视图的数据相似图 S 差异较大, 将它们强行融合不尽合理. 采用类似融合方案的 GFSC 在 ALOI 上也效果不佳.

③ 包含附加的 K -means 聚类的算法, 如 SC, AMG, GFSC 等, 结果方差较大. AMG 方法尤甚. 而 ASMV, MKMVC 和本文算法由于不包含附加的 K -means 步骤, 方差几乎为零. 这说明消除附加的 K -means 聚类步骤可以提高稳定性.

④ 总体上, 采用多核学习方案的 MKMVC 和本文方法在大部分数据集上都效果良好, 说明多核学习能有效表征各个视图的数据点之间的非线性关系. 而同样采用多核学习方案, 本文的后融合方案显著优于 MKMVC 模型的相似图融合方案.

⑤ 某些情况下, 单视图聚类得到的最佳结果偶尔优于部分多视图聚类算法, 表明要充分利用多个视图之间的互补性和一致性并不容易, 需仔细考量.

2) 时间性能分析

各种方法的运行时间如表 8 所示. 其中 SC_w(SC_b) 分别记录了 SC 算法在相应数据集的聚类效果最差(最好)的单视图上运行 10 次的平均时间. 由于 SC 算法在各个数据集上取得最佳或最差效果的单视图序号不一致, 所以表 8 中 SC_w(SC_b) 省略

了代表最差(最好)单视图序号(相应的序号可以从表 2~7 中查询).

Table 8 Running Time of Each Method on All Dataset

表 8 各种方法在所有数据集上的运行时间

Method	3sources	BBC	Caltech7	100leaves	ALOI	WebKB
SC_w	0.29	0.97	3.89	7.03	14.96	0.19
SC_b	0.27	1.03	3.71	6.91	15.51	0.21
MVCNMF	32.78	74.31	214.30	343.52	967.30	29.36
MKC	1.59	6.32	72.13	151.37	434.70	1.43
CRSC	0.57	1.37	18.52	32.57	97.30	0.52
RMSC	0.45	8.94	48.23	66.84	134.70	0.43
ASMV	4.47	14.13	16.32	21.81	67.32	5.31
AMG	0.43	15.31	18.37	20.32	65.17	0.56
MKMVC	0.61	23.41	87.41	189.32	607.30	0.77
GFSC	0.59	19.36	64.57	203.10	942.70	0.73
LMLFMC	0.64	27.93	86.82	169.79	495	0.85

本文模型最耗时的环节是利用 QPSM 优化工具求解 F . 该计算过程时间的复杂度为 $O(n^2ct)^{[31]}$. 其中 n 代表样本个数, c 代表类别数, t 为 QPSM 算法内部的迭代次数.

对样本个数多于 1 000 的数据集, 所有算法都采用 Matlab 自带的 GPU 运算实现. 从表 8 可知, 基于多核学习的 MKMVC 算法和本文算法耗时都比较大, 但考虑到聚类性能的显著提升, 多核学习的时间代价依然是有价值的. 但对比 MKMVC 算法和本文算法的运行时间可以看出, 虽本文模型因包含 QPSM 子问题, 理论上的时间复杂度高于 MKMVC, 但因采用邻域学习的方案, 所以运行时间仍与 MKMVC 基本相当, 甚至在较大的数据集上比 MKMVC 更快. 这一点从 AMG 和 GFSC 的运行时间对比同样可看出, 虽然两者的复杂度都为 $O(n^3)$, 但因 AMG 采用邻域学习方案, 而 GFSC 采用全局自表达方案, 所以 AMG 的运行时间显著少于 GFSC. 这些现象表明, 邻域学习因需考虑的数据点较少, 得到的相似矩阵也更为稀疏, 可以显著提高运算速度. 邻域学习的速度优势在样本较多的 ALOI 数据集上更加明显.

4.3 参数敏感性分析

本文算法包含 3 个参数 (μ, α, λ) . 其中 α 在指定邻居个数 k 后, 用附录 A 中的公式 A(10) 可直接求解, 无需调参. 以 3source 数据集为例, 图 1~4 显示了当 $k=25$ 时, 各指标对不同参数设置的变化情况.

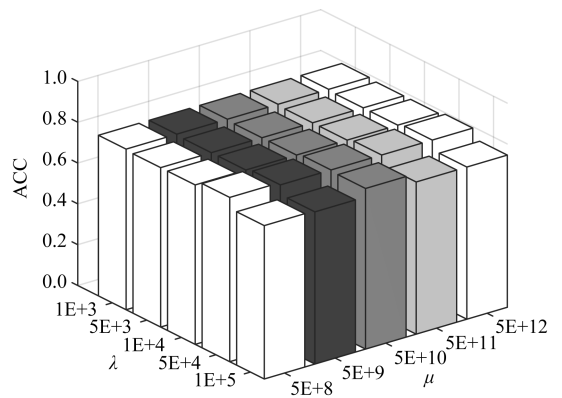


Fig. 1 ACC w.r.t. different parameter settings on 3sources

图 1 3sources 数据集上不同参数设置时的 ACC 指标

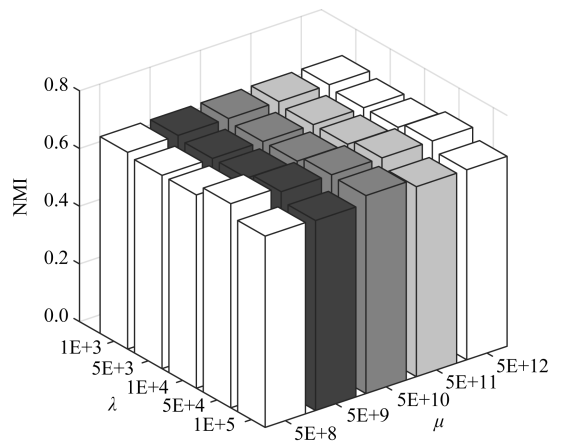


Fig. 2 NMI w.r.t. different parameter settings on 3sources

图 2 3sources 数据集上不同参数设置时的 NMI 指标

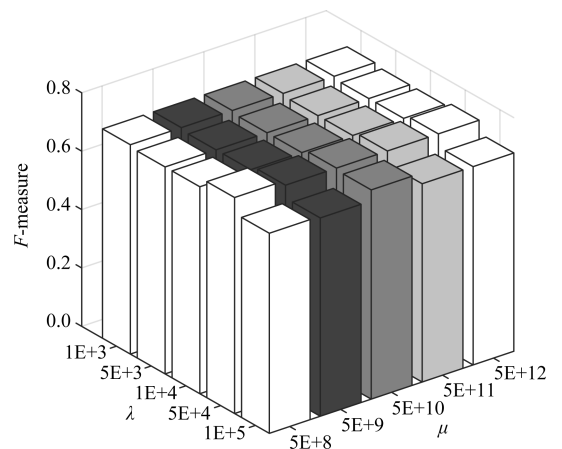


Fig. 3 F-measure w.r.t. different parameter settings on 3sources

图 3 3sources 数据集上不同参数设置时的 F-measure 指标

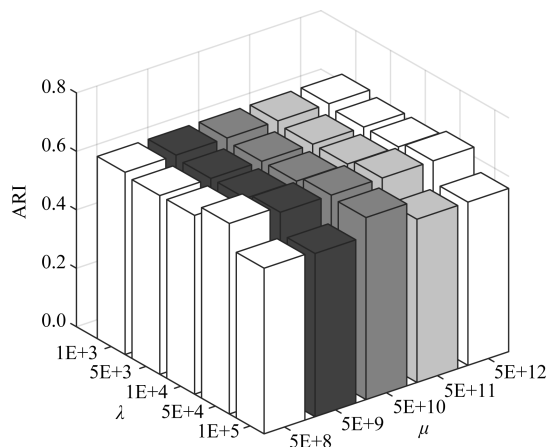


Fig. 4 ARI w.r.t. different parameter settings on 3sources

图4 3sources数据集上不同参数设置时的ARI指标

从图1~4可以看出,本文算法对于参数 μ 和 λ 都不敏感.事实上,所有数据集 μ 取值 $\{5E+8 \sim 5E+12\}$ 之间都能取得较好结果.算法对参数 λ 的稳定性稍次于 λ ,但当 $\lambda \in \{1E+3, 5E+3, 1E+4, 5E+4, 1E+5\}$ 时,大部分数据集都能取得较好结果.表2~7所报导的性能指标,在所有数据集上的都设置为 $\mu=5E+3, \lambda=5E+8$.因此本文算法对于较大范围的参数设置是稳定的.

此外,与所有邻域方法类似,本文需指定邻居个数 k .对所有邻域方法而言,最佳邻居个数 k 的确定目前仍是一个开放问题^[27].但在本文模型中,对文本数据集取 $k \approx 25$,对于图像数据集取 $k \approx 17$ 时,都能取得较好结果.事实上,表2~7所示结果,3sources, BBC, WebKB上都取 $k=25$,在Caltech101-7h和100leaves上取 $k=17$,在ALOI上取 $k=19$ 时获得. ALOI数据集取比另2个图像数据集稍多的邻居个数的原因在于,文献^[27]给出的启发性建议指出,邻居个数应该与 $\log(n)$ 正相关.此外,通过观察文本数据集,发现其特征取值更为稀疏,数据集中特征取值存在大量为0的位置,因而需要考虑更多的邻居个数才能描述数据的局部非线性结构.总之,本文模型性能对于邻居个数 k 的取值是有规律可循的,是稳定的.

5 总结

本文提出一种基于邻域多核学习的后融合多视图聚类算法.为充分挖掘数据间的非线性关系和减轻计算负荷,该算法采用邻域多核学习方案而不是

全局核空间自表达模型.考虑到不同视图的相似图之间的聚类性能差异,本文在类别指示矩阵层次而不是数据相似图的层次进行信息融合.最后,本文提出一种交替优化方案,将相似图的构建和聚类生成等子任务进行协同优化,从而避免所构建的相似图对于聚类任务并非最优.实验结果表明:本文算法的聚类性能优于若干现有的多视图聚类算法.

参 考 文 献

- [1] Yang Yan, Wang Hao. Multi-view clustering: A survey [J]. Big Data Mining and Analytics, 2018, 1(2): 83-107
- [2] Zhao Jing, Xie Xijiong, Xu Xin, et al. Multi-view learning overview: Recent progress and new challenges [J]. Information Fusion, 2017, 38(38): 43-54
- [3] Wang Hao, Yang Yan, Liu Bing. Gmc: Graph-based multi-view clustering [J]. IEEE Transactions on Knowledge and Data Engineering, DOI 10.1109/TKDE.2019.2903810
- [4] Hong Min, Jia Caiyan, Li Yafang, et al. Sample-weighted multi-view clustering [J]. Journal of Computer Research and Development, 2019, 56(8): 1677-1685 (in Chinese)
(洪敏, 贾彩燕, 李亚芳, 等. 样本加权的多视图聚类算法 [J]. 计算机研究与发展, 2019(8): 1677-1685)
- [5] Kang Zhao, Guo Zipeng, Huang Shudong, et al. Multiple partitions aligned clustering [C] //Proc of the 28th Int Joint Conf on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2019: 2701-2707
- [6] Nie Feiping, Tian Lai, Li Xuelong. Multiview clustering via adaptively weighted procrustes [C] //Proc of the 24th ACM Int Conf on Knowledge Discovery & Data Mining. New York: ACM, 2018: 2022-2030
- [7] Nie Feiping, Li Jing, Li Xuelong. Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification [C] // Proc of the 26th Int Joint Conf on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2016: 1881-1887
- [8] Kang Zhao, Shi Guoxin, Huang Shudong, et al. Multi-graph fusion for multi-view spectral clustering [J]. Knowledge Based Systems, 2020, 189: No.105102
- [9] Sun Shiliang, Shawe-Taylor J, Mao Liang. Pac-bayes analysis of multi-view learning [J]. Information Fusion, 2017, 35: 117-131
- [10] Huang Shudong, Kang Zhao, Tsang I, et al. Auto-weighted multi-view clustering via kernelized graph learning [J]. Pattern Recognition, 2019, 88: 174-184
- [11] Yu Kai, Zhang Tong, Gong Yihong. Nonlinear learning using local coordinate coding [C] //Proc of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2009: 2223-2231

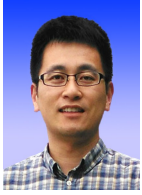
- [12] Wu Mingrui, Scholkopf B. A local learning approach for clustering [C] //Proc of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2006: 1529-1536
- [13] Zhou Zhihua, Machine Learning [M]. Beijing: Tsinghua University Press, 2016: 234-235 (in Chinese)
(周志华. 机器学习[M]. 北京: 清华大学出版社, 2016: 234-235)
- [14] Saha M. A graph based approach to multiview clustering [C] //Proc of Int Conf on Pattern Recognition and Machine Intelligence. Berlin: Springer, 2013: 128-133
- [15] Nie Feiping, Cai Guohao, Li Jing, et al. Auto-weighted multi-view learning for image clustering and semi-supervised classification [J]. IEEE Transactions on Image Processing, 2018, 27(3): 1501-1511
- [16] Zhan Kun, Chang Xiaojun, Guan Junpeng, et al. Adaptive structure discovery for multimedia analysis using multiple features [J]. IEEE Transactions on Systems, Man, and Cybernetics, 2019, 49(5): 1826-1834
- [17] Wang Yang, Zhang Wenjie, Wu Lin, et al. Unsupervised metric fusion over multiview data by graph random walk-based cross-view diffusion [J]. IEEE Transactions on Neural Networks, 2017, 28(1): 57-70
- [18] Yu Stella X, Shi Jianbo. Multiclass spectral clustering [C] //Proc of the 9th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2003: 313-319
- [19] Zelnikmanor L, Perona P. Self-tuning spectral clustering [C] //Proc of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2004: 1601-1608
- [20] Huang Jin, Nie Feiping, Huang Heng. Spectral rotation versus K -means in spectral clustering [C] //Proc of the 27th National Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2013: 431-437
- [21] Scholkopf B, Smola A J, Muller K. Nonlinear component analysis as a kernel eigenvalue problem [J]. Neural Computation, 1998, 10(5): 1299-1319
- [22] Zhang Changshui, Nie Feiping, Xiang Shiming. A general kernelization framework for learning algorithms based on kernel pca [J]. Neurocomputing, 2010, 73(4): 959-967
- [23] Langone R, Alzate C, Suykens J A K. Kernel spectral clustering with memory effect [J]. Physica A-statistical Mechanics and Its Applications, 2013, 392(10): 2588-2606
- [24] Kang Zhao, Peng Chong, Cheng Qiang. Twin learning for similarity and clustering: A unified kernel approach [C] //Proc of the 31th National Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2017: 2080-2086
- [25] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227
- [26] Du Liang, Zhou Peng, Shi Lei, et al. Robust multiple kernel K -means using L_{21} -norm [C] //Proc of the 29th National Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2015: 3476-3482
- [27] Von Luxburg U. A tutorial on spectral clustering [J]. Statistics and Computing, 2007, 17(4): 395-416
- [28] Fan Ky. On a theorem of weyl concerning eigenvalues of linear transformations I [J]. Proceedings of the National Academy of Sciences of the United States of America, 1949, 35(11): 652-655
- [29] Gonen M, Alpaydin E. Multiple kernel learning algorithms [J]. Journal of Machine Learning Research, 2011, 12: 2211-2268
- [30] Nie Feiping, Wang Xiaoqian, Huang Heng. Clustering and projected clustering with adaptive neighbors [C] //Proc of the 20th ACM SIGKDD Intl Conf on Knowledge Discovery & Data Mining. New York: ACM, 2014: 977-986
- [31] Nie Feiping, Zhang Rui, Li Xuelong. A generalized power iteration method for solving quadratic problem on the stiefel manifold [J]. Science in China Series F: Information Sciences, 2017, 60(11): 112101
- [32] Wen Zaiwen, Yin Wotao. A feasible method for optimization with orthogonality constraints [J]. Mathematical Programming, 2013, 142(1): 397-434
- [33] Cai Xiao, Nie Feiping, Huang Heng. Multi-view K -means clustering on big data [C] //Proc of the 20th Int Joint Conf on Artificial Intelligence. San Francisco: Margan Kaufmann, 2013: 2598-2604
- [34] Liu Jialu, Wang Chi, Gao Jing, et al. Multi-view clustering via joint nonnegative matrix factorization [C] //Proc of the 2013 SIAM Int Conf on Data Mining. Philadelphia, PA: SIAM 2013: 252-260
- [35] Kumar A, Rai P, Daume H. Co-regularized multi-view spectral clustering [C] //Proc of Advances in Neural Information Processing Systems, MA: MIT Press, 2011: 1413-1421
- [36] Xia Rongkai, Pan Yan, Du Lei, et al. Robust multi-view spectral clustering via low-rank and sparse decomposition [C] //Proc of the 28th National Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2014: 2149-2155
- [37] Ng A Y, Jordan M I, Weiss Y. On spectral clustering: Analysis and an algorithm [C] //Proc of Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2002: 849-856



Xia Dongxue, born in 1982. PhD candidate at the Southwest Jiaotong University, Chengdu, China. Her main research interests include data mining and multi-view learning.



Yang Yan, born in 1964. PhD, professor, PhD supervisor. Distinguished member of CCF. Her main research interests include artificial intelligence, big data analysis and mining, ensemble learning and multi-view learning.



Wang Hao, born in 1993. PhD. His main research interests include data mining, multi-view learning.



Yang Shuhong, born in 1977. PhD, associate professor. Member of CCF. His main research interests include machine learning and intelligent transportation system.

附录 A 求解正文问题式(12).

由 L_S 的性质,可知

$$\mathbf{F}^T \mathbf{L}_S \mathbf{F} = \sum_{i,j=1}^n \|\mathbf{F}_{i\cdot} - \mathbf{F}_{j\cdot}\|_2^2 s_{ij},$$

令 $d_{ij}^f = \|\mathbf{F}_{i\cdot} - \mathbf{F}_{j\cdot}\|_2^2$, 则式(12)可改写为

$$\min_{\mathbf{s}} \left\{ \sum_{i,j=1}^n [(\mu d_{ij} + \lambda d_{ij}^f) s_{ij} + \alpha s_{ij}^2] \right\},$$

$$\text{s.t. } s_{ij} \geq 0, \mathbf{1}^T \mathbf{S}_{i\cdot} = 1. \quad (\text{A1})$$

显然,上述以矩阵 \mathbf{S} 为变量的优化问题可对 \mathbf{S} 的每一行独立进行,令 $G_{ij} = \mu d_{ij} + \lambda d_{ij}^f$, 则式(A1)等价于:

$$\min_{\mathbf{s}} \left\{ \sum_{i,j=1}^n \frac{1}{2} \left\| \frac{\mathbf{G}_{i\cdot}}{2\alpha} + \mathbf{S}_{i\cdot} \right\|_2^2 \right\},$$

$$\text{s.t. } s_{ij} \geq 0, \mathbf{1}^T \mathbf{S}_{i\cdot} = 1. \quad (\text{A2})$$

式(A2)的拉格朗日函数为

$$\mathcal{L}(\mathbf{S}_{i\cdot}, \eta, \boldsymbol{\beta}) = \frac{1}{2} \left\| \frac{\mathbf{G}_{i\cdot}}{2\alpha} + \mathbf{S}_{i\cdot} \right\|_2^2 -$$

$$\boldsymbol{\beta}^T \mathbf{S}_{i\cdot} - \eta (\mathbf{1}^T \mathbf{S}_{i\cdot} - 1). \quad (\text{A3})$$

对 $\mathcal{L}(\mathbf{S}_{i\cdot}, \eta, \boldsymbol{\beta})$ 关于 $\mathbf{S}_{i\cdot}$ 求导,并令其为0,可得:

$$\mathbf{S}_{i\cdot} + \frac{\mathbf{G}_{i\cdot}}{2\alpha} - \boldsymbol{\beta} - \eta \mathbf{1} = 0. \quad (\text{A4})$$

据 KKT 条件有:

$$s_{ij} \geq 0 (\forall j=1, 2, \dots, n); \beta_j \geq 0 (\forall j=1, 2, \dots, n),$$

$$\beta_j s_{ij} = 0 (\forall j=1, 2, \dots, n); s_{ij} = -\frac{g_{ij}}{2\alpha} + \beta_j + \eta. \quad (\text{A5})$$

从而可得 s_{ij} :

$$s_{ij}^* = (-g_{ij}/2\alpha + \eta)_+. \quad (\text{A6})$$

若将矩阵 \mathbf{G} 的每一行按升序排序,因本文假定 \mathbf{S} 每一行仅有 k 个非零元素,即 $s_{ik}^* > 0$ 且 $s_{i(k+1)}^* = 0$, 则有:

$$\begin{cases} -g_{ik}/2\alpha + \eta > 0, \\ g_{i(k+1)}/2\alpha + \eta \leq 0. \end{cases} \quad (\text{A7})$$

同时,据约束 $\mathbf{1}^T \mathbf{S}_{i\cdot} = 1$,可得:

$$\eta = \frac{1}{k} + \frac{1}{2k\alpha} \sum_{j=1}^k g_{ik}. \quad (\text{A8})$$

综合式(A7)、(A8),可得:

$$\frac{kg_{ik}}{2} - \frac{1}{2} \sum_{j=1}^k g_{ik} < \alpha \leq \frac{kg_{i(k+1)}}{2} - \frac{1}{2} \sum_{j=1}^k g_{ik}. \quad (\text{A9})$$

因此,为确保邻接矩阵 \mathbf{S} 的每一行仅有 k 个非零元素,可将参数 α 设定为

$$\alpha = \frac{kg_{i(k+1)}}{2} - \frac{1}{2} \sum_{j=1}^k g_{ik}. \quad (\text{A10})$$

将式(A6)~(A9)代入式(A10),得到正式式(13)所示的 s_{ij}^+ 更新公式.