专题综述

基于低秩张量的多视图聚类相似性学习

中山大学 陈曼笙 王昌栋 赖剑煌

研究问题是基于低秩张量学习多视图样本间的相似性，并实现最终一致的聚类结果[1]。现有面向图的多视图聚类方法通常是使用多视图数据中隐藏的关系和复杂结构来实现令人深刻的聚类效果。然而，它们仍然存在以下两个常见问题：（1）他们以研究视图之间的共同表示或成对相关性为目标，忽略了多个视图间的全面性和更深层次的高阶相关性。（2）它们没有在统一的图构建中考虑特定视图表示的先验知识，并在统一的聚类框架中获得共识聚类指示矩阵。为了解决这些问题，我们提出了一种新颖的基于低秩张量的相似性学习方法用于多视图聚类(LTBPL)，在统一的框架中共同研究了多个低秩概率相似性矩阵和反映最终性能的共识聚类指示矩阵。具体来说，将多个相似性表示堆叠在一个低秩约束的张量中，以恢复它们的全面性和高阶相关性。同时，联合构建携带不同自适应置信度的特定视图表示和共识指示聚类矩阵的关系。在九个真实世界数据集的广泛实验表明了和最先进的聚类方法相比，LTBPL有明显的优越性。

一、研究背景

一个物品通常由来自多个视图的不同特征来表示，特别地不同特征之间是互补的，这直接推动了多视图学习的发展。多视图学习能够整合所有视图的不同特征，并利用它们之间的相关性去获得更精炼和更高层次的信息。多视图学习的成功来源于两个重要原则，即共识原则和互补性原则。其中，共识原则的目的是最大化多个视图之间的一致性；互补性原则意味着数据的某一视图包含了一些其他视图没有的信息。在这个工作中，我们关注于多视图聚类，其缺乏指导学习过程的真实类标。

近些年来，研究者们投入了许多的努力去设计多视图聚类算法。由于图的形式可以表征数据结构，现有的基于图的多视图聚类方法占多数，例如基于相似性的多视图谱聚类[2, 3, 4]，基于图的多视图子空间聚类[5,6]等等。其中，有几种常见的图构造方法，即k-最近邻图[7]，局部线性相似性图[8]、成对相似性图[9]和用子空间[10]学习的图。此外，基于张量核范数的张量奇异值分解(t-SVD)这个新兴策略，一些基于张量的多视图聚类方法被设计来发现多视图数据的空间结构和高阶信息[11,12]，这很好地改善了它们的聚类能力。但不幸的是，尽管这些方法取得了很大的成功，它们大多数旨在研究一个共同的表示或视图之间的成对相关性，导致了多视图数据之间全面性和更深层次高阶相关性的缺失，因此错过了重要的底层语义信息。此外，图的构建独立于聚类研究，且不关注学习相似性图趋向于聚类指示矩阵的先验信息，最终导致次优聚类效果。

针对上述问题，本文提出了一种新颖的基于低秩张量的相似性学习方法用于多视图聚类(LTBPL)，在一个统一的框架下联合研究低秩概率相似性矩阵和反映最终聚类结果的共识聚类指示矩阵。具体来说，在图学习的基础上，首先根据多视图样本点间的距离构造概率邻居图。为了全面探索所有视图之间的高阶相关性，多个视图的概率相似性矩阵被堆叠成张量，其中，本文利用基于 t-SVD 的加权张量核范数来恢复来自多个视图样本的潜在互补性和高阶相关性，在张量学习时考虑了矩阵不同奇异值之间的显著线索。此外，根据共识原则，本文联合学习了共识聚类指示矩阵和视图特定的概率相似性矩阵，其中利用多个视图概率相似性表示的不同

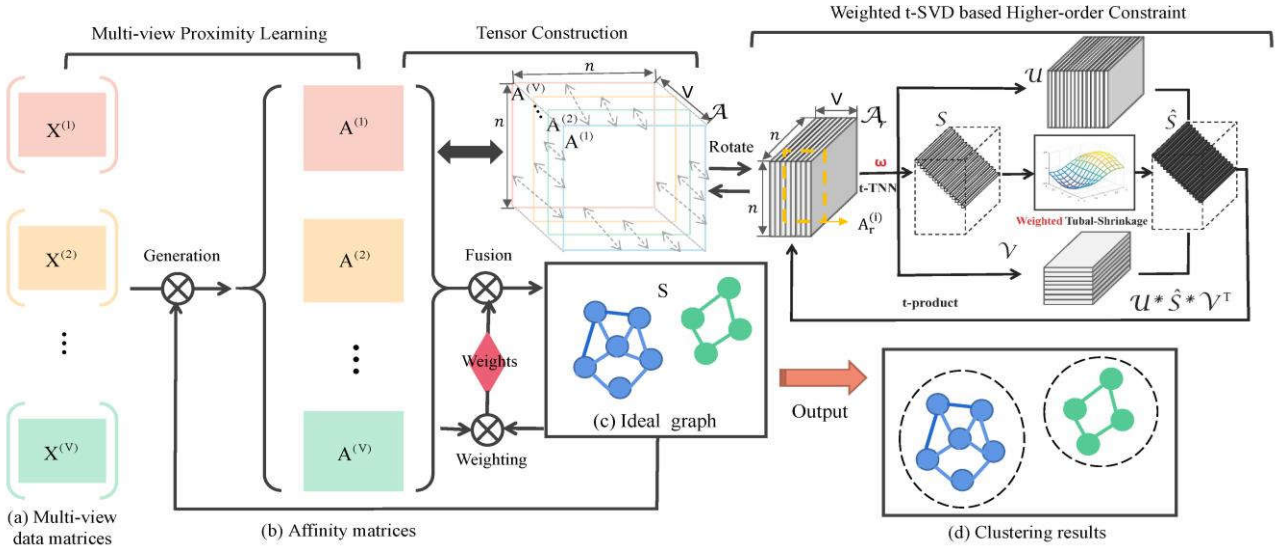

图 1 所提出 LTBPL 模型的示意图

置信度自适应地研究共识聚类指示矩阵。因此，所学习到的低秩概率相似性矩阵能够更好地表征数据结构潜在的互补和高阶相关性，而且反映最终聚类结果的共识聚类指示矩阵是同时通过了多个低秩概率相似性矩阵的不同贡献研究得到的。现将主要贡献概括如下：

- 一个新颖的框架被构造，以同时研究基于 t-SVD 加权张量核范数约束的低秩概率相似性矩阵和整合的共识聚类指示矩阵。
- 不同视图奇异值的先验信息通过基于 t-SVD 加权张量核范数能够被显示地考虑。同时，趋向最终共识聚类的不同低秩概率表征的贡献可以自适应学习得到。
- 在九个真实世界数据集上的广泛实验表明和最先进的多视图聚类方法相比，我们的方法有明显的优越性。

二、LTBPL方法介绍

1. 基于低秩张量的相似性学习

给定一个多视图数据集$X = \{X^{(1)}, \dots, X^{(V)}\}$，包含了V个视图，其中$X^{(v)} = \left[x_1^{(v)}, \dots, x_n^{(v)}\right] \in R^{d^v \times n}, \forall v = 1, \dots, V$表示第v个视图的特征空间。对于多视图聚类任务，探索数据的局部连通性是一种成功的策略，$x_i^{(v)} \in R^{d^v \times 1}$ 的邻居通常被描述为数据集中与$x_i^{(v)}$挨着的k个近邻样本点。特别地，在本文中，概率邻居简单地通过运用欧几里得距离作为距离度量来考虑，然后数据样本的相似性可以根据他们基于图学习的距离得到。具体而言，对于第v个视图的一个数据样本$x_i^{(v)}$，可以将所有数据样本点$\left[x_1^{(v)}, x_2^{(v)}, \dots, x_n^{(v)}\right]$作为连接到$x_i^{(v)}$的邻居，对应的概率为$a_{ij}^{(v)}$，其中当有一个较小的距离$||x_i^{(v)} - x_j^{(v)}||_2^2$时，可以得到一个较大的概率$a_{ij}^{(v)}$。因此，本文定义了一个基础的框架去学习相似性$a_{ij}^{(v)}$：

$$\min_{A^{(v)}} \sum_{v=1} \sum_{i,j=1}^{n} ||x_i^{(v)} - x_j^{(v)}||_2^2 a_{ij}^{(v)} + \alpha ||A^{(v)}||_F^2,$$
$$\text{s.t. } 0 \le a_{ij}^{(v)} \le 1, \left(a_i^{(v)}\right)^T \mathbf{1} = 1,$$

其中α是一个权衡参数，$a_i^{(v)} \in R^{n \times 1}$表示一个列向量，它的第j个元素是$a_{ij}^{(v)}$。第一项是被用于确定概率相似性，第二项引入正则化项去避免平凡解$A^{(v)} = I$。尽管取得了显著的效果，大多数现有的方法旨在研究一个共同的表示或视图之间的成对相关性，导致多视图数据间全面性和更深层次的高阶相关性的缺失，从而错过了重要的底层语义信息。 此外，它需要一个单独的后处理步骤以获得最终的聚类结果，并且无法在一个统一的框架中考虑不同视图的多个概率相似性矩阵和最终聚类指示矩阵的联系，从而导致随后次优的聚类性能。

针对上述问题，本文提出了一个新颖的基于低秩张量的多视图聚类相似性学习方法(LTBPL)，其中每个具有高阶相关性的视图特定的相似性矩阵和最终的聚类指示矩阵以相互作用的方式实现联合优化。为清晰起见，

所提出的 LTBPL 方法的流程图如图 1 所示。依据图示，从多个特征子集或者源头获取得到的数据样本$X^{(1)},\dots,X^{(V)}$首先作为输入。基于多视图相似性学习策略，可以得到每个视图对应的相似性矩阵$A^{(1)},\dots,A^{(V)}$。为了捕捉到不同视图中多个样本点之间的高阶相关性，本文运用了张量构造技术，其中被构造的张量 $\mathcal{A}\in R^{n\times n\times V}$ 是由多个相似性矩阵构成。然后，所提出 LTBPL 方法的模型表达可以进一步构造如下：

$$\min_{\mathcal{A},A^{(v)}} \sum_{v=1}\sum_{i,j=1}^{n} ||x_i^{(v)}-x_j^{(v)}||_2^2 a_{ij}^{(v)} + \alpha||A^{(v)}||_F^2 +\beta C(\mathcal{A}),$$

$$\text{s.t. } 0\le a_{ij}^{(v)}\le 1, \left(a_i^{(v)}\right)^T \mathbf{1}=1,\ \mathcal{A}=G(A^{(1)},\dots,A^{(V)}),$$

其中β是一个惩罚因子。$C(\cdot)$表示在构造张量 $\mathcal{A}$ 上的特定约束，$G(\cdot)$通过整合多个相似性矩阵$A^{(v)}$成一个三阶张量。具体地，本文采用了基于 t-SVD 加权张量核范数约束去恢复隐藏在多视图相似性矩阵里的高阶相关性，它的构造可以进一步改写如下：

$$\min_{\mathcal{A},A^{(v)}} \sum_{v=1}\sum_{i,j=1}^{n} ||x_i^{(v)}-x_j^{(v)}||_2^2 a_{ij}^{(v)} + \alpha||A^{(v)}||_F^2 +\beta||\mathcal{A}||_{\omega,*}$$

$$\text{s.t. } 0\le a_{ij}^{(v)}\le 1, \left(a_i^{(v)}\right)^T \mathbf{1}=1,\ \mathcal{A}=G(A^{(1)},\dots,A^{(V)}).$$

其中$||\cdot||_{\omega,*}$表示基于 t-SVD 加权张量核范数约束。特别地，通过加权张量核范数最小化，$\mathcal{A}$ 的所有奇异值会被不平等地正则化，且软阈值函数可以用不同的加权参数来收缩所有不同的奇异值。在进一步详细计算前，需要对构造的张量 $\mathcal{A}$ 进行旋转，以便更好地捕捉视图间的低秩属性，并显著降低计算复杂度，维度从$n\times n\times V$变化为$n\times V\times n$，其变换步骤可见于图 1。

2. 自适应加权的共识整合

基于改良的带有高阶信息的相似性矩阵，可以推导出直接反映最终聚类结果的共识理想相似性矩阵：

$$\min_S \sum_{v=1} ||S-A^{(v)}||_F^2,$$

$$\text{s.t. } 0\le s_{ij}\le 1, s_i^T\mathbf{1}=1, \text{rank}(L_s)=n-c.$$

在上述的公式中，$\text{rank}(L_s)$表示的是拉普拉斯矩阵$L_s=D_s+(S+S^T)/2$的秩，其中$D_s\in R^{n\times n}$是一个对角矩阵，它的第j个元素是$\sum_j(s_{ij}+s_{ji})/2$。

注意到上述模型平等地对待每个相似性矩阵去学习一致的图表达，这忽略了多个视图的不同贡献度，并导致最终次优的性能。因此，本文设计了一个更合理的自适应加权策略去整合多个相似性矩阵[13]，它的目标函数可以表达如下：

$$\min_S \sum_{v=1} \gamma^{(v)}||S-A^{(v)}||_F^2,$$

$$\text{s.t. } 0\le s_{ij}\le 1, s_i^T\mathbf{1}=1, \text{rank}(L_s)=n-c,$$

其中$\gamma^{(v)}$被定义如下：

$$\gamma^{(v)}=\frac{1}{||S-A^{(v)}||_F}.$$

明显地，可以看到$\gamma^{(v)}$依赖于S。如果第v个视图是良好的，对应的$||S-A^{(v)}||_F$应该是小的，那么权重$\gamma^{(v)}$应该是大的。反过来，一个较差的视图会被赋予较小的权重，这表明了本文自适应加权学习策略的意义。然而，解决上述模型的优化问题是十分困难的，因为L_s依赖于目标变量S，且秩约束$\text{rank}(L_s)=n-c$是非线性的。

依据文献[14]，让$\theta_i(L_s)$表示L_s的第i个最小特征值。由于L_s是半正定的，那么有$\theta_i(L_s)\ge 0$。给定一个足够大的 λ，上述的模型中的秩约束可以被去掉，并等同地改写为：

$$\min_{S,\gamma} \sum_{v=1} \gamma^{(v)}||S-A^{(v)}||_F^2 + 2\lambda\sum_{i=1}^{c}\theta_i(L_s),$$

$$\text{s.t. } 0\le s_{ij}\le 1, s_i^T\mathbf{1}=1.$$

当 λ 足够大且对于每个i有$\theta_i(L_s)\ge 0$，上述模型的最优解S会让第二项$\sum_{i=1}^{c}\theta_i(L_s)$接近于 0，从而满足秩约束$\text{rank}(L_s)=n-c$。额外地，受文献[15]的启发，可以得到以下的等式：

$$\sum_{i=1}^{c}\theta_i(L_s)=\min_{F^TF=I}\text{Tr}(F^TL_sF).$$

因此，关于共识相似性矩阵的学习模型可以重构为：

$$\min_{S,F,\gamma} \sum_{v=1} \gamma^{(v)}||S-A^{(v)}||_F^2 + 2\lambda\text{Tr}(F^TL_sF),$$

$$\text{s.t. } 0\le s_{ij}\le 1, s_i^T\mathbf{1}=1, F^TF=I.$$

最终，考虑到多视图相似性矩阵和共识相似性矩阵的联合学习，将所提出的 LTBPL 模型构造如下：

$$\min_{\mathcal{A},A^{(v)},S,F,\gamma} \sum_{v=1}\sum_{i,j=1}^{n} ||x_i^{(v)}-x_j^{(v)}||_2^2 a_{ij}^{(v)} + \alpha||A^{(v)}||_F^2 +\beta||\mathcal{A}||_{\omega,*} + \sum_{v=1}\gamma^{(v)}||S-A^{(v)}||_F^2 + 2\lambda\text{Tr}\left(F^TL_sF\right),$$

$$\text{s.t. } 0 \le a_{ij}^{(v)} \le 1, \left(a_i^{(v)}\right)^T \mathbf{1} = 1, \ \mathcal{A} = G(A^{(1)}, \ldots, A^{(V)}),$$

$$0 \le s_{ij} \le 1, s_i^T \mathbf{1} = 1, F^T F = I.$$

我们观察到，最终的共识图S和由低秩张量 $\mathcal{A}$ 约束得到的每个相似性矩阵$A^{(v)}$能够在统一框架中联合学习。

上述模型中的低秩张量正则化项是用于挖掘多个视图$A^{(v)} \in R^{n\times n}$之间的潜在全面性和高阶相关性的。一方面，张量旋转之后，旋转张量 $\mathcal{A}_r$ 的第i个正面切片 $\mathcal{A}_r^{(i)} \in R^{n\times V}$描述了在不同视图中n个样本点的关系。一个好的图$A^{(v)}$应该确保在不同视图里n个样本点之间的关系应该是一致的。考虑到不同视图通常揭示了不同的类结构，本文将张量多秩最小化约束施加于张量 $\mathcal{A}$，确保了每个 $\mathcal{A}_r^{(i)}$拥有空间低秩的结构，从而使得 $\mathcal{A}_r^{(i)}$可以很好地刻画多个视图之间的全面性信息。另一方面，和矩阵(二阶的张量)相比较，这里的三阶张量是高阶的。有了低秩的约束，高阶相关性(三阶的)可以通过张量挖掘得到，而矩阵只能捕捉到二阶的关系。因此，不同视图之间的潜在全面性和高阶相关性可以被模型中的低秩张量正则化项很好地挖掘得到。

三、实验结果

1. 数据集

表 1　九个真实世界数据集的统计数据

Datasets	Type	#Objects	View dimensions	#Classes
Yale	Image	165	4096, 3304, 6750	15
ORL	Image	400	4096, 3304, 6750	40
COIL-20	Image	1440	1024, 3304, 6750	20
UCI	Image	2000	240, 76, 6	10
Caltech-101	Image	1474	48, 40, 254, 1984, 512, 928	7
Notting-Hill	Image	4660	6750, 3304, 2000	5
Hdigit	Image	10000	784, 256	10
BBCSport	Document	544	3183, 3203	5
BBC4view	Document	685	4659, 4633, 4665, 4684	5

本文在九个真实世界数据集上对所提出的 LTBPL 模型进行了广泛的实验，以证实 LTBPL 的有效性和优越性。具体地，九个数据集的统计数据可见于表 1。所提出 LTBPL 方法的源代码可以通过以下的链接进行下载：https://github.com/ManshengChen/Code-for-LTBPL-master.

2. 对比实验

不同的聚类方法在九个真实数据集上得到的聚类结果分别报告于表 2、表 3 和表 4。"SC 1" 表示在数据集的第一个视图中执行谱聚类算法，类似地对于 "SC 2" 和 "SC 3" 等等。在不同的表中，我们用粗体强调了不同数据集上的最佳性能。

从这三个表中可以观察到，提出的 LTBPL 方法在所有基准数据集上几乎都达到了最好的聚类性能，尤其是在九个数据集中的七个上获得了聚类结构与真实标签的完全匹配(即全部 1)。例如，在 Yale 数据集上，LTBPL 明显优于第二最佳方法(UGLTL[16])，通过分别实现 ACC 和 NMI 的改进为 0.7% 和 0.8%。在 BBC4view 数据集上，LTBPL 明显优于第二最佳方法(WTNM[12])，通过分别实现 ACC 和 NMI 的改进为 0.44% 和 1.63%。

尤其是，可以观察到基于 SVD 张量核范数的方法，即 LTBPL、UGLTL[16]、ETMC[17]、tSMC[11]和 WTNM[12]，比其他的方法通常能够实现更好的聚类效果。这证实了张量核范数在捕获多视图数据高阶关系的有效性。尽管如此，本文所提出的 LTBPL 方法有更明显的优越性，其中每个视图的相似性矩阵和共识的聚类指示矩阵可以在统一的框架中联合学习得到。

此外，在 Yale 数据集上的可视化结果可见于图 2，可以看到，LTBPL 揭示了一个相当清晰的底层集群结构。

3. 消融实验

为研究低秩张量和自适应联接多个相似性矩阵和共识聚类指示矩阵策略的作用，本文进行了深入的消融实验。对于 LTBPL-t1，张量核范数项被去掉($\beta = 0$)，其他项保持不变。对于 LTBPL-t2，将学习到的多个相似性矩阵累加得到一致的相似性表征($\gamma^{(v)} = 0$)，并作为谱聚类算法的输入得到最终的聚类结果。在所有基准数据集上的对比结果可见于表 5。从表中可以看出，LTBPL 在所有的测试中都比 LTBPL-t1 和 LTBPL-t2 更优越。

四、总结

本文设计了一种新颖的基于低秩张量的多视图相似性学习方法(LTBPL)。多个相似性表征被堆叠成一个受 t-SVD 加权张量核范数约束的低秩张量，去挖掘多个视图之间的全面性和高阶相关性，其中多个视图奇异值

表 2　对比结果：在 Yale、ORL 和 COIL-20 数据集上通过不同方法得到的均值和标准差

Datasets	Yale				ORL				COIL-20			
Method	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI
SC 1	$0.538_{\pm 0.044}$	$0.586_{\pm 0.038}$	$0.383_{\pm 0.043}$	$0.341_{\pm 0.046}$	$0.650_{\pm 0.016}$	$0.798_{\pm 0.010}$	$0.528_{\pm 0.021}$	$0.516_{\pm 0.022}$	$0.655_{\pm 0.028}$	$0.756_{\pm 0.014}$	$0.598_{\pm 0.023}$	$0.959_{\pm 0.00}$
SC 2	$0.569_{\pm 0.033}$	$0.598_{\pm 0.024}$	$0.423_{\pm 0.031}$	$0.384_{\pm 0.034}$	$0.774_{\pm 0.025}$	$0.891_{\pm 0.013}$	$0.712_{\pm 0.031}$	$0.704_{\pm 0.032}$	$0.745_{\pm 0.024}$	$0.828_{\pm 0.012}$	$0.712_{\pm 0.023}$	$0.971_{\pm 0.00}$
SC 3	$0.640_{\pm 0.039}$	$0.657_{\pm 0.031}$	$0.489_{\pm 0.037}$	$0.454_{\pm 0.040}$	$0.704_{\pm 0.027}$	$0.842_{\pm 0.013}$	$0.611_{\pm 0.030}$	$0.602_{\pm 0.031}$	$0.691_{\pm 0.022}$	$0.792_{\pm 0.009}$	$0.654_{\pm 0.016}$	$0.965_{\pm 0.00}$
CoTr	$0.622_{\pm 0.003}$	$0.656_{\pm 0.004}$	$0.486_{\pm 0.005}$	$0.450_{\pm 0.005}$	$0.753_{\pm 0.005}$	$0.881_{\pm 0.003}$	$0.688_{\pm 0.007}$	$0.680_{\pm 0.007}$	$0.737_{\pm 0.003}$	$0.826_{\pm 0.002}$	$0.706_{\pm 0.004}$	$0.691_{\pm 0.00}$
RMSC	$0.610_{\pm 0.016}$	$0.648_{\pm 0.012}$	$0.473_{\pm 0.015}$	$0.437_{\pm 0.016}$	$0.758_{\pm 0.011}$	$0.884_{\pm 0.004}$	$0.698_{\pm 0.010}$	$0.690_{\pm 0.011}$	$0.754_{\pm 0.002}$	$0.831_{\pm 0.002}$	$0.716_{\pm 0.003}$	$0.702_{\pm 0.00}$
CSMSC	$0.766_{\pm 0.036}$	$0.782_{\pm 0.020}$	$0.645_{\pm 0.032}$	$0.621_{\pm 0.035}$	$0.816_{\pm 0.026}$	$0.917_{\pm 0.010}$	$0.774_{\pm 0.025}$	$0.768_{\pm 0.026}$	$0.732_{\pm 0.035}$	$0.832_{\pm 0.016}$	$0.694_{\pm 0.029}$	$0.677_{\pm 0.03}$
LTMSC	$0.737_{\pm 0.009}$	$0.760_{\pm 0.007}$	$0.618_{\pm 0.012}$	$0.593_{\pm 0.013}$	$0.791_{\pm 0.023}$	$0.902_{\pm 0.010}$	$0.739_{\pm 0.024}$	$0.732_{\pm 0.025}$	$0.706_{\pm 0.029}$	$0.809_{\pm 0.016}$	$0.668_{\pm 0.025}$	$0.650_{\pm 0.02}$
LMSC	$0.667_{\pm 0.017}$	$0.689_{\pm 0.015}$	$0.502_{\pm 0.022}$	$0.466_{\pm 0.023}$	$0.801_{\pm 0.033}$	$0.906_{\pm 0.020}$	$0.745_{\pm 0.046}$	$0.739_{\pm 0.047}$	$0.730_{\pm 0.027}$	$0.835_{\pm 0.016}$	$0.697_{\pm 0.025}$	$0.680_{\pm 0.02}$
MCIAS	$0.837_{\pm 0.039}$	$0.830_{\pm 0.026}$	$0.706_{\pm 0.042}$	$0.686_{\pm 0.046}$	$0.872_{\pm 0.015}$	$0.931_{\pm 0.007}$	$0.824_{\pm 0.016}$	$0.820_{\pm 0.016}$	$0.886_{\pm 0.025}$	$0.951_{\pm 0.007}$	$0.871_{\pm 0.024}$	$0.863_{\pm 0.02}$
MLAN	$0.703_{\pm 0.000}$	$0.717_{\pm 0.000}$	$0.547_{\pm 0.000}$	$0.515_{\pm 0.000}$	$0.727_{\pm 0.000}$	$0.838_{\pm 0.000}$	$0.509_{\pm 0.000}$	$0.494_{\pm 0.000}$	$0.775_{\pm 0.000}$	$0.855_{\pm 0.000}$	$0.740_{\pm 0.000}$	$0.726_{\pm 0.00}$
GMC	$0.654_{\pm 0.000}$	$0.689_{\pm 0.000}$	$0.480_{\pm 0.000}$	$0.441_{\pm 0.000}$	$0.632_{\pm 0.000}$	$0.857_{\pm 0.000}$	$0.359_{\pm 0.000}$	$0.336_{\pm 0.000}$	$0.791_{\pm 0.000}$	$0.940_{\pm 0.000}$	$0.794_{\pm 0.000}$	$0.781_{\pm 0.00}$
WTNM	$0.957_{\pm 0.034}$	$0.964_{\pm 0.017}$	$0.936_{\pm 0.034}$	$0.932_{\pm 0.036}$	$0.977_{\pm 0.015}$	$0.993_{\pm 0.004}$	$0.977_{\pm 0.015}$	$0.976_{\pm 0.015}$	$0.816_{\pm 0.001}$	$0.903_{\pm 0.000}$	$0.812_{\pm 0.000}$	$0.802_{\pm 0.00}$
tSMC	$0.913_{\pm 0.044}$	$0.919_{\pm 0.027}$	$0.857_{\pm 0.050}$	$0.848_{\pm 0.053}$	$0.974_{\pm 0.013}$	$0.992_{\pm 0.003}$	$0.974_{\pm 0.013}$	$0.973_{\pm 0.013}$	$0.825_{\pm 0.012}$	$0.902_{\pm 0.003}$	$0.817_{\pm 0.009}$	$0.808_{\pm 0.01}$
ETMC	$0.629_{\pm 0.018}$	$0.668_{\pm 0.015}$	$0.504_{\pm 0.020}$	$0.470_{\pm 0.022}$	$0.717_{\pm 0.011}$	$0.857_{\pm 0.005}$	$0.640_{\pm 0.013}$	$0.631_{\pm 0.013}$	$0.861_{\pm 0.017}$	$0.927_{\pm 0.008}$	$0.851_{\pm 0.015}$	$0.843_{\pm 0.01}$
UGLTL	$0.993_{\pm 0.000}$	$0.992_{\pm 0.000}$	$0.987_{\pm 0.000}$	$0.986_{\pm 0.000}$	$0.967_{\pm 0.000}$	$0.989_{\pm 0.000}$	$0.960_{\pm 0.000}$	$0.959_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.00}$
LTBPL	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.00}$

表 3　对比结果：在 UCI、Caltech-101 和 Notting-Hill 数据集上通过不同方法得到的均值和标准差

Datasets	UCI				Caltech-101				Notting-Hill			
Method	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI
SC 1	$0.617_{\pm 0.011}$	$0.585_{\pm 0.007}$	$0.506_{\pm 0.010}$	$0.451_{\pm 0.011}$	$0.356_{\pm 0.001}$	$0.165_{\pm 0.001}$	$0.313_{\pm 0.001}$	$0.634_{\pm 0.000}$	$0.694_{\pm 0.000}$	$0.663_{\pm 0.000}$	$0.673_{\pm 0.000}$	$0.856_{\pm 0.000}$
SC 2	$0.684_{\pm 0.000}$	$0.587_{\pm 0.001}$	$0.554_{\pm 0.000}$	$0.504_{\pm 0.001}$	$0.404_{\pm 0.002}$	$0.271_{\pm 0.001}$	$0.438_{\pm 0.000}$	$0.691_{\pm 0.000}$	$0.844_{\pm 0.000}$	$0.670_{\pm 0.000}$	$0.772_{\pm 0.000}$	$0.901_{\pm 0.000}$
SC 3	$0.546_{\pm 0.005}$	$0.489_{\pm 0.001}$	$0.427_{\pm 0.001}$	$0.362_{\pm 0.001}$	$0.328_{\pm 0.002}$	$0.217_{\pm 0.001}$	$0.334_{\pm 0.001}$	$0.644_{\pm 0.000}$	$0.740_{\pm 0.000}$	$0.623_{\pm 0.000}$	$0.673_{\pm 0.000}$	$0.859_{\pm 0.000}$
SC 4	—	—	—	—	$0.379_{\pm 0.014}$	$0.385_{\pm 0.010}$	$0.435_{\pm 0.010}$	$0.699_{\pm 0.005}$	—	—	—	—
SC 5	—	—	—	—	$0.355_{\pm 0.014}$	$0.307_{\pm 0.005}$	$0.396_{\pm 0.018}$	$0.677_{\pm 0.008}$	—	—	—	—
SC 6	—	—	—	—	$0.481_{\pm 0.001}$	$0.343_{\pm 0.001}$	$0.460_{\pm 0.001}$	$0.705_{\pm 0.000}$	—	—	—	—
CoTr	$0.840_{\pm 0.014}$	$0.796_{\pm 0.005}$	$0.779_{\pm 0.009}$	$0.754_{\pm 0.010}$	$0.437_{\pm 0.003}$	$0.423_{\pm 0.005}$	$0.473_{\pm 0.005}$	$0.326_{\pm 0.006}$	$0.843_{\pm 0.010}$	$0.781_{\pm 0.005}$	$0.823_{\pm 0.006}$	$0.773_{\pm 0.008}$
RMSC	$0.859_{\pm 0.018}$	$0.822_{\pm 0.009}$	$0.800_{\pm 0.014}$	$0.777_{\pm 0.016}$	$0.460_{\pm 0.000}$	$0.394_{\pm 0.000}$	$0.483_{\pm 0.000}$	$0.332_{\pm 0.000}$	$0.827_{\pm 0.000}$	$0.772_{\pm 0.000}$	$0.822_{\pm 0.000}$	$0.772_{\pm 0.000}$
CSMSC	$0.882_{\pm 0.000}$	$0.787_{\pm 0.001}$	$0.784_{\pm 0.001}$	$0.760_{\pm 0.001}$	$0.630_{\pm 0.010}$	$0.534_{\pm 0.021}$	$0.632_{\pm 0.017}$	$0.494_{\pm 0.018}$	$0.873_{\pm 0.000}$	$0.760_{\pm 0.000}$	$0.795_{\pm 0.000}$	$0.736_{\pm 0.000}$
LTMSC	$0.800_{\pm 0.006}$	$0.768_{\pm 0.007}$	$0.748_{\pm 0.009}$	$0.720_{\pm 0.010}$	$0.602_{\pm 0.000}$	$0.547_{\pm 0.000}$	$0.613_{\pm 0.000}$	$0.475_{\pm 0.000}$	$0.868_{\pm 0.000}$	$0.779_{\pm 0.000}$	$0.825_{\pm 0.000}$	$0.777_{\pm 0.000}$
LMSC	$0.856_{\pm 0.037}$	$0.783_{\pm 0.027}$	$0.762_{\pm 0.039}$	$0.736_{\pm 0.044}$	$0.564_{\pm 0.029}$	$0.448_{\pm 0.020}$	$0.564_{\pm 0.025}$	$0.410_{\pm 0.026}$	$0.913_{\pm 0.051}$	$0.833_{\pm 0.058}$	$0.866_{\pm 0.078}$	$0.829_{\pm 0.099}$
MCIAS	$0.976_{\pm 0.001}$	$0.946_{\pm 0.003}$	$0.953_{\pm 0.002}$	$0.948_{\pm 0.003}$	$0.742_{\pm 0.048}$	$0.535_{\pm 0.038}$	$0.735_{\pm 0.054}$	$0.591_{\pm 0.072}$	$0.523_{\pm 0.066}$	$0.362_{\pm 0.078}$	$0.460_{\pm 0.060}$	$0.294_{\pm 0.082}$
MLAN	$0.968_{\pm 0.000}$	$0.925_{\pm 0.000}$	$0.937_{\pm 0.000}$	$0.930_{\pm 0.000}$	$0.627_{\pm 0.004}$	$0.544_{\pm 0.003}$	$0.618_{\pm 0.000}$	$0.428_{\pm 0.003}$	$0.365_{\pm 0.000}$	$0.114_{\pm 0.000}$	$0.376_{\pm 0.000}$	$0.044_{\pm 0.000}$
GMC	$0.735_{\pm 0.000}$	$0.815_{\pm 0.000}$	$0.713_{\pm 0.000}$	$0.677_{\pm 0.000}$	$0.692_{\pm 0.000}$	$0.659_{\pm 0.000}$	$0.721_{\pm 0.000}$	$0.594_{\pm 0.000}$	$0.312_{\pm 0.000}$	$0.092_{\pm 0.000}$	$0.369_{\pm 0.000}$	$0.022_{\pm 0.000}$
WTNM	$0.996_{\pm 0.000}$	$0.990_{\pm 0.000}$	$0.993_{\pm 0.000}$	$0.992_{\pm 0.000}$	$0.685_{\pm 0.000}$	$0.668_{\pm 0.000}$	$0.702_{\pm 0.000}$	$0.587_{\pm 0.000}$	$0.983_{\pm 0.000}$	$0.956_{\pm 0.000}$	$0.975_{\pm 0.000}$	$0.968_{\pm 0.000}$
tSMC	$0.996_{\pm 0.000}$	$0.989_{\pm 0.000}$	$0.992_{\pm 0.000}$	$0.991_{\pm 0.000}$	$0.746_{\pm 0.000}$	$0.724_{\pm 0.002}$	$0.758_{\pm 0.001}$	$0.656_{\pm 0.001}$	$0.956_{\pm 0.000}$	$0.890_{\pm 0.000}$	$0.917_{\pm 0.000}$	$0.895_{\pm 0.000}$
ETMC	$0.933_{\pm 0.015}$	$0.961_{\pm 0.007}$	$0.939_{\pm 0.018}$	$0.932_{\pm 0.014}$	$0.514_{\pm 0.010}$	$0.535_{\pm 0.005}$	$0.559_{\pm 0.006}$	$0.425_{\pm 0.007}$	$0.951_{\pm 0.000}$	$0.911_{\pm 0.000}$	$0.924_{\pm 0.000}$	$0.898_{\pm 0.000}$
UGLTL	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$0.383_{\pm 0.000}$	$0.621_{\pm 0.000}$	$0.499_{\pm 0.000}$	$0.355_{\pm 0.000}$	$0.950_{\pm 0.000}$	$0.921_{\pm 0.000}$	$0.924_{\pm 0.000}$	$0.903_{\pm 0.000}$
LTBPL	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$0.945_{\pm 0.000}$	$0.894_{\pm 0.000}$	$0.953_{\pm 0.000}$	$0.924_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$

表 4　对比结果：在 Hdigit、BBCSport 和 BBC4view 数据集上通过不同方法得到的均值和标准差

Datasets	Hdigit				BBCSport				BBC4view			
Method	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI
SC 1	$0.526_{\pm 0.000}$	$0.468_{\pm 0.001}$	$0.412_{\pm 0.000}$	$0.345_{\pm 0.000}$	$0.846_{\pm 0.000}$	$0.672_{\pm 0.001}$	$0.760_{\pm 0.001}$	$0.688_{\pm 0.001}$	$0.581_{\pm 0.001}$	$0.413_{\pm 0.001}$	$0.504_{\pm 0.002}$	$0.357_{\pm 0.003}$
SC 2	$0.485_{\pm 0.009}$	$0.455_{\pm 0.007}$	$0.391_{\pm 0.008}$	$0.322_{\pm 0.008}$	$0.509_{\pm 0.001}$	$0.229_{\pm 0.002}$	$0.416_{\pm 0.000}$	$0.164_{\pm 0.001}$	$0.777_{\pm 0.000}$	$0.542_{\pm 0.001}$	$0.652_{\pm 0.000}$	$0.550_{\pm 0.001}$
SC 3	—	—	—	—	—	—	—	—	$0.636_{\pm 0.004}$	$0.439_{\pm 0.001}$	$0.523_{\pm 0.002}$	$0.384_{\pm 0.003}$
SC 4	—	—	—	—	—	—	—	—	$0.725_{\pm 0.000}$	$0.500_{\pm 0.000}$	$0.580_{\pm 0.000}$	$0.460_{\pm 0.000}$
CoTr	$0.910_{\pm 0.004}$	$0.826_{\pm 0.002}$	$0.845_{\pm 0.003}$	$0.828_{\pm 0.003}$	$0.902_{\pm 0.003}$	$0.809_{\pm 0.005}$	$0.875_{\pm 0.003}$	$0.837_{\pm 0.004}$	$0.543_{\pm 0.001}$	$0.386_{\pm 0.002}$	$0.464_{\pm 0.005}$	$0.272_{\pm 0.002}$
RMSC	$0.729_{\pm 0.008}$	$0.672_{\pm 0.010}$	$0.639_{\pm 0.010}$	$0.598_{\pm 0.011}$	$0.851_{\pm 0.028}$	$0.801_{\pm 0.015}$	$0.848_{\pm 0.020}$	$0.801_{\pm 0.027}$	$0.708_{\pm 0.019}$	$0.533_{\pm 0.005}$	$0.590_{\pm 0.009}$	$0.470_{\pm 0.011}$
CSMSC	$0.834_{\pm 0.000}$	$0.738_{\pm 0.000}$	$0.731_{\pm 0.000}$	$0.701_{\pm 0.000}$	$0.955_{\pm 0.000}$	$0.861_{\pm 0.000}$	$0.914_{\pm 0.000}$	$0.888_{\pm 0.000}$	$0.919_{\pm 0.000}$	$0.775_{\pm 0.000}$	$0.861_{\pm 0.000}$	$0.819_{\pm 0.000}$
LTMSC	$0.782_{\pm 0.000}$	$0.661_{\pm 0.000}$	$0.649_{\pm 0.000}$	$0.609_{\pm 0.000}$	$0.943_{\pm 0.000}$	$0.839_{\pm 0.000}$	$0.907_{\pm 0.000}$	$0.878_{\pm 0.000}$	$0.927_{\pm 0.000}$	$0.796_{\pm 0.000}$	$0.873_{\pm 0.000}$	$0.834_{\pm 0.000}$
LMSC	$0.802_{\pm 0.000}$	$0.796_{\pm 0.000}$	$0.758_{\pm 0.000}$	$0.730_{\pm 0.000}$	$0.920_{\pm 0.002}$	$0.839_{\pm 0.005}$	$0.901_{\pm 0.004}$	$0.870_{\pm 0.005}$	$0.874_{\pm 0.007}$	$0.680_{\pm 0.014}$	$0.790_{\pm 0.011}$	$0.726_{\pm 0.014}$
MCIAS	$0.728_{\pm 0.001}$	$0.831_{\pm 0.004}$	$0.763_{\pm 0.005}$	$0.734_{\pm 0.006}$	$0.891_{\pm 0.046}$	$0.818_{\pm 0.089}$	$0.883_{\pm 0.030}$	$0.846_{\pm 0.040}$	$0.860_{\pm 0.089}$	$0.705_{\pm 0.080}$	$0.800_{\pm 0.084}$	$0.739_{\pm 0.050}$
MLAN	$0.710_{\pm 0.000}$	$0.837_{\pm 0.000}$	$0.762_{\pm 0.000}$	$0.731_{\pm 0.000}$	$0.977_{\pm 0.000}$	$0.923_{\pm 0.000}$	$0.953_{\pm 0.000}$	$0.938_{\pm 0.000}$	$0.871_{\pm 0.000}$	$0.700_{\pm 0.000}$	$0.810_{\pm 0.000}$	$0.746_{\pm 0.000}$
GMC	$0.998_{\pm 0.000}$	$0.993_{\pm 0.000}$	$0.996_{\pm 0.000}$	$0.995_{\pm 0.000}$	$0.739_{\pm 0.000}$	$0.795_{\pm 0.000}$	$0.720_{\pm 0.000}$	$0.600_{\pm 0.000}$	$0.693_{\pm 0.000}$	$0.562_{\pm 0.000}$	$0.633_{\pm 0.000}$	$0.478_{\pm 0.000}$
WTNM	$0.998_{\pm 0.000}$	$0.996_{\pm 0.000}$	$0.997_{\pm 0.000}$	$0.997_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$0.995_{\pm 0.000}$	$0.983_{\pm 0.000}$	$0.993_{\pm 0.000}$	$0.991_{\pm 0.000}$
tSMC	$0.997_{\pm 0.000}$	$0.991_{\pm 0.000}$	$0.994_{\pm 0.000}$	$0.993_{\pm 0.000}$	$0.998_{\pm 0.000}$	$0.992_{\pm 0.000}$	$0.997_{\pm 0.000}$	$0.996_{\pm 0.000}$	$0.994_{\pm 0.000}$	$0.977_{\pm 0.000}$	$0.990_{\pm 0.000}$	$0.987_{\pm 0.000}$
ETMC	$0.917_{\pm 0.018}$	$0.962_{\pm 0.008}$	$0.932_{\pm 0.014}$	$0.924_{\pm 0.016}$	$0.964_{\pm 0.027}$	$0.976_{\pm 0.019}$	$0.972_{\pm 0.028}$	$0.963_{\pm 0.030}$	$0.906_{\pm 0.048}$	$0.899_{\pm 0.019}$	$0.911_{\pm 0.032}$	$0.884_{\pm 0.042}$
UGLTL	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$0.705_{\pm 0.000}$	$0.839_{\pm 0.000}$	$0.783_{\pm 0.000}$	$0.723_{\pm 0.000}$	$0.845_{\pm 0.000}$	$0.754_{\pm 0.000}$	$0.748_{\pm 0.000}$	$0.678_{\pm 0.000}$
LTBPL	$0.999_{\pm 0.000}$	$0.999_{\pm 0.000}$	$0.999_{\pm 0.000}$	$0.999_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$

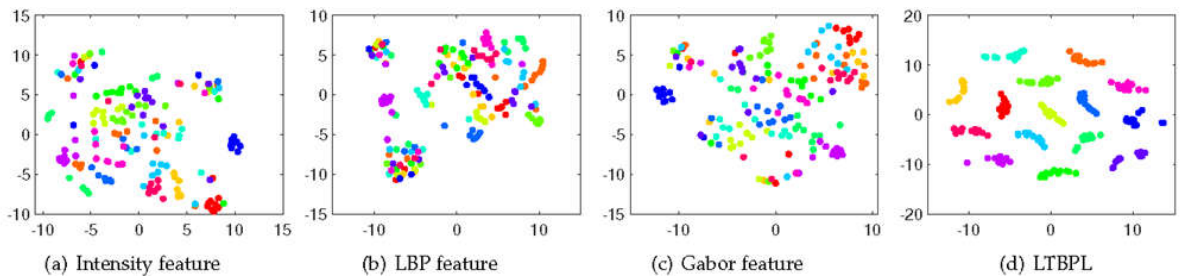

图 2　Yale 数据集上的可视化

表 5　消融实验：LTBPL 及其变体在 NMI 方面的比较结果

Variants	Yale	ORL	COIL-20	UCI	Caltech-101	Notting-Hill	Hdigit	BBCSport	BBC4view	Average
LTBPL	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$0.894_{\pm 0.000}$	$1.000_{\pm 0.000}$	$0.999_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	0.988
LTBPL-t1	$0.642_{\pm 0.000}$	$0.801_{\pm 0.000}$	$0.974_{\pm 0.000}$	$0.821_{\pm 0.000}$	$0.619_{\pm 0.000}$	$0.623_{\pm 0.000}$	$0.991_{\pm 0.000}$	$0.825_{\pm 0.000}$	$0.639_{\pm 0.000}$	0.770
LTBPL-t2	$0.916_{\pm 0.022}$	$0.987_{\pm 0.006}$	$0.967_{\pm 0.009}$	$0.998_{\pm 0.000}$	$0.664_{\pm 0.001}$	$0.970_{\pm 0.000}$	$0.995_{\pm 0.000}$	$1.000_{\pm 0.000}$	$0.988_{\pm 0.000}$	0.942

的先验知识通过赋予不同权重被显示地考虑。同时，视图对应的相似性表征和反映最终聚类的共识聚类指示矩阵能够通过不同的自适应置信度关联起来。在九个真实世界数据集的广泛实验表明了和最先进的聚类方法相比，所提出的 LTBPL 方法有明显的优越性。对于未来的工作，张量学习将被用于解决不完整多视图聚类里面具有挑战性的问题，其中视图中的样本或特征可能存在缺失。相关工作请参考中山大学团队在 IEEE TKDE 2022 等的论文[1, 19, 20]。

责任编委　崔海楠　王金甲

参考文献

[1] M.-S. Chen, C.-D. Wang, and J.-H. Lai, “Low-rank tensor based proximity learning for multi-view clustering.” IEEE Trans. Knowl. Data Eng.,2022.

[2] C. Tang, X. Zhu, X. Liu, M. Li, P. Wang, C. Zhang, and L. Wang, “Learning a joint affinity graph for multiview subspace clustering,” IEEE Trans. Multimedia, vol. 21, no. 7, pp. 1724–1736, 2019.

[3] H. Wang, Y. Yang, and B. Liu, “GMC: graph-based multi-view clustering,” IEEE Trans. Knowl. Data Eng., vol. 32, no. 6, pp. 1116–1129, 2020.

[4] A. Kumar, P. Rai, and H. Daum´e III, “Co-regularized multi-view spectral clustering,” in NIPS, 2011, pp. 1413–1421.

[5] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, “Exclusivity-consistency regularized multi-view subspace clustering,” in CVPR, 2017, pp. 1–9.

[6] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, and D. Xu, “Generalized latent multi-view subspace clustering,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 1, pp. 86–99, 2020.

[7] X. He and P. Niyogi, “Locality preserving projections,” in NIPS, 2003, pp. 153–160.

[8] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” Science, vol. 290, no. 5500, pp. 2323–2326, 2000.

[9] Z. Zhao, L. Wang, H. Liu, and J. Ye, “On similarity preserving feature selection,” IEEE Trans. Knowl. Data Eng., vol. 25, no. 3, pp.619–632, 2013.

[10] R. Vidal, “Subspace clustering,” IEEE Signal Processing Magazine, vol. 28, no. 2, pp. 52–68, 2011.

[11] Y. Xie, D. Tao,W. Zhang, Y. Liu, L. Zhang, and Y. Qu, “On unifying multi-view self-representations for clustering by tensor multi-rank minimization,” Int. J. Comput. Vis., vol. 126, no. 11, pp. 1157–1179, 2018.

[12] Q. Gao, W. Xia, Z. Wan, D. Xie, and P. Zhang, “Tensor-svd basedgraph learning for multi-view subspace clustering,” in AAAI, 2020, pp. 3930–3937.

[13] F. Nie, J. Li, and X. Li, “Self-weighted multiview clustering with multiple graphs,” in IJCAI, 2017, pp. 2564–2570.

[14] F. Nie, G. Cai, and X. Li, “Multi-view clustering and semi-supervised classification with adaptive neighbours,” in AAAI, 2017, pp. 2408-2414.

[15] K. Fan, “On a theorem of weyl concerning eigenvalues of linear transformations i,” Proceedings of the National Academy of Sciences, vol. 35, no. 11, pp. 652–655, 1949.

[16] J. Wu, X. Xie, L. Nie, Z. Lin, and H. Zha, “Unified graph and low-rank tensor learning for multi-view clustering,” in AAAI, 2020, pp. 6388–6395.

[17] J.Wu, Z. Lin, and H. Zha, “Essential tensor learning for multi-view spectral clustering,” IEEE Trans. Image Process., vol. 28, no. 12, pp. 5910–5922, 2019.

[18] X.- L. Li, M.-S. Chen, and C.-D. Wang, et al. Refining graph structure for incomplete multi-view clustering. IEEE Transactions on Neural Networks and Learning Systems, 2022.

[19] M.-S. Chen, J.-Q. Lin, X.-L. Li, B.-Y. Liu, C.-D. Wang, D. H, and J.-H. L, “Representation learning in multi-view clustering: A literature review,” Data Science and Engineering, 2022:1-17.

陈曼笙

中山大学计算机学院 2022 级博士研究生，导师为王昌栋副教授。在国际期刊和会议上发表了八篇论文，包括 IEEE TKDE、IEEE TCYB、IEEE TNNLS、Information Fusion、KDD、ACM MM、AAAI 和 DASFAA。主要研究方向是多视图聚类。

Email:chenmsh27@mail2.sysu.edu.cn

王昌栋

中山大学计算机学院副教授，博士生导师，师从中山大学赖剑煌教授和美国伊利诺大学-芝加哥校区 IEEE Fellow Philip S. Yu 教授。研究方向包括数据聚类、网络分析、推荐算法和大数据信息安全。以第一作者身份或者指导学生发表了 100 余篇 CCF B 类或中科院分区表 SCI 二区以上的学术论文，其中 IEEE/ACM Trans 超过 40 篇，A 类或一区论文 50 余篇。主持了包括广东省自然科学基金-杰出青年基金、广东特支计划“科技创新青年拔尖人才”、国家重点研发计划项目-子课题、国家自然科学基金-面上项目、CCF-腾讯犀牛鸟科研基金等 13 个项目。任人工智能权威期刊 JAIR 的副编辑。

Email: changdongwang@hotmail.com

赖剑煌

中山大学计算机学院教授、博士生导师。广东省信息安全技术重点实验室主任，视频图像智能分析与应用公安部重点实验室学术委员会主任。中国图象图形学学会副理事长、会士，自动化学报副主编，中国计算机学会杰出会员，中国计算机学会计算机视觉专业组副主任(第一、二届)。广东省图像图形学会理事长(第四、五届)，广东省人工智能与机器人学会副理事长、广东省安防协会人工智能专委会主任。IEEE 高级会员。已主持承担国家自然科学基金与广东联合重点项目，科技部科技支撑课题，国家自然科学基金、广东省前沿与关键技术创新专项等多项，获得广东省自然科学一等奖(2018 年)、中国图象图形学学会自然科学一等奖(2020 年)、广东省自然科学二等奖(2020 年)、广东省科学技术奖励二等奖(2016 年)等。已发表了 200 多篇学术论文，主要发表在 IEEE TPAMI、IJCV 、IEEE TIP、IEEE TNN、IEEE TCSVT、IEEE TSMC(Part B)、Pattern Recognition 等国际权威刊物以及 ICCV、 CVPR、 ICDM 等专业重要学术会议上。拥有 30 多项国家发明专利。

Email: stsljh@mail.sysu.edu.cn

热点追踪

噪声关联学习

四川大学 杨谋星 林义杰 黄振宇 彭玺

一、引言

深度神经网络的成功依赖于大规模且高质量的标记数据。作为标签的一种重要形式，数据点间的关联/对齐关系在跨模态检索、视觉定位、图像自动描述、目标重识别、图匹配、机器阅读、对比学习等应用和学习范式中至关重要。

在实际场景中，为获得关联的成对数据，通常采用人工标注或者互联网爬取的方式。然而，由于数据繁杂和人力资源受限，关联的准确性往往难以保证，数据点间不可避免地存在噪声关联(Noisy Correspondence)[1,2]。如图 1 所示，噪声关联的样本对通常分为两类，一类是假阳性样本对(False Positive Pairs，FP)，即本来没有关联的数据对被错误当做有关联的正样本对[1]，如生活中常见的图文不符、音画不同步、答非所问现象；另一类是假阴性样本对（False Negative Pair，FN)，即描述相同或相关目标的数据点被错误当做没有关联的负样本对[2]。

需要注意的是，噪声关联可以认为是噪声标签（Noisy Label）学习领域的一种新范式，其与注重于分类任务中错误类别标签的工作有着显著区别：i) 噪声关联指样本间的相关性/关系可能存在错误，而噪声标签主要强调样本的所属类别可能出错。从该角度出发，大部分需要成对样本作为输入的任务和应用都可能存在噪声关联，而噪声标签主要局限于传统的分类任务；ii) 噪声关联并不是非正即负，对于给定样本对，他们之间的关联其实是[0,1]之间的连续值。相比之下，在分类任务中，噪声标签的样本通常属于某个特定的类。综上，噪声关联学习丰富了噪声标签学习范式的内涵，并扩展了其外延。

图 1　噪声关联示意图，其中所有样本均选自多模态 Conceptual Captions 数据集 [3]。上图：由于该数据集是从互联网自动爬取得到的图文数据对，正样本对中不可避免地存在部分假阳性样本对。神经网络在优化过程中将拟合假阳性样本对而得到错误决策边界，最终导致性能下降。下图：给定锚点样本，除了其关联的正样本对，负样本集合中存在一些潜在相关的假阴性样本，它们与锚点可能存在完全对应、抽象对应和部分对应等情形。错误地将这些假阴性样本对当作负样本对不仅将失去关联样本对的多样性，还将导致模型被错误优化。

从 2021 年开始，一些学者认识到了噪声关联问题的重要性并开展了一系列研究。[2]最早意识到噪声关联

问题的重要性，并提出了对假阴性鲁棒的对比学习损失函数；受此启发，[1]正式揭示和定义噪声关联问题，并以跨模态匹配为验证场景，提出了对假阳性样本对鲁棒的跨模态检索方法；[4-10]进一步深入挖掘噪声关联在目标重识别、图匹配等应用场景下的特殊性，并设计了场景定制化的解决方案。在后续章节，我们将简要介绍这些工作。具体地，第二节将给出噪声关联的形式化定义，包含假阴性和假阳性关联；第三节将介绍噪声关联学习在不同场景下的研究现状；第四节将总结全文，并给出噪声关联学习研究的未来展望。

二、噪声关联学习

本节介绍噪声关联的形式化定义，主要包含对假阳性和假阴性关联的定义。

2.1 假阳性关联（False Positive）

给定正样本对集合$\{(x_i^{m_1}, x_i^{m_2}), c_i\}_{i=1}^N$，其中$(x_i^{m_1}, x_i^{m_2})$代表第$i$个正样本对，其通过以下两种方式获得：i）人工标注或互联网爬取的成对数据；ii）通过类别标签构建，即同类的样本作为正样本对。理想情况下，关联$c_i = 1$，即$x_i^{m_1}$和$x_i^{m_2}$都描述同一个或同类的目标。然而，实际应用中将不可避免得到一些假阳性样本对，它们实为无关联或弱关联的样本对，但关联c_i被错误标记为1。以常见的图文数据对为例，经常会出现字幕冗余、字幕欠完备，甚至是图文完全不符等情况，这些都将导致假阳性关联。需要注意的是，$(x_i^{m_1}, x_i^{m_2})$呈现的形式多样。例如，在跨模态检索任务中，x_i代表图像、文本等实例；在视觉定位中，x_i代表目标框、或单词等细粒度对象；在图匹配中，x_i代表图像块。同理m_1和m_2可以相等也可以不等，意味着数据可能来自同一或不同模态。例如，在跨模态任务中，$m_1 \neq m_2$；在单模态任务如行人重识别中，$m_1 = m_2$。特别的，在视频表征学习或者图像匹配等任务中，每个样本对可能包含 3 个甚至更多的样本，例如视频包含了图像、文本、音频等。综上，假阳性关联广泛存在于不同应用中，且在不同应用下可能存在不同的定制化解决方案。

2.2 假阴性关联（False Negative）

给定所构建的负样本对集合$\{(x_i^{m_1}, x_j^{m_2}), c_{ij} \mid i \neq j\}$，其中负样本对$(x_i^{m_1}, x_j^{m_2})$有以下两种来源：i）通过类别标签构建，即非同类的样本作为负样本对；ii）数据集或同批次（Batch）内随机采样获得。理想情况下，$x_i^{m_1}$和$x_j^{m_2}$将是对于不同类或不同目标的描绘，它们之间没有相关性，因而关联c_{ij}被记为 0。然而，实际应用中，上述两种构建方式均可能引入假阴性样本对。具体地，当样本的类别标签存在错误时，方式 1 不可避免地会将本属于同一类的样本错误作为负样本对。同理，方式 2 也可能将同批次内语义相似的样本对错误作为负样本对，特别是目前主流的大批次对比学习将引入更多假阴性样本对。

三、不同应用下的噪声关联学习

3.1 跨模态检索

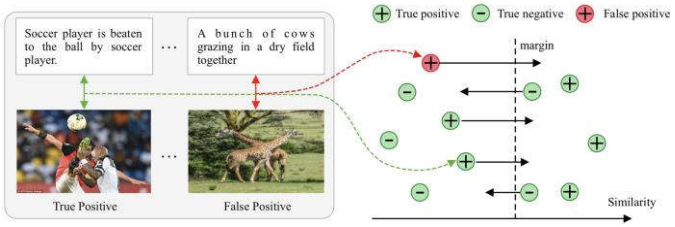

图 2　跨模态检索中的噪声关联。真阳性样本对正确地指导了跨模态匹配，但假阳性样本对则对训练的进了错误的监督。

跨模态检索大多依赖于正确匹配的跨模态数据，从而学习到一个可以衡量跨模态相似性的匹配模型。然而，在数据收集和标记过程中，常常引入噪声关联。如图 2 所示，给定的训练数据中图片和文本描述可能是错误匹配即假阳性关联的，这无疑会影响后续的跨模态匹配任务。为解决假阳性关联问题，[1]提出了基于神经网络记忆效应的噪声鉴别矫正方法，探明了神经网络对数据关联的拟合演化规律。该方法可自适应地识别噪声关联数据并对其关联进行矫正，结合所设计的鲁棒多模态匹配目标函数，最终实现假阳性关联鲁棒的多模态检索。

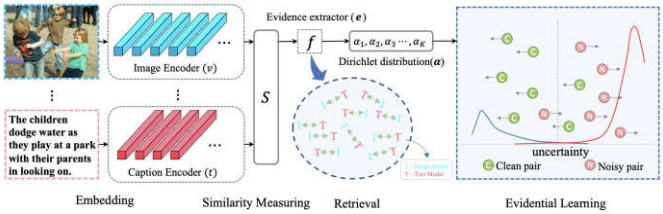

图 3　方案[4]将证据理论应用于跨模态检索中，可估计样本对的不确定度。

[4]将证据学习应用于跨模态检索学习，考虑了噪声关联在跨模态中的不确定性估计问题和难负样本选择问题，提出了一个广义的深度证据跨模态学习框架。该方案能以有效和高效的方式提供可信的检索，同时可直接应用于现有的跨模态检索方法以增强鲁棒性。

表 1　不同方法在 Conceptual Captions [3]子集上多模态检索召回率。

方法	Image → Text			Text → Image		
	R@1	R@5	R@10	R@1	R@5	R@10
SCAN[11]	30.5	55.3	65.3	26.9	53.0	64.7
SAF[12]	31.7	59.3	68.2	31.9	59.0	67.9
SGR[12]	11.3	29.7	39.6	13.1	30.1	41.6
NCR[1]	**39.5**	64.5	73.5	40.3	64.6	73.2
DECL[4]	39.0	**66.1**	**75.5**	**40.7**	**66.3**	**76.7**

如表 1 所示，处理噪声关联的方法[1,4]在互联网爬取的噪声多模态数据集上，取得了显著的检索性能提升，充分说明解决噪声关联在实际应用下具备重要意义。

3.2 多视图表示学习

[2]观察到了对比学习范式的随机负样本选择策略将不可避免地将相同语义、属于同类的样本对错误作为负样本对，即引入假阴性，最终导致模型性能退化。为此，该工作设计了一种假阴性鲁棒的对比损失函数，该损失函数具有良好的性质及相应数学证明，能够缓解或甚至避免错误拟合假阴性样本对。为验证该损失的有效性，[2]以多视图表示学习为应用场景，验证了所提出鲁棒对比学习的有效性，如图 4(b)所示。

3.3 跨模态行人重识别

给定一张可见光/红外光模态的行人照片，跨模态

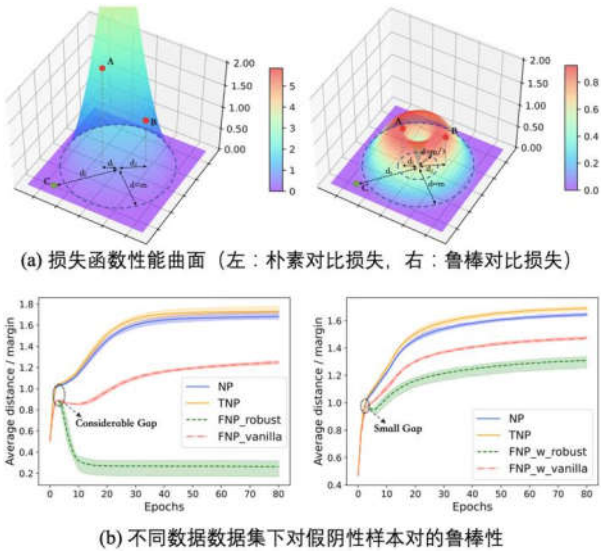

图 4　相比朴素对比损失函数（a 图左），[2]提出的鲁棒对比损失具备非单调优化性质，能够缓解或甚至避免错误拟合假阴性样本对。如 b 图所示，使用朴素对比损失将把假阴性样本当作负样本对，错误地增加样本对距离；使用鲁棒对比损失将缓解假阴性样本对距离的错误增加，甚至把假阴性样本对正确地作为正样本对优化。

行人重识别(VI-ReID)旨在从另一模态中匹配出同一行人的其他相片。目前主流的 VI-ReID 方案，需要在各模态进行判别性学习，同时依赖身份标注构建跨模态正负样本对，进一步执行跨模态相似性学习以提升性能。因此，现有 VI-ReID 范式严重依赖于身份标注的准确性。然而，监控系统中图像可识别性差，特别是丢失行人颜色信息的红外模态，精确标注所有行人的身份是不现实的。错误的行人身份标注将产生类别级的噪声标签，并进一步导致假阳性和假阴性噪声关联。

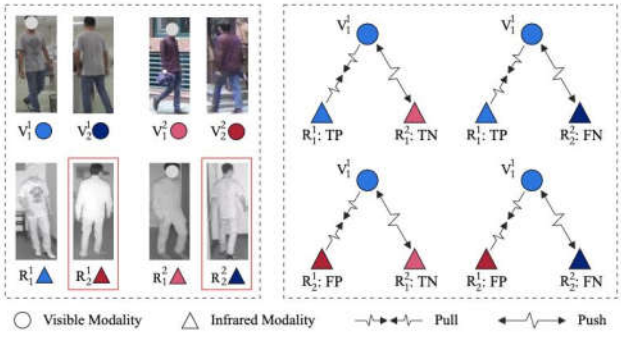

图 5　[5]指出 VI-ReID 任务会同时面临噪声标注（左）与噪声关联（右）问题。

论文[5]同时考虑了噪声标签与噪声关联问题，揭示了两种噪声之间的“孪生”关系，并针对性地提出了鲁棒的 VI-ReID 方案。其首先利用神经网络的记忆效应来估计身份标注的置信度；随后，基于置信度将跨模态正、负样本对分为不同子集并进一步校正其中的关联，最后利用所设计的双重鲁棒损失函数来实现对孪生噪声标签鲁棒的跨模态行人重识别。

3.4 图匹配

图匹配旨在寻找两个或多个存在复杂关系的集合间元素的对应关系。其最广为人知的应用场景是图像特征点的匹配，利用不同图像上匹配好的特征点，可以实现三维重建、目标追踪、运动结构理解等问题。类似分类问题中类别标签标注，图像中的关键点也是人工标注的。现实中的图像往往可见度低、图像间视角差异大，从而导致人工标注的关键点偏移、甚至错误，如图 6 所示。错误的关键点标注会导致图像间的特征点关联不正确，即噪声关联问题。不同于跨模态检索等应用中的噪声关联，图匹配任务会根据关键点构建一张图结构，并同时对齐两张或多张图中的点和边，此时噪声关联问题会同时存在于点与边的对应中。强迫图匹配网络拟合这种噪声，会显著降低其性能。

图 6　图匹配中的噪声关联问题示意，其中绿线表示正确关联，红线表示错误关联。由于可见度低、视角差异大，关键点的标注存在错误，导致噪声关联。

为同时解决噪声情况下点与边的匹配问题，论文[6]提出了一个鲁棒的图匹配损失函数，其将点的对齐与边的对齐形式化为一个二阶噪声关联问题。进一步地，该方法构造了一个动量网络以预测出点与点、边与边之间的匹配置信度，再按照置信度加权给匹配损失，实现了噪声关联下鲁棒的图匹配。

3.5 会话式机器阅读

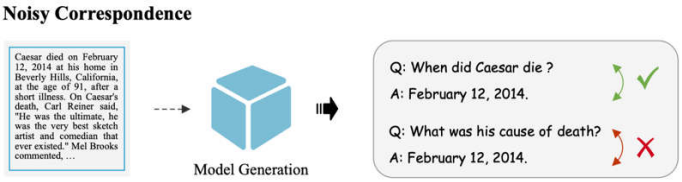

图 7　噪声关联数据可以是模型自身产生的两个数据点之间的不对齐。

会话式机器阅读理解（Machine Reading Comprehension，MRC）高度依赖于与给定文档相关的回答对，而对于无标签数据，通常会借助一个个预训练好的问题-答案生成器去自动生成伪标签数据，再去 finetune 一个预训练好的 MRC 模型。如图 7 所示，这样的方式不可避免的会遇到噪声关联问题，即构造的伪问答对是错误关联的。这显然极大损害了后续模型优化。

为此，[7]提出鲁棒的机器阅读理解域迁移方法，可以有效的缓解噪声问答对的产生。该方法包含了 QA Construction Model 和 MRC Model，前者用于在目标域构造问答对来 finetune 后者。为了解决噪声关联问题，该方案提出了一种强化自训练方法来将 MRC Model 对构造的问答对的评估作为反馈去优化 QA Construction Model，从而减少了噪声关联数据的产生。该方法被成功用于支付宝智能客服系统中，在“双十一”、“双十二”和“新春红包”等多项营销活动显著提升了智能问答的准确性。

3.6 多模态预训练

多模态预训练旨在从大规模图像-文本对中学习多模态表示，以改进下游的视觉-语言任务，例如图文检索、视觉问答、自然语言视觉推理、视觉定位等。大规模数据的预训练往往能够给下游任务提供很好的初始化和性能保障。

目前的多模态预训练所利用的数据规模从百万到十亿不等，这些海量的图像-文本数据均是通过互联网爬取得到的，包含大量的噪声关联数据。虽然预训练的数据规模足够，但是其中包含的噪声数据仍然会对模型的性能造成不利影响。为此，许多多模态预训练[8-10]尝试从噪声处理的角度来优化预训练,他们的实验也证

实了显式处理噪声关联能够提升下游任务的性能。

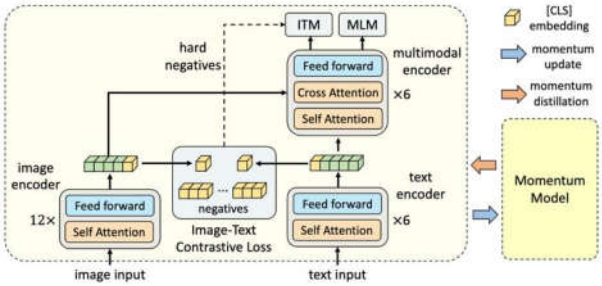

图 8　方法[8]利用动量模型作为教师，矫正对比学习（Image-text Contrastive）和掩码语言建模(Masked Language Model, ITM)中的噪声关联。

方法[8]和[10]通过蒸馏的方式，在多模态学习中同时减缓噪声关联的影响。[8]改进了对比学习和掩码语言建模技术，通过自蒸馏的方式来矫正原本的错误对应关系。[10]进一步提出渐进蒸馏方法，其在训练过程中选择部分样本（α%）按照给定的关联进行训练，同时剩余样本（1-α%）按照教师模型提供的相似度进行训练。随着训练进行，该方法逐渐依靠自身的预测来实现自我提升，不断地减少依靠给定监督信号的数据比例α，从而减缓对噪声的拟合。

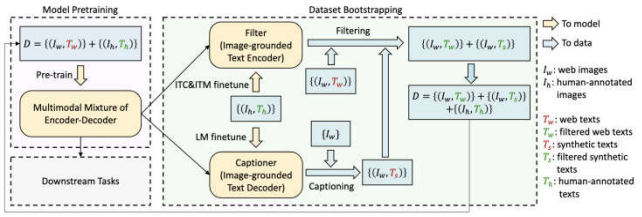

图 9　方法[9]通过对数据进行清理来去除噪声。其提出了一个标题生成器，用于为图像生成关联的标题；以及一个过滤器，用于去除噪声的图像-文本对。

相较于前两篇从鲁棒训练的角度解决噪声关联问题，[9]提出了一种新颖的数据清理方式，利用语言模型对原本的数据进行刷新从而解决噪声关联。在清洗过程中，该方法利用二分类损失判断图-文是否匹配，将不匹配的图文对通过语言模型实现关联文本的生成。最后利用清洗后的数据集在进行训练。

3.7 视听动作识别

视听动作识别旨在从视频片段中分辨不同的动作。目前视频听作识别通常使用多个模态的信息共同完成识别任务，此类方法高度依赖于数据的标签信息，要求视频信息和音频信息正确对应，才能完成动作识别。然而，在实际应用中，这样的条件常常无法满足。训练数据中常常包含噪声关联，即视频片段和音频片段不是对应的。

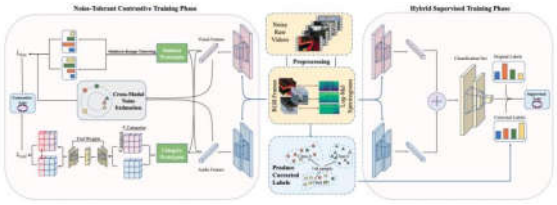

图 10　方法[13]利用噪声估计模块缓解噪声关联问题，并依赖跨模态的相似性修正噪声样本对。

为了解决这类问题，[13]提出了一种容噪学习框架，用于视听动作识别。简言之，[13]首先设计了一个跨模态的噪声估计模块去动态调整跨模态一致性，并用一个类别级的对比学习损失函数来进一步缓解噪声关联的负面影响。随后，[13]通过计算跨模态的相似性来修正噪声样本的关联信息，并将其作为一种补充信息进行模型训练。

四、总结与展望

本文介绍了噪声关联问题，调研了其在不同应用下的具象化存在以及对应的解决方案。噪声关联本质上是一种由于时空不同步所导致的数据错误关联现象，其广泛存在于不同应用和任务中。一旦使用噪声关联的数据去训练机器学习模型，即使是增大数据规模或模型容量也难以获得理想结果。

噪声关联学习赋予了传统噪声标签学习新内涵，可被视为噪声标签学习领域的一个新研究方向。目前针对该问题的研究还比较初步，后续有诸多改进点和探索点，例如：深入探索更多应用和任务下噪声关联的特殊性，设计应用定制化的解决方案；深入研究同时对假阳性和假阴性关联鲁棒的方法，避免模型过拟合假阳性样本对，同时增强正相关样本对的多样性；构建不同任务和应用下噪声关联的评估基准。

责任编委　储珺

参考文献

[1] Huang Z, Niu G, Liu X, et al. Learning with Noisy Correspondence for Cross-modal Matching, NeurIPS 2021.

[2] Yang M, Li Y, Huang Z, et al. Partially view-aligned representation learning with noise-robust contrastive loss, CVPR 2021.

[3] Sharma P, Ding N, Goodman S, et al. Conceptual captions: A cleaned, hypernymed, image alt-text dataset for automatic image captioning, ACL 2018.

[4] Qin Y, Peng D, Peng X, et al. Deep Evidential Learning with Noisy Correspondence for Cross-modal Retrieval, ACMMM 2022.

[5] Yang M, Huang Z, Hu P, et al. Learning with Twin Noisy Labels for Visible-Infrared Person Re-Identification, CVPR 2022.

[6] Lin Y, Yang M, Yu J, et al. Graph Matching with Bi-level Noisy Correspondence.

[7] Jiang L, Huang Z, Liu J, et al. Robust Domain Adaptation for Machine Reading Comprehension, AAAI 2023.

[8] Li J, Selvaraju R., Gotmare A., et al. Align before fuse: Vision and language representation learning with momentum distillation, NeurIPS 2021.

[9] Li J, Li D, Xiong C, et al. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation[J]. ICML 2022.

[10] Andonian A, Chen S, et al. Robust Cross-Modal Representation Learning with Progressive Self-Distillation, CVPR 2022.

[11] Lee K H, Chen X, Hua G, et al. Stacked cross attention for image-text matching, ECCV 2018.

[12] Diao H, Zhang Y, Ma L, et al. Similarity reasoning and filtration for image-text matching, AAAI 2021.

[13] Han H, Zheng Q, Luo M, et al. Noise-Tolerant Learning for Audio-Visual Action Recognition, arXiv:2205.

杨谋星

四川大学计算机学院博士生，研究方向：噪声关联学习，多模态学习。
Email: yangmouxing@gmail.com

林义杰

四川大学计算机学院博士生，研究方向：噪声关联学习，多模态学习。
Email: linyijie.gm@gmail.com

黄振宇

四川大学计算机学院博士生，研究方向：噪声关联学习，多模态学习。
Email: zyhuang.gm@gmail.com

彭玺

四川大学计算机学院教授，研究方向：机器学习理论及其在多媒体分析、计算机视觉、自然语言处理中的应用。
Email: pengx.gm@gmail.com

热点追踪

Distance Correlation 在深度学习中的应用

威斯康星大学麦迪逊分校 甄行践

一、引言

当我们比较两个（或多个）神经网络的时候，我们通常更在意其在某些测试集上的表现，比如准确度或AUROC。但是其实我们更值得在意的是，这个网络学到了什么信息。比如我们都知道，现在 ViT 在绝大多数情况下比 ResNet 的识别准确度更高，但是当我们真地考虑 ViT 学习到的信息量是否比 ResNet 多的时候，我们没有一种非常好的方式。换句话说，我们比较不同网络的时候，需要一种可以被理解、被解释的方式“做减法”。这个时候，我们从统计学中知道，correlation（相关性）是一种被广泛应用于比较两个随机变量的度量，甚至当随机变量增多时，我们可以用 partial correlation 或者 conditional correlation 去掉某些随机变量对于剩下的随机变量的影响[1]。

但我们该如何将 correlation 引入深度学习呢？深度网络中，什么是我们的随机变量呢？

二、方法

我们可以将深度网络抽象为特征提取器+分类器的组合，于是将原始图像输入特征提取器之后，我们可以得到对应的特征，我们将这些特征理解为随机变量。对于同一张图片，如果我们使用不同的神经网络（比如 ViT 和 ResNet），我们可以提取出两个特征向量（x 和 y）。我们想比较 x 和 y 之间的相关性。

如果使用传统的 correlation，最首要的问题就是 x 和 y 的维度需要相同。虽然我们可以用不同的方法把 x 和 y 投影到同一个空间（CCA），但是在训练过程中的 CCA 比较难收敛。这时，我们想到可以使用 distance correlation。

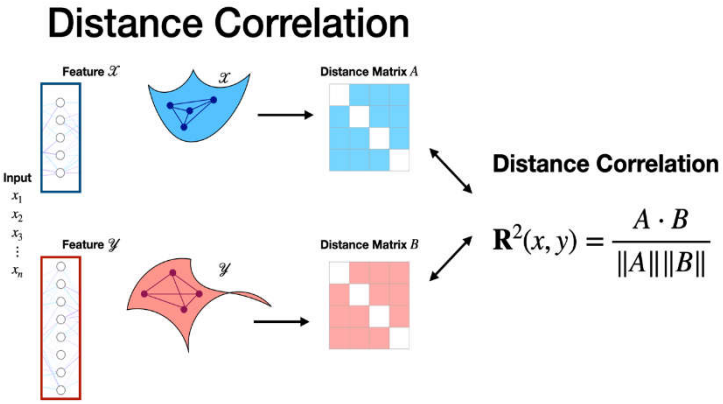

图 1　距离相关性

Distance correlation 与其说是比较 x 和 y 的相关性，不如说比较的是不同 x 之间的距离与对应的 y 之间的距离的相关性。比如我们有两张猫的图片，一张狗的图片，以及 ViT 和 ResNet 作为我们的特征提取器。直觉上，我们可以猜测两张猫的特征之间的距离在任何一种特征提取器之后都应该比较近，同时猫和狗的特征之间的距离都应该比较远。如果符合这种情况，两个特征提取器之间的相关性应当比较高；反之，如果这些特征之间的距离没有什么关系，那么相关性就应该比较低。这是 distance correlation 的基本想法，如图 1 所示，特别具体的证明请参照原文章[2,3]。

如果我们用 distance correlation，我们可以比较轻松地衡量两个神经网络之间的相关性。当我们想更进一步，比如我们想要衡量 ViT 比 ResNet 多学到了多少信息，我们可以将 ResNet 作为我们 correlation 的 condition，计算 partial distance correlation。

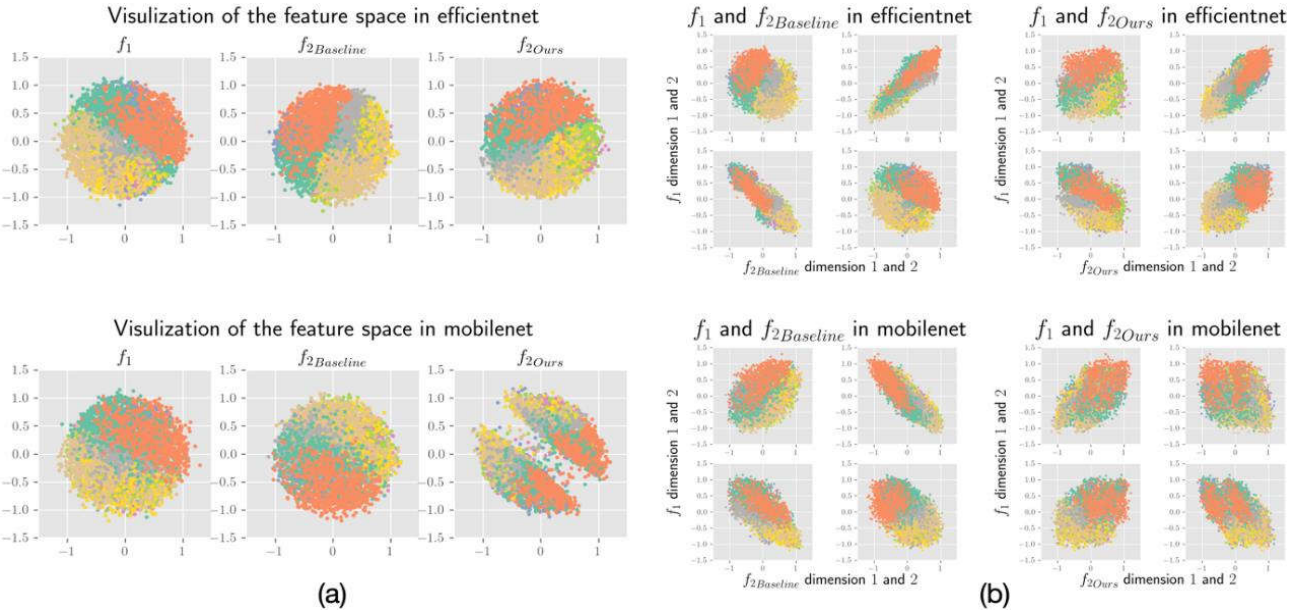

图 2　特征空间的 Picasso 可视化以及不同模型之间的相关性。（a）特征空间分布。（b）使用/不使用 DC 训练的 f_1 和 f_2 的特征空间之间的互相关。

三、应用

3.1. 提高组合网络鲁棒性（降低攻击图片的迁移能力）

现在的神经网络可能识别准确率很高，但是如果我们针对某种特定的神经网络，可以生成一些肉眼难以分辨但是神经网络无法正确识别的图像来攻击这种网络。一种非常简单直接的提高网络鲁棒性的方式是组合多个不同的神经网络。但是研究发现，在相似结构的神经网络之间，相同的攻击图片有很高的迁移概率，即相同的攻击图片针对不同但相似的神经网络，有很高的攻击成功概率。在这种情况下，即使我们组合了多个不同的神经网络，攻击图片仍然可以有较高的攻击成功几率。但是相对的，如果我们可以控制这些神经网络，使得他们学习的特征相互之间是独立的，我们可以假设攻击图片的迁移性会下降，这样组合的神经网络的鲁棒性可以得到提高。

我们比较轻松地发现，distance correlation 可以用于降低不同子网络之间的相关性，使得其彼此独立。

在文章中，我们更多地关注子网络之间的攻击迁移性。我们训练了 2 个相同结构的神经网络。第一种情况下我们不做任何操作，只让权重的初始值不同。第二种情况下我们在训练中，保持神经网络之间互相独立（最小化 distance correlation）。

如图 2 所示，我们检查了分别在使用 distance correlation 和不使用 distance correlation 训练之后，不同模型之间的相关性，可以看出使用 distance correlation 的模型随机变量之间相关度更低（更接近圆形）。

我们发现针对相同的攻击图片，独立网络之间的迁移性下降了 6%-9%不等。

3.2. Disentanglement（解开 latent vector 间的耦合）

对于生成网络，我们可以简单理解为把 latent vector 通过网络转换成图片。通常情况下我们不理解 latent vector 各维度之间的关系。但是假如我们知道某张人脸图片是亚裔少年男性，而我们希望将其转换为亚裔青年女性的图片，如果我们不理解 latent vector 或者对 latent vector 没有限制，是很难实现这种目标的。当我们想要控制 latent vector，比如前 16 维对应人种，接下来 16 维对应年龄，之后 16 维对应性别，之后的所

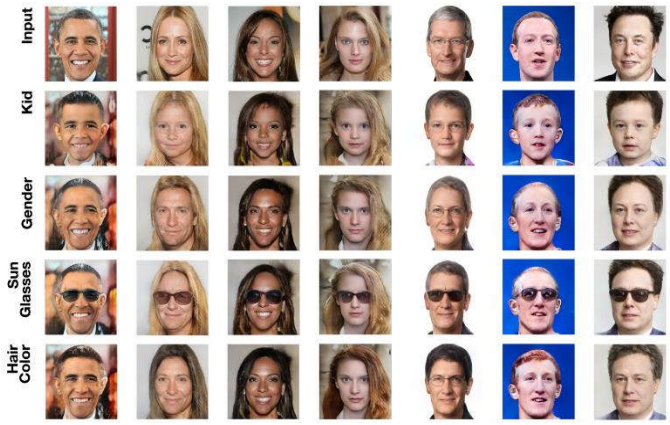

图 3　在 FFHQ 上的训练生成图像（这些结果仅使用 CLIP 的半监督数据集）

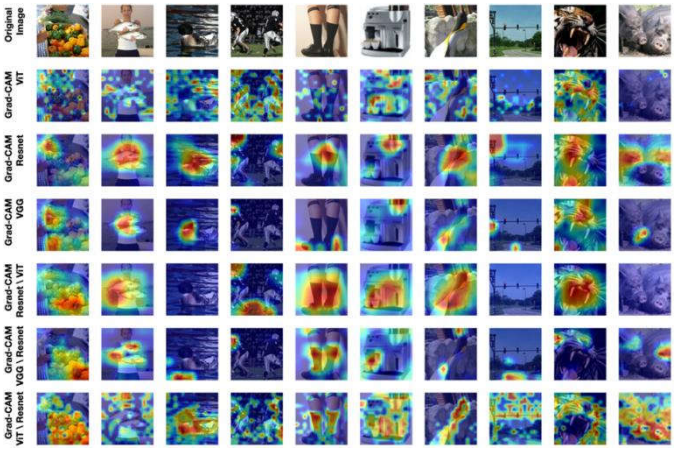

图 5　在 ImageNet 上使用 ViT、Resnet18 和 VGG16 获得 Grad-CAM 结果。

有信息都存在最后 128 维里（成为剩余信息），我们会希望剩余信息包含尽量少的人种年龄性别信息，换句话说，我们希望人种年龄性别对应的 latent vector 和剩余信息对应的 latent vector 相互独立。因此，我们可以很轻易地使用 distance correlation 来实现目标。

我们使用方法[4]，生成了一些高清的人脸图片，如图 3 所示，并且可以控制生成图片的人种年龄等信息。

3.3. 网络信息的比较

这是最重要的一部分实验，也是我们最独特的贡献点：回答“ViT 比 ResNet 的信息更多吗？”

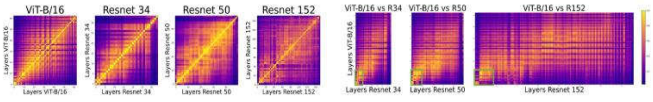

图 4　不同层之间相关性（a）左 4 是单模型中各层之间相似性。(b) 右 3 是 ViT 和 Resnets 层之间相似性。

首先我们可以使用 distance correlation 来比较同个网络不同层之间的相关性，以及不同网络不同层之间的相关性，如图 4 所示。更重要的是，我们可以使用 partial distance correlation 来回答上述问题。

我们首先使用预训练好的 BERT 来对 ImageNet 的 1000 个不同的类的名字进行 embedding，主要用于衡量不同类之间的相似程度（比如猫和老虎距离就应该比较近，而猫和飞机的距离就应该比较远。）

其次我们把 ResNet 提取出来的特征作为 condition，计算 ViT 提取的特征与文字特征之间的相关性。并且反过来将 ViT 作为 condition，计算 ResNet 提取的特征和文字特征的相关性。

实验结果如图 5 所示。结论是 ViT 在剔除 ResNet 的信息后的相关性高于 ResNet 剔除 ViT 的信息。

同时，我们也使用 Grad-CAM 来可视化检验剔除另一网络之后，当前网络更聚焦在图片的什么位置。我们发现 ViT 在剔除 ResNet 之后仍然可以“看清”图片的细微处，而 ResNet 在剔除 ViT 之后聚焦点比较散乱。

四、总结

我们将 distance correlation 和 partial distance correlation 引入深度学习，并且展示了三种完全不同的实验以证明其优势和潜力。对于 distance correlation 其他方向的挖掘，我们认为有很广阔的空间可以探索。比如最直接的 fairness，可以让网络学出的特征对于某些特定的 attribute 独立（比如收入预测独立于人种等等）；或者使用 partial distance correlation 替代 linear regression 来剔除网络信息等。

责任编委　王金甲

参考文献

[1] Xingjian Zhen, Zihang Meng, Rudrasis Chakraborty, and Vikas Singh: On the versatile uses of partial distance correlation in deep learning. European Conference on Computer Vision (ECCV), 2022.

[2] Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov: Measuring and testing dependence by correlation of distances. The Annals of Statistics 35(6), 2769–2794, 2007.

[3] Gábor J. Székely, and Maria L. Rizzo: Partial distance correlation with methods for dissimilarities. The Annals of Statistics 42(6), 2382–2412, 2014.

[4] Aviv Gabbay, Niv Cohen, and Yedid Hoshen: An image is worth more than a thousand words: Towards disentanglement in the wild. Advances in Neural Information Processing Systems 34, 9216–9228, 2021.

甄行践

UW-Madison 计算机科学博士生，师从 Vikas Singh。在 NeurIPS，ICCV，CVPR，AAAI，ECCV 等高水平会议上发表学术论文 7 篇，其中 2 篇获的 CVPR oral，1 篇获的 ECCV 最佳论文奖。
Email: xzhen3@wisc.edu