计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.2019.20180470
56(7): 1370-1382, 2019

# 共享和私有信息最大化的跨媒体聚类

闫小强　　叶阳东
(郑州大学信息工程学院　郑州　450001)
(iexqyan@zzu.edu.cn)

# Cross-Media Clustering by Share and Private Information Maximization

Yan Xiaoqiang and Ye Yangdong
(*School of Information Engineering*, *Zhengzhou University*, *Zhengzhou* 450001)

**Abstract**　Recently, the rapid emergence of cross media data with typical multi-source and heterogeneous characteristic brings great challenges to the traditional data analysis approaches. However, the most of existing approaches for cross media data heavily rely on the shared latent feature space to construct the relationships between multiple modalities, while ignoring the private information hidden in each modality. Aiming at this problem, this paper proposes a novel share and private information maximization (SPIM) algorithm for cross media data clustering, which leverages the shared and private information into the clustering process. Firstly, we present two shared information construction models: 1) Hybrid words (H-words) model. In this model, the low-level features in each modality are transformed into words or visual words co-occurrence vector, then a novel agglomerative information maximization is presented to build the hybrid word space for all modalities, which ensures the statistical correlation between the low-level features of multiple modalities. 2) Clustering ensemble (CE) model. This model adopts the mutual information to measure the similarity between the clustering partitions of different modalities, which ensures the semantic correlation of the high-level clustering partitions. Secondly, SPIM algorithm integrates the shared information of multiple modalities and the private information of individual modalities into a unified objective function. Finally, the optimization of SPIM algorithm is performed by a sequential "draw-and-merge" procedure, which guarantees the function converge to a local maximum. The experimental results on 6 cross media datasets show that the proposed approach compares favorably with the existing state-of-the-art cross-media clustering methods.

**Key words**　cross-media; multi-source heterogeneous; share and private information; information maximization; mutual information; clustering analyse

**摘　要**　近年来,具有典型多源异构特性的跨媒体数据的快速涌现给数据分析带来巨大挑战.然而,绝大多数现有跨媒体数据分析方法仅依赖模态间的共享信息发掘跨媒体数据中蕴含的模式结构,忽略各模态自身的重要信息.针对此问题,提出共享和私有信息最大化(share and private information maximization)的跨媒体聚类算法,通过兼顾跨媒体数据的共享和私有信息,以求得更加合理的聚类模

**收稿日期**:2018-06-27;**修回日期**:2019-02-13
**基金项目**:国家重点研发计划项目(2018YFB1201403);国家自然科学基金项目(61772475,61502434)
This work was supported by the National Key Research and Development Program of China (2018YFB1201403) and the National Natural Science Foundation of China (61772475, 61502434).
**通信作者**:叶阳东(ieydye@zzu.edu.cn)

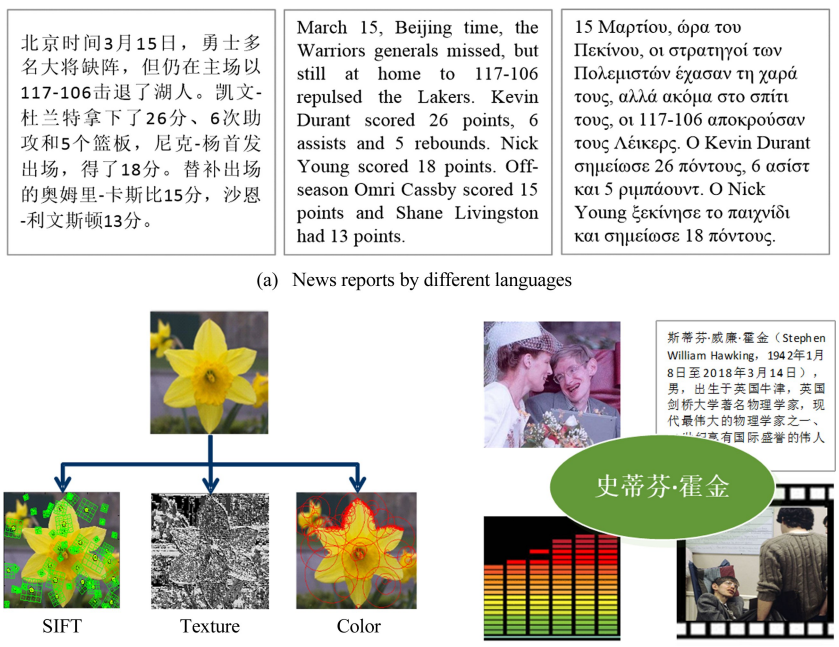
式.首先,提出 2 种跨媒体数据的共享信息构建模型:1)混合单词模型,该模型将各模态的底层特征转换为统一的词频向量表示,然后使用一种新的自凝聚信息最大化方法自底向上地构建多模态的混合单词空间,最大化地保持各模态底层特征的统计相似性;2)聚类集成模型,构建各模态自身的聚类划分,通过互信息度量各模态聚类划分间的信息量,抽取各模态的高层聚类划分之间的相关性.其次,提出基于信息论的目标函数,将跨媒体数据的共享和私有信息融合在同一目标函数中,在抽取聚类模式结构的过程中兼顾跨媒体数据的共享和私有信息.最后,采用顺序"抽取-合并"过程优化 SPIM 算法的目标函数,保证其收敛到局部最优解.在 6 种跨媒体数据上的实验结果表明 SPIM 算法的优越性.

**关键词** 跨媒体;多源异构;共享和私有信息;信息最大化;互信息;聚类分析

**中图法分类号** TP181

聚类方法按照"物以类聚"的原则将数据对象划分为不同的簇,并保持簇间数据元素间的距离尽可能地大、簇内数据元素间的距离尽可能地小,进而抽取数据对象中蕴含的模式结构.聚类分析无须借鉴数据的先验知识,仅根据数据的实际分布情况即可得到自然的数据划分,在认识数据中的不确定性和价值的隐蔽性方面具有重要的研究价值.然而,随着信息技术的迅猛发展和广泛应用,具有典型多源异构特性的跨媒体数据已经遍布生活的各个角落.跨媒体数据是指以不同模态、来源、空间等形式出现,但具有相似的高层语义的数据.如图 1 所示,相同的新闻可以用多种语言进行报道[1];同一幅图像可以在形状、纹理、颜色等特征空间上获得异构的描述[2];同一概念或事件可用图像、文本、视频、音频等不同类型的媒体共同表达[3];传统聚类方法在做数据分析时仅考虑单模态的数据信息,已经无法适应跨媒体数据的特征异构性.

跨媒体聚类(cross-media clustering, CMC)[4]是一种基于数据驱动的分析方法,旨在根据不同模态数据的分布相似性,同时聚类多个模态以揭示不同模态间的潜在关联.跨媒体聚类任务的核心问题在于捕捉多模态数据间的关联性,以实现信息的跨模态共享.针对此问题,较为直观的解决方法是寻找多模态数据的公共子空间.例如文献[5]提出基于典型关联分析(canonical correlation analysis, CCA)[6]的跨媒体聚类算法,该算法将多个模态的特征投影到低维的子空间上;文献[7]使用共享核嵌入、文献[8]使用高斯过程隐式变量模型学习多模态间的共享特征.然而,基于子空间的跨媒体聚类方法将各模态的特征映射到低维空间的同时,会破坏跨




(a) News reports by different languages

(b) The heterogeneous image features

(c) Different media types

Fig. 1 The typical cross-media data

图 1 典型的跨媒体数据

媒体数据的原始结构,导致一些重要信息的丢失.除了学习多模态共享子空间之外,近年来相关研究人员也提出了一些行之有效的跨媒体聚类策略.文献[9]使用层次模型[10](hierarchical model)自底向上地构建文本和视觉特征间的关联,进而在聚类过程中结合跨媒体特征;文献[11]首先将图像分割为区域的集合,此时,若将图像视为文档,则每个图像区域类似于文档中的单词,之后通过主题模型隐含狄利克雷分布(latent Drichlet allocation, LDA)[12]将文本与视觉信息转化为一种跨模态的向量表示;文献[2]提出多模态谱聚类算法,自动地结合图像数据的多种异构特征表示;文献[13]提出多视角联合矩阵分解方法,通过学习多视角数据的公共系数矩阵寻求多视角之间兼容的聚类划分;文献[14]提出鲁棒的多视角 $k$-means 方法,用来处理大规模数据的多种异构特征表示.然而,上述跨媒体数据聚类分析方法仅依赖各模态间的共享信息建立多模态数据的关联,忽略了各模态自身的私有信息,这显然与实际应用情况不符.

另外,机器学习领域中的集成聚类(consensus clustering, CC)方法可有效地处理多模态数据,引起了跨媒体研究人员的关注.集成聚类方法在处理跨媒体数据时,首先根据单模态的数据分布得到其自身的聚类划分(基聚类),之后按照特定的合并准则将不同模态的聚类划分进行合并,从而将多个模态的异构信息进行融合,得到最终的聚类划分.例如文献[15]设计基于相似簇、超图、集群 3 种一致性度量函数合并基聚类;文献[16]提出基于稀疏图表示和概率轨迹的聚类集成算法,同时根据基聚类的局部和全局信息获取最终的聚类划分;文献[17]在不考虑原始数据分布的情况下,通过局部密度估计方法对基聚类进行加权,进而区分基聚类的可依赖程度.然而,现有的集成聚类在处理跨媒体数据时忽略原始数据的特征分布,仅依赖基聚类构建最终的聚类划分,导致最终的聚类划分过度依赖基聚类的质量.

针对上述问题,本文提出共享和私有信息最大化(share and private information maximization, SPIM)的跨媒体聚类算法.如图 2 所示,该算法通过兼顾跨媒体数据间的共享信息和各媒体数据自身的私有信息进行聚类分析,以求得更加合理的聚类模式结构.首先,提出混合单词模型(hybrid words model, H-words)和聚类集成模型(clustering ensemble model, CE)构建跨媒体数据的 2 种共享信息,分别保持各模态底层特征的统计相似性和各模态的高层聚类划分间的相关性.其次,提出基于信息论的目标函数,将跨媒体数据的共享和私有信息融合在同一目标函数中.同时处理各模态自身的私有信息(原始特征)和聚类集成模型构建的共享信息(聚类划分),有助于克服集成聚类算法对基聚类的过度依赖.最后,采用顺序"抽取-合并"优化过程,保证 SPIM 算法的目标函数收敛到局部最优解.在 6 种跨媒体数据上的实验结果表明 SPIM 算法性能优于现有方法.本文的主要贡献总结为 3 个方面:

1) 提出共享和私有信息最大化的跨媒体聚类算法 SPIM,该算法通过兼顾跨媒体数据的共享和私有信息,以求得更加合理的模式结构.

2) 提出 2 种跨媒体数据的共享信息构建模型:混合单词模型和聚类集成模型,分别保持各模态底层特征的统计相似性和各模态的高层聚类划分间的相关性.

3) 提出基于信息论的目标函数,并采用顺序"抽取-合并"优化策略对该目标函数进行优化,保证其收敛到局部最优解.

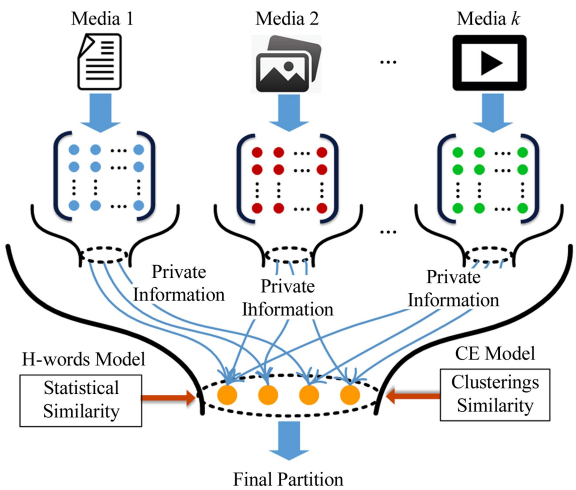



Fig. 2 The illustration of SPIM method

图 2 SPIM 方法示意图

本文工作与文献[18-19]中的多特征信息瓶颈(multi-feature information bottleneck, MfIB)算法、文献[20]中的多视角概念学习(multi-view concept learning, MCL)算法、文献[21]中的跨媒体社交图像聚类(cross-media social image clustering, CSIC)算法关联较为密切.其中 MfIB 算法是信息瓶颈(information bottlenede, IB)算法的扩展,该算法使用互信息量化聚类模式与多个特征间的信息量,然而,该算法要求多种特征表示来自同一数据分布,

无法处理跨模态的异构特征.MCL 算法和 CSIC 算法采用多模态数据的共享特征和独有特征进行跨模态分析.其中,MCL 算法学习多模态数据的概念性隐式空间,并将该隐式空间分解为共享和独有 2 个部分.然而,该算法是半监督算法,无法处理聚类问题.CSIC 算法将社交图像中视觉和社交标签信息的共享特征空间学习视为一个共轭词典学习问题,通过 $L_{1,\infty}$ 范数的正则项保证各模态的词典稀疏化,这种稀疏化使得各模态的独有特征得以保留.然而,该算法需要借助 WordNet 获得辅助的社交语义关系,仅适用图像和文本 2 种模态数据,无法处理更多模态.近年来,深度神经网络在跨媒体数据的一致性表征中取得了较好的结果,例如多层次跨模态关联学习[22]、多网络共享表征[23].深度神经网络需要大量已标注的训练数据,然而,数据标签信息的获取是费时费力的过程.

## 1 背景知识

IB 算法[24-25]是一种典型的基于信息最大化的数据分析方法[18-19,26].给定源变量 $\boldsymbol{X}$ 与特征变量 $\boldsymbol{Y}$ 之间的联合概率分布 $p(\boldsymbol{X},\boldsymbol{Y})$,IB 算法力图寻求源变量 $\boldsymbol{X}$ 的最优压缩表示 $\boldsymbol{T}$,同时使压缩变量 $\boldsymbol{T}$ 最大化地保存特征变量 $\boldsymbol{Y}$ 中蕴含的关于源变量 $\boldsymbol{X}$ 的信息量.如图 3 所示,源变量 $\boldsymbol{X}$ 与其特征变量 $\boldsymbol{Y}$ 之间的信息通过压缩变量 $\boldsymbol{T}$ 进行保存.IB 算法可形式化描述为

$$\mathcal{R}(D)=\min_{\{p(\boldsymbol{t}|\boldsymbol{x}):I(\boldsymbol{T};\boldsymbol{Y})\geqslant D\}} I(\boldsymbol{X};\boldsymbol{T}), \tag{1}$$

其中,$p(\boldsymbol{t}|\boldsymbol{x})$是源变量 $\boldsymbol{X}$ 到压缩变量 $\boldsymbol{T}$ 之间的编码方案,$I(\boldsymbol{X};\boldsymbol{T})$是变量之间的互信息,$D$ 是 $\boldsymbol{X}$ 到 $\boldsymbol{T}$ 之间所有可能的编码方案.从式(1)可知,IB 算法是在信息保存程度满足 $I(\boldsymbol{T};\boldsymbol{Y})\geqslant D$ 的条件下,寻找能够最小化 $I(\boldsymbol{X};\boldsymbol{T})$的一个编码方案 $p(\boldsymbol{t}|\boldsymbol{x})$.文献[22]给出 IB 算法的目标函数:

$$\mathcal{L}_{\max}[p(\boldsymbol{t}|\boldsymbol{x})]=I(\boldsymbol{T};\boldsymbol{Y})-\beta^{-1}\cdot I(\boldsymbol{T};\boldsymbol{X}), \tag{2}$$

其中,$\beta$ 是拉格朗日乘数因子,用来平衡数据对象的压缩与相关信息的保存.在聚类任务中,簇的个数 $M$ 往往远远小于原始数据对象的数量,即 $M\ll|\boldsymbol{X}|$,这意味着源变量 $\boldsymbol{X}$ 与其压缩变量 $\boldsymbol{T}$ 之间存在大幅度压缩.因此,实际应用中通常将 $\beta$ 设置为无穷大.这种做法的有效性在多种数据类型的聚类任务中得到验证,例如文本[25]、图像[18]、视频[26].因此,IB 算法的目标函数可改写为

$$\mathcal{L}_{\max}[p(\boldsymbol{t}|\boldsymbol{x})]=I(\boldsymbol{T};\boldsymbol{Y}), \tag{3}$$

其中,互信息 $I(\boldsymbol{T};\boldsymbol{Y})$度量压缩变量 $\boldsymbol{T}$ 与特征变量 $\boldsymbol{Y}$ 之间的信息量.

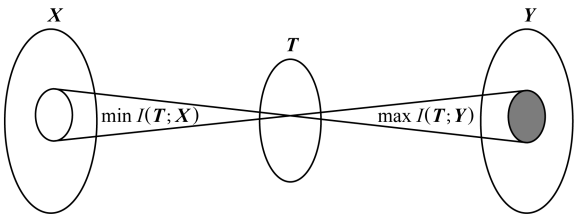



Fig. 3 The model of IB algorithm

图 3 IB 算法的模型图

## 2 共享和私有信息最大化的跨媒体聚类

本文提出共享和私有信息最大化的跨媒体聚类算法 SPIM,该算法旨在兼顾跨模态数据间的共享信息和各模态自身的私有信息进行跨媒体聚类分析,以求得更加合理的模式结构.为了清晰地描述,本节首先给出 SPIM 算法的问题定义.

**定义 1.** 使用源变量 $\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)$表示跨媒体数据对象的集合,使用私有变量 $\boldsymbol{Y}^1,\boldsymbol{Y}^2,\cdots,\boldsymbol{Y}^k$ 表示不同模态数据自身的私有信息(即各模态自身的特征表示,又称作源变量 $\boldsymbol{X}$ 的特征变量),使用共有变量 $\boldsymbol{S}$ 表示模态间的共享信息,其中,$n$ 是数据对象的数量,$k$ 是跨媒体数据的模态数量.SPIM 算法的目标是寻求源变量 $\boldsymbol{X}$ 到压缩变量 $\boldsymbol{T}$ 的一种最优压缩表示 $p(\boldsymbol{t}|\boldsymbol{x})$,在压缩过程中使压缩变量 $\boldsymbol{T}$ 同时最大化地保存与各模态自身的私有变量 $\boldsymbol{Y}^1,\boldsymbol{Y}^2,\cdots,\boldsymbol{Y}^k$ 和共有变量 $\boldsymbol{S}$ 间的信息.根据式(3),可给出 SPIM 目标函数:

$$\mathcal{L}_{\max}[p(\boldsymbol{t}|\boldsymbol{x})]=\sum_{i=1}^{k}I(\boldsymbol{T};\boldsymbol{Y}^i)+\lambda\cdot I(\boldsymbol{T};\boldsymbol{S}), \tag{4}$$

其中,$\lambda$ 是控制私有信息与共享信息之间侧重程度的平衡参数.当 $\lambda=0$ 且 $k=1$ 时,SPIM 算法回归至 IB 算法,因此,IB 算法可视为 SPIM 算法的特例.

给定数据对象与其特征表示的联合概率分布,则压缩变量与各模态自身的私有变量之间的互信息 $\sum_{i=1}^{k}I(\boldsymbol{T};\boldsymbol{Y}^i)$ 是可计算的.为计算 $I(\boldsymbol{T};\boldsymbol{S})$,接下来首先给出 2 种跨媒体数据的共享信息构建模型.

### 2.1 混合单词模型

词库模型[27]是一种常用的数据表示方法,该方法可将跨媒体的各模态数据转换为单词或视觉单词出现频率的向量形式,例如一幅城市场景的图像可

由高楼、街道、红绿灯等视觉单词出现频率表示;一则体育新闻可由比分、队员、场地等文字出现频率表示.在聚类任务中同时考虑不同模态数据的词频表示能在一定程度上刻画多模态间的关联性,但不同模态数据的词频向量在尺度上具有明显的差异,且存在较大的样本冗余.因此,本文提出混合单词H-words模型,首先将各模态的底层特征转换为统一的词频向量表示,然后使用自凝聚信息最大化方法自底向上地构建多模态的混合单词空间,抽取模态间的共享信息,最大化地保持各模态底层特征的统计相似性.

给定源变量 $\boldsymbol{X}$ 及其 $k$ 个模态的私有变量 $\boldsymbol{Y}^1,\boldsymbol{Y}^2,\cdots,\boldsymbol{Y}^k$,其中,私有变量 $\boldsymbol{Y}^i$ 的值域为 $\boldsymbol{Y}^i=(\boldsymbol{y}_1^i,\boldsymbol{y}_2^i,\cdots,\boldsymbol{y}_{m_i}^i)$,$m_i$ 为第 $i$ 个模态的特征维度,$\boldsymbol{y}_{x_j}^i$ 表示特征 $\boldsymbol{y}_j^i$ 在数据 $\boldsymbol{x}_j$ 中出现的统计次数.因此,可以得到 $k$ 个联合概率分布 $p(\boldsymbol{X},\boldsymbol{Y}^1),p(\boldsymbol{X},\boldsymbol{Y}^2),\cdots,p(\boldsymbol{X},\boldsymbol{Y}^k)$.本文定义目标函数对混合单词空间 $\widetilde{\boldsymbol{Y}}$ 进行求解:

$$\mathcal{F}_{\max}[p(\tilde{\boldsymbol{y}}\mid\boldsymbol{y}^i)]=I(\widetilde{\boldsymbol{Y}};\boldsymbol{X})-\sum_{i=1}^{k}I(\widetilde{\boldsymbol{Y}};\boldsymbol{Y}^i),\quad(5)$$

其中,$p(\tilde{\boldsymbol{y}}\mid\boldsymbol{y}^i)$是私有变量 $\boldsymbol{Y}^i$ 到混合单词空间 $\widetilde{\boldsymbol{Y}}$ 的映射.本文采用文献[26]中自底向上的层次模型对式(5)求解,该模型首先将每个数据元素视为单独的簇,即 $|\widetilde{\boldsymbol{Y}}|=\sum_{i=1}^{k}|\boldsymbol{Y}^i|$,其中 $|\widetilde{\boldsymbol{Y}}|$ 是混合单词空间的维度,$|\boldsymbol{Y}^i|$ 是第 $i$ 个模态的特征个数.之后通过不断合并最相似的数据元素,可逐步降低 $|\widetilde{\boldsymbol{Y}}|$ 的值,去除冗余特征.由于在自底向上的合并过程中最大化地保持各模态特征间的互信息,因此,我们称此过程为自凝聚信息最大化.假设 $\boldsymbol{y}_h,\boldsymbol{y}_g$ 是私有变量 $\boldsymbol{Y}^i$ 中的2个特征向量,则 $\boldsymbol{y}_h,\boldsymbol{y}_g$ 合并前后式(5)的变化量计算为

$$\Delta\mathcal{F}_{\max}(\boldsymbol{y}_h,\boldsymbol{y}_g)=\mathcal{F}_{\max}^{\text{bef}}-\mathcal{F}_{\max}^{\text{aft}},\quad(6)$$

其中,$\mathcal{F}_{\max}^{\text{bef}},\mathcal{F}_{\max}^{\text{aft}}$分别是将特征向量 $\boldsymbol{y}_h,\boldsymbol{y}_g$ 合并前后式(5)的值.

**定义2.** 合并 $\boldsymbol{y}_h,\boldsymbol{y}_g$ 生成簇 $\tilde{\boldsymbol{y}}$ 的概率计算为

$$p(\tilde{\boldsymbol{y}})=p(\boldsymbol{y}_h)+p(\boldsymbol{y}_g),\quad(7)$$

$$p(\tilde{\boldsymbol{y}}\mid\boldsymbol{y}^i)=\frac{p(\boldsymbol{y}_h)}{p(\tilde{\boldsymbol{y}})}p(\boldsymbol{y}^i\mid\boldsymbol{y}_h)+\frac{p(\boldsymbol{y}_g)}{p(\tilde{\boldsymbol{y}})}p(\boldsymbol{y}^i\mid\boldsymbol{y}_g).\quad(8)$$

给出自凝聚信息最大化的详细执行过程:

1) 将每个特征点初始化为1个单独簇.

2) 计算所有特征对合并引起式(5)的改变量 $\Delta\mathcal{F}_{\max}(\boldsymbol{y}_h,\boldsymbol{y}_g)$.

3) 合并满足 $\arg\min\Delta\mathcal{F}_{\max}(\boldsymbol{y}_h,\boldsymbol{y}_g)$的特征对.

4) 根据定义2更新 $p(\tilde{\boldsymbol{y}}),p(\tilde{\boldsymbol{y}}\mid\boldsymbol{y}^i)$,直至达到事先给定的簇个数.

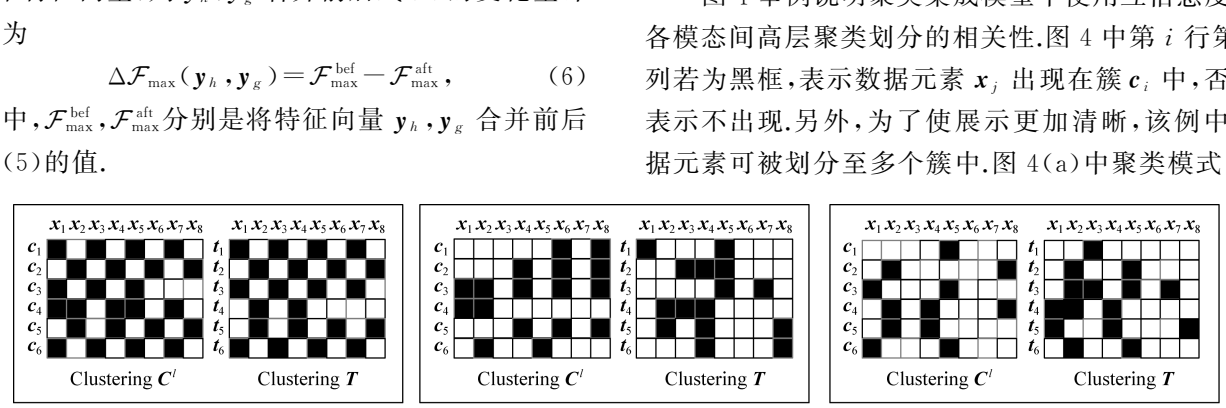
### 2.2 聚类集成模型

为了进一步挖掘各模态间的关联关系,本文提出聚类集成CE模型,首先为各模态数据构建各自的聚类划分,然后使用互信息度量各模态高层聚类划分间的相关性,进而捕捉到多个模态的异构信息.

SPIM算法的目标是寻求源变量 $\boldsymbol{X}$ 中的压缩变量 $\boldsymbol{T}$,假设 $\boldsymbol{T}$ 中有 $M$ 个簇 $\boldsymbol{T}=(\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_M)$.根据源变量 $\boldsymbol{X}$ 的 $k$ 个模态构建辅助聚类 $\boldsymbol{C}=(\boldsymbol{C}^1,\boldsymbol{C}^2,\cdots,\boldsymbol{C}^k)$,并假设第 $l$ 个模态的聚类划分 $\boldsymbol{C}^l$ 中同样具有 $M$ 个簇 $\boldsymbol{C}^l=(\boldsymbol{c}_1^l,\boldsymbol{c}_2^l,\cdots,\boldsymbol{c}_M^l)$.假设 $n_i$ 为 $\boldsymbol{C}^l$ 中被划分至簇 $\boldsymbol{c}_i^l$ 中的数据元素个数;$n_j$ 为聚类划分 $\boldsymbol{T}$ 中划分到簇 $\boldsymbol{t}_j$ 中数据元素个数;$n_{ij}$为同时属于簇 $\boldsymbol{c}_i^l$ 和簇 $\boldsymbol{t}_j$ 中数据元素个数,则聚类划分 $\boldsymbol{C}^l$ 和 $\boldsymbol{T}$ 之间的概率分布可计算为

$$\begin{aligned}p(\boldsymbol{c}_i^l,\boldsymbol{t}_j)&=\frac{n_{ij}}{n},\\ p(\boldsymbol{c}_i^l)&=\frac{n_i}{n},\\ p(\boldsymbol{t}_j)&=\frac{n_j}{n}.\end{aligned}\quad(9)$$

图4举例说明聚类集成模型中使用互信息度量各模态间高层聚类划分的相关性.图4中第 $i$ 行第 $j$ 列若为黑框,表示数据元素 $\boldsymbol{x}_j$ 出现在簇 $\boldsymbol{c}_i$ 中,否则表示不出现.另外,为了使展示更加清晰,该例中数据元素可被划分至多个簇中.图4(a)中聚类模式 $\boldsymbol{C}^l$




(a) The mutual information is 0.5200 (b) The mutual information is 0.2524 (c) The mutual information is 0.0703

Fig. 4 The mutual information between two clusters of different modalities

图4 不同模态的聚类划分间的互信息

和 $\boldsymbol{T}$ 高度相似,得到最高的互信息;图 4(b)中 $\boldsymbol{C}^l$ 和 $\boldsymbol{T}$ 的相关性减弱,互信息也相应降低;图 4(c)中 $\boldsymbol{C}^l$ 和 $\boldsymbol{T}$ 的相似度最弱,得到最低的互信息.因此,使用互信息可有效地度量跨模态高层聚类划分之间的相关性.

### 2.3 SPIM 算法的目标函数

根据定义 1 可知,SPIM 算法的目标是寻求源变量 $\boldsymbol{X}$ 到压缩变量 $\boldsymbol{T}$ 的一种最优压缩表示 $p(\boldsymbol{t}\mid\boldsymbol{x})$,在压缩过程中使压缩变量 $\boldsymbol{T}$ 最大化地保存与各模态自身的私有变量 $\boldsymbol{Y}^1,\boldsymbol{Y}^2,\cdots,\boldsymbol{Y}^k$ 和共有变量 $\boldsymbol{S}$ 的信息量,其中共有变量由混合单词模型和聚类集成模型共同求得,因此,SPIM 算法的目标函数可改写为

$$\mathcal{L}_{\max}[p(\boldsymbol{t}\mid\boldsymbol{x})]=\sum_{i=1}^{k}I(\boldsymbol{T};\boldsymbol{Y}^i)+\lambda\cdot[I(\boldsymbol{T};\widetilde{\boldsymbol{Y}})+\sum_{i=1}^{k}I(\boldsymbol{T};\boldsymbol{C}^i)],\tag{10}$$

其中,$\sum_{i=1}^{k}I(\boldsymbol{T};\boldsymbol{Y}^i)$ 度量压缩变量与各模态自身的私有变量之间的信息量;$I(\boldsymbol{T};\widetilde{\boldsymbol{Y}})$表示压缩变量与跨模态共享特征之间的信息量;$\sum_{i=1}^{k}I(\boldsymbol{T};\boldsymbol{C}^i)$ 度量压缩变量与跨模态的高层聚类划分之间的信息量;$\lambda$ 是控制私有信息与共享信息之间的侧重程度的平衡参数.

### 2.4 SPIM 算法目标函数的优化

本节使用顺序“抽取-合并”策略优化 SPIM 算法的目标函数,求解源变量 $\boldsymbol{X}$ 到压缩变量 $\boldsymbol{T}$ 之间的编码方案 $p(\boldsymbol{t}\mid\boldsymbol{x})$.顺序“抽取-合并”优化方法包含 3 个步骤:

1) 将源变量 $\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n)$ 随机划分至 $M$ 个簇 $\boldsymbol{T}=(\boldsymbol{t}_1,\boldsymbol{t}_2,\cdots,\boldsymbol{t}_M)$;

2) 顺序地将每个数据元素 $\boldsymbol{x}$ 从当前簇 $\boldsymbol{t}^{\text{old}}$ 中抽取出来,作为单独簇$\{\boldsymbol{x}\}$;

3) 计算将单独簇$\{\boldsymbol{x}\}$合并至其他簇中时 SPIM 算法目标函数中信息的损失量,并选取使得信息损失最小的簇 $\boldsymbol{t}^{\text{new}}$进行合并.

顺序抽取合并策略的核心问题是在迭代过程中选取合适的簇 $\boldsymbol{t}^{\text{new}}$对单独簇$\{\boldsymbol{x}\}$进行合并.使用 $\mathcal{L}^{\text{old}}$ 表示将 $\boldsymbol{x}$ 从簇 $\boldsymbol{t}^{\text{old}}$ 中抽取之前式(10)的值;使用 $\mathcal{L}^{\text{bef}}$ 表示抽取 $\boldsymbol{x}$ 之后目标函数的值;使用 $\mathcal{L}^{\text{aft}}$ 表示将单独簇$\{\boldsymbol{x}\}$合并至 $\boldsymbol{t}^{\text{new}}$之后式(10)的值,$\boldsymbol{t}^{\text{new}}$满足 $\boldsymbol{t}^{\text{new}}=\arg\min\Delta\mathcal{L}=\mathcal{L}^{\text{bef}}-\mathcal{L}^{\text{aft}}$,其中,$\Delta\mathcal{L}$ 是$\{\boldsymbol{x}\}$合并前后式(10)的改变量,在此称之为合并代价,其计算过程为

$$\begin{aligned}\Delta\mathcal{L}&=\mathcal{L}^{\text{bef}}-\mathcal{L}^{\text{aft}}=\sum_{i=1}^{k}[I(\boldsymbol{T}^{\text{bef}};\boldsymbol{Y}^i)-I(\boldsymbol{T}^{\text{aft}};\boldsymbol{Y}^i)]+\lambda\cdot[I(\boldsymbol{T}^{\text{bef}};\widetilde{\boldsymbol{Y}})-I(\boldsymbol{T}^{\text{aft}};\widetilde{\boldsymbol{Y}})+\\&\lambda\cdot\sum_{i=1}^{k}[I(\boldsymbol{T}^{\text{bef}};\boldsymbol{C}^i)-I(\boldsymbol{T}^{\text{aft}};\boldsymbol{C}^i)]=\\&\sum_{i=1}^{k}\Delta I_i^{\text{private}}+\lambda\Delta I^{\text{com}}+\lambda\sum_{i=1}^{k}\Delta I_i^{\text{clustering}},\end{aligned}\tag{11}$$

其中,$\Delta I_i^{\text{private}}$,$\Delta I^{\text{com}}$,$\Delta I_i^{\text{clustering}}$ 分别是目标函数中 $I(\boldsymbol{T};\boldsymbol{Y}^i)$,$I(\boldsymbol{T};\widetilde{\boldsymbol{Y}})$,$I(\boldsymbol{T};\boldsymbol{C}^i)$的合并代价.我们首先计算 $\Delta I_i^{\text{private}}$,假设单独簇$\{\boldsymbol{x}\}$被合并到簇 $\boldsymbol{t}$ 形成新簇 $\tilde{\boldsymbol{t}}$,则:

$$\begin{cases}p(\tilde{\boldsymbol{t}})=p(\boldsymbol{x})+p(\boldsymbol{t}),\\p(\boldsymbol{y}^i\mid\tilde{\boldsymbol{t}})=\dfrac{p(\boldsymbol{x})}{p(\tilde{\boldsymbol{t}})}p(\boldsymbol{y}^i\mid\boldsymbol{x})+\dfrac{p(\boldsymbol{t})}{p(\tilde{\boldsymbol{t}})}p(\boldsymbol{y}^i\mid\boldsymbol{t}).\end{cases}\tag{12}$$

根据互信息的定义,可得:

$$\begin{aligned}\Delta I_i^{\text{private}}&=I(\boldsymbol{T}^{\text{bef}};\boldsymbol{Y}^i)-I(\boldsymbol{T}^{\text{aft}};\boldsymbol{Y}^i)=\\&p(\boldsymbol{t})\sum_{\boldsymbol{y}^i\in\boldsymbol{Y}^i}p(\boldsymbol{y}^i\mid\boldsymbol{t})\mathrm{lb}\frac{p(\boldsymbol{y}^i\mid\boldsymbol{t})}{p(\boldsymbol{y}^i)}+\\&p(\boldsymbol{x})\sum_{\boldsymbol{y}^i\in\boldsymbol{Y}^i}p(\boldsymbol{y}^i\mid\boldsymbol{x})\mathrm{lb}\frac{p(\boldsymbol{y}^i\mid\boldsymbol{x})}{p(\boldsymbol{y}^i)}-\\&p(\tilde{\boldsymbol{t}})\sum_{\boldsymbol{y}^i\in\boldsymbol{Y}^i}p(\boldsymbol{y}^i\mid\tilde{\boldsymbol{t}})\mathrm{lb}\frac{p(\boldsymbol{y}^i\mid\tilde{\boldsymbol{t}})}{p(\boldsymbol{y}^i)}.\end{aligned}$$

把式(12)代入 $\Delta I_i^{\text{private}}$ 中可得:

$$\begin{aligned}\Delta I_i^{\text{private}}&=I(\boldsymbol{T}^{\text{bef}};\boldsymbol{Y}^i)-I(\boldsymbol{T}^{\text{aft}};\boldsymbol{Y}^i)=\\&p(\boldsymbol{x})\sum_{\boldsymbol{y}^i\in\boldsymbol{Y}^i}p(\boldsymbol{y}^i\mid\boldsymbol{x})\mathrm{lb}\frac{p(\boldsymbol{y}^i\mid\boldsymbol{x})}{p(\boldsymbol{y}^i)}+\\&p(\boldsymbol{t})\sum_{\boldsymbol{y}_i\in\boldsymbol{Y}_i}p(\boldsymbol{y}^i\mid\boldsymbol{t})\mathrm{lb}\frac{p(\boldsymbol{y}^i\mid\boldsymbol{t})}{p(\boldsymbol{y}^i)}-\\&\sum_{\boldsymbol{y}^i\in\boldsymbol{Y}^i}p(\boldsymbol{x})p(\boldsymbol{y}^i\mid\boldsymbol{x})\mathrm{lb}\frac{p(\boldsymbol{y}^i\mid\tilde{\boldsymbol{t}})}{p(\boldsymbol{y}^i)}-\\&\sum_{\boldsymbol{y}^i\in\boldsymbol{Y}^i}p(\boldsymbol{t})p(\boldsymbol{y}^i\mid\boldsymbol{t})\mathrm{lb}\frac{p(\boldsymbol{y}^i\mid\tilde{\boldsymbol{t}})}{p(\boldsymbol{y}^i)}=\\&p(\boldsymbol{x})\sum_{\boldsymbol{y}^i\in\boldsymbol{Y}^i}p(\boldsymbol{y}^i\mid\boldsymbol{x})\mathrm{lb}\frac{p(\boldsymbol{y}^i\mid\boldsymbol{x})}{p(\boldsymbol{y}^i\mid\tilde{\boldsymbol{t}})}+\\&p(\boldsymbol{t})\sum_{\boldsymbol{y}^i\in\boldsymbol{Y}^i}p(\boldsymbol{y}^i\mid\boldsymbol{t})\mathrm{lb}\frac{p(\boldsymbol{y}^i\mid\boldsymbol{t})}{p(\boldsymbol{y}^i\mid\tilde{\boldsymbol{t}})}=\\&[p(\boldsymbol{x})+p(\boldsymbol{t})]\cdot JS_\pi[p(\boldsymbol{Y}^i\mid\boldsymbol{x}),p(\boldsymbol{Y}^i\mid\boldsymbol{t})],\end{aligned}$$

其中,$\pi=\{\pi_1,\pi_2\}=\{\dfrac{p(\boldsymbol{x})}{p(\boldsymbol{x})+p(\boldsymbol{t})},\dfrac{p(\boldsymbol{t})}{p(\boldsymbol{x})+p(\boldsymbol{t})}\}$,$JS_\pi$ 为 2 个概率分布之间的 Jensen-Shannon 距离.同理可知 $I(\boldsymbol{T};\widetilde{\boldsymbol{Y}})$的合并代价 $\Delta I^{\text{com}}$:

$$\Delta I^{\text{com}}=I(\boldsymbol{T}^{\text{bef}};\widetilde{\boldsymbol{Y}})-I(\boldsymbol{T}^{\text{aft}};\widetilde{\boldsymbol{Y}})=[p(\boldsymbol{x})+p(\boldsymbol{t})]\cdot JS_\pi[p(\widetilde{\boldsymbol{Y}}\mid\boldsymbol{x}),p(\widetilde{\boldsymbol{Y}}\mid\boldsymbol{t})].\tag{13}$$

在将$\{\boldsymbol{x}\}$合并至簇 $\boldsymbol{t}$ 时,根据聚类集成模型中的式(9)可得到 $p(\boldsymbol{T}^{\text{bef}},\boldsymbol{C}^i)$,$p(\boldsymbol{T}^{\text{aft}},\boldsymbol{C}^i)$,因此可计算

将$\{\boldsymbol{x}\}$合并前后由$I(\boldsymbol{T};\boldsymbol{C}^i)$引起的合并代价.综上所述,可计算将$\{\boldsymbol{x}\}$合并至簇$\boldsymbol{t}$的合并代价:

$$\Delta\mathcal{L}=\sum_{i=1}^{k}\Delta I_i^{\text{private}}+\lambda\Delta I^{\text{com}}+\lambda\sum_{i=1}^{k}\Delta I_i^{\text{clustering}}. \quad (14)$$

给出SPIM算法的实现过程:

**算法1.** SPIM算法.

输入:源变量$\boldsymbol{X}$与多个模态的私有变量$\boldsymbol{Y}^1,\boldsymbol{Y}^2,\cdots,\boldsymbol{Y}^k$之间的联合概率分布$p(\boldsymbol{X},\boldsymbol{Y}^1),p(\boldsymbol{X},\boldsymbol{Y}^2),\cdots,p(\boldsymbol{X},\boldsymbol{Y}^k)$;簇的个数$M$;平衡参数$\lambda$.

输出:$\boldsymbol{X}$到$\boldsymbol{T}$的编码方案$p(\boldsymbol{t}|\boldsymbol{x})$.

① 根据混合词库模型生成多个模态的混合单词空间$\widetilde{\boldsymbol{Y}}$,并通过式$(|\boldsymbol{Y}^1|+|\boldsymbol{Y}^2|+\cdots+|\boldsymbol{Y}^k|)/k$确定$\widetilde{\boldsymbol{Y}}$的维度;

② 根据聚类集成模型构建多个模态自身的聚类划分$\boldsymbol{C}^1,\boldsymbol{C}^2,\cdots,\boldsymbol{C}^k$;

③ 将源变量$\boldsymbol{X}$随机划分为$M$个簇;

④ Repeat:

⑤ For every $\boldsymbol{x}\in\boldsymbol{X}$

⑥ 将$\boldsymbol{x}$从当前簇$\boldsymbol{t}^{\text{old}}$中抽取出来,作为单独簇$\{\boldsymbol{x}\}$;

⑦ 合并$\{\boldsymbol{x}\}$至簇$\boldsymbol{t}^{\text{new}}$,其中$\boldsymbol{t}^{\text{new}}=\arg\min\Delta\mathcal{L}$,并根据式(14)计算合并代价;

⑧ End For

⑨ Until $p(\boldsymbol{t}|\boldsymbol{x})$不再发生变化.

### 2.5 算法分析

#### 2.5.1 收敛性分析

**定理1.** SPIM算法可在有限迭代次数内收敛到局部最优解.

证明. 首先证明每次抽取合并过程都能够增加式(10)的值.使用$\mathcal{L}^{\text{old}}$表示将$\boldsymbol{x}$从其原始簇$\boldsymbol{t}^{\text{old}}$中抽取之前目标函数的值,使用$\mathcal{L}^{\text{bef}}$和$\mathcal{L}^{\text{aft}}$表示合并单独簇$\{\boldsymbol{x}\}$至$\boldsymbol{t}^{\text{new}}=\arg\min\Delta\mathcal{L}$前后的目标函数值.合并过程有2种情况:1)$\boldsymbol{t}^{\text{new}}=\boldsymbol{t}^{\text{old}}$,意味着将$\{\boldsymbol{x}\}$合并至其原本所在的簇$\boldsymbol{t}^{\text{old}}$,此时,数据分布没有发生改变,则$\mathcal{L}^{\text{bef}}=\mathcal{L}^{\text{aft}}$;2)$\boldsymbol{t}^{\text{new}}\neq\boldsymbol{t}^{\text{old}}$,意味着将$\{\boldsymbol{x}\}$合并至其他簇$\boldsymbol{t}^{\text{new}}$,因为$\boldsymbol{t}^{\text{new}}=\arg\min\Delta\mathcal{L}$,故将$\{\boldsymbol{x}\}$合并至新簇$\boldsymbol{t}^{\text{new}}$的合并代价$\Delta(\boldsymbol{x},\boldsymbol{t}^{\text{new}})$一定小于将其合并至原始簇$\boldsymbol{t}^{\text{old}}$的合并代价$\Delta\mathcal{L}(\boldsymbol{x},\boldsymbol{t}^{\text{old}})$,即$\Delta\mathcal{L}(\boldsymbol{x},\boldsymbol{t}^{\text{new}})<\Delta\mathcal{L}(\boldsymbol{x},\boldsymbol{t}^{\text{old}})$.又因为$\Delta\mathcal{L}(\boldsymbol{x},\boldsymbol{t}^{\text{old}})=\mathcal{L}^{\text{bef}}-\mathcal{L}^{\text{old}}$,$\Delta\mathcal{L}(\boldsymbol{x},\boldsymbol{t}^{\text{new}})=\mathcal{L}^{\text{bef}}-\mathcal{L}^{\text{aft}}$,故$\mathcal{L}^{\text{aft}}>\mathcal{L}^{\text{old}}$.因此,每次抽取合并过程都有$\mathcal{L}^{\text{aft}}\geqslant\mathcal{L}^{\text{old}}$.

证明SPIM算法的目标函数值有上界.压缩变量$\boldsymbol{T}$是源变量$\boldsymbol{X}$的一种压缩表示,则必有$I(\boldsymbol{T};\boldsymbol{Y}^i)\leqslant I(\boldsymbol{X};\boldsymbol{Y}^i)$和$I(\boldsymbol{T};\widetilde{\boldsymbol{Y}})\leqslant I(\boldsymbol{X};\widetilde{\boldsymbol{Y}})$,当且仅当$|\boldsymbol{X}|=|\boldsymbol{Y}|$时,$I(\boldsymbol{T};\boldsymbol{Y}^i)\leqslant I(\boldsymbol{X};\boldsymbol{Y}^i)$和$I(\boldsymbol{T};\widetilde{\boldsymbol{Y}})\leqslant I(\boldsymbol{X};\widetilde{\boldsymbol{Y}})$中的等号成立,此时$\boldsymbol{X}$到$\boldsymbol{T}$不存在压缩.另外,假设源变量$\boldsymbol{X}$的正确划分为$\boldsymbol{C}$,则$I(\boldsymbol{T};\boldsymbol{C}^i)\leqslant I(\boldsymbol{T};\boldsymbol{C})$.综上所述,SPIM算法目标函数值的上界为$\sum_{i=1}^{k}I(\boldsymbol{X};\boldsymbol{Y}^i)+\lambda\cdot[I(\boldsymbol{X};\widetilde{\boldsymbol{Y}})+\sum_{i=1}^{k}I(\boldsymbol{T};\boldsymbol{C})]$.因此SPIM算法可在有限迭代次数内收敛到局部最优解. 证毕.

#### 2.5.2 复杂度分析

预处理中步骤①可在$O(|\widetilde{\boldsymbol{Y}}|^2)$的时间内构建多模态公共特征空间.预处理中步骤②的时间复杂度由构建基聚类所采用的聚类算法决定,本文使用顺序信息瓶颈(sequectial information bottleneck, sIB)算法,其时间复杂度为$O(n\log n)$.初始化过程可在$O(1)$时间内实现.在SPIM算法主循环中,抽取合并过程需要计算合并代价,时间复杂度为$O(|\boldsymbol{X}|(|\boldsymbol{Y}^1|+|\boldsymbol{Y}^2|+\cdots+|\boldsymbol{Y}^k|+|\widetilde{\boldsymbol{Y}}|))$,因此,SPIM算法的时间复杂度为$O(M|\boldsymbol{X}|(|\boldsymbol{Y}^1|+|\boldsymbol{Y}^2|+\cdots+|\boldsymbol{Y}^k|+|\widetilde{\boldsymbol{Y}}|))$,其中$M$是最终聚类划分中簇的个数.

## 3 实验与性能分析

### 3.1 数据集

本文在6种跨媒体数据集上验证SPIM算法的有效性:

1) Wikipedia数据集[28].该数据集包含2 866个文本图像对共计10个类,每幅图像的共生文本至少有70个单词.对于图像,本文采用文献[28]提供的128维SIFT[29]特征构建BoVW视觉特征表示;对于文本,首先构建500维BoW表示,然后通过LDA抽取100个话题的概率分布作为文本特征表示.

2) Pascal Sentence数据集[30].该数据集包含1 000个文本图像对共计20个类.对于图像,本文抽取1 024维的SIFT BoVW视觉特征表示;对于文本,首先构建300维BoW表示,然后通过LDA抽取100个话题的概率分布作为文本特征表示.

3) Pascal VOC 2007数据集[31].该数据集包含9 963个文本图像对共计20个类.本文采用文献[32]提供的798维基于标签排序的文本特征和776维的BoVW视觉特征.

4) X-Media[33-34]是针对检索任务[35]构建的跨媒体数据集,该数据集包含文本、图像、视频、音频、3维模型等多种模态的数据,共计20个类.我们使用5 000个文本、图像对评估算法性能.关于该数据集

的特征描述,本文采用文献[33]提供的10维LDA文本特征、128维BoVW图像特征.

5) Reuters多语种数据集[1].该数据集由5种语言的新闻文档组成,包括西班牙语、意大利语、德语、法语和英语,每种语言的文档被分为6类:C15,CCAT,E21,ECAT,GCAT,M11.本文随机从每种语言类中挑取500个文档,并使用BoW模型抽取1 000个关键词,为每个语种构建1 000维BoW表示.

6) HMDB数据集[36].该数据集由51类共计6 849个人体动作视频序列组成,主要来源电影片段、网络视频等.本文提取视频序列3种异构描述子[37]:梯度直方图(histogram of oriented gradient, HOG)、光流直方图(histogram of optical flow, HOF)和空间时间特征(space-time interest points, STIP),分别构建视频序列的1 000维BoVW表示.

### 3.2 评价指标

为了公正地对聚类结果进行评估,本文使用2种指标[15]评估算法的聚类性能:

1) 聚类精度(clustering accuracy, $ACC$):

$$ACC=\frac{1}{n}\sum_{i=1}^{n}\delta(\boldsymbol{l}_i,map(\boldsymbol{t}_i)),\tag{15}$$

其中,$\boldsymbol{l}_i$ 和 $\boldsymbol{t}_i$ 分别表示数据对象的真实划分和聚类划分,$n$ 是数据集的大小.$\delta(\boldsymbol{x},\boldsymbol{y})$为狄克拉函数,当 $\boldsymbol{x}==\boldsymbol{y}$ 时,$\delta(\boldsymbol{x},\boldsymbol{y})=1$,否则 $\delta(\boldsymbol{x},\boldsymbol{y})=0$.$map(\boldsymbol{t}_i)$是聚类划分 $\boldsymbol{t}_i$ 与真实划分之间的映射函数.

2) 标准化互信息(normalized mutual information, $NMI$):

$$NMI=\frac{I(\boldsymbol{T};\boldsymbol{C})}{\max[H(\boldsymbol{T}),H(\boldsymbol{C})]},\tag{16}$$

其中,$\boldsymbol{T}$ 和 $\boldsymbol{C}$ 分别表示数据对象的聚类划分和真实划分,$I(\boldsymbol{T};\boldsymbol{C})$是 $\boldsymbol{T}$ 和 $\boldsymbol{C}$ 间的互信息,$H(\boldsymbol{T})$和 $H(\boldsymbol{C})$分别表示聚类划分和真实划分的信息熵.$ACC$,$NMI$ 的值越大,聚类结果越好.

### 3.3 对比方法

为验证SPIM算法在跨媒体聚类任务中的有效性,本文将其与7种算法进行对比:

1) $k$-means算法.经典的单模态聚类算法.

2) IB算法[22].对各模态分别使用IB算法进行聚类,我们在所有图表中公布各模态中最好的聚类结果.

3) Concate-IB算法.将各模态的特征直接相连,然后使用IB算法进行聚类.

4) 典型关联分析(canonical correlation analysis, CCA)算法[6].采用CCA算法学习2个模态间的共享信息.对于Reuters和HMDB数据集,本文选择2种表现最优的模态作为CCA算法的输入.

5) CSPA(cluster-based similarity partitioning algorithm)算法[15].该算法是一种典型的聚类集成算法,首先根据基聚类的co-association矩阵生成图,其中顶点是数据对象,边的权重的数据对象间的co-association值,然后使用METIS算法进行聚类.

6) PTGP(probability trajectory graph partitioning)算法[16].该算法是基于稀疏图表示和概率轨迹的聚类集成算法,能够同时根据基聚类的局部和全局信息获取最终的聚类划分.

7) LWGP(locally weighted evidence accumulation)算法[17].该算法在不考虑原始数据分布的情况下,通过局部密度估计方法对基聚类进行加权,进而区分基聚类的可依赖程度.

表1给出算法的时间复杂度对比.其中,CCA算法的 $O(Md^2)$项为计算2个模态的协方差矩阵的时间复杂度,$O(d^3)$项是特征分解的时间复杂度,$d=\max(|\boldsymbol{Y}^i|,|\boldsymbol{Y}^j|)$是任意2个模态的特征维度的最大值.CSPA算法采用METIS算法[38]对基聚类的生成图进行划分,其时间复杂度为 $O(|\boldsymbol{X}|^2)$.LWGP算法首先构建数据对象 $\boldsymbol{X}$ 与基聚类之间的二分图(bipartite graph),之后使用TCut算法[39]图划分.与LWGP算不同,PTGP算法首先对数据对象 $\boldsymbol{X}$ 进行初步聚类生成中间簇 $\boldsymbol{X}^{\text{mcluster}}$,构建中间簇 $\boldsymbol{X}^{\text{mcluster}}$与基聚类的二分图,最后使用TCut算法进行二分图划分,其中 $|\boldsymbol{X}^{\text{mcluster}}|$是中间簇的数量.

**Table 1 The Complexities of Different Baselines**

**表1 不同算法的时间复杂度对比**

| Baselines | Time Complexity |
|---|---|
| CCA | $O(Md^2)+O(d^3)$, $d=\max(\|\boldsymbol{Y}^i\|,\|\boldsymbol{Y}^j\|)$ |
| CSPA | $O(\|\boldsymbol{X}\|^2)$ |
| PTGP | $O(2M(1+k)\|\boldsymbol{X}^{\text{mcluster}}\|+kM^{3/2})$ |
| LWGP | $O(2M(1+k)\|\boldsymbol{X}\|+kM^{3/2})$ |
| Our SPIM | $O(M\|\boldsymbol{X}\|(\|\boldsymbol{Y}^1\|+\|\boldsymbol{Y}^2\|+\cdots+\|\boldsymbol{Y}^k\|+\|\tilde{\boldsymbol{Y}}\|))$ |

### 3.4 实验结果分析

为了验证数据不同模态表示能力的差异,图5给出IB算法在跨媒体数据不同模态上的聚类结果.图5中的实验结果表明:

1) IB算法在Wikipedia,Pascal Sentence和Pascal VOC 07数据集的文本、图像2种模态上的聚类结果具有较大的差异,且IB算法在文本模态上的聚类性能明显优于图像模态.这说明在跨模态的

图像聚类任务中仅依赖图像信息很难获得较好的聚类结果,丰富的语义关系能够带来更加准确的语义相关性.因此,兼顾语义和视觉等模态信息有助于提升跨模态聚类任务的聚类性能.

2) IB 算法在多语种数据集 Reuters 的各语种上和视频数据集 HMDB 的各异构特征空间上得到相近的聚类结果.这说明数据对象往往具有多个侧面的描述,例如相同的新闻可以用多种语言进行报道;包含相同人体动作的视频序列可以在梯度、光流、时空等特征空间上获得异构的描述.不同侧面的特征信息反映数据不同的内在特性,有效地组织和使用特征之间的互补作用以便确保高质量的聚类结果.

表 2 和表 3 展示了各种方法的对比结果.对于 CCA 算法,本文首先按照相关系数排序,并选取前 $d$ 对典型得分作为特征向量进行聚类.我们在表 2 和表 3 中公布的是 CCA 算法聚类结果稳定后的值.为保证 CSPA,PTGP,LWGP 这 3 种集成聚类算法中基聚类间的差异性,我们对数据对象的各模态分别使用 IB 算法运行 20 次,然后使用聚类集成算法对所有模态的聚类成员进行合并,得到最终的聚类划分.根据原作者的建议,本文将 PTGP 算法中节点数量 $k$ 设置为 $\sqrt{|\boldsymbol{X}|}/2$,将 LWEA 算法中控制聚类不确定性(cluster uncertainty)的参数 $\theta$ 设置为 0.4.另外,除 CCA 算法外,其他对比算法及本文提出的 SPIM 算法均受随机初始化的影响,因此,本文将这些算法运行 10 次,公布其平均结果(*mean*)及标准差(*std*).从表 2 和表 3 可知 3 点实验现象.

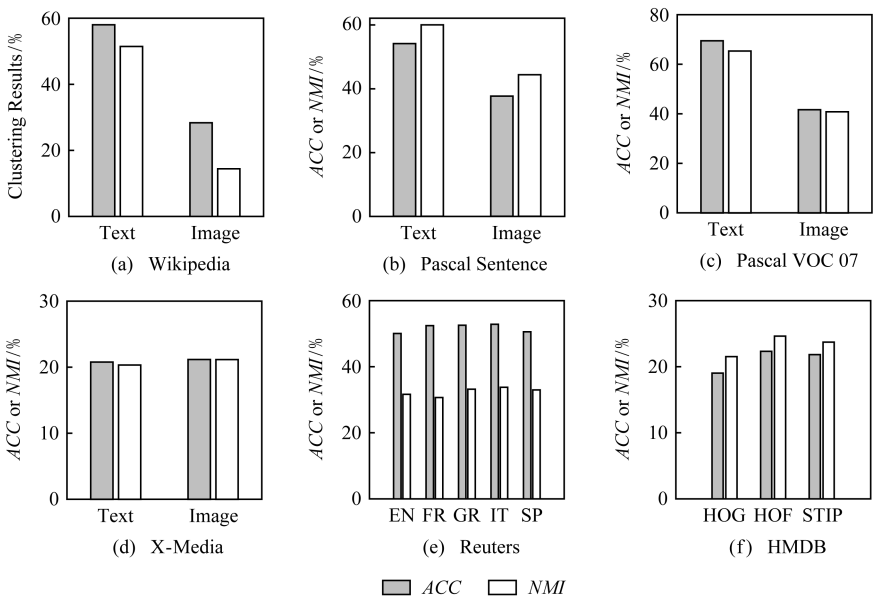



Fig. 5 Clustering results of IB on different media sources

图 5 IB 算法在不同媒体源上的聚类结果

**Table 2 The *ACC* Comparison Results on 6 Cross-Media Datasets**

**表 2 在 6 种跨媒体数据集上的 *ACC* 对比结果** %

| Datasets | $k$-means | IB | Concate-IB | CCA | CSPA | PTGP | LWEA | SPIM |
|---|---|---|---|---|---|---|---|---|
| | *mean*±*std* | *mean*±*std* | *mean*±*std* | | *mean*±*std* | *mean*±*std* | *mean*±*std* | *mean*±*std* |
| Wikipedia | 45.22±2.4 | 58.27±4.8 | 28.58±0.9 | 59.67 | 60.09±5.1 | 62.07±1.0 | 62.86±0.8 | **65.50±2.3** |
| Pascal Sentence | 52.93±3.8 | 54.10±2.9 | 41.75±2.8 | 55.20 | 55.15±2.0 | 57.83±0.8 | 59.67±0.7 | **63.02±1.5** |
| Pascal VOC 07 | 64.99±3.5 | 69.53±3.2 | 62.77±2.3 | 69.09 | 69.84±5.3 | 72.89±0.9 | 73.64±0.8 | **76.33±1.2** |
| X-Media | 20.55±0.6 | 21.16±0.5 | 21.31±0.5 | 23.58 | 18.75±0.7 | 24.31±1.1 | 21.78±0.6 | **25.16±1.9** |
| Reuters | 53.16±1.4 | 53.12±3.1 | 53.43±3.4 | 50.93 | 59.92±0.2 | 56.61±3.6 | 55.28±4.1 | **60.59±3.6** |
| HMDB | 18.46±0.9 | 22.31±2.2 | 23.35 ±1.3 | 25.34 | 20.82±0.5 | 25.75±0.9 | 26.63±1.4 | **29.42±0.8** |
| Average | 42.55 | 46.41 | 38.53 | 47.30 | 47.42 | 49.91 | 49.97 | **53.34** |

Note: The best clustering results in each case is boldfaced.

**Table 3 The *NMI* Comparison Results on 6 Cross-Media Datasets**

**表 3 在 6 种跨媒体数据集上的 *NMI* 对比结果** %

| Datasets | *k*-means | IB | Concate-IB | CCA | CSPA | PTGP | LWEA | SPIM |
|---|---|---|---|---|---|---|---|---|
| | *mean*±*std* | *mean*±*std* | *mean*±*std* | | *mean*±*std* | *mean*±*std* | *mean*±*std* | *mean*±*std* |
| Wikipedia | 45.50±1.7 | 51.68±1.3 | 14.57±0.4 | 52.32 | 48.12±2.3 | 52.98±0.2 | 53.01±0.3 | **54.97±1.6** |
| Pascal Sentence | 59.88±1.9 | 59.99±1.6 | 47.88±1.8 | 60.89 | 61.99±1.4 | 62.12±0.4 | 63.75±0.3 | **66.01±1.8** |
| Pascal VOC 07 | 65.16±1.5 | 65.38±1.2 | 65.29±1.0 | 66.28 | 71.05±2.4 | 70.41±0.5 | 71.92±0.2 | **75.21+1.5** |
| X-Media | 20.12±0.5 | 21.24±0.2 | 21.23±0.2 | 23.28 | 21.45±0.8 | 23.94±0.4 | 20.03±0.3 | **26.54±2.2** |
| Reuters | 33.94±1.2 | 43.29±3.1 | 44.33±3.4 | 46.77 | 48.80±0.1 | 47.35±4.9 | 46.79±4.0 | **51.79±2.8** |
| HMDB | 26.84±1.0 | 33.74±2.2 | 34.15 ±2.0 | 36.14 | 37.83±0.6 | 37.25±0.6 | 35.41±0.8 | **41.13±1.1** |
| Average | 41.91 | 45.89 | 37.89 | 47.61 | 48.21 | 49.01 | 48.49 | **52.61** |

Note: The best clustering results in each case is boldfaced.

1) 将各模态的特征直接相连不能有效地提升聚类性能.例如相比 IB 算法在单一模态上的最优结果,Concate-IB 算法在 Wikipedia,Pascal Sentence,Pascal VOC 07 数据集上的 *ACC* 和 *NMI* 值出现下降.说明简单地连接多模态的特征信息不能稳定地提升算法的聚类质量.

2) 结合多模态信息的算法在聚类任务上的表现优于单模态聚类.从表 2 和表 3 可知,CCA,CSPA,PTGP,LWEA 算法在本文使用的 6 种跨媒体数据上的平均聚类结果均优于 *k*-means,IB 等单模态聚类算法.这种现象验证了多模态之间的互补作用能够有效地提升聚类结果.

3) 相比于其他单模态和跨模态聚类算法,SPIM 算法在本文使用的 6 种跨媒体数据集上的聚类结果均有提升.

### 3.5 参数分析

SPIM 算法使用参数 $\lambda$ 控制共享信息与私有信息间的侧重程度,因此,本节通过实验来观察不同参数值对聚类结果的影响.从图 6 可以看出,当 $\lambda$ 取值较小时,SPIM 算法得到较低的 *ACC* 值.随着 $\lambda$ 值的增大,SPIM 算法聚类效果逐渐变好,此时共享信息与私有信息的互补作用得以体现.随着 $\lambda$ 值进一步增大,各模态的私有信息与共享信息达到平衡,SPIM 算法的聚类结果也在一定程度上趋于稳定.注意,本文公布的 SPIM 算法的聚类结果对应参数取值为 $\lambda=60$,聚类结果如图 6 中星号所示.

### 3.6 收敛性分析

SPIM 算法的目标函数只能收敛到一个局部最优解,因此有必要对其收敛性进行经验性分析.图 7 给出 SPIM 算法每次迭代后式(10)的值.从图 7 可以

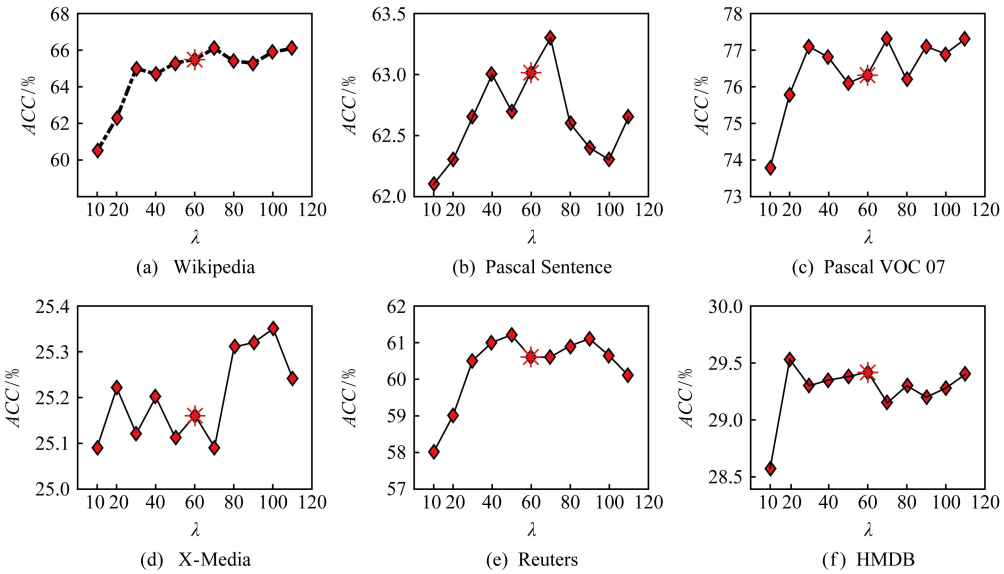



Fig. 6 The impact of $\lambda$ on the performance of SPIM algorithm

图 6 平衡参数 $\lambda$ 对 SPIM 算法聚类结果的影响

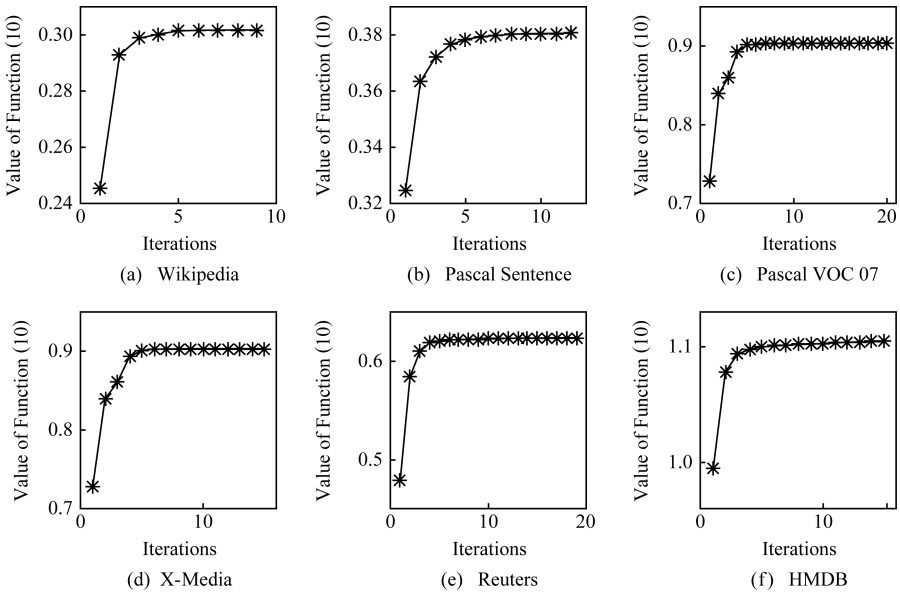



Fig. 7 The iteration number of SPIM on different datasets

图 7 SPIM 算法在不同数据集上达到收敛时的迭代次数

看出,随着算法的运行,式(10)的值先是迅速上升,然后上升的幅度趋于平缓,最终得到目标函数的局部最优值,说明该算法具有较好的收敛性.

### 3.7 模型简化测试

SPIM 算法通过混合单词模型和聚类集成模型对跨模态数据中各模态间的关联进行建模,分别保持各模态底层的统计相似性和各模态的高层聚类划分间的相关性.本节针对混合单词模型和聚类集成模型设计单独的实验,以验证单个模型的有效性.根据式(10)可知,SPIM 算法在仅考虑单一混合单词模型和聚类集成模型时的目标函数可分别简化为

$$\mathcal{L}_{\max}[p(\boldsymbol{t}\mid\boldsymbol{x})]=\sum_{i=1}^{k}I(\boldsymbol{T};\boldsymbol{Y}^i)+\lambda\cdot I(\boldsymbol{T};\tilde{\boldsymbol{Y}}),\quad(17)$$

$$\mathcal{L}_{\max}[p(\boldsymbol{t}\mid\boldsymbol{x})]=\sum_{i=1}^{k}I(\boldsymbol{T};\boldsymbol{Y}^i)+\lambda\cdot\sum_{i=1}^{k}I(\boldsymbol{T};\boldsymbol{C}^i).\quad(18)$$

图 8 和表 4 给出分别考虑单一混合单词模型和聚类集成模型时 SPIM 算法的聚类精度和标准化互信息.从图 8 和表 4 可知:

1) SPIM 算法在仅考虑单一模型时的聚类结果优于 IB 算法在单个模态中最优的聚类结果,验证 SPIM 算法中单一模型的有效性;

2) 在同时使用混合单词模型和聚类集成模型对共享信息建模时,SPIM 算法的聚类结果进一步提升.例如,表 4 中 SPIM 算法在 6 种跨媒体数据集

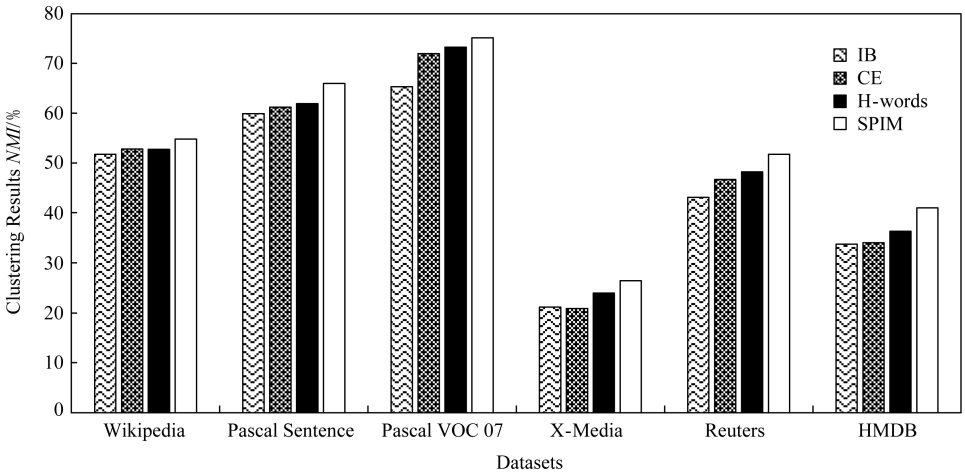



Fig. 8 Model ablation test of SPIM algorithm in terms of *NMI*

图 8 SPIM 算法在 *NMI* 指标上的模型简化测试

上的平均结果均优于 IB 算法、聚类集成模型和混合词库模型.

**Table 4 Model Ablation test of SPIM Algorithm in Terms of *ACC***

**表 4 SPIM 算法在 *ACC* 指标上的模型简化测试 %**

| Datasets | IB | CE | H-words | SPIM |
|---|---|---|---|---|
| | *mean*±*std* | *mean*±*std* | *mean*±*std* | *mean*±*std* |
| Wikipedia | 58.27±4.8 | 62.18±4.3 | 61.99±1.7 | **65.50±2.3** |
| Pascal Sentence | 54.10±2.9 | 56.74±2.8 | 55.41±3.7 | **63.02±1.5** |
| Pascal VOC 07 | 69.53±3.2 | 70.23±2.3 | 72.38±1.6 | **76.33±1.2** |
| X-Media | 21.16±0.5 | 20.58±0.4 | 23.76±0.8 | **25.16±1.9** |
| Reuters | 53.12±3.1 | 57.79±3.4 | 58.09±2.2 | **60.59±3.6** |
| HMDB | 22.31±2.2 | 23.32±2.4 | 26.14±1.3 | **29.42±0.8** |
| Average | 46.41 | 48.47 | 49.63 | **53.34** |

Note: The best clustering results in each case is boldfaced.

## 4 结 论

本文提出共享和私有信息最大化的跨媒体聚类算法 SPIM,该算法能够使用跨模态数据间的共享信息和各模态自身的私有信息进行聚类分析.首先,提出 2 种跨媒体数据的共享信息构建模型,分别保持各模态底层特征的统计相似性和各模态的高层聚类划分间的相关性,其次,提出基于信息论的目标函数,将跨媒体数据的共享和私有信息融合在同一目标函数中.最后,采用顺序"抽取-合并"优化过程,保证 SPIM 的目标函数收敛到局部最优解.在 6 种跨媒体数据上的实验结果表明:本文提出的 SPIM 算法的优越性

## 参 考 文 献


[1] Kumar A, Daume H. A co-training approach for multi-view spectral clustering [C] //Proc of the 28th Int Conf on Machine Learning. New York: ACM, 2011: 393-400

[2] Cai Xiao, Nie Feiping, Huang Heng, et al. Heterogeneous image feature integration via multi-modal spectral clustering [C] //Proc of the 24th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2011: 1977-1984

[3] Zhang Hong, Wu Fei, Zhuang Yueting. Cross-media correlation reasoning and retrieval [J]. Journal of Computer Research and Development, 2008, 45(5): 869-876 (in Chinese)
(张鸿, 吴飞, 庄越挺. 跨媒体相关性推理与检索研究[J]. 计算机研究与发展, 2008, 45(5): 869-876)

[4] Zhang Lei, Zhao Yao, Zhu Zhenfeng. Advances in semantically shared subspace learning for cross-media data [J]. Chinese Journal of Computers, 2017, 40(6): 1394-1421 (in Chinese)
(张磊, 赵耀, 朱振峰. 跨媒体语义共享子空间学习研究进展[J]. 计算机学报, 2017, 40(6): 1394-1421)

[5] Chaudhuri K, Kakade M, Livescu K, et al. Multi-view clustering via canonical correlation analysis [C] //Proc of the 26th Int Conf on Machine Learning. New York: ACM, 2009: 129-136

[6] Hardoon D R, Szedmak S, Shawe-Taylor J. Canonical correlation analysis: An overview with application to learning methods [J]. Neural Computation, 2004, 16(12): 2639-2664

[7] Sigal L, Memisevic R, Fleet D J. Shared kernel information embedding for discriminative inference [C] //Proc of the 22nd IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2009: 2852-2859

[8] Carl H, Phipip P, Lawrence N D. Gaussian process latent variable models for human pose estimation [C] //Proc of the 4th Machine Learning for Multimodal Interaction. Berlin: Springer, 2007: 132-143

[9] Barnard K, Forsyth D. Learning the semantics of words and pictures [C] //Proc of the 8th Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2001: 408-415

[10] Hofmann T. Learning and representing topic—A hierarchical mixture model for word occurrence in document databases [C] //Proc of the Workshop on Learning from Text and the Web. Pittsburgh, PA: CMU, 1998

[11] Barnard K, Duygulu P, Forsyth D, et al. Matching words and pictures [J]. Journal of Machine Learning Research, 2003, 3(2): 1107-1135

[12] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3(1): 993-1022

[13] Gao Jing, Han Jiawei, Liu Jailu, et al. Multi-view clustering via joint nonnegative matrix factorization [C] //Proc of the 13th SIAM Int Conf on Data Mining. Philadelphia, PA: SIAM, 2013: 252-260

[14] Cai Xiao, Nie Feiping, Huang Heng. Multi-view $k$-means clustering on big data [C] //Proc of the 23rd Int Joint Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2013: 2598-2604

[15] Strehl A, Ghosh A. Cluster ensembles-a knowledge reuse framework for combining multiple partitions [J]. The Journal of Machine Learning Research, 2003 (3): 583-617

[16] Huang Dong, Lai Jianhuang, Wang Changdong. Robust ensemble clustering using probability trajectories [J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(5): 1312-1326

[17] Huang Dong, Wang Changdong, Lai Jianhuang. Locally weighted ensemble clustering [J]. IEEE Transactions on Cybernetics, 2017, 48(5): 1460-1473

[18] Lou Zhengzheng, Ye Yangdong, Yan Xiaoqiang. The multi-feature information bottleneck with application to unsupervised image categorization [C] //Proc of the 23rd Int Joint Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2013: 1508-1515

[19] Yan Xiaoqiang, Ye Yangdong, Lou Zhengzheng. Unsupervised video categorization based on multivariate information bottleneck method [J]. Knowledge-Based Systems, 2015, 84(C): 34-45

[20] Luo Peng, Peng Jinye, Guan Ziyu, et al. Multi-view semantic learning for data representation [J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(11): 3016-3028

[21] Zhao Qi, Li Zongmin, Cross-modal social image clustering [J]. Chinese Journal of Computers, 2018, 41(1): 98-111 (in Chinese)
(赵其鲁, 李宗民. 跨模态社交图像聚类[J]. 计算机学报, 2018, 41(1): 98-111)

[22] Peng Yuxin, Qi Jinwei, Huang Xin, et al. CCL: Cross-modal correlation learning with multi-grained fusion by hierarchical network [J]. IEEE Transactions on Multimedia, 2018, 20(2): 405-420

[23] Peng Yuxin, Huang Xin, Qi Jinwei. Cross-media shared representation by hierarchical learning with multiple deep networks [C] //Proc of the 25th Int Joint Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2016: 3846-3853

[24] Tishby N, Pereira F, Bialek W. The information bottleneck method [C] //Proc of the 37th Allerton Conf on Communication, Control and Computing. Piscataway, NJ: IEEE, 1999: 368-377

[25] Slonim N. The information bottleneck: Theory and applications [D]. Hebrew, IL: Hebrew University, 2002

[26] Yan Xiaoqiang, Hu Shizhe, Ye Yangdong. Multi-task clustering of human actions by sharing information [C] // Proc of the 29th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 6401-6409

[27] Svetlana L, Cordelia S, Jean P. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories [C] //Proc of the 19th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2006: 2169-2178

[28] Rasiwasia N, Pereira J C, Coviello E, et al. A new approach to cross-modal multimedia retrieval [C] //Proc of the 18th ACM Int Conf on Multimedia. New York: ACM, 2010: 251-260

[29] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110

[30] Rashtchian C, Young P, Hodosh M, et al. Collecting image annotations using Amazon's Mechanical Turk [C] //Proc of the Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk. Stroudsburg, PA: ACL, 2010: 139-147

[31] Everingham M, Gool L V, Williams C K I, et al. The pascal visual object classes (VOC) challenge [J]. International Journal of Computer Vision, 2010, 88(2): 303-338

[32] Hwang S J, Grauman K. Accounting for the relative importance of objects in image retrieval [C] //Proc of the British Machine Vision Conf. Berlin: Springer, 2010: 1-12

[33] Peng Yuxin, Zhai Xiaohua, Zhao Yunchao, et al. Semi-supervised cross-media feature learning with unified patch graph regularization [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2016, 26(3): 583-596

[34] Zhai Xiaohua, Peng Yuxin, Xiao Jianguo. Learning cross-media joint representation with sparse and semisupervised regularization [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2014, 24(6): 965-978

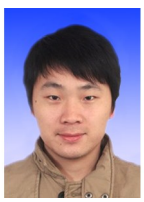
[35] Peng Yuxin, Huang Xin, Zhao Yunchao. An overview of cross-media retrieval: Concepts, methodologies, benchmarks and challenges [J]. IEEE Transactions on Circuits & Systems for Video Technology, 2017, 28(9): 2372-2385

[36] Kuehne H, Jhuang H, Garrote E, et al. HMDB: A large video database for human motion recognition [C] //Proc of the 13th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2011: 2556-2563

[37] Laptev I, Marszalek M, Schmid C, et al. Learning realistic human actions from movies [C] //Proc of the 21st IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2008: 1-8

[38] Karypis G, Kumar V. A fast and high quality multilevel scheme for partitioning irregular graphs [J]. SIAM Journal on Scientific Computing, 1998, 20(1): 359-392

[39] Chang Shifu, Wu Xiaoming, Li Zhenguo. Segmentation using superpixels: A bipartite graph partitioning approach [C] // Proc of the 25th IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2012: 789-796



**Yan Xiaoqiang**, born in 1989. PhD. Member of CCF. His main research interests include machine learning, computer vision, pattern recognition and data mining.

**Ye Yangdong**, born in 1962. PhD, professor, PhD supervisor at Zhengzhou University. Senior member of CCF. His main research interests includes machine learning, pattern recognition, knowledge engineering and intelligent system.