

# 一种双重加权的多视角聚类方法

胡世哲 娄铮铮 王若彬 闫小强 叶阳东

(郑州大学信息工程学院 郑州 450001)

**摘 要** 在大数据时代下,如何利用多源异构数据中的互补信息来识别数据中的簇模式,是多视角聚类研究中的热点问题.然而,大多数现有的多视角聚类方法只在基于内容的特征表示上(如基于  $k$ -means 的加权多视角聚类方法)或基于上下文的相似度表示上(如基于谱聚类的加权多视角聚类方法)学习和施加权重,未能同时考虑这两种表示以充分地表达出视角内部固有的信息.另外,大多数加权多视角聚类方法需要引入额外的参数以控制视角权重的分布,但是,在没有任何先验知识的前提下,很难人为选择出恰当的权重控制参数.针对上述问题,提出了一种双重加权的多视角聚类算法 DWMVC.它通过互信息自动学习视角权重,并将这些权重施加到基于内容和基于上下文的多视角数据表示上,以便于充分利用两种数据表示下的视角互补信息.构造了一个基于信息瓶颈的目标函数,在压缩这两种数据表示的同时最大限度地保留着相关特征和相似度信息.最后,设计了一种顺序的优化方法,以保证模型收敛到局部最优解.在多种多视角数据集上的实验结果表明,该方法优于目前先进的单视角和多视角聚类方法.

**关键词** 多视角聚类;权重学习;信息瓶颈;互信息;词袋模型  
**中图法分类号** TP301 **DOI号** 10.11897/SP.J.1016.2020.01708

## Dual-Weighted Multi-View Clustering

HU Shi-Zhe LOU Zheng-Zheng WANG Ruo-Bin YAN Xiao-Qiang YE Yang-Dong

(School of Information Engineering, Zhengzhou University, Zhengzhou 450001)

**Abstract** In the Big Data era, how to utilize the complementary information among multi-source heterogeneous data, such as multi-lingual text, multi-feature image and multi-view human action video, to discover the underlying cluster patterns is one of the hot topics in multi-view clustering, the aim of which is to learn and leverage the complementary information among different views to discover a consistent clustering partition. However, there are two challenging issues when we resort to current multi-view clustering methods. (1) Inadequate utilization of intrinsic information resided in each view. Most of the existing multi-view clustering methods learn and apply weights on only either the content-based bag-of-words feature representations (such as  $k$ -means based weighted multi-view clustering methods) or the context-based similarity representations (such as spectral based weighted multi-view clustering methods), which fail to consider both of them to fully represent the intra-view intrinsic information. Thus, it may indirectly do harm to the final multi-view clustering performance. (2) Needing to introduce more parameters for controlling weight distribution. Most of these weighted multi-view clustering methods need to introduce extra parameters to control the weight distribution. But to select an optimal weight parameter by hand is quite difficult without giving any prior knowledge. Hence, directly applying these weighted multi-view clustering methods to complex practical scenarios on pattern recognition may not obtain

收稿日期: 2019-08-31; 在线发布日期: 2020-02-07. 本课题得到国家自然科学基金(61772475)、国家重点研发计划“先进轨道交通”重点专项(2018YFB1201403)和国家自然科学基金青年基金(61906172)资助. 胡世哲, 博士研究生, 中国计算机学会(CCF)会员(94305G), 主要研究领域为模式识别、信息理论. E-mail: ieshizhehu@gmail.com. 娄铮铮, 博士, 讲师, 中国计算机学会(CCF)会员, 主要研究领域为机器学习、模式识别. 王若彬, 硕士研究生, 主要研究领域为机器学习、数据挖掘. 闫小强, 博士, 中国计算机学会(CCF)会员, 主要研究领域为机器学习、计算机视觉. 叶阳东(通信作者), 博士, 教授, 中国计算机学会(CCF)高级会员, 主要研究领域为机器学习、知识工程、智能系统. Email: yeyd@zzu.edu.cn.

satisfactory clustering result. To deal with the above two problems, a novel dual-weighted multi-view clustering (DWMVC) method is proposed in this paper. Specifically, the view weights are learned automatically by the concept of mutual information and then are simultaneously imposed on the content-based and context-based multi-view data representations in order to enjoy the best of both worlds, especially the complementary information resided in the two kinds of multi-view data representations. The motivation behind this idea is that, the content-based feature representations and context-based similarity representations depict the characteristics of multi-view data from different point of view. Specifically, feature representations describe the properties of individual data samples, which reflect the intrinsic information of each sample from each view. While similarity representations describe the affinity or closeness among pairwise data samples in each view, which reflect the underlying correlations among samples. Therefore, we apply the automatically learned weights on both of the representations to fully utilize the complementary information among views so that the overall data clustering performance can be enhanced. The proposed method is formulated into an information bottleneck based objective function, which can simultaneously compress the two kinds of data representations (including content-based feature and context-based similarity representations) while preserving the relevant information maximally. Finally, a new sequential “draw and merge” method is designed to solve the optimization problem, which ensures that it can converge to a local maximal objective value since discovering the global optimal value of the objective function is a NP-hard problem. Experimental results on various kinds of multi-view datasets, including multi-lingual text dataset, multi-feature image and human action video dataset, and multi-modal action video datasets, demonstrate the superiority and effectiveness of the proposed DWMVC method in comparison to the state-of-the-art single-view and multi-view clustering methods.

**Keywords** multi-view clustering; weight learning; information bottleneck; mutual information; bag of words model

## 1 引 言

在大数据时代，海量多源异构数据不断涌现，呈现着大规模、多样化、复杂性和相关性等特点。其中，多视角数据是最为常见的一种，主要描述同一事物在不同侧面下的数据信息。例如，如图 1 所示 (1) 多语言文档：相同“谢谢”的表达可以用中文、英文和韩文等语言来表示；(2) 多特征图像：

同一个图像数据集下使用不同的特征加以刻画，如形状、纹理和颜色等；(3) 多模态人体行为视频：使用 Kinect 摄像机捕捉人体“鼓掌”行为，则可以获得 RGB、深度和骨骼等模态的表示。以上例子可以看出，多视角数据中包含着丰富的视角互补信息或者类别一致性信息。通过合理挖掘并使用这些视角相关信息，聚类算法可能会取得满意的聚类效果。

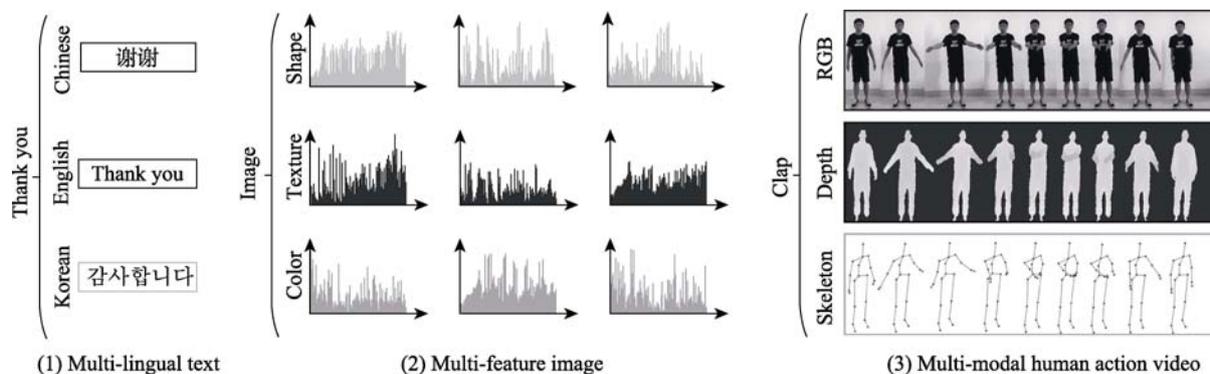


图 1 不同类型的多视角数据

然而, 由于缺少视角关系发现机制, 传统的单视角聚类算法很难被用于解决多视角聚类问题. 直觉上, 我们可以把各个视角顺次拼接在一起, 然后使用传统的单视角算法进行聚类. 然而, 不同视角的特性不一, 简单的融合不能充分地挖掘视角间的互补信息, 往往难以得到理想的结果. 更糟糕的是, 在这种方法下的聚类实验结果有时甚至会低于传统单视角下对应的结果(见实验部分的表3和表4). 为了解决这些问题, 基于不同策略的多视角聚类算法<sup>[1]</sup>相继被提出, 旨在利用多视角间的相关信息得到更为准确有效的聚类结果. 常见的多视角聚类算法包含基于共享子空间<sup>[2]</sup>、基于聚类融合<sup>[3]</sup>以及基于权重<sup>[4]</sup>的多视角聚类算法. 其中, 加权策略的多视角方法以其差别视角信息利用的优势总能取得较好的聚类结果, 引起了大量研究者的注意.

迄今为止, 关于加权多视角聚类的研究相对较多. 较早的研究[4]提出一种基于核的权重多视角聚类算法, 其首先将视角表示为核矩阵, 然后加权每个视角中的信息来进行聚类. 文献[5-6]分别引入拉普拉斯秩约束图和自适应邻接点赋值, 进而提出一种自动加权的多视角聚类算法, 其优点在于最终学习到的最优图可以直接用于求得聚类结果, 无需使用  $k$ -means 等后处理方法. 为了能够处理大规模多视角数据, 文献[7-8]分别提出一种基于  $k$ -means 和基于谱聚类的加权多视角聚类算法, 均得到了可观的运行效率. 为了能够处理高维或含有噪声的多视角数据, 文献[9-12]分别基于不同策略提出包含特征学习的加权多视角聚类算法, 旨在学习到每个视角权重的同时, 挑选出各个视角中最有代表性的特征以提高聚类性能.

尽管以上的加权多视角聚类算法取得了不错的聚类结果, 但是仍然存在着以下两个问题:

(1) 视角内部固有信息表达不足. 当前多视角聚类算法中使用的视角表示主要有两种方式, 一种是基于内容的特征表示(如多视角  $k$ -means 聚类算法中的“词袋模型”表示), 另一种是基于上下文的相似度表示(如多视角谱聚类算法). 然而, 任意一种表示都不能充分表达出视角内部蕴含的信息, 从而间接地影响最终的多视角聚类结果.

(2) 需要更多的额外参数. 当前大部分加权的多视角聚类算法在求权重的过程中, 需要引入更多额外的参数以控制视角参数的分布. 然而, 在没有任何先验知识的情况下, 很难人为选择出合适的视角权重控制参数. 因此, 将该类多视角聚类算法直接应用于复杂的实际场景中难以获得可观的效果.

针对以上问题, 本文提出一种全新的双重加权多视角聚类算法 (Dual-Weighted Multi-View Clustering, DWMVC), 如图2所示, 该算法通过综合考虑自动加权的多视角特征信息和多视角相似度表示信息进行聚类分析, 以得到更加准确的聚类结果. 其背后的动机在于, 基于内容的特征表示和基于上下文的相似度表示分别从不同角度刻画了单个视角中数据的特性信息. 特征表示刻画的是每个孤立的数据样本, 反映了数据自身蕴含的固有信息. 而相似度表示刻画的是所有数据样本之间的亲密度, 反映了数据之间的关联信息. 因此, 我们对每个视角的两种数据表示双重加权, 以充分利用两方面的视角互补信息, 从而提高算法的聚类性能. 具体来讲, 首先使用互信息度量两种多视角表示以求得不同表示下的视角权重, 分别用以刻画各种视角特征信息和上下文信息的重要性. 其次, 构建基于信息瓶颈理论的目标函数, 平衡两种加权的多视角表示中包含的互补信息, 从而有利于挖掘出令人满意的数据模式结构. 最终, 提出一种顺序的优化方法对 DWMVC 算法进行问题求解, 以保证得到局部最优解. 在五种不同的多视角数据集上的实验结果表明了所提 DWMVC 算法优于当前先进的单视角和多视角聚类算法.

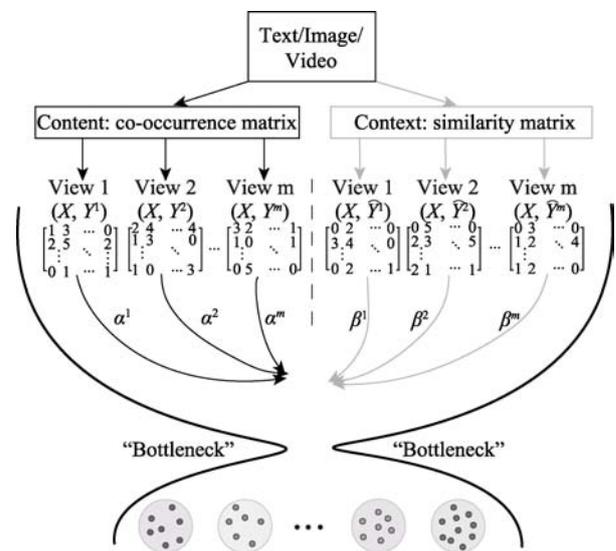


图2 DWMVC 算法框架图

本文的主要贡献包括3个方面:

(1) 提出一种双重加权的多视角聚类算法 DWMVC, 该算法通过自动加权基于内容的多视角特征表示和基于上下文的相似度表示, 以更全面地挖掘和利用多视角间的互补信息;

(2) 互信息度量方法被应用到多视角数据表示中来自动学习不同视角的权重, 该度量方法能够从根

本上衡量数据  $X$  与每个视角表示 ( $Y^i$  或  $\bar{X}^i$ ) 之间的内在关系, 从而反映每个视角的重要性 (即权重);

(3) 提出一种顺序的优化策略对目标函数进行问题求解, 确保得到局部最优解.

## 2 相关工作

### 2.1 互信息

互信息<sup>[13, 14]</sup>一般用来描述两个随机变量之间的相关性. 对于两个离散随机变量  $X$  和  $Y$ , 如果  $(x, y) \sim p(x, y)$ , 则变量  $X$  与  $Y$  之间的互信息  $I(X; Y)$  可以被定义为

$$I(X; Y) = \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

其中  $p(x, y)$  是变量  $X$  和  $Y$  的联合概率分布, 而  $p(x) = \sum_y p(x, y)$  和  $p(y) = \sum_x p(x, y)$  分别是对应的边缘概率分布. 互信息反映了随机变量  $X$  与  $Y$  之间的共有信息量, 即两者之间相互包含的程度. 当变量  $X$  和  $Y$  相互独立时, 联合分布  $p(x, y)$  等于各自边缘分布的乘积, 也即  $p(x, y) = p(x)p(y)$ , 此时两者之间的互信息  $I(X; Y) = 0$ .

### 2.2 信息瓶颈

IB 方法<sup>[15]</sup>是由 Tishby 等人于提出的一种源自于率失真理论的数据分析方法. 该方法在进行数据分析的过程中, 试图通过一个压缩变量  $T$  寻求源变量  $X$  的紧凑表示形式, 并使  $T$  尽可能多地保存关于相关变量  $Y$  的信息. 从互信息的角度讲, 其旨在将源变量  $X$  压缩到一个“瓶颈” $T$  中, 使得源变量  $X$  和压缩变量  $T$  之间的互信息  $I(X; T)$  最小化, 同时保证相关信息的保存量  $I(T; Y)$  最大化. 基于此特性, IB 理论近年来被用于解释深度神经网络的有效性, 同时从原理方面给出了深入的分析<sup>[16, 17]</sup>.

IB 方法可形式化描述为

$$R(D) \equiv \min_{\{p(t|x): I(T; Y) \geq D\}} I(X; T) \quad (2)$$

其中  $p(t|x)$  表示变量  $X$  到  $T$  的编码方案, 揭示了数据对象集合  $X$  所包含的模式结构. 在满足压缩变量  $T$  所保存相关变量  $Y$  中的信息量  $I(T; Y) \geq D$  的条件下, 寻求所有可能编码方案中能够使压缩信息  $I(X; T)$  最小的一个编码方案  $p(t|x)$ . 文献[15]采用拉格朗日

乘法将式(2)表示为如下的目标函数:

$$\Phi_{\min}[p(t|x)] = I(X; T) - \gamma I(Y; T) \quad (3)$$

其中  $\gamma$  为拉格朗日因子, 一般来说  $\gamma \in (0, \infty)$ , 用于平衡对信息的压缩和保留. 其模型化表示如图 3 所示, 1) 数据压缩: 虚线箭头表示源变量  $X$  被压缩为变量  $T$ , 实线表示源变量  $X$  和相关变量  $Y$  之间已知的联合概率分布. 2) 信息保留: 压缩变量  $T$  保留相关变量  $Y$  中的信息.

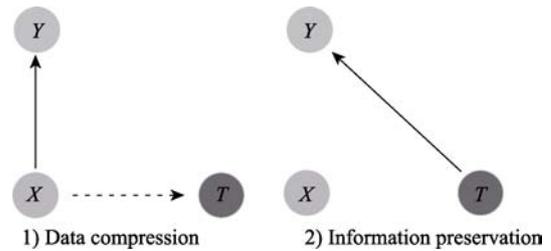


图 3 IB 方法模型

若两边同时除以  $-\gamma$ , 可以得到目标函数的最大化表示:

$$\Phi_{\max}[p(t|x)] = I(Y; T) - \gamma^{-1} I(X; T) \quad (4)$$

从公式(4)可看出, 当  $\gamma \rightarrow 0$  时, IB 方法更加关注源变量  $X$  的压缩; 当  $\gamma \rightarrow \infty$  时, IB 方法更加注重相关信息  $Y$  的保存.

以 IB 方法在聚类中的使用为例, 其将聚类过程视为从源变量  $X$  到变量  $T$  的压缩, 同时, 压缩变量  $T$  能够最大化保留相关变量 (如特征变量)  $Y$  的信息. 以图 4 为例, 我们使用 IB 方法将输入的 5 个数据样本  $X=(x_1, x_2, \dots, x_5)$  聚类为 2 个簇  $T=(t_1, t_2)$ . 在此, 我们采用硬聚类的方式, 即最终的聚类标签  $p(t|x) = 0$  或者  $p(t|x) = 1$ . (1) 使用“词袋模型”对输入的数据进行预处理 (如图 5 所示), 以得到数据的共现矩阵  $(X, Y)$ . (2) 将其标准化为联合概率分布的特征表示  $p(X, Y)$ . (3) IB 方法将前两个数据压缩至第一个簇  $t_1$ , 其聚类标签为  $p(t_1|x_1) = p(t_1|x_2) = 1$ , 对应的簇中心为  $p(y|t_1) = [p(y|x_1) + p(y|x_2)] / 2$ , 而将后三个元素压缩至第二个簇  $t_2$ , 其聚类标签为  $p(t_2|x_3) = p(t_2|x_4) = p(t_2|x_5) = 1$ , 对应的簇中心为  $p(y|t_2) = [p(y|x_3) + \dots + p(y|x_5)] / 3$ . 最终, IB 方法将数据维度为 5 的源变量  $X$  压缩到数据维度为 2 的压缩变量  $T$ , 同时变量  $T$  也尽可能保留维度为 4 的特征变量  $Y$ .

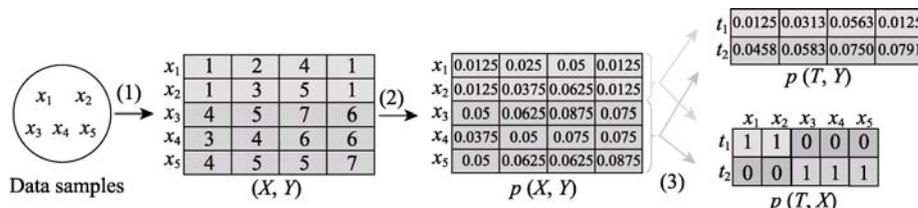


图 4 IB 方法在聚类场景下的使用

IB 方法的优化过程<sup>[15]</sup>可由如下自包含的公式所得:

$$\begin{cases} p = (t|x) = \frac{p(t)}{Z(x, \gamma)} e^{-\gamma D_{KL}[p(y|x)||p(y|t)]} \\ p = (y|t) = \frac{1}{p(t)} \sum_x p(x, y) p(t|x) \\ p(t) = \sum_x p(t|x) p(x) \end{cases} \quad (5)$$

其中,  $D_{KL}$  表示 KL (Kullback Leibler) 距离,  $Z(x, \gamma)$  为归一化函数.

### 3 双重加权的多视角聚类算法

在本节中, 我们首先给出多视角聚类算法 DWMVC 的问题定义, 然后展示视角权重自动学习策略, 最后介绍 DWMVC 算法的目标函数以及相应的优化方法.

#### 3.1 问题定义

定义 1. 给定多视角数据, 令源变量  $X=(x_1, x_2, \dots, x_n)$  代表包含  $n$  个数据的集合, 特征变量  $\{Y^i\}_{i=1}^m$  代表  $m$  个视角数据的基于内容的特征表示, 相似度变量  $\{\hat{X}^i\}_{i=1}^m$  表示不同视角数据的基于上下文的相似度表示. DWMVC 算法旨在最大化压缩源变量  $X$  至压缩变量  $T$  以得到最佳的压缩表示  $p(t|x)$ , 同时压缩变量  $T$  最大化地保留与特征变量  $\{Y^i\}_{i=1}^m$  和相似度变量  $\{\hat{X}^i\}_{i=1}^m$  之间的信息. 在此, 我们用  $c$  表示多视角数据的类个数. 而此时压缩变量  $T$  的数据维度即为类个数  $c$ .

基于内容的特征表示: 以多语言文本的“英语”视角为例, 简要介绍使用“词袋模型”构建的基于内容的数据特征表示  $\{p(X, Y^i)\}_{i=1}^m$  的构建过程, 如图

5 所示. (1) 从多语言文本中获得“英语”视角数据; (2) 忽略文本中语序、语法和句法, 依次从每个文本中提取所有单词; (3) 基于各个文档中提取的单词, 去掉重复的单词后构建为一个词典, 即 bag of words; (4) 以词典中的每个单词为索引, 计算其在每个文档中出现的次数, 即词频表示, 然后使用直方图予以展现; (5) 将各个文档的直方图表示转换为向量表示, 并堆叠在一起形成共现矩阵; (6) 共现矩阵  $(X, Y^i)$  最终被标准化为联合概率分布表示 (也即是数据特征表示)  $p(x, y^i)$ , 其计算公式为

$$p(x, y^i) = \frac{N_{y^i \in x}}{|Q|} \quad (6)$$

其中,  $N_{y^i \in x}$  是单词  $y^i$  在文档  $x$  中出现的次数,  $|Q|$  为单词在所有文档中出现的总次数, 即共现矩阵  $(X, Y^i)$  所有元素之和.

基于上下文的相似度表示: 基于内容的多视角表示仅考虑视角内部各个数据样本的特征表示, 并未顾及样本的上下文信息, 可能不会得到令人满意的聚类结果. 因此, 本文引进多视角数据  $X=(x_1, x_2, \dots, x_n)$  的基于上下文的相似度表示. 其基本思想是, 距离较远的数据点之间相似度较小, 而距离较近的数据点之间相似度较大. 一般地, 我们使用数据点之间的高斯距离度量来获得相似度矩阵, 即

$$S_{ij} = \exp\left(-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}\right) \quad (7)$$

其中参数  $\sigma$  用来控制数据点的领域宽度. 为了和多视角数据  $X$  区分开来, 本文使用相似度变量  $\{\hat{X}^i\}_{i=1}^m$  来表示所有视角的基于上下文的相似度表示.

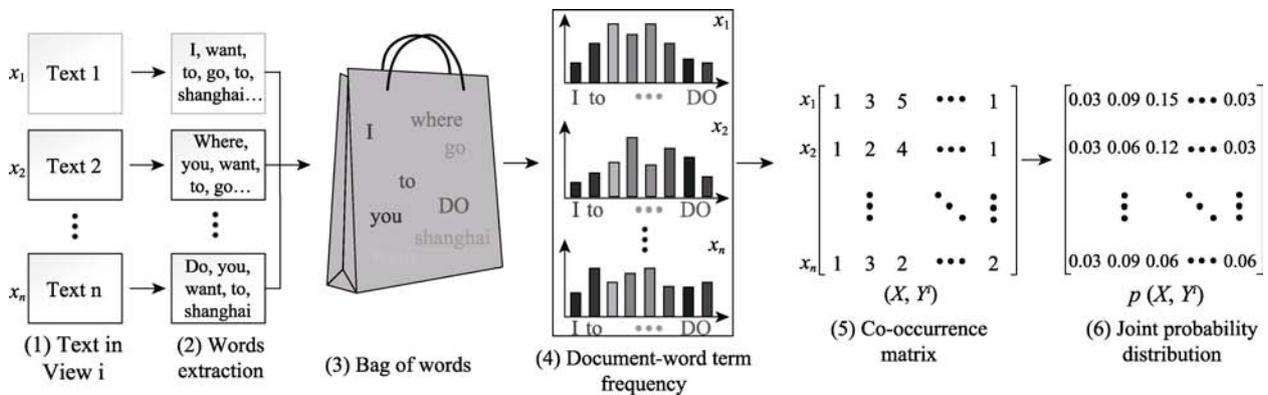


图 5 “词袋模型”特征表示的构建

#### 3.2 双权重学习策略

对于多视角数据而言, 每个视角所提供信息的重要性大为迥异. 以多特征图像数据为例, 不同的

特征对数据的刻画角度不同, 因此不可避免地存在优点和缺点. 颜色特征旨在描述图像中事物的表面颜色分布, 即各种颜色在图像中的比重, 较适用于

描述无需考虑事物空间分布的图像，因此很难刻画图像中特定的物体或者对象。纹理特征较擅长描述疏密度差异较大的图像，很难用于刻画疏密度相差不大的图像。从以上例子可以看出，尽管不同视角的贡献度不同，但是视角之间存在着天然的互补性，因此，对不同视角求权重并将其加权整合可看作是一种非常有效的多视角聚类策略。然而，在没有任何先验知识的情况下，人为地设定视角权重具有很大的盲目性，不能准确地反映各个视角的重要性。

在本文中，我们相继构建两种多视角数据的表示模型，即基于内容的特征表示和基于上下文的相似度表示，以充分利用视角内部的固有信息，进而提出一种全新的双权重自动学习策略以全面挖掘视角之间的互补信息。我们之前提到过，互信息是信息论中一种有效的信息度量方式，旨在刻画两个随机变量之间的相关性。受此启发，本节中我们使用互信息理论来学习每个视角的权重，其可行性和合理性在于：在多视角数据的表示（特征表示或者相似度表示）中，数据  $X$  与某个视角的数据表示  $Y^i$  或  $\hat{X}^i$  的互信息越大，则表明该视角的贡献度越大，重要性越高，相应的视角权重也越大。另外，由于使用互信息度量无需引入额外的参数或者假设，因此，该度量方式非常适合于多视角聚类复杂模型中。下面我们定义的形式给出视角权重的详细求解策略。

**定义 2.** 给定多视角数据  $X=(x_1, x_2, \dots, x_n)$  及其数据表示  $\{(X, U)\}_{i=1}^m$ （如特征表示或者相似度表示），其中， $m$  表示数据的视角数量。每个视角的权重求解可被定义为

$$w^i = \frac{I(X; U^i)}{\sum_{i=1}^m I(X; U^i)} \quad (8)$$

根据以上定义，我们可以得到多视角数据的基于内容的特征表示  $\{(X, Y^i)\}_{i=1}^m$  中各个视角的权重为

$$a^i = \frac{I(X; Y^i)}{\sum_{i=1}^m I(X; Y^i)} \quad (9)$$

类似地，我们可以得到多视角数据的基于上下文的相似度表示  $\{(X, \hat{X}^i)\}_{i=1}^m$  中各个视角的权重为

$$\beta^i = \frac{I(X; \hat{X}^i)}{\sum_{i=1}^m I(X; \hat{X}^i)} \quad (10)$$

双权重自动学习不仅可以利用各个视角中数据样本内部的信息，而且可以整合视角中数据样本的上下文（即样本之间）信息，有利于挖掘并利用多视角数据中的综合互补信息。

### 3.3 目标函数

传统的 IB 方法仅能处理单视角聚类问题，其在

压缩数据的同时最大化保留数据中的相关信息以发现令人满意的聚类结果。为了使 IB 方法能够更好地利用视角内部的固有信息和视角之间的互补信息，本文提出一种基于 IB 方法的双重加权的多视角聚类算法 DWMVC，如图 6 所示，1) 数据压缩：表示变量之间的压缩关系，其中虚线箭头表示源变量  $X$  被压缩为压缩变量  $T$ ，实线表示源变量  $X$  与特征变量  $\{Y^i\}_{i=1}^m$  和相似度变量  $\{\hat{X}^i\}_{i=1}^m$  之间存在已知的联合概率分布。2) 信息保留：实线表示压缩变量  $T$  最大化保留特征变量和相关变量中的信息。其目标函数可被形式化表示为

$$\Phi[p(t|x)] = \sum_i^m a^i [I(T; Y^i) - \gamma^{-1} I(T; X)] + \lambda \sum_i^m \beta^i [I(T; \hat{X}^i) - \gamma^{-1} I(T; X)] \quad (11)$$

其中，第一项为在多视角数据特征表示上的互补信息加权融合， $a^i$  表示在特征表示下的第  $i$  个视角的权重；第二项为在多视角数据相似度表示上的互补信息加权融合， $\beta^i$  表示在相似度表示下的第  $i$  个视角的权重。平衡参数  $\lambda$  调节两种多视角数据表示的侧重程度。

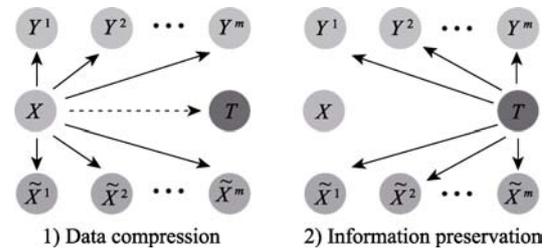


图 6 DWMVC 模型

### 3.4 优化算法

为了优化 DWMVC 算法的目标函数，提出一种行之有效的顺序优化方法。首先，随机划分源变量  $X=(x_1, x_2, \dots, x_n)$  至  $c$  个簇。然后，我们顺序地执行以下两个步骤：(1) 提取：从当前簇  $t^{old}$  中提取出数据样本  $x$ ，使其成为独立簇  $\{x\}$ ；(2) 融合：为保证总的类别（即簇）数量不变，需要计算独立簇  $\{x\}$  与其他各个簇的融合代价，并将其最终融合至代价最小的簇  $t^{new}$  中，其中  $t^{new} = \arg \min \Delta \Phi$ ， $\Delta \Phi$  即为融合代价，表示融合过程中目标函数  $\Phi$  的信息损失。

在以上顺序优化的过程中，我们分别用  $\Phi^1$ ， $\Phi^2$ ， $\Phi^3$  来表示提取之前、提取后且融合之前、融合之后的目标函数(11)的值。因此，融合代价  $\Delta \Phi$  为

$$\Delta \Phi = \Phi^2 - \Phi^3 \quad (12)$$

其具体的计算过程可表示为

$$\Delta\Phi = \Phi^2 - \Phi^3 = \sum_i^m a^i \{ [I(T^2; Y^i)] - I(T^3; Y^i) \} - \gamma^{-1} [I(T^2; X) - I(T^3; X)] + \lambda \sum_i^m \beta^i \{ [I(T^2; \hat{X}^i) - I(T^3; \hat{X}^i)] - \gamma^{-1} [I(T^2; X) - I(T^3; X)] \} = \sum_i^m a^i (\Delta I_i^{f1} - \gamma^{-1} \Delta I_i^{f2}) + \lambda \sum_i^m \beta^i (\Delta I_i^{g1} - \gamma^{-1} \Delta I_i^{g2})$$

首先, 我们计算融合代价  $\Delta I_i^{f1}$ , 其可以展开为

$$\Delta I_i^{f1} = I(T^2; Y^i) - I(T^3; Y^i) = p(x) \sum_{y^j \in Y^i} p(y^j|x) \log \frac{p(y^j|x)}{p(y^j)} + p(t) \sum_{y^j \in Y^i} p(y^j|t) \log \frac{p(y^j|t)}{p(y^j)} - p(\tilde{t}) \sum_{y^j \in Y^i} p(y^j|\tilde{t}) \log \frac{p(y^j|\tilde{t})}{p(y^j)} \quad (14)$$

假定独立簇  $\{x\}$  被融合至某个簇  $t$  中形成新的簇  $\tilde{t}$ , 则可以得到:

$$\begin{cases} p(\tilde{t}) = p(x) + p(t) \\ p(y^j|\tilde{t}) = \frac{p(x)}{p(\tilde{t})} p(y^j|x) + \frac{p(t)}{p(\tilde{t})} p(y^j|t) \end{cases} \quad (15)$$

因此, 将公式(13)代入  $\Delta I_i^{f1}$ , 可以得到:

$$\begin{aligned} \Delta I_i^{f1} &= p(t) \sum_{y^j} p(y^j|t) \log \frac{p(y^j|t)}{p(y^j)} + p(x) \sum_{y^j} p(y^j|x) \log \frac{p(y^j|x)}{p(y^j)} - \sum_{y^j} p(t) p(y^j|t) \log \frac{p(y^j|t)}{p(y^j)} - \sum_{y^j} p(x) p(y^j|x) \log \frac{p(y^j|x)}{p(y^j)} - \sum_{y^j} p(t) p(y^j|t) \log \frac{p(y^j|\tilde{t})}{p(y^j)} + \sum_{y^j} p(x) p(y^j|x) \log \frac{p(y^j|\tilde{t})}{p(y^j)} \\ &= p(t) \sum_{y^j} p(y^j|t) \log \frac{p(y^j|t)}{p(y^j|\tilde{t})} + p(x) \sum_{y^j} p(y^j|x) \log \frac{p(y^j|x)}{p(y^j|\tilde{t})} = [p(x) + p(t)] \cdot JS_{\pi} [p(y^j|x), p(y^j|t)] \end{aligned} \quad (16)$$

其中,  $JS_{\pi} [p(y^j|x), p(y^j|t)]$  为 JS 距离,  $\pi = \{\pi_1, \pi_2\} = \left\{ \frac{p(x)}{p(\tilde{t})}, \frac{p(t)}{p(\tilde{t})} \right\}$ .

同样的推导分析, 我们可以得到其它的融合代价:

$$\begin{cases} \Delta I_i^{f2} = [p(x) + p(t)] \cdot JS_{\pi} [p(x), p(x|t)] \\ \Delta I_i^{g1} = [p(x) + p(t)] \cdot JS_{\pi} [p(\hat{x}^i|x), p(\hat{x}^i|t)] \\ \Delta I_i^{g2} = [p(x) + p(t)] \cdot JS_{\pi} [p(x), p(x|t)] \end{cases} \quad (17)$$

根据以上分析, 我们给出 DWMVC 算法的主要实现步骤:

**算法 1.** DWMVC 算法.

输入: 源变量与特征变量和相似度变量之间的联合概率分布, 即  $\{p(X, Y^i)\}_{i=1}^m$  与  $\{p(X, \hat{X}^i)\}_{i=1}^m$ ; 簇的数量

$c$ ; 拉格朗日因子  $\gamma$ ; 平衡参数  $\lambda$ ;

输出: 聚类划分  $p(t|x)$ .

(1) 在  $\{p(X, Y^i)\}_{i=1}^m$  和  $\{p(X, \hat{X}^i)\}_{i=1}^m$  上使用互信息度量方式来自动获得各个视角的权重;

(2) 随机划分源变量  $X$  至  $c$  个簇;

(3) Repeat

(4) For each  $x \in X$

(5) 提取: 从当前簇  $t^{old}$  中提取出  $x$ , 使其成为独立簇  $\{x\}$ ;

(6) 融合: 依据公式(11)计算独立簇  $\{x\}$  与其他各个簇的融合代价, 并将其最终融合至代价最小的簇  $t^{new}$  中;

(7) End For

(8) Until 聚类划分  $p(t|x)$  不再变动.

### 3.5 算法复杂度分析

在 DWMVC 算法中, 初始化过程可以在  $O(n)$  时间内实现. 在算法的主循环过程中, 第五步和第六步需要顺序地执行提取和融合操作, 时间复杂度为  $O(|X|(|Y^1| + |Y^2| + \dots + |Y^m|))$ . 因此, DWMVC 的时间复杂度为  $O(lc|X|(|Y^1| + |Y^2| + \dots + |Y^m|))$ , 其中,  $l$  为算法迭代至收敛所需的次数,  $c$  是类别的数量. 对于特定数据集, 簇的数量为某一常数, 最终, 算法的时间复杂度为  $O(l|X|(|Y^1| + |Y^2| + \dots + |Y^m|))$ .

## 4 与其他工作的关系

在本节中, 着重阐释与一些相关工作的关系. 文献[18]提出一种加权的多特征 IB 方法, 旨在整合利用图像不同特征的互补信息. 然而, 各个特征的权重需要人为设定, 而且该算法仅利用了多视角数据的特征表示. 文献[19]提出了一种基于聚类融合的多视角 IB 方法, 旨在发现视角间聚类划分的一致性. 但是, 本文 DWMVC 算法聚焦于加权策略, 旨在挖掘视角之间的互补信息. 文献[20]提出了一种无冗余的多视角 IB 方法, 主要适应于在已知一个或多个聚类结果的前提下, 再求出一个与已知聚类划分不同的聚类结果. 然而, 本文主要解决一般意义下的多视角聚类, 即整合多视角数据中的各种信息以获得更加准确的聚类划分. 文献[21]主要关注多视角分类, 然而, 本文旨在处理无监督的多视角分类 (即多视角聚类) 问题.

另外, 文献[9-12]分别基于不同策略提出包含特征学习的加权多视角聚类算法, 旨在学习出每个视角权重时, 也对各个视角的特征进行加权以挑选出各个视角中最有代表性的特征来提高聚类性能. 尽管此类方法和本文 DWMVC 方法都采用双权重学习

的方式，但是它们也有较大的差别，主要体现在以下两个方面：（1）此类算法的两个权重主要描述的是单一的多视角特征表示上的视角和特征，而我们对两种不同的多视角表示（特征表示和相似度表示）上的双重加权。（2）以上算法都需要引入额外的参数来控制权重分布，而本文所提算法不需要此类参数设置。

最后，文献[22-24]在数据随机生成的多个子空间上发现不同的聚类划分，然后综合考虑这些划分的差异性，最终得到一致性的聚类结果。尽管这些方法与本文所提的DWMVC都利用了数据的多角度信息，但是仍然存在如下两个不同点：（1）此类随机子空间聚类集成方法主要是在单视角数据上进行聚类分析，而本文处理的是多视角聚类问题。（2）以上方法主要采用了聚类集成的策略，综合考虑多个不同划分的差异性以获得最终的一致性聚类划分。而本文使用了加权的策略，旨在挖掘多个视角数据的互补信息来提高聚类性能。

## 5 实验与结果

### 5.1 评价标准

为了对比各个聚类算法的有效性，本文使用聚类中常用的两种指标<sup>[25]</sup>来评估各算法的聚类性能。对两种评估指标来讲，所求值越大，则表明聚类效果越好。

（1）聚类准确度(Clustering Accuracy, Acc):

$$Acc = \frac{\sum_{i=1}^m \delta(r_i, \text{map}(p_i))}{m} \quad (18)$$

其中  $m$  为给定数据集中数据对象的总量； $p_i$  和  $r_i$  分别表示数据对象的聚类划分和真实标签； $\delta(x, y)$  为狄拉克函数，当  $x=y$  时  $\delta(x, y)=1$ ，否则该值为 0； $\text{map}(p_i)$  为一个排列映射函数，将聚类得到的聚类划分  $p_i$  映射到与之等价的真实标签上，其中映射关系可以通过 Kuhn-Munkres 算法得到最优解。

（2）标准化互信息 (Normalized Mutual Information, NMI):

$$NMI = \frac{I(V; C)}{\max(H(V), H(C))} \quad (19)$$

其中， $V$  和  $C$  分别表示数据对象的聚类划分和真实标签， $I(V; C)$  为  $V$  和  $C$  之间的互信息， $H(V)$  和  $H(C)$  分别表示  $V$  和  $C$  的信息熵。

### 5.2 数据集

本文将在多语言文档、多特征图像和多模态人体行为视频等 5 个多视角数据集上测试我们提出的

DWMVC 算法。下面我们从数据大小和特点等方面分别介绍这些数据集，其简要信息如表 1 所示。

Reuters<sup>[26]</sup>。该多语言文档数据集包含了五种不同语言(英语、法语、德语、西班牙语和意大利语)编写的文档，每种语言包含 1200 个文档。我们从中选取前三种语言作为测试数据集，其中每一种语言作为一个视角。

COIL20<sup>[27-30]</sup>。该数据集共有 20 个对象，包含 1440 个图像。这些数据通过以下方式捕捉得到：每个对象在转盘上旋转时以 5 度的间隔拍摄 72 张图像。我们提取图像的三种视觉特征，即 SURF<sup>[31]</sup>、TPLBP<sup>[33]</sup> 和 SIFT<sup>[32]</sup> 特征，每种特征可看为一个视角。

NUS22<sup>[28]</sup>。该数据集包含 269648 张图片和相关标签，共计 81 个对象类别。我们从中随机抽取 22 个常见类别，共有 10155 个图像，并使用 SIFT 和 TPLBP 特征加以描述。

MSRDA3D<sup>[30]</sup>。该数据集是一个日常活动视频数据集，包含了三种不同的模态数据，即 RGB、深度和骨骼数据。我们选取其中两个视角：深度图和 RGB 视频，共包含 16 种活动类型。每种类型以两种不同的姿势进行活动，且大多数活动都涉及到人与对象的交互，因此，该视频数据集颇具挑战性。

HMDB<sup>[29]</sup>。该数据集为一个关于人体行为的数据集，其主要来自于电影，共包括 6766 个视频剪辑，分为 51 种行为类别，每个类别至少包含 101 个剪辑。我们使用 HoF 和 HoG 两种特征<sup>[34]</sup>加以描述，其中，每一种特征视为一个视角。

表 1 数据集简介

Dataset	Type	View	Class	Number*Feature
Reuters	Text	3	6	1200*2000
COIL20	Image	3	20	1440*1000
NUS22	Image	3	22	10155*1000
MSRDA3D	Video	2	16	320*500
HMDB	Video	2	51	6766*1000

### 5.3 对比算法与实验设置

为了验证所提 DWMVC 算法的有效性和优越性，本文与以下三类聚类算法进行对比，其主要特点如表 2 所示。

单视角聚类算法：（1） $k$ -means (KM) 算法，属于基于特征表示的聚类算法；（2）Ncut 算法<sup>[35]</sup>，即 Normalized Cuts 算法，属于基于图表示的聚类算法，以数据样本间的相似度矩阵为输入，着重从数据样本的上下文信息进行聚类分析；（3）IB 算法。

拼接视角聚类算法：(4) All-KM 算法；(5) All-Ncut 算法；(6) All-IB 算法。这些方法主要通过简单拼接各个视角的表示，然后分别使用  $k$ -means、Ncut 和 IB 算法聚类。

多视角聚类算法：(7) MVIB (Multi-view Information Bottleneck) 算法<sup>[19]</sup>，该算法通过融合不同视角下的聚类假设，从而发现出最终的一致性聚类结果。(8) Coreg (Co-regularized MVC) 算法<sup>[36]</sup>，该算法在单视角谱聚类的基础上，增加了不同视角间的约束，通过对不同视角的聚类划分进行联合正则化来实现多视角数据之间的聚类结果一致性。(9) RMKMC (Robust Multi-view  $k$ -means clustering) 算法<sup>[7]</sup>，该算法针对大规模多视角聚类问题，提出了一种自动加权的多视角  $k$ -means 聚类方法。

(10) MfIB (Multi-feature Information Bottleneck) 算法<sup>[18]</sup>，该算法通过人为设置权重，然后对多个特征变量进行加权以得到合理的数据划分。(11) SwMC (Self-weighted Multiview Clustering)<sup>[5]</sup> 算法，该算法通过拉普拉斯秩约束图来近似表示不同置信度下的各个视角的图中心。通过引入超参数，进一步提出了一种完全自我加权的多视角聚类算法。

(12) TWCOKM (Two-level Weighted Collaborative KM) 算法<sup>[12]</sup>，该算法以协作方式来设计目标函数，有效地发现嵌入在多个视角中的公共结构，并同时保留了视角的多样性和异构信息。此外通过使用加权策略，使得每个视角的重要性得到了充分的考虑。(13) MLAN (Multi-view Learning with Adaptive Neighbors)<sup>[6]</sup> 算法，该算法自动为每个视角分配合适的权重，并在合理的秩约束下，将得到的最优图直接划分为特定的簇。

对于单视角聚类算法而言，我们选择各个视角中最好的聚类结果。对于多视角对比算法，我们按照作者原文中给的参数设定分别运行，并选取最佳参数所对应的聚类结果。对于本文提出的 DWMVC 算法，其共有两个参数， $\gamma$  和  $\lambda$ 。在本文的聚类任务中，簇的个数  $c$  远小于数据样本的数量，即  $c \ll n$ ，这说明变量  $X$  到变量  $T$  之间的压缩是巨大的。因此，我们可将  $\gamma$  设置为无穷大。事实上，这种做法的有效性已在文本<sup>[37]</sup>、图像<sup>[38]</sup>和视频<sup>[39]</sup>上的聚类中得到了验证。对于平衡参数  $\lambda$ ，其取值范围为  $\{0.01, 0.02, \dots, 0.09\}$ ，对应的参数敏感性分析将在第 5.5 小节呈现。为了缓解以上算法初始随机聚类划分的影响，我们在各个数据集上运行 10 次，以求得 Acc 和 NMI 的

平均值和方差。

表 2 对比算法的特点 (SC 表示谱聚类；F 表示特征表示；S 表示相似度表示)

Method	Basis	Represent	Weighted	Weight Parameter
KM	KM	F	-	-
Ncut	SC	S	-	-
IB	IB	F	-	-
MVIB	IB	F	No	-
Coreg	SC	S	No	-
RMKMC	KM	F	Yes	Yes
MfIB	IB	F	Yes	No
SwMC	SC	S	Yes	No
TWCOKM	KM	F	Yes	Yes
MLAN	SC	S	Yes	Yes
DWMVC	IB	F+S	Yes	No

#### 5.4 实验和结果分析

在本节中，我们在五个数据集上进行实验来验证我们提出的 DWMVC 算法的有效性。表 3 和 4 展示了各种聚类方法的实验结果，经过分析表明：

(1) 简单地将各个视角连接在一起并不能够稳定地提升单视角聚类算法的性能。例如，相对于  $k$ -means、Ncut 和 IB 等单视角聚类算法在部分多视角数据集 (如 COIL20) 上的视角最佳结果，All-KM、All-Ncut 和 All-IB 算法的聚类结果都有明显下降。这是由于简单地将多视角拼接在一起并使用单视角聚类算法，视角间的互补信息没有被合理地挖掘利用。

(2) 相对于单/拼接视角聚类算法，DWMVC 算法在各个数据集上聚类表现比较突出。这主要是因为所提算法能够将不同重要性的视角信息最优化整合并加以利用，从而发现出更准确的聚类划分。

(3) 本文提出的 DWMVC 算法在各种类型的多视角数据集上均明显优于对比的多视角聚类方法。相对于其它方法，DWMVC 算法的优势主要体现在以下两个方面。一方面，通过综合考虑特定视角的两种数据表示，即基于内容的特征表示和基于上下文的相似度表示，可以充分利用特定视角内部数据样本的固有信息和数据样本之间的关联信息；另一方面，对每个视角的两种数据表示双重加权，以充分利用两方面的视角互补信息，从而提高算法的聚类性能。

表 3 在各个数据集上的 Acc (mean±std) 实验结果

Datasets	Reuters	COIL20	NUS22	MSRDA3D	HMDB
KM	24.10±7.1	52.5±4.6	12.84±0.5	30.71±2.6	9.5±0.4
Ncut	42.86±0.1	74.53±1.2	12.90±0.1	33.62±1.5	10.66±0.1
IB	44.22±3.3	69.35±3.6	14.53±0.3	31.75±2.7	9.62±0.3
All-KM	25.64±7.0	45.17±4.9	12.79±0.6	31.65±1.4	9.9±0.2
All-Ncut	40.75±0.1	46.27±0.3	14.86±0.2	34.56±0.9	11.23±0.2
All-IB	40.55±5.7	77.52±6.0	6.50±0.1	30.41±2.5	9.31±0.1
MVIB	48.47±4.1	65.46±10.7	11.21±0.6	37.78±2.0	8.03±0.2
Coreg	42.19±0.7	63.23±1.6	10.7±0.1	32.66±0.7	10.89±0.1
RMKMC	45.79±2.9	53.22±4.3	13.52±0.5	35.81±1.5	11.09±0.4
MfIB	48.63±4.3	84.99±3.6	18.20±0.4	35.81±3.5	11.67±0.3
SwMC	18.30±0.1	46.63±2.5	6.6±0.1	26.25±2.0	9.2±0.1
TWCOKM	22.93±8.2	30.24±2.7	12.53±0.5	35.22±3.3	9.74±0.9
MLAN	21.33±2.9	87.22±2.3	7.61±0.4	37.81±5.8	12.30±0.1
DWMVC	52.82±3.1	89.29±6.3	19.57±0.4	39.50±3.4	12.98±0.2

表 4 在各个数据集上的 NMI (mean±std) 实验结果

Datasets	Reuters	COIL20	NUS22	MSRDA3D	HMDB
KM	10.42±9.6	64.18±2.6	8.6±0.4	37.34±1.8	15.24±0.4
Ncut	28.04±0.1	83.83±0.4	7.4±0.1	38.55±1.0	16.22±0.1
IB	26.10±2.2	82.17±1.2	13.22±0.2	39.79±2.1	15.99±0.4
All-KM	7.8±6.0	59.17±4.3	7.3±0.4	35.28±2.1	14.0±0.4
All-Ncut	23.31±0.1	58.03±0.3	9.39±0.2	39.74±0.4	16.83±0.3
All-IB	25.73±5.0	91.42±2.6	2.24±0.0	41.32±2.3	6.34±0.1
MVIB	27.54±1.8	75.55±6.4	5.85±0.4	44.77±1.2	11.97±0.3
Coreg	27.45±0.3	83.49±0.3	6.9±0.1	43.62±0.3	18.38±0.1
RMKMC	29.00±2.5	73.27±2.0	9.0±0.3	41.80±1.5	17.94±0.3
MfIB	29.69±3.4	91.85±2.3	13.52±0.3	43.40±2.1	19.19±0.2
SwMC	3.53±0.4	59.52±3.1	1.46±0.1	32.18±1.8	2.8±0.1
TWCOKM	7.94±8.7	39.71±4.1	8.49±0.3	40.80±1.6	13.43±7.1
MLAN	8.6±0.0	94.35±1.1	2.30±0.3	42.99±1.1	8.97±0.1
DWMVC	31.99±2.2	94.53±3.1	14.54±0.3	48.20±2.3	21.07±0.3

## 5.5 参数分析

DWMVC 算法利用参数  $\lambda$  来平衡特征表示和相似度表示对双重加权多视角聚类的作用, 因此本文通过实验评估  $\lambda$  的值对算法性能的影响. 从图 7 中可以看出: (1) 当  $\lambda$  取值较小时, DWMVC 算法更加侧重于特征表示而忽略了自身相似度特征对多视角聚类的作用. 此时, DWMVC 算法在大部分数据集上得到了较低的 Acc 值; (2) 随着  $\lambda$  值的不断增加, DWMVC 算法的 Acc 值在大部分数据集呈现上升趋势. 当  $\lambda$  增加到一定程度时, 即取值在 [0.06, 0.07, 0.08] 时, 相似度表示对于特征表示的互补作用增强至最佳效果, DWMVC 算法取得了令人满意的局部

稳定聚类结果; (3) 当  $\lambda$  取值过大时, 由于过多考虑相似度表示的影响而忽略了自身特征表示对多视角聚类的作用, DWMVC 算法在大部分数据集上的性能表现较差.

## 5.6 收敛性分析

图 8 给出了 DWMVC 算法在多视角数据集上每次迭代后该算法的目标函数值. 从图 8 可以看出随着算法运行迭代次数的增加, 目标函数的值迅速提升, 而后增长速度趋于稳定, 且每次迭代都确保了该数值的增加. 在经过多次迭代后, 算法最终得到了局部最优值, 从而验证了该算法具有很好的收敛性.

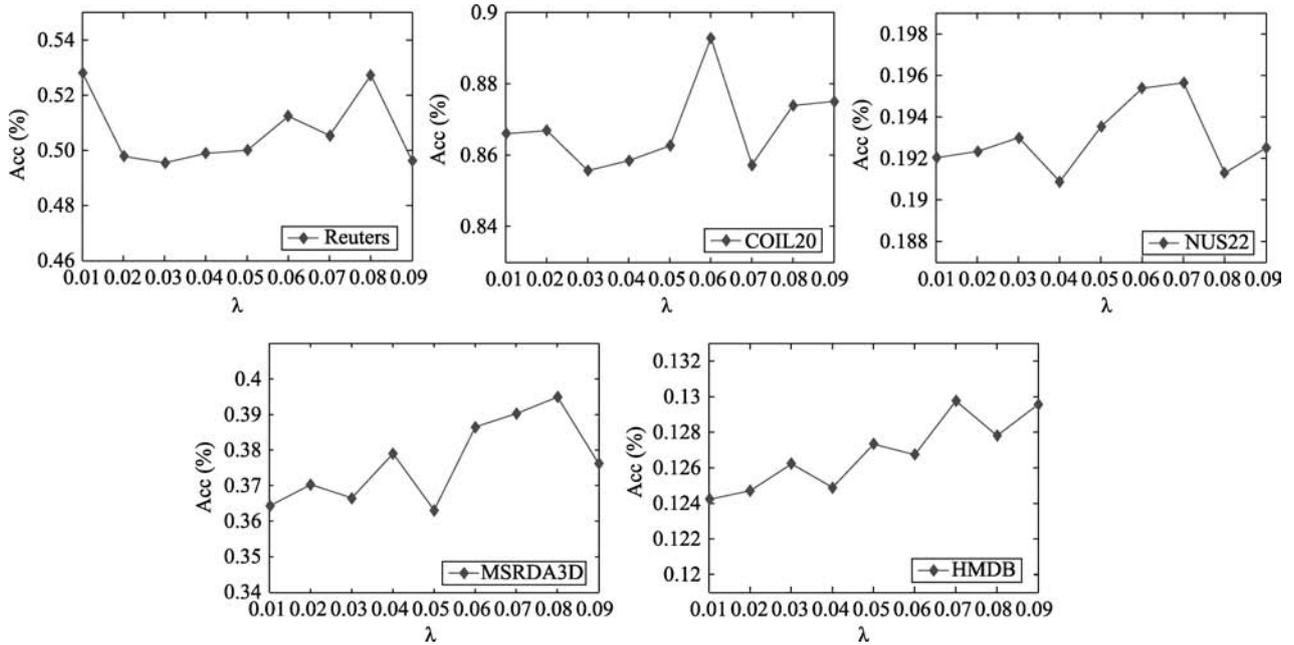


图 7 在各个数据集上的参数敏感性分析

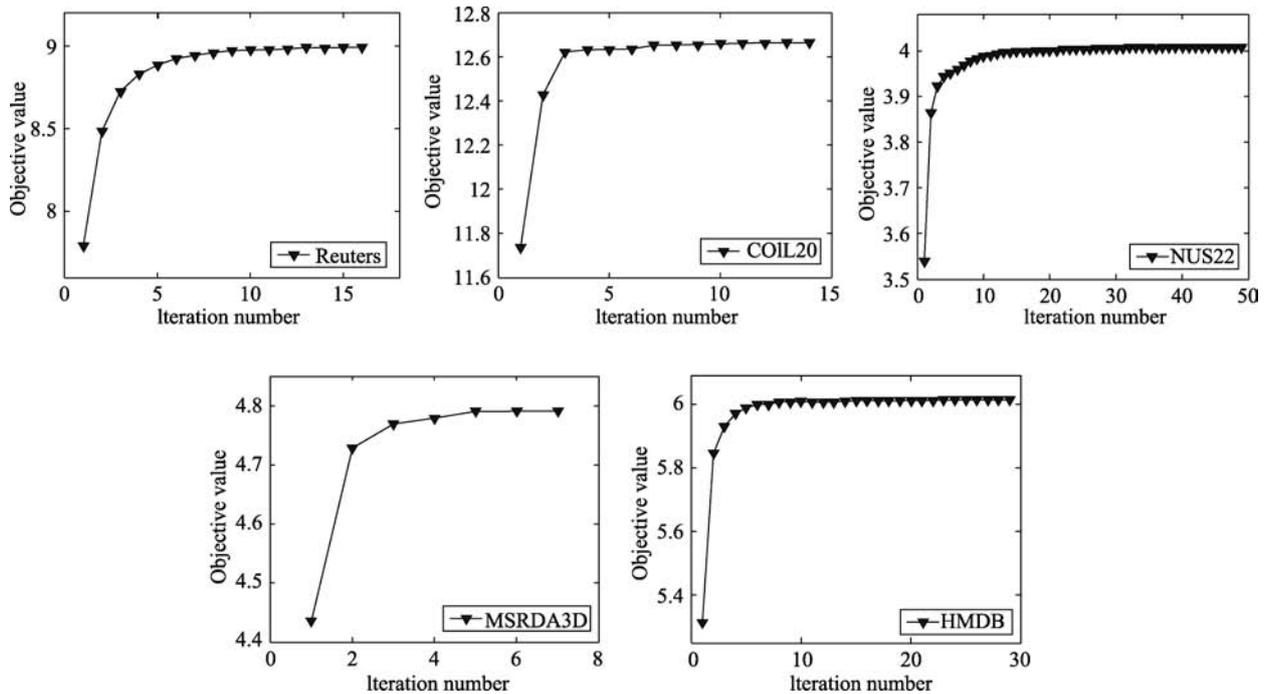


图 8 在各个数据集上的收敛性分析

## 6 总 结

本文提出了一种双重加权的多视角聚类算法 DWMVC, 该算法对基于内容的多视角特征表示和基于上下文的多视角相似度表示进行双重加权, 以充分利用视角间的互补信息发现数据中潜在的模式结构. 算法的优点在于, 通过互信息理论学习视角权重, 无需引入额外的参数, 使得方法在面向实际应用时更加切实可行. 在多视角文本、图像和视频

上的实验结果表明, 本文提出的 DWMVC 算法优于现有先进的单视角和多视角聚类算法.

在未来的工作中, 我们考虑将特征学习策略融入到所提多视角聚类算法中, 进而使其能够高效地处理高维数据. 另外, 在大数据的背景下, 部分视角中的数据可能会存在丢失现象, 因而会给多视角聚类带来巨大的阻碍和挑战, 这也将是我们下一步着重研究的工作重点.

## 参 考 文 献

- [1] Bickel S, Scheffer T. Multi-view clustering//Proceedings of the IEEE International Conference on Data Mining. Brighton, UK, 2004: 19-26
- [2] Chaudhuri K, Kakade S M, Livescu K, et al. Multi-view clustering via canonical correlation analysis//Proceedings of the Annual International Conference on Machine Learning. Montreal, Canada, 2009: 129-136
- [3] Zhan Kun, Nie Feiping, Wang Jing, et al. Multiview consensus graph clustering. IEEE Transactions on Image Processing, 2019, 28(3): 1261-1270
- [4] Tzortzis G, Likas A. Kernel-based weighted multi-view clustering//Proceedings of the IEEE International Conference on Data Mining. Brussels, Belgium, 2012: 675-684
- [5] Nie Feiping, Li Jing, Li Xuelong. Self-weighted multiview clustering with multiple graphs//Proceedings of the International Joint Conference on Artificial Intelligence. Melbourne, Australia. 2017: 2564-2570
- [6] Nie Feiping, Cai Guohao, Li Jing, et al. Auto-weighted multi-view learning for image clustering and semi-supervised classification. IEEE Transactions on Image Processing, 2018, 27(3): 1501-1511
- [7] Cai Xiao, Nie Feiping, Huang Heng. Multi-view k-means clustering on big data//Proceedings of the International Joint Conference on Artificial Intelligence. Beijing, China, 2013: 2598-2604
- [8] Li Yeping, Nie Feiping, Huang Heng, et al. Large-scale multi-view spectral clustering via bipartite graph//Proceedings of the AAAI Conference on Artificial Intelligence. Austin, USA, 2015: 2750-2756
- [9] Wang Hua, Nie Feiping, Huang Heng. Multi-view clustering and feature learning via structured sparsity//Proceedings of the International Conference on Machine Learning. Atlanta, USA, 2013: 352-360
- [10] Chen Xiaojun, Xu Xiaofei, Huang J Z, et al. TW-k-means: automated two-level variable weighting clustering algorithm for multiview data. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(4): 932-944
- [11] Xu Yumeng, Wang Changdong, Lai Jianhuang. Weighted multi-view clustering with feature selection. Pattern Recognition, 2016, 53: 25-35
- [12] Zhang Guangyu, Wang Changdong, Huang Dong, et al. TW-Co-k-means: Two-level weighted collaborative k-means for multi-view clustering. Knowledge Based Systems, 2018, 150: 127-138
- [13] Cover T M, Thomas J A. Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Seventh Edition. USA: Wiley-Blackwell, 2006
- [14] Lv Qingwen, Chen Wufan. Image segmentation based on mutual information, Chinese Journal of Computers, 2006(02): 296-301 (吕庆文, 陈武凡. 基于互信息量的图像分割. 计算机学报, 2006(02): 296-301)
- [15] Tishby N, Pereira F C, Bialek W. The information bottleneck method//Proceedings of the ALLerton Conference on Communication, Control and Computing. Illionis, USA, 1999: 368-377
- [16] Tishby N, Zaslavsky N. Deep learning and the information bottleneck principle//Proceedings of the IEEE Information Theory Workshop. Jerusalem, Israel, 2015: 1-5.
- [17] Amjad R A, Geiger B C. Learning representations for neural network-based classification using the information bottleneck principle. IEEE Transactions on Pattern Analysis and Machine Intelligence, DOI: 10.1109/TPAMI. 2019. 2909031
- [18] Lou Zhengzheng, Ye Yangdong, Yan Xiaoqiang. The multi-feature information bottleneck with application to unsupervised image categorization//Proceedings of the International Joint Conference on Artificial Intelligence. Beijing, China, 2013: 1508-1515
- [19] Gao Yan, Gu Shiwen, Li Jianhua, et al. The multi-view information bottleneck clustering//Proceedings of the International Conference on Database Systems for Advanced Applications. Bangkok, Thailand, 2007: 912-917
- [20] Lou Zhengzheng, Ye Yangdong, Liu Ruina. Non-redundant multi-view clustering based on information bottleneck. Journal of Computer Research and Development, 2013, 50(09): 1865-1875 (娄铮铮, 叶阳东, 刘瑞娜. 基于 IB 方法的无冗余多视角聚类. 计算机研究与发展, 2013, 50(09): 1865-1875)
- [21] Xu Chang, Tao Dacheng, Xu Chao. Large-margin multi-view information bottleneck. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(8): 1559-1572
- [22] Yu Zhiwen, Zhu Xianjun, Wong Hau-San, et al. Distribution based Cluster Structure Selection. IEEE Transactions on Cybernetics, 2017, 47(11): 3554-3567
- [23] Yu Zhiwen, Li Le, Liu Jiming, et al. Adaptive noise immune cluster ensemble using affinity propagation. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(12): 3176-3189
- [24] Yu Zhiwen, Chen Hantao, You Jane, et al. Adaptive fuzzy consensus clustering framework for clustering analysis of cancer data. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2015, 12(3): 568-582
- [25] Lou Zhengzheng, Ye Yangdong. Symmetric information bottleneck method on maximization inter-correlated mutual information, Chinese Journal of Computers, 2016, 39(08): 1515-1527 (娄铮铮, 叶阳东. 基于最大化交叉互信息的对称 IB 算法. 计算机学报, 2016, 39(08): 1515-1527)
- [26] Amini M R, Usunier N, Goutte C, et al. Learning from multiple partially observed views-an application to multilingual text categorization//Proceedings of the Neural Information Processing Systems. Vancouver, Canada. 2009: 28-36
- [27] Nene S A, Nayar S K, Murase H. Columbia object image library (COIL-20). Technical Report CUCS-005-96, Columbia University, 1996
- [28] Chua Tat-Seng, Tang Jinhui, Hong Richang, et al. NUS-WIDE: a real-world web image database from National University of Singapore //Proceedings of the ACM International Conference on Image and Video Retrieval. Santorini Island, Greece, 2009: 368-375
- [29] Kuehne H, Jhuang H, Garrote E, et al. HMDB: A large video database for human motion recognition//Proceedings of the IEEE International Conference on Computer Vision. Barcelona, Spain, 2011: 2556-2563

- [30] Wang Jiang, Liu Zicheng, Wu Ying, et al. Mining actionlet ensemble for action recognition with depth cameras//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012: 1290-1297
- [31] Bay H, Tuytelaars T, Van Gool L, et al. SURF: speeded up robust features//Proceedings of the European Conference on Computer Vision. Graz, Austria, 2006: 404-417
- [32] Lindeberg T. Scale invariant feature transform. Scholarpedia, 2012, 7(5)
- [33] Wolf L, Hassner T, Taigman Y. Descriptor based methods in the wild//Proceedings of the European Conference on Computer Vision Workshop on Faces in Real-Life Images. Marseille, France, 2008: 1-14
- [34] Laptev I. On Space-time interest point. International Journal of Computer Vision, 2005, 64(2-3): 107-123
- [35] Shi Jianbo, Malik J. Normalized cuts and image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905
- [36] Kumar A, Rai P, Daume H. Co-regularized multi-view spectral clustering//Proceedings of the Annual Conference on Neural Information Processing Systems. Granada, Spain, 2011: 1413-1421
- [37] Slonim N. The information bottleneck: Theory and applications [Ph.D. Dissertation]. Hebrew, Israel: Hebrew University, 2002
- [38] Hu Shizhe, Yan Xiaoqiang, Ye Yangdong. Multi-task Image Clustering through Correlation Propagation. IEEE Transactions on Knowledge and Data Engineering. DOI: 10.1109/TKDE.2019.2937026
- [39] Yan Xiaoqiang, Hu Shizhe, Ye Yangdong. Multi-task clustering of human actions by sharing information//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 4049-4057

**HU Shi-Zhe**, Ph.D. candidate.

His main research interests include pattern recognition and information theory.



**LOU Zheng-Zheng**, Ph.D.,

lecturer. His main research interests include machine learning and pattern recognition.

**WANG Ruo-Bin**, master candidate. His main

research interests include machine learning and data mining.

**YAN Xiao-Qiang**, Ph.D. His main research

interests include machine learning and computer vision.

**YE Yang-Dong**, Ph.D, professor. His main research

interests include machine learning, knowledge engineering and intelligent system.

## Background

With the rapid growth of multi-source heterogeneous data in the Big Data era, such as multi-lingual text, multi-feature image and multi-modal video, the clustering on these data, i.e., multi-view clustering, has been a challenging task in the field of machine learning. Multi-view clustering is an important method for pattern discovery on complex data. However, most of these existing multi-view clustering methods learn and apply weights on only either the content-based feature representations or the context-based similarity representations, failing to consider both of them to fully represent the intra-view intrinsic information. Moreover, most of these methods need to introduce extra parameters to control the weight distribution, and the parameter selection is quite difficult without prior knowledge.

In this work, a novel dual-weighted multi-view

clustering (DWMVC) method is proposed. First, we utilize Bag-of-words model and Gaussian kernel function to construct the content-based and context-based representations of the given multi-view data, respectively. Then, the mutual information theory is adopted to measure the importance of each view to obtain the view weights. Finally, an information theory based objective function is formulated, which aims to compresses the two kinds of data representations while preserving the relevant information as much as possible. Experimental results on different kinds of multi-view data shows the proposed method outperforms several state-of-the-art methods.

This work has been supported by the National Natural Science Foundation of China (61772475), National Key R&D Program of China (2018YFB1201403) and National Science Foundation for Young Scientists of China (61906172).