

doi: 10.12052/gdutxb.200120

半监督两个视角的多示例聚类模型

蔡 昊, 刘 波

(广东工业大学 自动化学院, 广东 广州 510006)

摘要: 提出了一种新的半监督两个视角的多示例聚类模型, 整合文本视角和图像视角解决了伴有少量标签的多示例图像聚类问题。提出的模型首先嵌入概念分解和多示例核成为一个整体, 学习每个视角的关联矩阵和两个视角所共享的聚类指示矩阵。而后, 应用 $l_{2,1}$ 范数学习最优的关联矩阵和聚类指示矩阵。进一步地, 为了增加包之间的判别力, 提出的模型强迫相同标签包的聚类指示向量间的相似性趋于1, 不同标签包的指示向量间的相似性趋于0。最后, 给出一种迭代更新算法优化提出的模型。实验结果表明, 提出的模型优于现有的多示例聚类模型。

关键词: 多示例学习; 多视角学习; 概念分解; 多示例核函数

中图分类号: TP391.4 文献标志码: A 文章编号: 1007-7162(2021)03-0022-07

A Semi-supervised Two-view Multiple-Instance Clustering Model

Cai Hao, Liu Bo

(School of Automation, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: A novel semi-supervised two-view multi-instance clustering model is proposed, which bands text-view with image-view and solves the multi-instance image clustering problem with a small amount of label. Firstly, the proposed model embeds Concept Factorization and multi-instance kernel into a joint framework, which learns the association matrix of each view and the cluster indicator matrix shared by both views. Then, a $l_{2,1}$ -norm is applied to learn the optimal association matrix and cluster indicator matrix. Furthermore, to enhance the discriminability between bags, the proposed model enforces the similarity of the cluster indicators for the bag with the same label to approximate 1 and the similarity with different labels to 0. Finally, an iterative updating algorithm is derived to solve the proposed model. The experimental results show that the proposed model is superior to other multi-instance clustering models.

Key words: multi-instance learning; multi-view learning; concept factorization; multi-instance kernel function

在大数据时代,许多应用领域均包含大量的图像数据^[1-2],图像聚类能够有效地处理这些图像数据,因此得到了广泛的关注。目前大量的图像聚类方法^[3-4]已经被提出。例如,Ren等^[3]提出了一种基于深度密度的图像聚类框架,其首先使用深度卷积自编码去提取图像的低维度特征,而后使用基于密度的聚类方法进行聚类。Yang等^[4]为图像聚类提出了一种对偶约束的非负矩阵分解算法,其中一个约束用来保持图像标签特征,而另一个约束用来增强图像表示的稀疏性。

近年来,多示例学习也在图像处理中引起了广泛的关注^[5-6]。针对于多示例图像聚类,Zhang等^[6]提出了一种大边缘的多示例图像聚类框架,他们首先识别最相关的示例,而后划分这些示例进入几个不同的组。此外,在图像聚类中,人们可以获得除了图像本身之外的其他信息。例如文本信息,其能够完整地描述相应图像的内容。在现实生活中,很容易收集到大量的图像信息和文本信息,并构造它们成两个视角的数据集去解决图像聚类问题。进一步地,也可以为少量的图像添加标签信息,以提高聚类的性能。

收稿日期: 2020-09-17
基金项目: 国家自然科学基金资助项目(61876044)
作者简介: 蔡昊(1992-),男,博士研究生,主要研究方向为数据挖掘,E-mail: caiyh9658@163.com
通信作者: 刘波(1978-),男,教授,博士,主要研究方向为机器学习、数据挖掘,E-mail: csbliu@gmail.com

本文提出了一种半监督两个视角的多示例聚类模型,其将文本视角引入图像视角去解决具有少量标签的图像聚类问题。该模型首先嵌入概念分解^[7]和多示例核成为一个整体,学习每个视角的关联矩阵和两个视角所共享的聚类指示矩阵;随后,模型引入 $l_{2,1}$ 范数去学习最优的关联矩阵和聚类指示矩阵。进一步地,为了提高包之间的判别力,模型强迫具有相同标签包的聚类指示向量间的相似性趋于1,不同标签包的指示向量间的相似性趋于0。本文的主要贡献如下:

(1) 提出了一个新的模型,即半监督两个视角的多示例图像聚类模型。该模型引入 $l_{2,1}$ 范数提高了包与聚类中心的关联度,同时也获得了理想的聚类指示矩阵。

(2) 基于已知的标签信息,提出的模型强迫包的聚类指示向量间的相似性趋于1或0,提高了包之间的判别力,有助于进一步区分包。

(3) 真实数据集上的实验结果显示:与已有的模型相比,提出的模型能够获得一个更好的聚类结果。

1 相关工作

1.1 多视角学习

通常而言,多视角学习一般先从不同的视角中学习多个特征,而后引入一个联合框架来融合这些特征。现有的多视角学习方法主要基于共识原则和互补原则^[8]。共识原则认为不同视角之间存在一致性信息,这种信息应该得到最大化。而互补原则认为每个视角均包含其他视角所不具备的信息,应该使用多个视角去更全面地描述数据对象。基于共识原则,Zhou等^[9]为每一个视角构建一个完整的图,而后通过自动加权方法强迫所有构造的图趋于一个共识图。基于互补原则,Cao等^[10]利用希尔伯特-斯密特独立性准则去实现不同视角间的多样性,从而增强多视角表示之间的互补性;Wang等^[11]提出了一种多样性的非负矩阵分解算法,其定义一个多样性项,该项迫使不同视角表示两两正交,从而实现视角间的互补。对于两种原则的组合,Liu等^[12]提出了部分共享潜因子学习算法,该算法主要学习一个潜在表示,该表示是由多个视角所共享的一致信息和每个视角的互补信息所组成。

1.2 多示例学习

多示例学习是一种弱监督学习方法,其训练数

据以包的形式存在,而包由多个示例组成。数据的标签信息与整个包相关联,而包中示例标签是未知的。现有的多示例学习方法主要基于包水平和示例水平^[13]。基于包水平的方法通常将每个包视为一个整体,而后从每个包中提取目标概念来预测包的标签。相反,基于示例水平的方法尝试识别关键示例,通过预测关键示例的标签来获得包标签。在基于包水平的方法中,Melki等^[14]提出了多示例表示支持向量机。它学习一个包的表示选择器,其能够选择出对分类影响较大的示例,并将其作为包的表示卷入到支持向量机中去寻找最优的分离超平面。在基于示例水平的方法中,一个典型的例子是多示例支持向量机 (Multiple-instance Support Vector Machine, mi-SVM^[15]),通过对训练包中的示例进行分类,mi-SVM得到一个最优的分离超平面,该超平面可以在每个正的训练包中至少分离出一个正示例。因此,当一个未知包通过该超平面获得一个正示例时,未知包被预测为正。

2 半监督两个视角的多示例聚类模型

本节首先提出半监督两个视角的多示例聚类模型,而后给出一个迭代更新算法去优化这个目标模型,最后,为提出的目标模型引入两个多示例核函数。

2.1 目标模型

首先给出两个视角的 m 个包的数据集 $\mathbf{B}^v = \{\mathbf{B}_1^v, \dots, \mathbf{B}_m^v\}_{v=1}^2$,其中 $\mathbf{B}_i^v = \{\mathbf{b}_{i1}^v, \dots, \mathbf{b}_{im_i}^v\}$ 为第 v 个视角中的第 i 个包,且包含 m_i 个示例。假设前 l 个包的标签信息已知,余下包的标签信息未知。本文的目标是将这 m 个包划分为 c 组。考虑一个映射函数 ϕ ,其能够有效地压缩一个包矩阵成一个向量,即 $\mathbf{B}_i^v \rightarrow \phi(\mathbf{B}_i^v)$ 或 $\mathbf{B}^v \rightarrow \phi(\mathbf{B}^v) = (\phi(\mathbf{B}_1^v), \phi(\mathbf{B}_2^v), \dots, \phi(\mathbf{B}_m^v))$ 。通过使用概念分解算法^[7],两个视角的多示例聚类公式被构造为

$$\sum_{v=1}^2 \|\phi(\mathbf{B}^v) - \phi(\mathbf{B}^v) \mathbf{W}^v (\mathbf{H}^v)^T\|_F^2 \quad (1)$$

其中 $\mathbf{W}^v$ 为关联矩阵, $\mathbf{H}^v$ 为聚类指示矩阵;由于两个视角具有相同的聚类结构,故两个视角的聚类指示矩阵是相同的,即 $\mathbf{H}^1 = \mathbf{H}^2 = \mathbf{H}$ 。式(1)可以转化为

$$\sum_{v=1}^2 \|\phi(\mathbf{B}^v) - \phi(\mathbf{B}^v) \mathbf{W}^v \mathbf{H}^T\|_F^2 \quad (2)$$

具体地说,对于第 v 个视角,式(2)可写为

$$\phi(\mathbf{B}_i^v) \approx \sum_c \mathbf{H}_{ic} \mathbf{R}_c^v = \sum_c \mathbf{H}_{ic} \sum_i \mathbf{W}_{ic}^v \phi(\mathbf{B}_i^v) \quad (3)$$

其中 $\mathbf{R}_c^v$ 属于第 c 个聚类中心点, $\mathbf{W}_{ic}^v$ 属于第 i 个包与第 c 个聚类中心点的关联度; 显然地, 这里存在大量的包与第 c 个聚类中心点无关, 即 $\mathbf{W}^v$ 的第 c 列包含大量的0, 因此 $\mathbf{W}^v$ 的第 c 列是稀疏的; 进一步地, 因 $\mathbf{W}^v$ 的每一列均存在上述类似现象, 故 $\mathbf{W}^v$ 是列稀疏的。又因为每一个包只属于一个类, 与其他的类不相干, 所以聚类指示矩阵 $\mathbf{H}$ 的每一行中也存在大量的0, 即 $\mathbf{H}$ 是行稀疏的。这里引入 $l_{2,1}$ 范数去实现稀疏性, 即

$$\sum_{v=1}^2 \|\phi(\mathbf{B}^v) - \phi(\mathbf{B}^v)\mathbf{W}^v\mathbf{H}^T\|_F^2 + \alpha \left\{ \sum_{v=1}^2 \|(\mathbf{W}^v)^T\|_{2,1} + \|\mathbf{H}\|_{2,1} \right\} \quad (4)$$

此外, 少量的标签信息也很容易获得。如果两个包具有相同的标签信息, 则它们的聚类指示向量应该相同或者高度相似; 反之, 标签信息不同, 则聚类指示向量应该不同或者极其不相似。为了方便, 这里使用内积去权衡相似性, 即相同标签包的聚类指示向量间的内积应趋于1, 不同标签包的指示向量间的内积应趋于0。具体公式表达如下:

$$\|\mathbf{P} \circ (\mathbf{H}\mathbf{H}^T - \mathbf{Q})\|_F^2 \quad (5)$$

其中 $\circ$ 表示矩阵间的点乘; $\mathbf{Q}$ 是相似矩阵, 基于已知的标签信息, $\mathbf{Q}$ 通过式(6)定义。

$$\mathbf{Q}_{ij} = \begin{cases} 1, & i = j \text{ 或者 } \mathbf{B}_i^v \text{ 与 } \mathbf{B}_j^v \text{ 标签相同} \\ 0, & \text{其他} \end{cases} \quad (6)$$

而 $\mathbf{P}$ 是选择矩阵, 其能够选择出前 l 个已知标签包的聚类指示向量, 并强迫它们的内积趋于1或0; 此外, $\mathbf{P}$ 也选择出所有包的聚类指示向量, 并强迫向量自身内积趋于1; $\mathbf{P}$ 通过式(7)定义。

$$\mathbf{P}_{ij} = \begin{cases} 1, & i = j \text{ 或者 } 1 \leq i, j \leq l \\ 0, & \text{其他} \end{cases} \quad (7)$$

最后, 组合式(4)和式(5), 获得最终的目标方程:

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}} & \sum_{v=1}^2 \|\phi(\mathbf{B}^v) - \phi(\mathbf{B}^v)\mathbf{W}^v\mathbf{H}^T\|_F^2 + \alpha \left\{ \sum_{v=1}^2 \|(\mathbf{W}^v)^T\|_{2,1} + \|\mathbf{H}\|_{2,1} \right\} + \beta \|\mathbf{P} \circ (\mathbf{H}\mathbf{H}^T - \mathbf{Q})\|_F^2 \\ \text{s.t.} & \quad \mathbf{W}^v, \mathbf{H} \geq 0, \alpha, \beta \geq 0 \end{aligned} \quad (8)$$

其中 α 为稀疏约束参数, β 为相似性约束参数。

2.2 模型优化

由于式(8)的变量 $\mathbf{W}^v$ 和 $\mathbf{H}$ 均是未知的, 因此目标模型无法求得全局最优解。但是可以使用一个迭代更新算法去求其局部最优解。通过定义 $\mathbf{K}^v = \phi(\mathbf{B}^v)^T \phi(\mathbf{B}^v)$, 式(8)被重写为

$$\begin{aligned} \mathcal{O} = & \sum_{v=1}^2 \text{Tr}((\mathbf{I} - \mathbf{W}^v\mathbf{H}^T)^T \phi(\mathbf{B}^v)^T \phi(\mathbf{B}^v)(\mathbf{I} - \mathbf{W}^v\mathbf{H}^T)) + \\ & \alpha \left\{ \sum_{v=1}^2 \|(\mathbf{W}^v)^T\|_{2,1} + \|\mathbf{H}\|_{2,1} \right\} + \\ & \beta \text{Tr}((\mathbf{P} \circ (\mathbf{H}\mathbf{H}^T - \mathbf{Q}))(\mathbf{P} \circ (\mathbf{H}\mathbf{H}^T - \mathbf{Q}))^T) = \\ & \sum_{v=1}^2 (\text{Tr}(\mathbf{K}^v) - 2\text{Tr}((\mathbf{W}^v)^T \mathbf{K}^v \mathbf{H}) + \text{Tr}((\mathbf{W}^v)^T \mathbf{K}^v \mathbf{W}^v \mathbf{H}^T \mathbf{H})) + \\ & \alpha \left\{ \sum_{v=1}^2 \|(\mathbf{W}^v)^T\|_{2,1} + \|\mathbf{H}\|_{2,1} \right\} + \beta \text{Tr}((\mathbf{P} \circ (\mathbf{H}\mathbf{H}^T))((\mathbf{H}\mathbf{H}^T) \circ \mathbf{P}^T)) - \\ & 2\text{Tr}((\mathbf{P} \circ (\mathbf{H}\mathbf{H}^T))(\mathbf{Q}^T \circ \mathbf{P}^T)) + \text{Tr}((\mathbf{P} \circ \mathbf{Q})(\mathbf{Q}^T \circ \mathbf{P}^T)) \end{aligned} \quad (9)$$

分别为 $\mathbf{W}^v \geq 0$ 和 $\mathbf{H} \geq 0$ 引入拉格朗日乘子 λ^v 和 γ , 可获得式(9)的拉格朗日方程:

$$\mathcal{L} = \mathcal{O} - \sum_{v=1}^2 \text{Tr}(\lambda^v (\mathbf{W}^v)^T) - \text{Tr}(\gamma \mathbf{H}^T) \quad (10)$$

而后求 $\mathcal{L}$ 关于 $\mathbf{W}^v$ 和 $\mathbf{H}$ 的偏导, 可获得如下方程:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^v} = -2\mathbf{K}^v \mathbf{H} + 2\mathbf{K}^v \mathbf{W}^v \mathbf{H}^T \mathbf{H} + 2\alpha \mathbf{W}^v \mathbf{D}_1^v - \lambda^v \quad (11)$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{H}} = & \sum_{v=1}^2 (-2\mathbf{K}^v \mathbf{W}^v + 2\mathbf{H}(\mathbf{W}^v)^T \mathbf{K}^v \mathbf{W}^v) + 2\alpha \mathbf{D}_2 \mathbf{H} + \\ & 4\beta(\mathbf{P} \circ (\mathbf{H}\mathbf{H}^T) \circ \mathbf{P}^T) \mathbf{H} - 4\beta(\mathbf{P} \circ \mathbf{Q}^T \circ \mathbf{P}^T) \mathbf{H} + \\ & 8\beta(\mathbf{P} \circ \mathbf{P})(\mathbf{H} \circ \mathbf{H} \circ \mathbf{H}) - 8\beta(\mathbf{P} \circ \mathbf{P})(\mathbf{H} \circ \mathbf{H} \circ \mathbf{H}) - \gamma \end{aligned} \quad (12)$$

其中 $\mathbf{D}_1$, $\mathbf{D}_2$ 为对角阵, 其对角元素为

$$(\mathbf{D}_1^v)_{ii} = \frac{1}{2\|(\mathbf{W}^v)^T_i\|_2}, (\mathbf{D}_2)_{ii} = \frac{1}{2\|\mathbf{H}_i\|_2} \quad (13)$$

通过使用KKT条件 $\lambda_{ij}^v \mathbf{w}_{ij}^v = 0$, $\gamma_{ij} \mathbf{h}_{ij} = 0$, 可获得

$$(\mathbf{K}^v \mathbf{W}^v \mathbf{H}^T \mathbf{H} + \alpha \mathbf{W}^v \mathbf{D}_1^v)_{ij} \mathbf{w}_{ij}^v - (\mathbf{K}^v \mathbf{H})_{ij} \mathbf{w}_{ij}^v = 0 \quad (14)$$

$$\begin{aligned} & \left(\sum_{v=1}^2 \mathbf{H}(\mathbf{W}^v)^T \mathbf{K}^v \mathbf{W}^v + \alpha \mathbf{D}_2 \mathbf{H} + 2\beta(\mathbf{P} \circ (\mathbf{H}\mathbf{H}^T) \circ \mathbf{P}^T) \mathbf{H} + \right. \\ & \left. 4\beta(\mathbf{P} \circ \mathbf{P})(\mathbf{H}\mathbf{H} \circ \mathbf{H} \circ \mathbf{H}) \right)_{ij} \mathbf{h}_{ij} - \left(\sum_{v=1}^2 \mathbf{K}^v \mathbf{W}^v + \right. \\ & \left. 2\beta(\mathbf{P} \circ \mathbf{Q}^T \circ \mathbf{P}^T) \mathbf{H} + 4\beta(\mathbf{P} \circ \mathbf{P})(\mathbf{H} \circ \mathbf{H} \circ \mathbf{H}) \right)_{ij} \mathbf{h}_{ij} = 0 \end{aligned} \quad (15)$$

随后, 可获得如下更新规则:

$$\mathbf{w}_{ij}^v \leftarrow \mathbf{w}_{ij}^v \frac{(\mathbf{K}^v \mathbf{H})_{ij}}{(\mathbf{K}^v \mathbf{W}^v \mathbf{H}^T \mathbf{H} + \alpha \mathbf{W}^v \mathbf{D}_1^v)_{ij}} \quad (16)$$

$$\mathbf{h}_{ij} \leftarrow \mathbf{h}_{ij} \frac{\gamma_{ij}}{\Gamma_{ij}} \quad (17)$$

其中

$$\begin{aligned} \mathbf{r} &= \sum_{v=1}^2 \mathbf{K}^v \mathbf{W}^v + 2\beta(\mathbf{P} \circ \mathbf{Q}^T \circ \mathbf{P}^T) \mathbf{H} + 4\beta(\mathbf{P} \circ \mathbf{P})(\mathbf{H} \circ \mathbf{H} \circ \mathbf{H}) \\ \mathbf{r} &= \sum_{v=1}^2 \mathbf{H}(\mathbf{W}^v)^T \mathbf{K}^v \mathbf{W}^v + \alpha \mathbf{D}_2 \mathbf{H} + 2\beta(\mathbf{P} \circ (\mathbf{H} \mathbf{H}^T) \circ \mathbf{P}^T) \mathbf{H} + \\ &4\beta(\mathbf{P} \circ \mathbf{P})(\mathbf{H} \circ \mathbf{H} \circ \mathbf{H}) \end{aligned}$$

通过迭代的更新式(16)和式(17),目标模型(8)能够被优化。

2.3 多示例核

本节介绍两个多示例核方法,它们能够嵌入到目标模型(8)中。

2.3.1 混合模型核

该核方法^[16]首先为每个包 $\mathbf{B}_s = \{\mathbf{b}_i\}_{i=1}^m$ 定义其高斯混合模型 $\{(\mathbf{A}_j, \omega_j)\}_{j=1}^K$ 的聚合后验:

$$\phi(\mathbf{B}_s) = \sum_{i=1}^m \left(\frac{\omega_1 \text{pr}(\mathbf{b}_i | \mathbf{A}_1)}{\sum_{j=1}^K \omega_j \text{pr}(\mathbf{b}_i | \mathbf{A}_j)}, \dots, \frac{\omega_K \text{pr}(\mathbf{b}_i | \mathbf{A}_K)}{\sum_{j=1}^K \omega_j \text{pr}(\mathbf{b}_i | \mathbf{A}_j)} \right) \quad (18)$$

借助于式(18),两个包 $\mathbf{B}_s, \mathbf{B}_t$ 的多示例核函数被定义为

$$K(\mathbf{B}_s, \mathbf{B}_t) = \langle \phi(\mathbf{B}_s), \phi(\mathbf{B}_t) \rangle \quad (19)$$

其中 $\langle \cdot, \cdot \rangle$ 是内积运算。

2.3.2 极大极小核

该核方法^[17]首先定义包的映射函数 ϕ :

$$\phi(\mathbf{B}_s) = (\min_{b \in \mathbf{B}_s} \mathbf{b}_1, \dots, \min_{b \in \mathbf{B}_s} \mathbf{b}_d, \max_{b \in \mathbf{B}_s} \mathbf{b}_1, \dots, \max_{b \in \mathbf{B}_s} \mathbf{b}_d) \quad (20)$$

$$\phi(\mathbf{B}_t) = (\min_{b \in \mathbf{B}_t} \mathbf{b}_1, \dots, \min_{b \in \mathbf{B}_t} \mathbf{b}_d, \max_{b \in \mathbf{B}_t} \mathbf{b}_1, \dots, \max_{b \in \mathbf{B}_t} \mathbf{b}_d) \quad (21)$$

通过使用 $\phi(\mathbf{B}_s)$ 和 $\phi(\mathbf{B}_t)$,极大极小核^[17]被定义为

$$K(\mathbf{B}_s, \mathbf{B}_t) = (\langle \phi(\mathbf{B}_s), \phi(\mathbf{B}_t) \rangle + 1)^p \quad (22)$$

3 实验

3.1 实验基线及设置

对于提出的两个视角的多示例聚类模型,它能够使用两个多示例核函数,即混合模型核与极大极小核。将两个核的模型分别命名为Ker1和Ker2。为了验证Ker1和Ker2的有效性,本文使用以下4种基线作为比较:

(1) 基线1为BAMIC1^[18],其使用最小Hausdorff距离度量计算两个包中示例的最小距离,而后采用k-Medoids算法去划分包。定义最小Hausdorff距离度量为

$$\min H(\mathbf{A}, \mathbf{B}) = \min_{a \in \mathbf{A}, b \in \mathbf{B}} \|a - b\| \quad (23)$$

(2) 基线2为BAMIC2^[18],其使用最大Hausdorff距离度量计算两个包之间的示例距离,最后采用k-Medoids算法去划分包。定义最大Hausdorff距离度量为

$$\max H(\mathbf{A}, \mathbf{B}) = \max_{a \in \mathbf{A}} \{ \max_{b \in \mathbf{B}} \|a - b\|, \max_{b \in \mathbf{B}} \{ \max_{a \in \mathbf{A}} \|b - a\| \} \} \quad (24)$$

(3) 基线3为BAMIC3^[18],其使用平均Hausdorff距离度量计算两个包之间的示例距离,最后采用k-Medoids算法去划分包。定义平均Hausdorff距离度量为

$$\text{ave } H(\mathbf{A}, \mathbf{B}) = \frac{\sum_{a \in \mathbf{A}} \min_{b \in \mathbf{B}} \|a - b\| + \sum_{b \in \mathbf{B}} \min_{a \in \mathbf{A}} \|b - a\|}{|\mathbf{A}| + |\mathbf{B}|} \quad (25)$$

(4) 基线4为unKer1和unKer2,其是提出模型的两种无监督的比较方法,它们没有使用标签信息,而是直接初始化 $\mathbf{P} = \mathbf{Q} = \mathbf{I}$,基线4的目的是验证半监督学习的优越性。

上述所有的基线均根据原始文献的建议设置参数。对于Ker1和Ker2,随机抽取每一类的5%的标签信息作为监督信息,其余标签信息未知。Ker1和Ker2中参数 α, β 的范围被设置为[0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000]。在Ker1中,设置参数 $K = 100$;对于Ker2,参考文献^[17]设置参数 $p = 5$ 。为了避免随机性,基于不同的初始化值,算法运行10次,并记录相应的平均聚类精度。聚类精度定义为

$$\text{ACC} = \frac{\sum_{i=1}^N \delta(\mathbf{t}_i, \text{map}(\mathbf{r}_i))}{N} \quad (26)$$

其中, N 表示图像包的总数目, $\mathbf{t}_i$ 为真实标签, $\mathbf{r}_i$ 为算法所学习到的聚类标签; $\text{map}(\mathbf{r}_i)$ 为映射函数,其目的是映射 $\mathbf{r}_i$,使之能与 $\mathbf{t}_i$ 匹配; $\delta(x, y)$ 为 δ 函数,当 $x = y$ 时, $\delta(x, y) = 1$,否则 $\delta(x, y) = 0$ 。

3.2 数据集

NUS-WIDE数据集^[19]是由新加坡国立大学所创立的图像数据集,其包含有269 648张图片,每张图片均有对应的文本描述。实验是从NUS-WIDE数据集中选择图像和相应的文本去构建2个不同的数据集。其一为NUS-WIDE1数据集,该数据集包含6种混合图像(actor, car_racing, fruit, insect, leopard, tunnel),共有2 605张图片。其二为NUS-WIDE2数据集,该数据集包含6种不同的花(即chrysanthemums, lily, orchid, poppies, rose, tulip),共有2 522张图片。

对于上述数据集,每一个文本被分割成一个文

本包,每一张图像也被分割成一个图像包。3种不同的分割方法^[20]分割图像并获得图像包,即K均值分割(K-means Segmentation, K-meansSeg)、局部二值模式(Local Binary Patterns, LBP) 和尺度不变特征变换(Scale Invariant Feature Transform, SIFT)。

3.3 实验结果

表1和表2分别列出了不同模型在NUS-WIDE1数据集和NUS-WIDE2数据集的聚类精度,其中Avg为平均精度。对于Ker1和Ker2,表中列出的是模型在 $\alpha = 0.01, \beta = 0.01$ 的聚类精度。从表1和表2能够

表 1 在NUS-WIDE1数据集中各模型的聚类精度对比
Table 1 The clustering accuracy comparisons of models on NUS-WIDE1 dataset

%

	Class	BAMIC1	BAMIC2	BAMIC3	unKer1	unKer2	Ker1	Ker2
K-meansSeg	c=2	57.40	60.26	59.86	65.93	59.07	73.69	61.00
	c=3	40.37	47.21	46.42	49.15	42.57	52.17	46.35
	c=4	31.88	34.93	34.57	35.20	32.10	39.15	36.36
	c=5	26.44	28.01	29.15	28.31	28.84	32.33	31.62
	c=6	22.61	25.26	25.15	24.07	22.37	26.17	28.93
	Avg	35.74	39.13	39.03	40.53	36.99	44.70	40.85
LBP	c=2	62.80	76.92	73.48	71.52	77.36	75.14	81.27
	c=3	50.85	45.83	57.73	52.79	56.78	56.22	67.18
	c=4	34.75	38.22	42.41	44.57	49.60	50.54	55.34
	c=5	29.65	34.23	37.24	38.96	40.36	43.68	45.66
	c=6	22.88	28.89	29.60	33.74	33.41	38.56	37.62
	Avg	40.19	44.82	48.09	48.32	51.50	52.83	57.41
SIFT	c=2	59.17	56.90	54.67	63.15	61.93	69.15	64.81
	c=3	44.33	47.15	40.99	50.90	45.25	55.64	50.49
	c=4	32.80	33.22	32.93	34.85	35.83	38.94	41.46
	c=5	27.81	27.70	29.54	30.89	31.99	33.75	35.53
	c=6	23.36	24.25	25.72	24.16	25.53	26.84	28.59
	Avg	37.49	37.84	36.77	40.79	40.11	44.86	44.18

表 2 在NUS-WIDE2数据集中各模型的聚类精度对比
Table 2 The clustering accuracy comparisons of models on NUS-WIDE2 dataset

%

	Class	BAMIC1	BAMIC2	BAMIC3	unKer1	unKer2	Ker1	Ker2
K-meansSeg	c=2	51.33	52.13	51.67	54.39	50.46	57.19	54.17
	c=3	35.23	36.45	36.16	36.34	35.70	39.36	37.83
	c=4	28.46	29.21	30.19	32.12	29.74	34.85	31.29
	c=5	23.53	24.14	26.49	26.27	24.41	29.95	26.59
	c=6	20.48	21.47	22.24	23.42	21.13	26.87	23.63
	Avg	31.81	32.68	33.35	34.51	32.29	37.64	34.70
LBP	c=2	53.25	58.64	56.55	57.96	61.88	62.66	66.32
	c=3	35.46	43.20	43.40	41.37	44.81	43.45	49.23
	c=4	27.34	33.39	33.71	33.10	35.41	37.54	39.54
	c=5	23.51	29.57	30.31	28.61	29.26	31.10	33.42
	c=6	19.83	24.77	25.66	24.68	24.59	27.32	29.30
	Avg	31.88	37.91	37.93	37.14	39.19	40.14	43.56
SIFT	c=2	56.21	54.33	57.67	53.25	52.91	62.48	56.91
	c=3	36.72	36.78	39.29	38.98	38.13	43.15	39.78
	c=4	29.22	27.35	31.00	29.97	28.61	32.36	33.21
	c=5	25.01	24.53	26.28	25.77	24.72	28.16	26.55
	c=6	20.81	21.01	22.80	21.71	20.08	23.87	23.17
	Avg	33.59	32.80	35.41	33.94	32.89	38.00	35.92

观察到提出的模型在聚类精度上是优于其他模型的。以NUS-WIDE1数据集为例,在K-meansSeg中,Ker1比其他的模型在平均精度上至少提高了4.17%;在LBP中,Ker2比其他的模型在平均精度上至少提高了5.91%;在SIFT中,Ker1比其他的模型在平均精度上至少提高了4.07%。这是因为提出的模型使用了标签信息,其监督着模型的学习进程,从而有效地提高了聚类的性能。

3.4 参数敏感性分析

本节着重构造实验研究参数 α, β 对模型性能的影响。对于模型Ker1和Ker2,参数的选择范围为[0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 1000]。在不同的正则化参数下,实验在NUS-WIDE1数据集内的两类

数据上执行,对应的实验结果被展示在图1和图2中。从图1中,能够观察到Ker1在参数 α, β 的大多数情况下的聚类精度均超过了65%。从图2中,能观察到Ker2的聚类精度在K-meansSeg和SIFT中随着参数 α, β 变化保持稳定;对于LBP,当 $\alpha \leq 100, \beta \leq 1$ 时,Ker2的聚类精度大多数超过了75%。此外,从图1和图2中,还可以观察到参数 β 设置不大于1更合适。总之,Ker1和Ker2的聚类性能对于参数 α, β 的调整是相对鲁棒的。

3.5 收敛性分析

提出的模型采用迭代更新规则发现目标方程的局部极小值。为了研究模型的收敛性,有必要可视化模型(Ker1和Ker2)在更新规则下的收敛曲线。图3和图4分

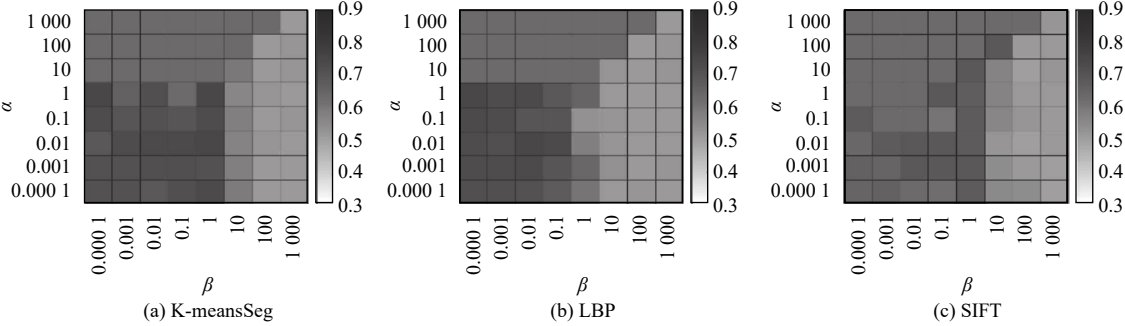


图 1 NUS-WIDE1数据集内的两类数据的Ker1参数敏感性实验

Fig.1 Parameter sensitivity of Ker1 on the two-class data of NUS-WIDE1

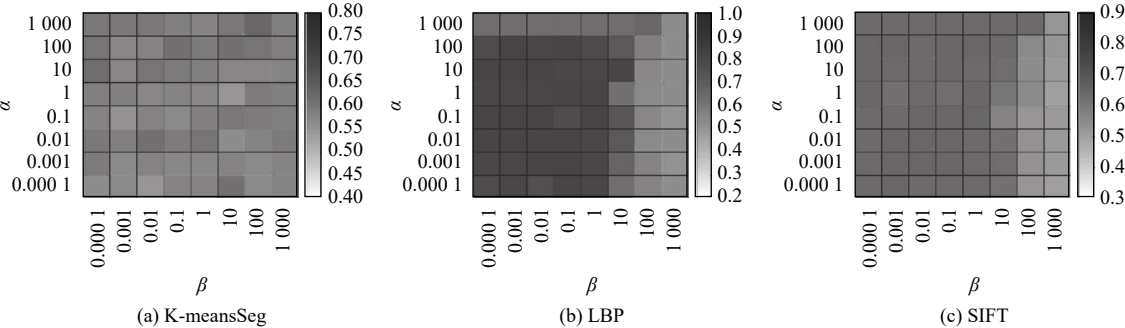


图 2 NUS-WIDE1数据集内的两类数据的Ker2参数敏感性实验

Fig.2 Parameter sensitivity of Ker2 on the two-class data of NUS-WIDE1

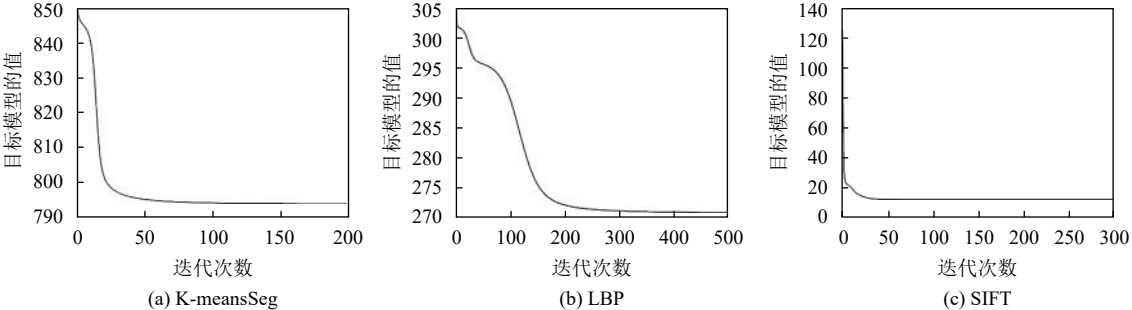


图 3 NUS-WIDE1数据集内的两类数据的Ker1的收敛曲线

Fig.3 Convergence curve of Ker1 on the two-class data of NUS-WIDE1

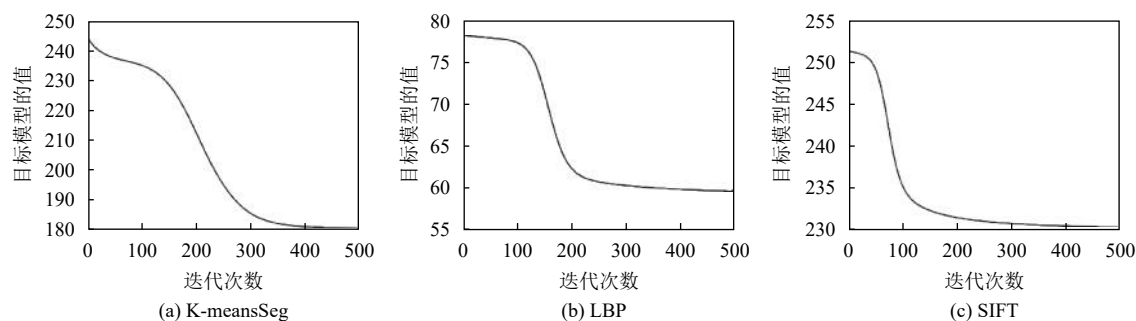


图4 NUS-WIDE1数据集内的两类数据的Ker2的收敛曲线

Fig.4 Convergence curve of Ker2 on the two-class data of NUS-WIDE1

别展示了提出的模型在NUS-WIDE1数据集内的两类数据上的收敛曲线。从图3和图4中,能够观察到Ker1和Ker2随着迭代次数的增加是逐渐收敛的。

4 结语

本文提出了一种新的半监督两个视角的多示例聚类模型,其将文本视角和图像视角结合,有效地解决带有少量标签多示例图像聚类问题。通过嵌入概念分解和多示例核函数为一个整体,该模型为每个视角学习了一个关联矩阵,同时也获得了被两个视角所共享的聚类指示矩阵。而后,通过在关联矩阵和聚类指示矩阵上引入 $l_{2,1}$ 范数,模型不仅提高了包与聚类中心的关联度,也获得了最优的聚类指示矩阵。随后,基于已知的标签信息,模型强迫相同标签包的聚类指示向量的相似性趋于1,不同标签的指示向量相似性趋于0,这有效地提高了包之间的判别力。最后,一个迭代更新算法被提出,有效地优化了提出的模型。在真实数据集的实验结果表明,提出的模型在聚类精度上优于现有的多示例聚类模型。

参考文献:

- [1] TIAN M W, YAN S R, TIAN X X, *et al.* Research on image recognition method of bank financing bill based on binary tree decision [J]. *Journal of Visual Communication and Image Representation*, 2019, 60: 123-128.
- [2] WANG P, ZHANG P F, LI Z W. A three-way decision method based on gaussian kernel in a hybrid information system with images: an application in medical diagnosis [J]. *Applied Soft Computing*, 2019, 77: 734-749.
- [3] REN Y Z, WANG N, LI M X, *et al.* Deep density-based image clustering [J]. *Knowledge-Based Systems*, 2020, 197(7): 105841.
- [4] YANG Z Y, ZHANG Y, XIANG Y, *et al.* Non-negative matrix factorization with dual constraints for image clustering [J]. *IEEE Transactions on Systems Man & Cybernetics Systems*, 2018, 50(7): 1-10.
- [5] 黎启祥, 肖燕珊, 郝志峰, 等. 基于抗噪声的多任务多示例学习算法研究[J]. *广东工业大学学报*, 2018, 35(3): 47-53.
- LI Q X, XIAO Y S, HAO Z F, *et al.* An algorithm based on multi-instance anti-noise learning [J]. *Journal of Guangdong University of Technology*, 2018, 35(3): 47-53.
- [6] ZHANG D, WANG F, SI L, *et al.* Maximum margin multiple instance clustering with applications to image and text clustering [J]. *IEEE Transactions on Neural Networks*, 2011, 22(5): 739-751.
- [7] XU W, GONG Y H. Document clustering by concept factorization[C]//*Proceedings of the International ACM Sigir Conference on Research and Development in Information Retrieval*. Sheffield: ACM, 2004: 202-209.
- [8] YANG Y, WANG H. Multi-view clustering: a survey [J]. *Big Data Mining & Analytics*, 2018, 1(2): 3-27.
- [9] ZHOU W, WANG H, YANG Y. Consensus graph learning for incomplete multi-view clustering[C]//*Proceedings of the 23rd Pacific-asia Conference on Knowledge Discovery and Data Mining*. Macau: ACM, 2019: 529-540.
- [10] CAO X C, ZHANG C Q, FU H Z, *et al.* Diversity-induced multi-view subspace clustering[C]//*Proceedings of the IEEE Computer Vision and Pattern Recognition*. Boston: IEEE, 2015: 586-594.
- [11] WANG J, TIAN F, YU H C, *et al.* Diverse non-negative matrix factorization for multi-view data representation [J]. *IEEE Transactions on Cybernetics*, 2018, 48(9): 1-13.
- [12] LIU J, JIANG Y, LI Z C, *et al.* Partially shared latent factor learning with multiview data [J]. *IEEE Transactions on Neural Networks*, 2015, 26(6): 1233-1246.
- [13] CARBONNEAU M A, CHEPLYGINA V, GRANGER E, *et al.* Multiple instance learning: a survey of problem characteristics and applications [J]. *Pattern Recognition*, 2017, 77: 329-353.
- [14] MELKI G, CANO A, VENTURA S. Mirsvm: multi-instance support vector machine with bag representatives [J]. *Pattern Recognition*, 2018, 79: 228-241.
- [15] ANDERWS S, TSOCHANTARIDIS I, HOFMANN T. Support vector machines for multiple-instance learning[C]//*Proceedings of the Neural Information Processing Systems*. Vancouver: Nips, 2003: 577-584.