

基于超图切割的半监督学习和聚类算法

艾明

(郑州大学 计算机与人工智能学院, 郑州 450001)

摘要: 对超图切割上的半监督学习和聚类算法进行了研究; 通过对超图切割和超边展开法及其切割函数的讨论, 引入了超图上的总变异作为超图切割的洛瓦兹扩展, 并在此基础上提出了一组正则化函数, 它对应于图上的拉普拉斯型正则化; 基于正则化函数族提出了半监督学习方法, 并基于平衡超图切割提出了谱聚类方法; 为了求解这两个学习问题, 将它们转化为求解凸优化问题, 并为此提出了一种主要组成部分为近端映射的可扩展算法, 从而实现半监督学习和聚类; 仿真实验结果表明, 提出的基于超图切割实现的半监督学习和聚类方法相比于经典的超边展开法和其他图切割方法有更好的标准偏差和聚类误差性能。

关键词: 超图展开; 图切割; 正则化函数; 半监督学习; 谱聚类; 标准偏差; 聚类误差

Semi-supervised Learning and Clustering Algorithms Based on Hypergraph Cutting

AI Ming

(School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China)

Abstract: Semi-supervised learning and clustering algorithms on hypergraph cutting are conducted a research; By discussing hypergraph cutting and hyperedge expansion methods as well as its cutting function, the total variation on hypergraph is introduced as a Lovasz extension of hypergraph cutting. Based on this, this paper puts forward a set of regularization functions related to the Laplacian regularization on the graph, presents a semi-supervised learning method based on regularization function family, and proposes a spectral clustering method based on balanced hypergraph cutting; In order to solve these two learning problems, they are transformed into solving the convex optimization problem, and a scalable algorithm whose main component is proximal mapping is proposed to realize the semi-supervised learning and clustering; Simulation results show that the proposed semi-supervised learning and clustering method based on hypergraph cutting has a better standard deviation and clustering error performance than the classical hyperedge expansion and other graph cutting methods.

Keywords: hypergraph expansion; graph cutting; regularization function; semi-supervised learning; spectral clustering; standard deviation; clustering error

0 引言

基于图和超图的学习在机器学习中得到了很好的发展, 是处理编码成对关系数据的标准方法^[1]。超图是图的自然扩展, 它允许建模数据中的高阶关系, 在诸如计算机视觉^[2-3]、生物信息学^[4-5]和信息检索^[6-7]等多种应用领域已经获得认可, 有助于提高学习性能。

基于超图学习的现有方法可以分为两大类。第一类是将张量方法用于聚类, 作为图的矩阵(谱)方法的高阶扩展。文献[8]为解决张量不更新问题, 在概率模型上给出了优化张量的方法并提出了优化张量超图匹配算法, 从而得到更高的匹配准确率; 文献[9]将图分割问题重构为一种特殊的非凸优化问题, 将非凸问题扩展到超图方法中, 并将其与基于定向双边滤波器组的张量场相结合, 提出了一种求解全局收敛的非凸问题的有效最小化算法, 从而实

现图像的分割; 尽管张量方法在数学上很有吸引力, 但它们仅限于所谓的 k -一致超图, 即每个超边恰好包含 k 个顶点。因此, 它们不能建模混合高阶关系; 第二类方法可以处理任意的超图。为了采用现有针对图的方法处理超图, 文献[10]采用超边展开(CE, clique expansion)方法, 针对具有分类边界标签(或不同交互类型)的聚类超图问题提出了一种计算框架。计算框架基于一个组合目标函数, 它与图上的聚类相关, 但能够设计出更高效的算法, 也可以无缝地推广到超图。当只有两种标签类型时, 可以采用基于最小切割算法。在超过两种标签类型时, 采用基于线性规划松弛的快速逼近算法, 具有理论上的聚类质量保证; 文献[11]提出了一种基于超图 Markov 链松弛的聚类学习方法。该算法的基本思想是用 Markov 过程形式化描述超图并开始随机游走, 在超图 Markov 链松弛过程中, 通过随机转移矩阵和扩散映射找到数据集有意义的几何分布, 然后

收稿日期: 2023-12-12; 修回日期: 2024-01-23。

基金项目: 河南省省科技攻关项目(232102211033)。

作者简介: 艾明(1982-), 女, 大学本科, 讲师。

引用格式: 艾明. 基于超图切割的半监督学习和聚类算法[J]. 计算机测量与控制, 2024, 32(5): 260-266.

基于互信息的目标函数进行聚类数目的自动收敛。实验结果表明，该算法在准确率上优于简单图谱聚类算法和标准超图谱聚类算法；文献 [12] 研究了基于图的聚类算法的统计特性，这些算法依赖于平衡图切割的优化，并详细讨论了 Cheeger 切割的优化；文献 [13] 表明，图的最小切割问题是组合优化中最基本的问题之一，而超图提供了建模多路关系的灵活性。然而，当从图推广到超图时，“切超边”的概念就变得模糊，因为超边的节点可以用几种方式分割。作者通过将超图中的两个终端节点以一种最小化分裂超边处惩罚之和的方式分离，提出了一种超图切割框架，确定了一类自然的基于基数的超边分裂函数，并证明了当且仅当分裂函数是子模块时，一般的超图切割问题可以简化为一个可处理的图切割问题，通过保留其切割属性的图来表示超图是可能的；文献 [14] 研究了如何采用图拉普拉斯来解决完全非线性的 Cheeger 切割问题，以及比例切割优化任务。这两个问题都与总变异最小化有关，并推导出了相关的 Ginzburg-Landau 函数；文献 [15] 针对目前基于监督学习的专利文本分类方法存在着依赖大量的训练样本等问题，将基于超图的半监督学习方法引入到专利文本分类中，提出了一种基于 k 近邻策略的专利文本概率超图构建方法，拓展了超图理论的应用领域；文献 [16] 提出了一种基于超图 Markov 链松弛的聚类学习方法。算法的基本思想是用 Markov 过程形式化描述超图并开始随机游走，然后提出基于互信息的目标函数进行聚类数目的自动收敛；文献 [17] 提出了一种新的基于扩散的超图聚类算法，算法解决了一个基于二次超图切割的目标，使得能够计算各种基于基数的超图切函数，找到更好的聚类；为了学习多视图数据样本流形和特征流形之间存在的高阶关系，文献 [18] 提出了一种新的半监督多视图聚类方法—对偶超图正则化部分共享非负矩阵分解。为了提高获得的低维特征的识别能力，在模型中加入了半监督回归项，在捕获数据复杂流形结构时有效利用标签信息。

这些图展开或切割仅在超点或超边的级别上进行，或者整个超图学习完全依赖于边的连通性，必然会损失超图的部分信息，因此缺少了共现数据的对称性，导致高维数据的信息丢失。

对于半监督学习和聚类来说，关键的元素是切割函数。本文首先讨论了超图切割与通过超边展开法和星展开法引起的切割的区别；其次，引入超图上的总变异作为超图切割的洛瓦兹 (Lovasz) 扩展。并在此基础上，提出了一组正则化函数，它们在超图上执行更平滑的函数，对应于图上的拉普拉斯型正则化；然后，基于正则化函数族提出了半监督学习方法，并基于平衡超图切割提出了谱聚类方法。在这两个学习问题中，必须求解凸优化问题，为此提出了一种主要组成部分为近端映射的可扩展算法来实现半监督学习和聚类；仿真实验结果表明，本文提出的基于超图切割实现的半监督学习和聚类方法相比于经典的超边展开法和其他图切割方法在标准偏差和聚类误差方面更有优势。

1 基于超图切割的半监督学习和聚类算法

1.1 传统的超图展开算法

传统的超图展开算法中，两个最经典的算法是超边展开 (CE, clique expansion) 和星展开 (SE, star expansion)。这两个超图展开算法都有各自的弊端。以如图 1 所示的论文作者合作关系网络为例：假设超点表示作者，每一条超边表示一篇论文。使用超边展开超图后，无法知道哪几个作者合作发表了一篇论文。而使用星展开后，原本的超图则转化为了异质结构，而大多数图学习算法是基于同质图设计的。因此，这两种传统超图展开算法在许多应用场合并不足够优越。

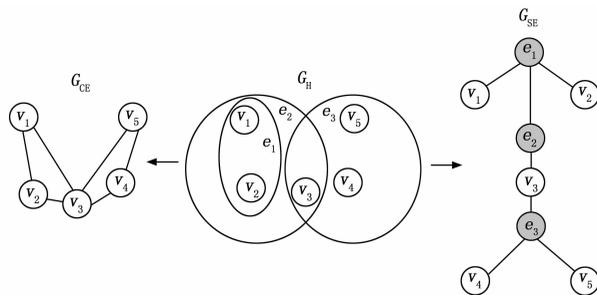


图 1 两种超图展开

而在半监督学习和聚类中，大量基于图的算法或显式或隐式地基于切割。因此，下面首先讨论超图切割和相应的近似。然后通过图引入超图的总变异作为超图切割的 Lovasz 扩展。

1.2 超图、图和切割

超图允许建模关系，这些关系不仅像图中那样是成对的，而且包含多个顶点。本文考虑加权无向超图 $H = (V, E, w)$ ，其中 V 是顶点集，且 $|V| = n, E$ 是超边集，且 $|E| = m$ ，每个超边 $e \in E$ 对应顶点的一个子集，即对应 2^V 的一个元素。向量 $w \in R^m$ 包含每个超边 e 的非负权值 w_e 。为分析简便，下面也用字母 H 表示关联矩阵 $H \in R^{|V| \times |E|}$ ，它表示对于 $i \in V$ 和 $e \in E$ ：

$$H_{i,e} = \begin{cases} 1 & \text{如果 } i \in e \\ 0 & \text{否则} \end{cases} \quad (1)$$

一个顶点 $i \in V$ 的度定义为：

$$d_i = \sum_{e \in E} w_e H_{i,e} \quad (2)$$

一条边 e 的基可写为：

$$|e| = \sum_{j \in V} H_{j,e} \quad (3)$$

需要强调的是，这里并没有强加超图是 k -一致的这一限制条件，即每个超边恰好包含 k 个顶点。

所考虑的超图类包含一组无向加权图，它等价于一组 2-一致超图。超图的总变异思想来自于图切割和总变异函数之间的对应关系。因此，本文采用具有权值矩阵 W 的加权图 $G = (V, W)$ 上的切割的定义。令 $\bar{C} = V/C$ 表示 V 中 C 的补，则对于一个划分 (C, \bar{C}) 来说，切割定义为：

$$cut_G(C, \bar{C}) = \sum_{i,j \in C, j \in \bar{C}} w_{ij} \quad (4)$$

切割的这种标准定义自然适用于超图 H ：

$$cut_H(C, \bar{C}) = \sum_{e \in E: e \cap C \neq \emptyset, e \cap \bar{C} \neq \emptyset} w_e \quad (5)$$

因此，超图的切割函数恰好就是同时在 C 和 \bar{C} 中有顶点的超边的权值之和。它不会偏向于超边的特定切割方式，也就是说，分别在 C 和 \bar{C} 中的超边有多少顶点。这说明超边中的顶点属于一起。

对于超边展开 (CE, clique expansion) 法，将每个超边 $e \in H$ 用一个全连通子图替换，其中该子图中的每条边的权值为 $w_e / |e|$ ，从而得到切割函数 cut_{CE} ：

$$cut_{CE}(C, \bar{C}) = \sum_{e \in E: e \cap C \neq \emptyset, e \cap \bar{C} \neq \emptyset} \frac{w_e}{|e|} |e \cap C| |e \cap \bar{C}| \quad (6)$$

注意，与超图切割式 (5) 相比， cut_{CE} 的值依赖于每个超边的切割方式，因为项 $|e \cap C| |e \cap \bar{C}|$ 使得权值依赖于划分。如果只分离了一个顶点，则获得最小的权值，而如果超边的划分是最平衡的，则获得最大的权值。与超图切割相比，这导致切割偏向于从超边分离单个顶点，这是大多数应用不希望的属性。下面用图 2 的示例来说明这一点，图中对于边权值 $w_1 = w_4 = 10$ ， $w_2 = w_5 = 0.1$ 和 $w_3 = 0.6$ ，最小超图切割为 (C_1, \bar{C}_1) ，这是完美平衡的。尽管多切割一条超边且不平衡， (C_2, \bar{C}_2) 是超边展开近似的最佳切割，即最小超图切割 (cut_H) 得到一个平衡的划分，而最小超边展开切割 (cut_{CE}) 不仅切割一个额外的超边，而且还是不平衡的，这是由于它倾向于分离超边的单个节点。反对超边展开的另一个观点是计算的复杂性。对于大型超边，超边展开导致 (几乎) 全连通图，这使得计算速度变慢，对于大型超图是行不通的。

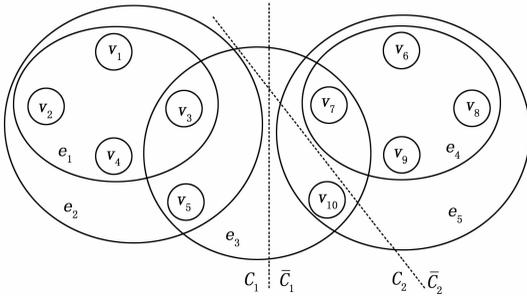


图 2 最小超图切割 cut_H 与最小超边展开切割 cut_{CE}

超图的星展开法近似讨论非常类似于超边展开法。文献 [13] 表明，一般情况下，不存在具有相同顶点集 V 的图，其每个划分 (C, \bar{C}) 有与超图切割相同的切割值。

需要注意的是，对于加权 3-一致超图，总是有可能找到一个对应的图，使得图的任何切割都等于超图的对应切割。

命题 1.1: 设 $H = (V, E, w)$ 是一个加权 3-一致超图，则对于 $W \in R^{|V| \times |V|}$ 的图 G 来说，定义的

$$W = \frac{1}{2} H \text{diag}(w) H^T \quad (7)$$

定义了图 $G = (V, W)$ 的权值矩阵，其中 G 的每个切割都有与 H 的对应超图切割相同的值。

1.3 超图的总变异

本节来确定超图的总变异。关键技术是洛瓦兹 (Lovasz) 扩展，它将集合函数 (看作为 2^V 上的映射) 扩展为 $R^{|V|}$ 上的函数。

定义 1.1: 令 $\hat{S}: 2^V \rightarrow R$ 是一个集合函数，且 $\hat{S}(\emptyset) = 0$ 。令 $f \in R^{|V|}$ ， V 的顺序使得 $f_1 \leq f_2 \leq \dots \leq f_n$ ，定义 $C_i = \{j \in V \mid j > i\}$ ，则 \hat{S} 的 Lovasz 扩展 $S: R^{|V|} \rightarrow R$ 为：

$$S(f) = \sum_{i=1}^n f_i (\hat{S}(C_{i-1}) - \hat{S}(C_i)) = \sum_{i=1}^{n-1} \hat{S}(C_i) (f_{i+1} - f_i) + f_1 \hat{S}(V) \quad (8)$$

对于集合 $C \subset V$ 的特征函数有：

$$S(1_C) = \hat{S}(C) \quad (9)$$

Lovasz 扩展 S 是一个凸函数，当且仅当 \hat{S} 是子模的。对于图 $G = (V, W)$ ，图的总变异定义为图切割的 Lovasz 扩展，即 $TV_G: R^{|V|} \rightarrow R$ ， $TV_G(f) = \frac{1}{2} \sum_{i,j=1}^n w_{ij} |f_i - f_j|$ 。

命题 1.2: 超图 $H = (V, E, w)$ 的总变异 $TV_H: R^{|V|} \rightarrow R$ 定义为超图切割的 Lovasz 扩展 $\hat{S}(C) = cut_H(C, \bar{C})$ ，它是如下的一个凸函数：

$$TV_H(f) = \sum_{e \in E} w_e (\max_{i \in e} f_i - \min_{j \in e} f_j) = \sum_{e \in E} w_e \max_{i,j \in e} |f_i - f_j| \quad (10)$$

注意，如果 H 是 2-一致的 (标准图)，则超图切割的总变异就简化为图的总变异。超图的总变异与稀疏性诱导群规范之间存在着某种关系，即对于每条边 $e \in E$ ，定义差分算子 $D_e: R^{|V|} \rightarrow R^{|V| \times |V|}$ ：

$$(D_e f)_{ij} = \begin{cases} f_i - f_j & \text{如果 } i, j \in e \\ 0 & \text{否则} \end{cases} \quad (11)$$

TV_H 可以写成：

$$TV_H(f) = \sum_{e \in E} w_e \|D_e f\|_\infty \quad (12)$$

这可以看作是在梯度水平上诱导出组稀疏结构，组是超边，因此通常是重叠的。这可能导致图上的弹性网扩展到超图。

在半监督学习 (SSL, semi-supervised learning) 中，采用图的总变异作为正则化函数会对少量标记点导致“尖峰”解。因此，希望采用正则化函数来增强解的更多平滑性。对于图来说，这可以通过采用正则化函数族 $\Omega_{G,p}: R^{|V|} \rightarrow R$ 来得到：

$$\Omega_{G,p}(f) = \frac{1}{2} \sum_{i,j=1}^n w_{ij} |f_i - f_j|^p \quad (13)$$

对于 $p=2$ ，得到图拉普拉斯的正则化函数，这是一大类关于图方法的基础。与图类似，可在超图上定义一个对应的函数族。

定义 1.2: 对于超图 $H = (V, E, w)$ ，当 $p \geq 1$ 时，正则化函数 $\Omega_{H,p}: R^{|V|} \rightarrow R$ 定义为：

$$\Omega_{H,p}(f) = \sum_{e \in E} w_e (\max_{i \in e} f_i - \min_{j \in e} f_j)^p \quad (14)$$

引理 1.1: 函数 $\Omega_{H,p}: R^{|V|} \rightarrow R$ 是凸的。

证明：注意到， $\Omega_{H,1} = TV_H(f)$ 。如果 H 为图且 $p \geq 1$ ，则 $\Omega_{H,p}$ 简化为拉普拉斯正则化 $\Omega_{G,p}$ 。对于集合的特征函数 $f = 1_C$ ，它使得 $\Omega_{H,p}(1_C) = cut_H(C, \bar{C})$ 成立。因此，超图切割及其近似（如超边展开和星展开）之间的差异分别延续到 $\Omega_{H,p}$ 和 $\Omega_{G,p}$ ，由于拉普拉斯正则化 $\Omega_{G,p}$ 是凸的，故函数 $\Omega_{H,p}: R^{|\mathcal{V}|} \rightarrow R$ 是凸的。

1.4 基于正则化函数族的半监督学习

利用上一节中得到的正则化函数，可以写出超图上的两类半监督学习公式。给定标签集 L ，构造向量 $Y \in R^n$ ，其元素 Y_i 为：

$$Y_i = \begin{cases} 0 & \text{如果 } i \notin L \\ \{-1, 1\} \text{ 中的标签} & \text{如果 } i \in L \end{cases} \quad (15)$$

我们提出求解：

$$f^* = \operatorname{argmin}_{f \in R^{|\mathcal{V}|}} \frac{1}{2} \|f - Y\|_2^2 + \lambda \Omega_{H,p}(f) \quad (16)$$

式中， $\lambda > 0$ 为正则化参数。后面将讨论如何在 $p=1$ 和 $p=2$ 的情况下高效地求解这个凸优化问题。注意，除了采用平方损失之外，还可以采用其他损失函数。然而，正则化器的目标是收缩函数，采用标签集 $\{-1, 1\}$ ，以使得 $f^* \in [-1, 1]^{|\mathcal{V}|}$ 。因此，在区间 $[-1, 1]$ 上，平方损失表现出与其他基于边界的损失函数非常相似的行为。一般来说，建议采用 $p=2$ ，因为它对应于图的拉普拉斯型正则化。对于图来说，已知 $p=1$ 对少量标记点会产生尖峰解。这是由于切割出的标记点导致的切割比产生一个平衡分割要小得多。然而，在只有少量超边的情况下，这种效应要小得多。

1.5 基于平衡超图切割的谱聚类

1.2 节中讨论了超图切割式 (5) 和超图的超边展开的图切割式 (6) 之间的差异，并在图 1 中给出了一个简单的示例，这些切割得到了截然不同的结果。对于超图来说，比例切割和归一化切割分别计算为：

$$RCut(C, \bar{C}) = \frac{cut_H(C, \bar{C})}{|C| |\bar{C}|} \quad (17)$$

$$NCut(C, \bar{C}) = \frac{cut_H(C, \bar{C})}{vol(C) vol(\bar{C})} \quad (18)$$

式中， $vol(C)$ 和 $vol(\bar{C})$ 分别表示集合 C 和 \bar{C} 中包含边的总权重，结合了不同的平衡标准。与图的归一化切割相比，归一化超图切割不允许对线性特征问题进行松弛（谱松弛）。

在谱聚类中采用归一化切割的标准谱松弛是松散的，而一个严格和精确的松弛可以用非线性特征问题来表述。尽管非线性特征问题是非凸的，但可以相当高效地计算非线性特征向量。然而，研究表明，精确松弛的潜在非最优解在实践中往往比广义松弛的最优解表现得更好。因此，本文将这种方法扩展到超图，考虑一般平衡超图切割 $Bcut(C, \bar{C})$ ：

$$Bcut(C, \bar{C}) = \frac{cut_H(C, \bar{C})}{\hat{S}(C)} \quad (19)$$

式中， $\hat{S}: 2^{\mathcal{V}} \rightarrow R^+$ 是一个非负对称集合函数，即 $\hat{S}(C) = \hat{S}(\bar{C})$ 。对于归一化切割来说，有 $\hat{S}(C) = vol(C) vol(\bar{C})$ ，而对于 Cheeger 切割来说，有 $\hat{S}(C) = \min\{vol(C), vol(\bar{C})\}$ 。以下结果表明，平衡超图切割具有精确松弛为连续非线性特征问题的性质。

定理 1.1：设 $H = (V, E, w)$ 为一个有限加权超图， $S: R^{|\mathcal{V}|} \rightarrow R$ 为对称非负集函数 $\hat{S}: 2^{\mathcal{V}} \rightarrow R$ 的 Lovasz 扩展，则有：

$$\min_{f \in R^{|\mathcal{V}|}} \frac{\sum_{e \in E} w_e \frac{(\max_{i \in e} f_i - \min_{j \in e} f_j)}{S(f)}}{S(f)} = \min_{C \subset V} \frac{cut_H(C, \bar{C})}{\hat{S}(C)} \quad (20)$$

进一步，令 $f \in R^{|\mathcal{V}|}$ 并定义 $C_t = \{i \in V \mid f_i > t\}$ ，则：

$$\min_{t \in R} \frac{cut_H(C_t, \bar{C}_t)}{\hat{S}(C_t)} \leq \frac{\sum_{e \in E} w_e \frac{(\max_{i \in e} f_i - \min_{j \in e} f_j)}{S(f)}}{S(f)} \quad (21)$$

证明：定理的最后部分表明，在任何 $f \in R^{|\mathcal{V}|}$ 的所有集合中，最优阈值化（将 $f \in R^V$ 变成一个分割）只能得到更好或相等的平衡超图切割，所以问题仍然是如何使比值 $Q(f) = TV_H(f) / S(f)$ 最小化。正如文献 [14] 中所讨论的那样，每个 Lovasz 扩展 S 可以写成凸正 1-齐次函数的差 $S = S_1 - S_2$ 。此外，如命题 1.2 所示，总变量 TV_H 是凸的。因此，必须最小化凸函数与凸函数的差（DC, difference of convex）的非负比，应用下面的算法 1—RatioDCA 算法来实现。主要部分是凸内部问题。在本文的情形中， $R_1 = TV_H$ ， $R_2 = 0$ ，因此内部问题如下：

$$\min_{\|u\|_{\infty} \leq 1} \{TV_H(u) + \lambda^k (S_2(u) - \langle u, s_1(f^k) \rangle)\} \quad (22)$$

为简单起见，将问题限制为子模块平衡函数，在这种情况下， S 是凸的，因此 $S_2 = 0$ 。对于一般情形的证明，可参见文献 [14]。

注意，比例/归一化切割和 Cheeger 切割的平衡函数都是子模块的。结果表明，式 (22) 的内部问题与半监督学习公式 (16) 非常相似。1.6 节将讨论这两个问题的有效解。

算法 1：RatioDCA—1 齐次 DC 函数的非负比的最小化

- 1) 目标: $Q(f) = \frac{R_1(f) - R_2(f)}{S_1(f) - S_2(f)}$
- 2) 初始化: $\|f^0\| = 1, \lambda^0 = Q(f^0)$
- 3) 重复
- 4) $s_1(f^k) \in \partial S_1(f^k), r_2(f^k) \in \partial R_2(f^k)$
- 5) $f^{k+1} = \operatorname{argmin}_{\|u\|_{\infty} \leq 1} \{R_1(u) - \langle u, r_2(f^k) \rangle + \lambda^k (S_2(u) - \langle u, s_1(f^k) \rangle)\}$
- 6) $\lambda^{k+1} = (R_1(f^{k+1}) - R_2(f^{k+1})) / (S_1(f^{k+1}) - S_2(f^{k+1}))$
- 7) 直至 $|\lambda^{k+1} - \lambda^k| / \lambda^k < \epsilon$
- 8) 输出: 特征值 λ^{k+1} 和特征向量 f^{k+1}

1.6 基于超图切割的总实现算法

要求解的半监督学习问题式 (16) 与 RatioDCA 的内部问题式 (22) 有共同的结构，它们是两个凸函数的和：其中一个是新的正则化函数 $\Omega_{H,p}$ ，另一个是用 G 表示的数据项，如表 1 所示，其中指标函数定义为：

$$I_{\|\cdot\|_{\infty} \leq 1} = \begin{cases} 0 & \text{如果 } \|x\|_{\infty} \leq 1 \\ +\infty & \text{否则} \end{cases} \quad (23)$$

表 1 SSL 式(16)(左)和 RatioDCA 的内部问题式(22)(右)的数据项和近端映射

$G(f) = \frac{1}{2} \ f - Y\ _2^2$	$G(f) = -\langle s_1(f^k), f \rangle + l_{\ \cdot\ _1}(f)$
$prox_{g(f)}(\tilde{x}) = \frac{1}{1+\tau}(\tilde{x} + \tau Y)$	$prox_{g(f)}(\tilde{x}) = \frac{\tilde{x} + \tau s_1(f^k)}{\max\{1, \ \tilde{x} + \tau s_1(f^k)\ _2\}}$

下面提出采用原始对偶算法 (PDA, primal-dual algorithm)^[19-20]来求解这些问题。其主要思想是对目标函数中的每个凸项迭代地求解一个近端问题。近端映射 $prox_g$ 关于映射 $g: R^n \rightarrow R$ 定义为:

$$prox_g(\tilde{x}) = \arg \min_{x \in R} \left\{ \frac{1}{2} \|x - \tilde{x}\|_2^2 + g(x) \right\} \quad (24)$$

该算法的关键思想是可以有效地求解近端问题,从而使整个算法快速收敛。从表 1 可见,对于两个 G ,近端问题都有一个明确的解。对于 $\Omega_{H,p}$,区分两种情形 $p=1$ 和 $p=2$ 。

1.6.1 $\Omega_{H,1}$ 的 PDA

令 m_e 为超边 $e \in E$ 中的顶点数,则有:

$$\lambda_{\Omega_{H,1}}(f) = F(Kf) = \sum_{e \in E} (F_{(e,1)}(K_e f) + F_{(e,2)}(K_e f)) \quad (25)$$

式中,矩阵 $K_e \in R^{m_e \times n}$ 的行是 $i \in e$ 的第 i 个标准单位向量,且函数 $F_{(e,j)}: R^m \rightarrow R$ 定义为:

$$F_{(e,1)}(\alpha^{(e,1)}) = \lambda w_e \max(\alpha^{(e,1)}) \quad (26)$$

$$F_{(e,2)}(\alpha^{(e,2)}) = -\lambda w_e \min(\alpha^{(e,2)}) \quad (27)$$

与函数 G 相比,PDA 需要 $F_{(e,j)}$ 的共轭函数的近端映射,即:

$$F_{(e,1)}^* = l_{S_{\lambda w_e}} \quad (28)$$

$$F_{(e,2)}^* = l_{-S_{\lambda w_e}} \quad (29)$$

式中, $S_{\lambda w_e} = \{x \in R^m; \sum_{i=1}^{m_e} x_i = \lambda w_e, x_i \geq 0\}$ 是 R^m 中按比例缩放的单纯形。 $F_{(e,1)}$ 和 $F_{(e,2)}$ 的近端问题的解分别为单纯形上的正交投影,在这里分别为 $P_{S_{\lambda w_e}}$ 和 $P_{-S_{\lambda w_e}}$ 。根据本文所提出的近端映射,PDA 即为以下的算法 2。

算法 2: $\Omega_{H,1}$ 的 PDA

1) 初始化: $f^{(0)} = \bar{f}(0) = 0, \theta \in [0,1], \sigma, \tau > 0$ 且 $\sigma < 1/2$

$$\max_{i=1, \dots, n} \{c_i\}$$

2) 重复

- 3) $\alpha^{(e,1)^{(k+1)}} = P_{S_{\lambda w_e}}(\alpha^{(e,1)^{(k)}} + \sigma K_e \bar{f}^{(k)}), e \in E$
- 4) $\alpha^{(e,2)^{(k+1)}} = P_{-S_{\lambda w_e}}(\alpha^{(e,2)^{(k)}} + \sigma K_e \bar{f}^{(k)}), e \in E$
- 5) $f^{(k+1)} = \text{prog}_{g(f)}(f^{(k)} - \tau \sum_{e \in E} K_e^T (\alpha^{(e,1)^{(k+1)}} + \alpha^{(e,2)^{(k+1)}}))$
- 6) $\bar{f}^{(k+1)} = f^{(k+1)} + \theta(f^{(k+1)} - f^{(k)})$
- 7) 直至相对对偶间隙 $< \epsilon$
- 8) 输出: $f^{(k+1)}$

其中值 $c_i = \sum_{e \in E} H_{i,e}$ 为顶点 i 所在的超边的数目。这里必须指出的是,算法使问题得到了解耦,从某种意义上说,在每次迭代中都求解子问题,即分别处理函数 $G, F_{(e,1)}$ 和 $F_{(e,2)}$,因而可高效求解。

1.6.2 $\Omega_{H,2}$ 的 PDA

如上面一样定义矩阵 K_e 。此外,对每条超边 $e \in E$ 引入函数:

$$F_e(\alpha^e) = \lambda w_e (\max(\alpha^e) - \min(\alpha^e))^2 \quad (30)$$

因此,可以写出 $\Omega_{H,2}(f) = \sum_{e \in E} F_e(K_e f)$ 。注意,这里共轭函数 F_{*e} 不是指标函数,因而通过 F_e 的近端问题来求解相应的近端问题。更具体地说,利用

$$prox_{F_e}(\tilde{\alpha}^e) = \tilde{\alpha}^e - prox_{1/F_e}(\tilde{\alpha}^e) \quad (31)$$

并采用以下关于式(31)的右端的近端问题的结果。

命题 1.3: 对于任意 $\sigma > 0$ 和任意 $\tilde{\alpha}^e \in R^m$, 近端映射

$$prox_{1/F_e}(\tilde{\alpha}^e) = \arg \min_{\alpha^e \in R^m} \left\{ \frac{1}{2} \|\alpha^e - \tilde{\alpha}^e\|_2^2 + \frac{1}{\sigma} \lambda w_e (\max(\alpha^e) - \min(\alpha^e))^2 \right\} \quad (32)$$

可以用 $O(m_e \log m_e)$ 算术运算来计算。

注意,这里的复杂性是由于对输入向量 $\tilde{\alpha}^e$ 进行了排序,它的结构与算法 2 相同,唯一的区别是现在对于每个超边求解式(31)。

2 实验结果

2.1 数据集

表 2 所示为进行比较所采用的数据集,包括植物园、菌菇类、覆盖类和 20 个新闻组,且给出了分类特征,其中,权值为 1 的超边是由具有相同分类特征值的所有数据点创建。对于覆盖类,将数字特征量化到大小相等的 10 个箱中,并创建了两个数据集,每个数据集包含原始数据集的两个类(4,5)和(6,7);由于所有数据集都包含大型超边,所以超边展开对所有数据集导致一个接近完全连通的图。对于覆盖类(6,7),权值矩阵需要超过 10 GB 的内存,而原始超图仅需要 4 MB。

表 2 用于仿真实验的数据集

属性	植物园	菌菇类	覆盖类(4,5)	覆盖类(6,7)	20 个新闻组
类数目	7	2	2	2	4
V	101	8 124	12 240	37 877	16 242
E	42	112	104	123	100
$\sum_{e \in E} e $	1 717	170 604	146 880	454 522	65 451
CE 的 E	10 201	6 599 936	14 300 809	134 821 915	5 328 464

2.2 仿真实验结果

2.2.1 半监督学习 (SSL)

这里主要跟文献 [10] 的超边展开法进行对比。文献 [10] 提出采用由超边展开产生的归一化拉普拉斯 L_{CE} 诱导的正则化矩阵:

$$L_{CE} = I - D_{CE}^{-1/2} H W' H^T D_{CE}^{-1/2} \quad (33)$$

式中, I 表示单位阵, D_{CE} 是一个对角矩阵, 其元素为 $d_{CE}(i) = \sum_{e \in E} H_{i,e} w_e / |e|$, $W' \in R^{|E| \times |E|}$ 为对角矩阵, 其元素为 $w'(e) = w_e / |e|$, SSL 问题表述为:

$$\begin{cases} \arg \min_{f \in R^{1 \times V}} \{ \|f - Y\|_2^2 + \lambda \langle f, L_{CE} f \rangle \} \\ \lambda > 0 \end{cases} \quad (34)$$

该式的优点在于可以通过线性系统得到解。然而, 明显的缺点是 L_{CE} 是一个潜在的非常密集的矩阵, 因此, 在最坏的情况下需要 $|V|^2$ 内存, 相比之下, 本文方法需要 $2 \sum_{e \in E} |e| + |V|$ 内存。对于最大的示例 (覆盖类 (6, 7)), 由于内存问题导致文献 [10] 的超边展开法不能应用。当获得 10^{-6} 的相对对偶差距时, 终止全部实验。在实验中, 对不同数量的标记点进行 10 次试验。对于两种方法, 参数 λ 从集合 10^{-k} 中选择, 其中 $k = \{0, 1, 2, 3, 4, 5, 6\}$ 通过 5 倍交叉验证。

两种方法得到的误差和标准偏差如表 3 所示, 其中第一行为标记点编号; 从表 3 可见, 本文的基于 $\Omega_{H,p}$ ($p = 1, 2$) 的 SSL 方法在除 20 个新闻组外的所有数据集上, 始终优于文献 [10] 的超边展开法。然而, 20 个新闻组是一个非常难处理的数据集, 因为 16, 242 个数据点中只有 10, 267

个是不同的, 这导致最小可能误差为 9.6%。基于成对相互作用的方法 (如超边展开) 可以更好地处理这类标签噪声, 因为这种数据集的大型超边累积了标签噪声。在所有其他数据集上, 本文提出的结合超图切割和超图总变异的方法会得到更好的结果。正如预期的那样, 本文的平方 TV 函数 (即 $p=2$) 的表现略优于总变异 ($p=1$), 尽管差异很小。因此, 当 $\Omega_{H,2}$ 简化为基于图 Laplacian 的标准正则化时, 将 $\Omega_{H,2}$ 用于超图上的 SSL 更好。

2.2.2 聚类

本文采用归一化超图切割作为聚类目标, 对于两个以上的聚类, 递归地对超图进行分割, 直到达到所需的聚类数; 对于比较, 与基于拉普拉斯 L_{CE} (超边展开) 的归一化谱聚类方法 [10] 和文献 [13] 的基于基数的超边分裂函数获得的结果进行比较, 比较指标为 3 种方法的聚类误差以及获得的归一化切割。

从表 4 首先可以看到, 在聚类误差方面, 本文提出的优化归一化超图切割方法优于或等同于超边展开法 (20 个新闻组除外, 原因与 2.2.1 中相同)。特别是在菌菇类数据集的情况下, 改进是最显著的, 而在植物园数据集的情况下, 本文的聚类误差略高。但本文方法总能获得更小的归一化超图切割; 同样, 不能在覆盖类 (6, 7) 上运行文献 [10] 的方法, 因为权值矩阵非常密集。其次, 与文献 [13] 的基于基数的超边分裂函数获得的结果比较表明, 采用超图结构确实有用。尽管如此, 我们认为在超图的构建方面还有改进的空间, 这是我们未来的研究方向。

表 3 对于不同数量的标记点, 运行 10 次的 SSL 的测试误差和标准偏差比较

植物园	20	40	60	80	100	120	160	200
文献[10]方法	35.1±17.2	30.3±7.9	40.7±14.2	29.7±8.8	32.9±16.8	27.6±10.8	25.3±14.4	28.1±16.2
本文的 $\Omega_{H,1}$	2.9±3.0	1.4±2.2	2.2±2.1	0.7±1.0	0.7±1.5	0.9±1.4	1.9±3.0	1.8±1.2
本文的 $\Omega_{H,2}$	2.3±1.9	1.5±2.4	2.9±2.3	0.9±1.4	0.8±1.7	1.2±1.8	1.6±2.9	1.7±1.3
菌菇类	20	40	60	80	100	120	160	200
文献[10]方法	15.5±12.8	10.9±4.4	9.5±2.7	10.3±2.0	9.0±4.5	8.8±1.4	8.8±2.3	9.3±1.0
本文的 $\Omega_{H,1}$	19.5±10.5	10.8±3.7	7.4±3.8	5.6±1.9	5.7±2.2	5.4±2.4	4.9±3.8	5.6±3.8
本文的 $\Omega_{H,2}$	18.4±7.4	9.8±4.5	9.9±5.5	6.4±2.7	6.3±2.5	4.5±1.8	4.4±2.1	3.0±0.6
覆盖类(4,5)	20	40	60	80	100	120	160	200
文献[10]方法	18.9±4.6	18.3±5.2	17.2±6.7	16.6±6.4	17.6±5.2	18.4±5.1	19.2±4.0	20.4±2.9
本文的 $\Omega_{H,1}$	21.4±0.9	17.6±2.6	12.6±4.3	7.6±3.5	6.2±3.8	4.5±3.6	2.6±1.6	1.5±1.3
本文的 $\Omega_{H,2}$	20.7±2.0	16.1±4.1	10.9±4.9	5.9±3.7	4.6±3.4	3.3±3.1	2.2±1.8	1.0±1.1
覆盖类(6,7)	20	40	60	80	100	120	160	200
本文的 $\Omega_{H,1}$	40.6±8.9	6.4±10.4	3.6±3.2	3.3±2.5	1.8±0.8	1.3±0.9	0.9±0.4	1.2±0.9
本文的 $\Omega_{H,2}$	25.2±18.3	4.3±9.6	2.1±2.0	2.2±1.4	1.4±1.1	1.0±0.8	0.7±0.4	1.1±0.8
20 个新闻组	20	40	60	80	100	120	160	200
文献[10]方法	45.5±7.5	34.4±3.1	31.5±1.4	29.8±4.0	27.0±1.3	27.3±1.5	25.7±1.4	25.0±1.3
本文的 $\Omega_{H,1}$	35.7±6.1	31.4±6.1	23.2±5.7	26.2±3.7	22.4±3.3	20.9±3.2	16.1±1.5	14.7±3.6
本文的 $\Omega_{H,2}$	25.0±4.8	28.8±6.0	15.9±5.9	20.4±3.0	18.3±2.7	18.5±2.2	15.0±2.1	10.0±2.0

表 4 不同数据集上对于不同方法的聚类误差和归一化切割的比较

数据集	聚类误差			超图切割			CE 切割		
	本文方法	文献[10]方法	文献[13]方法	本文方法	文献[10]方法	文献[13]方法	本文方法	文献[10]方法	文献[13]方法
菌菇类	10.98	32.25	48.2	0.001 1	0.001 3	0.001 4	0.699 1	0.705 3	1.263 6
植物园	16.83	15.84	5.94	0.673 9	0.678 4	1.636 0	5.131 5	5.170 3	12.453 3
20 个新闻组	47.77	33.20	66.38	0.017 6	0.030 3	0.103 1	2.384 6	1.849 2	0.010 3
覆盖类(4,5)	22.44	22.44	22.44	0.001 8	0.002 2	0.015 2	0.740 0	0.966 1	1.876 7
覆盖类(6,7)	8.16	—	45.85	8.18e-4	—	0.004 1	0.688 2	—	1.323 3

3 结束语

超图可以对数据中的高阶关系进行编码,因而是一种非常灵活的建模工具。目前的学习方法要么基于超图通过图的近似,要么基于仅适用于特殊条件的张量方法。本文提出了一种新的超图学习框架,它充分利用了超图结构,关键元素是基于超图总变异的正则化泛函族和基于平衡超图的切割,使得它们能够很好地应用于半监督学习和聚类分析,以提高学习误差性能和聚类误差性能;在机器学习中,谱聚类是最流行的基于图的聚类方法之一,因为它可以应用于任何基于图的数据或相似信息可用的数据,从而可以构建邻域图;谱聚类通常基于组合归一化/比值图切割问题的松弛,这实际上是不可行的;相比之下,稀疏图的特征向量计算很容易扩展到大图。本文对这一结果进行了扩展,这允许对不同的图切割标准进行灵活建模。

图的最佳平衡切割是计算机科学中的一个重要问题。这种方法可用于诸如并行计算、稀疏矩阵重排序、图像分割和聚类等方面的最小化通信开销,这也是本文的不足和未来需要进一步研究的问题。

参考文献:

[1] 胡秉德,王新根,王新宇,等. 超图学习综述:算法分类与应用分析 [J]. 软件学报, 2022, 33 (2): 498-523.

[2] LV X, WANG L, ZHANG Q L, et al. Video object co-segmentation from noisy videos by a multi-level hypergraph model [C] //Piscataway, NJ: IEEE, 2018: 2207-2211.

[3] 吕欣,王乐,郑南宁. 一种基于多层超图模型的多视频目标协同分割方法 [P]. CN109191485A. 2019-01-11.

[4] 于祥春. 机器学习在计算机视觉和癌症生物信息学中的若干关键问题研究 [D]. 长春:吉林大学, 2019.

[5] WANG X S, LIU J, CHENG Y H, et al. Dual hypergraph regularized PCA for biclustering of tumor gene expression data [J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 31 (12): 2293-2303.

[6] 于琨,孙波,海本斋. 基于超图排序和组稀疏最优化的推荐系统 [J]. 计算机工程与设计, 2018, 39 (7): 1996-2001.

[7] 马慧芳,刘芳,夏琴,等. 基于加权超图随机游走的文献关键词提取算法 [J]. 电子学报, 2018, 46 (6): 1410-1414.

[8] 周君. 基于优化张量的超图匹配算法研究 [D]. 北京:北

京交通大学, 2019.

[9] ZHU H, ZHANG J K, XU G X, et al. Tensor field graph-cut for image segmentation: a non-convex perspective [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31 (3): 1103-1113.

[10] AMBURG I, VELDT N, BENSON A R. Clustering in graphs and hypergraphs with categorical edge labels [C] // New York: ACM, 2020: 706-717.

[11] 郭鹏,李仁发,胡慧. 一种基于超图 Markov 链松弛的聚类学习方法 [J]. 计算机科学, 2019, 46 (s1): 452-456.

[12] TRILLOS N G, MURRAY R, THORPE M. From graph cuts to isoperimetric inequalities: convergence rates of cheeger cuts on data clouds [J]. Spectral Theory, 2020, 4 (1): 1-43.

[13] VELDT N, BENSON A R, KLEINBERG J. Hypergraph cuts with general splitting functions [J]. Computer Science-Data Structures and Algorithms, 2020, 1 (9): 1-49.

[14] MERKURJEV E, BERTOZZI A, YAN X, et al. Modified cheeger and ratio cut methods using the ginzburg-landau functional for classification of high-dimensional data [J]. Inverse Problems, 2017, 33 (7): 74003-74029.

[15] 刘桂锋,汪满容,刘海军. 基于概率超图半监督学习的专利文本分类方法研究 [J]. 情报杂志, 2016, 35 (9): 187-191, 173.

[16] 郭鹏,李仁发,胡慧. 一种基于超图 Markov 链松弛的聚类学习方法 [J]. 计算机科学, 2019, 46 (z1): 452-456.

[17] LIU M, VELDT N, SONG H, et al. Strongly local hypergraph diffusions for clustering and semi-supervised learning [C] //New York: ACM, 2021: 2092-2103.

[18] ZHANG D P, LUO Y H, YU Y Y, et al. Semi-supervised multi-view clustering with dual hypergraph regularized partially shared non-negative matrix factorization [J]. Science China (Technological Sciences), 2022, 65 (6): 1349-1365.

[19] MANSOORI F, WEI E. FlexPD: a flexible framework of first-order primal-dual algorithms for distributed optimization [J]. IEEE Transactions on Signal Processing, 2021, 69: 3500-3512.

[20] WANG K, HAN D. On the linear convergence of the general first order primal-dual algorithm [J]. Journal of Industrial and Management Optimization, 2022, 18 (5): 3749-3770.