

热点追踪

可信冲突多视角学习算法

西安电子科技大学 徐偲 司徒俊 管子玉 赵伟

本文解读了西安电子科技大学团队发表在AAAI 2024并获得最佳论文奖的工作。多视角学习旨在通过联合来自不同视角的数据信息，有效地利用多视角数据的共性和特性，从而更精准地揭示事物内在的本质特征。随着多视角学习向更多应用领域不断拓展，越来越多的学习任务面临更为真实的复杂开放环境。相比于传统“实验室”数据环境，复杂开放环境中具有更多的未知性，其中重要一点就是收集到的多视角数据往往存在着视角信息冲突的情况。以往解决这类问题的主要策略包括删除或替换冲突视角，但实际应用需要对冲突样本进行可靠的决策。针对视角冲突的多视角数据，本文提出了一种新的可靠冲突多视角学习问题，该问题要求模型为冲突的多视角数据提供决策结果和可靠性。因此，本文提出了一种可信冲突多视角学习算法，旨在利用视角内的不确定性信息和视角间的一致性信息来提高模型的学习能力，并通过一个简单有效的平均池化层聚合冲突观点。最后，本文对该方法进行了分析和验证，证明了该方法可以准确地建模多视角公共可靠度和视角特定可靠度之间的内在联系。

一、引言

人工智能系统通常通过多视角数据来感知和理解世界。例如，自动驾驶汽车系统通过多个传感器（如摄像头、激光雷达）感知周围环境；推荐系统从用户的多视角生成的内容（如文本和图片）中捕获用户的偏好。整合多个视角的一致和互补信息可以获得更全面的数据描述，从而促进聚类^[1-3]，检索^[4]和推荐^[5,6]等各种任务。以往关于多视角学习^[7-9]的大部分研究总是假设不同视角的数据是严格对齐的。例如，在分类任务中，不

同的视角始终属于基本事实类别。然而，在现实环境中，这种假设可能并不成立。图1可视化了用户多视角生成内容的一个案例：文本和图像表示的食品类别存在冲突。因此，这种不同视角信息的冲突使得大多数多视角学习方法不可避免地退化甚至失败。



图1 视角冲突的多视角数据

当前解决该问题的主要方法是消除冲突数据样本。最初的思路是将冲突数据视为异常值。该类方法通常包括3个步骤：1) 测量视角间的一致性；2) 识别异常值，即：不同视角之间存在着的显著不一致的样本；3) 去除异常值，构建干净的数据集。最近，一些多视角学习方法致力于学习原始数据的对齐关系，并相应地构建新的数据样本。例如，将图1中的文本将被替换为另一个对齐样本的视角，从而解决原样本中的冲突问题。

然而，现实世界的应用程序通常需要为冲突样本做出决策，而不仅仅是消除它们。例如，推荐系统需要从用户的冲突多视角评论中预测用户的偏好。考虑到冲突样本的决策可能是不可靠的，我们需要模型能够回答“该决策可靠吗？”因此，本文提出了一个新的问题——可靠的冲突多视角学习 (Reliable Conflictive Multi-view Learning, RCML) 问题，该问题要求模型对冲突多视角数据提供决策结果和附加的信度。

本文实现了一个可信冲突多视角学习算法——证据冲突多视角学习(Evidential Conflictive Multi view Learning, ECML)方法, 整个网络框架如图 2 所示, 它采用两个阶段来识别整合多视角信息, 其中第一阶段以特定视角信息作为输入, 采用证据神经网络来提取证据并得到特定视角的狄利克雷分布观点, 第二阶段是利用冲突观点的聚合策略机制来整合多视角观点信息。本文的主要贡献总结如下: (1) 认识到在处理冲突的多视角数据时显式地提供决策结果和相关可靠性的重要性。(2) 提出了一种冲突观点的聚合策略和冲突度量, 并从理论上证明了它可以准确地建模多视角共同可靠度和视角特定可靠度。(3) 实验结果验证了提出的模型在六个基准数据集上取得了领先的性能。这直接证明了该模型在冲突的多视角学习任务中的有效性和优越性, 为该领域提供了更全面、更高效的解决方案。

二、相关工作

1. 冲突的多视角学习

多视角学习通过整合多个互补视角的信息, 能够有效地提升模型的泛化性能和适应性, 但之前的研究大多假设不同视角的数据是严格对齐的。因此, 在面对不同视角信息不一致的情况时, 大多数现有的多视角学习方法不可避免地退化甚至失败。目前针对冲突的多视角学习问题, 现有的方法主要分为两类: 多视角异常点检测方法和部分视角对齐的多视角学习方法。

多视角异常点检测方法旨在检测特定环境中具有异常行为的异常点。Gao 等人^[10]提出的水平异常检测算法是解决该问题的第一个方法, 该算法构造一个集合相似矩阵并通过谱嵌入计算共识表示。另一种代表性的方法是由 Alvarez 等人^[11]提出的亲和传播方法。它计算每个视角中每个样本与其相邻样本之间的差异来检测异常点。Li 等人^[12]提出了一种多视角低秩分析框架, 该框架执行交叉视角低秩分析, 并采用精心设计的标准计算每个样本的异常值得分。Zhao 等人^[13]提出了一种共识正则化离群值检测方法, 该方法将所有视角的指示矩阵近似为具有共识正则化的共享矩阵以消除异常的影响。Li 等人^[14]提出了一种能够从任意数量的数据视角中检测异常值的方法, 该方法学习所有视角数据的潜在判别表示, 并基于潜在判别表示定义了一个异常值评分函数。Hou 等人^[15]提出了一种快速的低秩潜在表示用于大规模多视角异常点检测, 该方法利用特定于视角的判别潜在表示来探索样本的判别成分, 并利用深度编码器架构实现样本的异常分数。

在多视角学习领域, 大多数现有工作的成功在很大程度上依赖于视角一致性的假设。换言之, 一致性假设要求来自不同视角的数据必须严格对齐。然而, 在实际操作中, 数据很容易在收集或传输过程中出现错位, 从而导致部分视角对齐问题。因此, 人们提出了一些解决这一问题的传统方法。Lampert 等人^[16]提出了弱配对

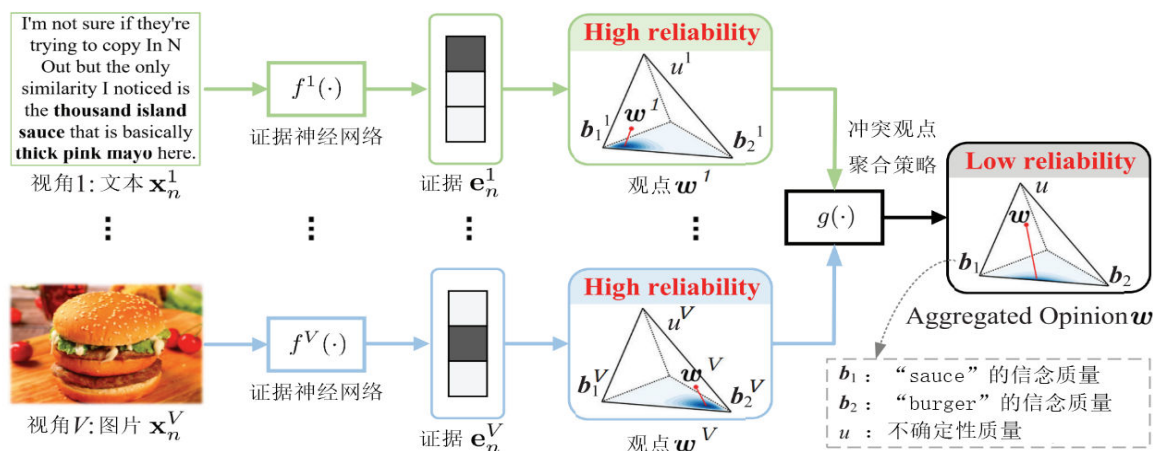


图 2 网络架构图。首先, 利用特定视角的神经网络收集证据, 即对每个类别的支持量。然后, 该方法形成由所有类别的信念质量和不确定性组成的特定观点。最后, 通过冲突观点聚合策略来整合观点。

最大协方差分析来克服弱配对数据的局限性。Zhang 等人^[17]提出了少量的视角间约束代替映射来获取视角间的相互信息。目前, 捕获数据高维特征的深度部分视角对齐方法尚处于起步阶段。Huang 等人^[18]提出了一种称为部分视角对齐聚类的多视角聚类方法, 该方法实现了匈牙利算法的可微代理来建立未对齐数据的对齐关系。Yang 等人^[19]提出了使用噪声鲁棒对比损失同时学习数据高层表示和对齐数据的方法。Zhang 等人^[20]提出了一种自适应视角对齐和特征增强网络, 该网络使用自适应视角对齐模块和自聚焦机制来计算对齐矩阵。

现有的方法虽然可以解决部分的冲突多视角学习问题, 但仍然存在着以下限制: 1) 多视角异常点检测方法不能很好地检测出由模型本身引起的离群值, 也不能给出一个直观的度量离群值程度的指标, 并且现有的多视角异常点检测方法致力于学习原始数据的对齐关系, 并据此构造新的数据样本, 忽略了冲突样本的信息。2) 目前解决部分视角对齐问题的方法是存在问题的。现有的该类方法将其他视角对齐到一个视角上, 这一操作改变了原始的数据分布, 改变了数据集的整体结构和样本样本的真实信息。总之, 以前的方法的目的是为了消除冲突的样本, 而实际的应用程序通常需要为这些冲突的样本做出决策分析。

2. 感知不确定性深度学习

深度神经网络已经在各种任务中取得了显著的成功, 但其往往无法捕获其决策的不确定性, 特别是对于低质量数据。不确定性可分为任意不确定性(数据不确定性)和认知不确定性(模型不确定性)。目前, 不确定性量化的深度学习^[21]可以分为: 单一确定性方法^[22,23], 贝叶斯方法^[24-26], 集成方法^[27]和测试时间增强方法^[28]。证据深度学习(Evidential deep learning, EDL)^[22]是单一确定性方法的代表性方法, 其根据单个神经网络计算特定类别的证据。近年来, 研究者将 EDL 扩展到多视角学习领域。开创性工作可信多视角分类(TMC)^[29]利用 Dempster-Shafer 证据理论对不同视角的观点进行融合。随后, 多种观点聚合方法^[30-33]被研究者们提了出来。这些方法的一个重要特征是将另一种观点整合到原来的观点中, 得到的不确定性质量会减小。但是, 当面对

现实中复杂的多视角数据时, 尤其是在纳入不可靠或冲突的观点时, 这些方法可能会出现错误。而且, 本文认为最终融合后观点的不确定性应该与被融合的视角质量相关, 而不是一味地减小。

三、网络框架

1. 问题定义

假设多视角数据集为 $\{\{x_n^v\}_{v=1}^V, y_n\}_{n=1}^N$ 有 $N = \bar{N} + \tilde{N}$ 个样本, 每个样本有 V 个视角。其中有 \bar{N} 个正常样本, \tilde{N} 个冲突样本, 如图 3 所示。 $y_n \in \{0,1\}^K$ 是 y_n 对应的 one-hot 编码, K 为类别数。RCML 的目标是为测试样本准确地预测类别 y_n , 并提供附加的预测不确定性 $u_n \in [0,1]^1$ 来度量决策可靠性 $1 - u_n$ 。

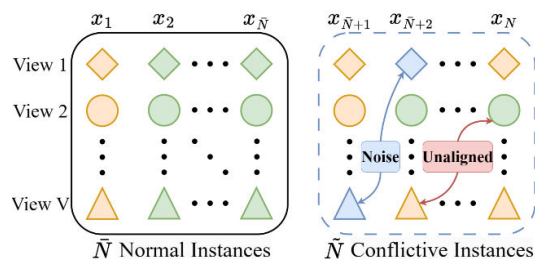


图 3 数据定义。不同的形状代表不同的视角, 黄色和绿色代表两种不同的类别, 蓝色代表该视角含有噪声。一个冲突的样本含了噪声数据和不对齐视角数据。

2. 特定视角的证据深度学习

大多数现有的深度多视角学习方法通常依赖于在深度神经网络上使用 Softmax 层来进行分类。然而, 这些基于 Softmax 的深度神经网络在准确估计预测不确定性方面存在局限性。这是因为 Softmax 只提供了预测分布的单点估计, 在出现错误预测时, 仍会产生高置信度的预测输出。本文通过引入证据深度学习来解决 Softmax 高置信度预测问题。

对于一个 K 分类问题, 假设某个样本的一个特定视角 (x_n^v, y_n) 所对应的多项式观点为 $w(b, u, a)$, 其中 $b = (b_1, \dots, b_K)$ 是信念质量向量, u 是不确定性质量, $a = (a_1, \dots, a_K)$ 是每个类别的先验概率。主观逻辑^[34]要求信念质量和不确定性质量是非负的, 且两者元素之和为 1, 即

$$u + \sum_{k=1}^K b_k = 1$$

其中, $u \geq 0$, $b_k \geq 0$ 。概率是多项式观点的投影, 如下所示:

$$p_k = b_k + a_k u.$$

通常, 先验概率 \mathbf{a} 是根据先验知识人为手动设置的。例如, 一种常见的方法是假定每个类别的先验概率相等, 即: $a_k = 1/K$ 。

根据主观逻辑, 将狄利克雷分布作为最终的类别分布的先验分布。它可以模拟二阶不确定性, 而 Softmax 层中的概率值只能捕捉一阶不确定性。具体的狄利克雷分布的概率密度函数 $D(\mathbf{p}|\boldsymbol{\alpha})$ 为:

$$D(\mathbf{p}|\boldsymbol{\alpha}) = \begin{cases} \frac{1}{B(\boldsymbol{\alpha})} \prod_{k=1}^K p_k^{\alpha_k-1} & \text{for } \mathbf{p} \in \mathcal{S}_K \\ 0 & \text{otherwise} \end{cases}$$

其中, $\mathbf{p} = (p_1, \dots, p_K)$ 是每个类别的概率, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$ 是狄利克雷分布的参数, \mathcal{S}_K 是 K 维单纯性, 具体如下所示:

$$\mathcal{S}_K = \left\{ \mathbf{p} \mid \sum_{k=1}^K p_k = 1, p_k > 0 \right\},$$

$B(\boldsymbol{\alpha})$ 是 K 维的 Beta 分布。

在本文中, 证据指的是使用证据神经网络 $f(\cdot)$ 从特定视角输入中收集的支持分类的指标 $\mathbf{e} = f(\mathbf{x}_n^v)$, 进而可以获得狄利克雷分布的参数 $\boldsymbol{\alpha} = \mathbf{e} + \mathbf{1}$ 。在狄利克雷分布和多项式观点之间存在一个映射, 具体如下:

$$b_k = \frac{\alpha_k - 1}{S}, u = \frac{K}{S}$$

其中, $S = \sum_{k=1}^K \alpha_k$ 。

根据上述公式可以发现, 整体的证据量越少, 不确定性就越高。另外, 第 k 个类别的期望概率的计算公式如下:

$$p_k = \frac{\alpha_k}{S} = \frac{e_k + 1}{S}.$$

3. 冲突观点的聚合策略

冲突的多视角数据主要分为两部分, 一部分是噪声

数据, 一部分是未对齐数据。其中噪声数据会显示出很高的不确定性, 因此需要在聚合阶段减少它们的影响; 而未对齐视角将提供高度冲突的观点, 并且它们的不确定性很低, 这表明某些视角是不可靠的。在这种情况下, 我们很难判断哪些视角的信息是可靠的。事实上, 多视角学习结果的不确定性不应该随着视角数量的增加而减少, 而应该与待融合的视角的质量有关, 特别是当两种视角的学习结果冲突时。为了解决这一问题, 本节提出了一种新的冲突观点的聚合方法。具体定义如下:

定义 1 冲突观点的聚合方法 假设 $\mathbf{w}^A = (\mathbf{b}^A, u^A, \mathbf{a}^A)$, $\mathbf{w}^B = (\mathbf{b}^B, u^B, \mathbf{a}^B)$ 分别是同一个样本中的视角 A 和视角 B 的观点, $\mathbf{w}^{A \diamond B} = (\mathbf{b}^{A \diamond B}, u^{A \diamond B}, \mathbf{a}^{A \diamond B})$ 表示视角 A 和视角 B 观点的聚合观点, 详细的计算方式如下:

$$\mathbf{w}^{A \diamond B} = \mathbf{w}^A \diamond \mathbf{w}^B = (\mathbf{b}^{A \diamond B}, u^{A \diamond B}, \mathbf{a}^{A \diamond B})$$

$$b_k^{A \diamond B} = \frac{b_k^A u^B + b_k^B u^A}{u^A + u^B},$$

$$u^{A \diamond B} = \frac{2u^A u^B}{u^A + u^B}, \mathbf{a}^{A \diamond B} = \frac{\mathbf{a}^A + \mathbf{a}^B}{2}.$$

这种聚合是通过使用多项式观点和狄利克雷分布之间的双射映射将证据映射到信念质量来实现的。从本质上讲, 该聚合规则确保了聚合后的观点的质量与被聚合观点的质量成正比。具体地说, 当聚合一个高度不确定的观点时, 新观点的不确定性大于原来的观点; 反之, 则小于原观点。冲突观点的聚合方法可以简化为证据的平均, 后面中将给出更详细的解释。

根据定义 1, 可以用以下规则融合不同观点来获得最终的共同观点:

$$\mathbf{w} = \mathbf{w}^1 \diamond \mathbf{w}^2 \diamond \dots \diamond \mathbf{w}^V.$$

根据上述融合规则, 可以得到最终的多视角联合观点 \mathbf{w} , 从而得到最终的各类概率和总体不确定性。

4. 冲突度

所提出的方法还致力于: 1) 在训练阶段(使用正常样本)确保模型在不同视角下的一致性; 2) 对视角信息冲突的程度有一个直观的感受。因此, 本文引入了根据观点熵建立的冲突度(Conflict Degree)度量, 具体定义如下:

定义 2 冲突度 假设 w^A , w^B 分别是同一个样本中视角 A 和视角 B 的观点, 那么视角 A 和视角 B 的冲突度 $c(w^A, w^B)$ 为:

$$c(w^A, w^B) = c_p(w^A, w^B) \cdot c_c(w^A, w^B),$$

其中 $c_p(w^A, w^B)$ 是视角 A 和视角 B 之间的投影距离, $c_c(w^A, w^B)$ 是视角 A 和视角 B 之间的联合确定性。两者可以通过以下公式计算获得:

$$c_p(w^A, w^B) = \frac{\sum_{k=1}^K |p_k^A - p_k^B|}{2},$$

$$c_c(w^A, w^B) = (1 - u^A)(1 - u^B).$$

根据上述定义可以看出, 该指标保证了两件事:(1) 当观察到相同的预测概率分布时, $c = 0$ 表明观点不冲突;(2) 当绝对观点存在但预测概率分布不同时, $c = 1$ 。特别地, 当 $c_c = 0$ 时, 它表示在一个或两个视角中存在不可靠的条件。另一方面, 当 $c_c = 1$ 时, 表示两者的观点都被认为是可信的, 也就是说它们的不确定性为零。

5. 讨论与分析

在本节中, 将从理论上分析 ECML 的优势, 特别是冲突多视角数据的冲突观点聚合策略。具体细节如下:

命题 1 冲突观点的聚合方法可以简化为证据的平均。

证明 1 假设 $w^A = (b^A, u^A, a^A)$, $w^B = (b^B, u^B, a^B)$ 分别是同一个样本中的视角 A 和视角 B 的观点, $w = (b, u, a)$ 是视角 A 和视角 B 观点的聚合后的观点。同样的 e_k^A , e_k^B 和 e_k 分别是视角 A、视角 B 和聚合后观点对应的第 k 个类别的证据。

$$\begin{aligned} e_k &= b_k S = \frac{b_k K}{u} \\ &= \frac{b_k^A u^B + b_k^B u^A}{u^A + u^B} \cdot \frac{u^A + u^B}{2u^A u^B} \cdot K \\ &= \frac{K}{2} \cdot \frac{\frac{K e_k^A}{S^A S^B} + \frac{K e_k^B}{S^A S^B}}{\frac{K}{S^A} \cdot \frac{K}{S^B}} \\ &= \frac{e_k^A + e_k^B}{2} \end{aligned}$$

基于这一命题, 在多视角融合阶段, 该方法建立了简单有效的平均池化融合层来实现冲突观点聚合。

命题 2 对于冲突观点的聚合策略, 如果新观点的不确定性小于原观点的不确定性, 则聚合后观点的不确定质量小于原观点; 反之大于原观点。

证明 2 假设 w^o , w^a 分别是同一个样本中的两个观点, w^o 是原观点, w^a 是待聚合的新观点, w 是聚合后的观点。同样的 u^o , u^a 和 u 分别是对应的不确定性。

$$u = \frac{2u^o u^a}{u^o + u^a} = \frac{1}{\frac{1}{2} \left(1 + \frac{u^o}{u^a}\right)} \cdot u^o.$$

由此可得聚合后观点的不确定性与待聚合观点的不确定性之间的关系:

$$\begin{cases} u < u^o, & u^a < u^o \\ u = u^o, & u^a = u^o \\ u > u^o, & u^a > u^o \end{cases}$$

现有的大多数可信多视角学习方法都是基于以下假设: “将另一种观点整合到原观点中, 得到的不确定性质量会减小。” 本文认为这是不合理的, 因为: 1) 当整合一个可靠的视角时, 融合过程应该减少整体的不确定性; 2) 当纳入不可靠或冲突的观点时, 融合应增加不确定性。此外, 现有的方法往往忽略了由于数据不一致或模型跨不同视角的性能变化引起的从不同观点收集的观点之间发生冲突的可能性。

为了更加清楚的解释这个问题, 本节以一个现实中的场景为例。假设现在有两个观察者 A 和 B, 他们观察从一个盒子里抽出的彩色球。这些球有四种颜色: 黑、白、红、绿。假如观察者 B 是色盲, 难以区分红色和绿色的球, 而能够区分其他颜色的组合。另一方面, 观察者 A 有正常的视力, 能识别出正确的颜色。当一个红球被选中时, 观察者 A 通常会认为它是红色的, 而观察者 B 可能会认为它是绿色的。此时, 观察者 A 和 B 对同一物体产生了不同的观点, 但是两者都坚信自己的判断。假设最初不知道其中一个观察者是否是色盲, 那么他们的观点都会被认为是同样可靠的。然而, 现有的融合方法在结合两个观察者的观点后, 会错误地降低整体的不确定性。在这个例子中, 观察者可以被看作是收集多视角数据的传感器, 他们的判断能力可以被视为每个视角所对应的模型的分辨能力。由于, 信息收集方式的错误或者

模型本身能力的不足，最终导致了不同的视角会产生冲突的多视角信息。综上所述，在视角冲突的情况下，应该平等地对待来自每个视角的信息，考虑到观点冲突的可能性，而不能简单地降低不确定性。

6. 损失函数

Softmax 函数为样本的类概率提供了一个点估计，但不提供相关的不确定性。而多项式观点或等价的狄利克雷分布可以用来模拟类概率的概率分布。因此，可以通过设计并训练神经网络，以形成对应的多项式观点。具体而言，将传统基于神经网络的分类器的 Softmax 层替换为激活函数层(如: Relu)，以确保网络输出非负值，将其作为证据向量 e ，从而得到狄利克雷分布的参数。

首先，对于特定的单视角来说，在传统的基于神经网络的分类器中，通常采用交叉熵损失函数。因此，本文对交叉熵损失函数进行调整以适用于本文提出的方法，具体如下：

$$\begin{aligned} L_{ce}(\alpha) &= \int [\sum_{j=1}^K -y_j \log(p_j)] \frac{1}{B(\alpha)} \prod_{j=1}^K p_j^{\alpha_j-1} dp \\ &= \sum_{j=1}^K y_j (\psi(\sum_{j=1}^K \alpha_j) - \psi(\alpha_j)) \end{aligned}$$

对一批训练样本的损失可以通过对批次中每个样本的损失之和来计算。在训练过程中，模型可以发现数据中的模式，并根据这些模式为特定的类标签生成证据，以最小化总体损失。例如，当模型发现手写数字图像上存在一个大的圆形图案时，类别零的证据会增加。虽然上述损失函数可以保证每个样本的正确标签比其他类别产生更多的证据，但不能保证错误标签产生更少的证据。换句话说，如果样本不能被正确分类，那么所有证据都应倾向于零。对于均匀的狄利克雷分布，即： $S = K$ ，通过计算可以发现其所对应的证据量为 0，也就是说整体不确定性 $u = 1$ 。因此，通过引入 Kullback-Leibler (KL) 散度作为正则化来惩罚无法正确分类的样本。这个正则化项为：

$$\begin{aligned} &KL[D(\mathbf{p}|\tilde{\alpha}) \parallel D(\mathbf{p}|\mathbf{1})] \\ &= \log \left(\frac{\Gamma(\sum_{k=1}^K \tilde{\alpha}_k)}{\Gamma(K) \prod_{k=1}^K \Gamma(\tilde{\alpha}_k)} \right) + \sum_{k=1}^K (\tilde{\alpha}_k - 1) [\psi(\tilde{\alpha}_k) - \psi(\tilde{S})] \end{aligned}$$

其中， $\tilde{\alpha} = \mathbf{y} + (\mathbf{1} - \mathbf{y}) \odot \alpha$ 是调整后的狄利克雷分布参数， $\tilde{S} = \sum_{k=1}^K \tilde{\alpha}_k$ ， $\Gamma(\cdot)$ 是伽马函数， $\lambda_t = \min(1.0, t/10) \in$

$[0,1]$ 是退火系数， t 是当前训练的次数。

因此，给定观点对应的狄利克雷分布的参数 α ，则对应的单个观点的损失函数为：

$$L_{acc}(\alpha) = L_{ce}(\alpha) + \lambda_t KL[D(\mathbf{p}|\tilde{\alpha}) \parallel D(\mathbf{p}|\mathbf{1})].$$

其次，采用了最小化观点冲突程度的方法来保证在训练过程中不同观点之间结果的一致性。对于第 n 个样本 $\{x_n^v\}_{v=1}^V$ ，其一致性损失的计算公式如下：

$$L_{con}^n = \frac{1}{V-1} \sum_{p=1}^V (\sum_{q \neq p}^V c(\mathbf{w}_n^p, \mathbf{w}_n^q)).$$

其中 V 是总视角数量， \mathbf{w}_n^p 是第 n 个样本第 p 个视角的观点， $c(\mathbf{w}_n^p, \mathbf{w}_n^q)$ 是视角 p 和视角 q 的冲突度。

最后，最终的损失函数如下所示：

$$L = \frac{1}{N} \sum_{n=1}^N (L_{acc}(\alpha_n) + \beta \sum_{v=1}^V L_{acc}(\alpha_n^v) + \gamma L_{con}^n).$$

其中 α_n^v 和 α_n 分别是第 n 个样本的第 v 个视角和最终视角的狄利克雷分布的参数， β 和 γ 是两个超参数。

四、实验结果与分析

为了验证所提出模型的有效性，本研究在六个多视角数据集上进行了实验，这些数据集的详细信息如下：

(1) HandWritten^[35] 包含了 2000 个从“0”到“9”的手写数字样本，每个类有 200 个样本，数据集提供了 6 种类型的特征：Pix、Fou、Fac、ZER、KAR 和 MOR。

(2) CUB^[36] 是一个包含 200 个类别的鸟类数据集，一共 11788 个样本。每个样本都含有一张图片和一段文本描述，使用 GoogleNet 提取图片特征和 doc2vec 提取文本特征作为两个视角。

(3) HMDB^[37] 是一个大规模的人类动作识别数据集，包含来自 51 个动作类别的 6718 个样本，提取 HOG 和 MBH 特征作为该数据集的两个视角。

(4) Scene15^[38] 包括来自 15 个室内和室外场景类别的 4485 张图像，分别提取了三种类型的特征 GIST、PHOG 和 LBP 作为不同的视角。

(5) Caltech101^[39] 包含了 8677 张图片，一共 101 个类别。本节选取了前十个类别，并使用 DECAF 和

VGG19 模型提取两种深度特征作为不同的视角。

(6) PIE^[40]包含 680 个样本，一共 68 个类，分别提取强度、LBP 和 Gabor 特征作为 3 个视角。

数据集	样本数	类别数	视角维度
HandWritten	2000	10	240/76/216/47/64/6
CUB	11788	10	1024/300
HMDB	6718	51	1000/1000
Scene15	4485	15	20/59/40
Caltech101	8677	101	4096/4096
PIE	680	68	484/256/279

表 1 多视角数据集

本文将所提出的方法与以下两大类方法进行比较：
基于特征融合的多视角学习方法：

(1) DCCAE^[41]是经典的多视角学习方法，使用归一化谱聚类计算样本的类隶属度和使用广义特征值方法计算特征融合的映射，以利用同一类样本之间的相关信息，从而寻找数据的共同表征。

(2) CPM-Nets^[42]是一种多视角特征融合方法，其重点是学习一种通用的表示来处理不同视角之间的复杂关联。

(3) DUA-Nets^[23]是一种不确定性感知方法，它利用反向网络将来自不同观点的内在信息整合为统一的表示。

基于决策融合的多视角学习方法：

(4) TMC^[29]是多视角不确定性感知方法的先驱，它解决了不确定性估计问题，产生了可靠的分类结果。

(5) TMDL-OA^[31]是一种基于证据深度学习的 SOTA 多视角决策融合方法，该方法提出了一致性度量损失来实现可靠的学习结果。

为了创建具有视角冲突样本的测试集，本文进行以下转换：(1) 对于噪声视角，选取一定比例的样本，添加不同标准差的高斯噪声来制造噪声数据。(2) 对于不对齐的视角，选取一定比例的样本，并随机修改某个视

角的信息，使得该视角信息与其他视角描述的不一致。本章实验是基于 PyTorch 深度学习框架实现的，并对每种方法进行了 10 次运行，统计平均值和标准差。

1. 对比试验

为了模型的有效性，在正常测试集和冲突测试集上对所提出的方法的性能进行验证，使用分类准确率作为模型性能的评价指标。根据以下方式创建冲突的多视角数据集：(1) 对于噪声视角，选取一半的样本，添加标准差为 1 的高斯噪声来制造噪声数据。(2) 对于不对齐的视角，选取一半的样本，并修改第一个视角的信息，使得该视角信息与其他视角描述的不一致。实验结果如表 2 和表 3 所示。

可以观察到，本文提出的方法在原始数据集上的分类准确率优于所有基线方法。在 HMDB 数据集上，与第二好的(TMDL-OA)模型相比，ECML 的准确率提高了大约 2.64%；在 PIE 数据集上，与第二好的(TMDL-OA)模型相比，ECML 的准确率提高了大约 2.38%。在原始测试集上，ECML 优于所有其他对比方法，其原因可能是由于一致性损失的加入，增强了模型的学习能力，在消融研究中也证实了该猜测。

根据实验结果可以观察到，在冲突测试集上进行测试时，所有方法的准确率均明显降低。尽管如此，由于冲突观点的聚合策略，ECML 显示出对特定视角的冲突的高效的感知能力，从而在所有数据集中产生了令人印象深刻的结果。这也突出了 ECML 对于冲突多视角数据的有效性。在 PIE 数据集上，与第二好的(TMDL-OA)模型相比，ECML 的准确率提高了大约 15.84%；Scene15 数据集上，与第二好的(TMDL-OA)模型相比，ECML 的准确率提高了大约 8.55%此外，从实验结果中还发现了一些现象：(1) 基于决策融合的方法优于基于特征融合的方法。这可能是由于面对视角冲突，当使用基于特征融合的方法时，融合后的特征可能会消失，导致最终决策失败。当使用基于决策融合的方法时，每个单独视角的决策都是正常的。尽管最终的决定可能是错误的，但它们比基于特征融合的方法要好。(2) 不同数据集的性能差异很大。原因可能是有些数据集(PIE, Scene15)在不同类别之间的特征差异很大，而有些数据集(CUB,

数据集	DCCAE	CPM-Nets	DUA-Nets	TMC	TMDL-OA	Ours
HandWritten	95.45 ± 0.35	94.55 ± 1.36	98.10 ± 0.32	98.51 ± 0.13	99.25 ± 0.45	99.40 ± 0.00
CUB	85.39 ± 1.36	89.32 ± 0.38	80.13 ± 1.67	90.57 ± 2.96	95.43 ± 0.20	98.50 ± 2.75
HMDB	49.12 ± 1.07	63.32 ± 0.43	62.73 ± 0.23	65.17 ± 2.42	88.20 ± 0.58	90.84 ± 1.86
Scene15	55.03 ± 0.34	67.29 ± 1.01	68.23 ± 0.11	67.71 ± 0.30	75.57 ± 0.02	76.19 ± 0.12
Caltech101	89.56 ± 0.41	90.35 ± 2.12	93.43 ± 0.34	92.80 ± 0.50	94.63 ± 0.04	95.36 ± 0.38
PIE	81.96 ± 1.04	88.53 ± 1.23	90.56 ± 0.47	91.85 ± 0.23	92.33 ± 0.36	94.71 ± 0.02

表 2 模型在正常多视角测试集上的分类准确率 (%)

数据集	DCCAE	CPM-Nets	DUA-Nets	TMC	TMDL-OA	Ours
HandWritten	82.85 ± 0.38	83.34 ± 1.07	87.16 ± 0.34	92.76 ± 0.15	93.05 ± 0.05	94.40 ± 0.05
CUB	63.57 ± 1.28	68.82 ± 0.17	60.53 ± 1.17	73.37 ± 2.16	74.43 ± 0.26	76.50 ± 1.15
HMDB	29.62 ± 1.79	42.62 ± 1.43	43.53 ± 0.28	47.17 ± 0.15	67.62 ± 0.28	70.84 ± 1.19
Scene15	25.97 ± 2.86	29.63 ± 1.12	26.18 ± 1.31	42.27 ± 1.61	48.42 ± 1.02	56.97 ± 0.52
Caltech101	60.90 ± 2.32	66.54 ± 2.89	75.19 ± 2.34	90.16 ± 2.50	90.63 ± 2.05	92.36 ± 1.48
PIE	26.89 ± 1.10	53.19 ± 1.17	56.45 ± 1.75	61.65 ± 1.03	68.16 ± 0.34	84.00 ± 0.14

表 3 模型在冲突多视角测试集上的分类准确率 (%)

Caltech101)的差异相对较小。特征差异较大的数据集经过冲突处理后,冲突程度较高。ECML可以解决这种情况,而传统的方法不能。另外,性能也可能与视角的数量有关。

2. 冲突度量可视化分析

为了更加直观地感受 ECML 方法中提出的冲突度量,本节基于 HandWritten 数据集对冲突度量进行可视化操作。为了构建冲突样本,本节修改了数据集中第一个视角中的内容,使得该视角与其他视角中的内容

不对齐。下图显示了 HandWritten 数据集上六个视角的冲突程度。颜色越深代表冲突的程度越高。

图 4 清楚地表明了 ECML 可以有效地捕获和量化视角之间的冲突程度,这一发现进一步验证了 ECML 的可靠性。

3. 不确定性估计分析

为了进一步评估 ECML 的不确定性量化能力,本节可视化了基于 CUB 数据集的正常和冲突测试集的分布。为了构建冲突测试集,引入了标准偏差 $\sigma = 0.1$ 、1、5、10 的高斯噪声随机加入正常的测试集中。实验结果如图 5 所示。

结果表明,当噪声强度较低($\sigma = 0.1$)时,冲突数据集的不确定性分布曲线与正常数据集的分布曲线基本一致。然而,随着噪声强度的增加,冲突数据集的不确定性也随之增加。这一发现表明,不确定性与数据集的质量相关,并且验证了本章提出的方法在不确定性量化方面的有效性。

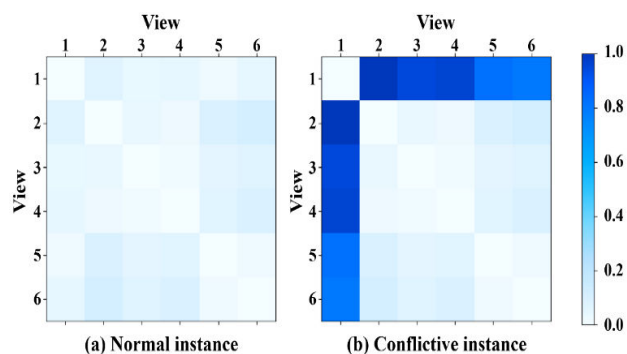


图 4 冲突度可视化分析

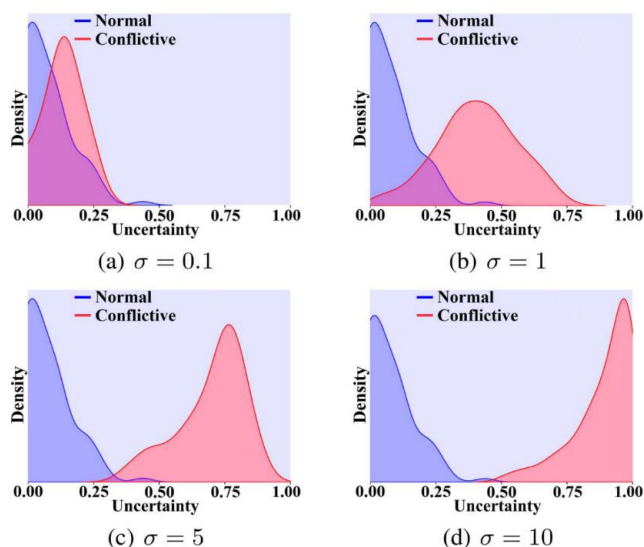


图5 不确定性量化分析

4. 消融实验

为了验证模型中一致性损失的有效性，本节在 Scene15 和 PIE 数据集上进行了消融实验，并与以下变体进行比较：

- (1) ECML-RC：该变体在训练过程中去除了 consistency loss
- (2) ECML-V1：该变体只使用视角一的信息
- (3) ECML-V2：该变体只使用视角二的信息
- (4) ECML-V3：该变体只使用视角三的信息

实验根据以下方式创建冲突的多视角数据集：(1) 对于噪声视角，选取一半的样本，添加标准差为 1 的高斯噪声来制造噪声数据。(2) 对于不对齐的视角，选取一半的样本，并修改第一个视角的信息，使得该视角信息与其他视角描述的不一致。实验结果如图 6 所示。

从实验结果可以看出：(1) 采用冲突观点的聚合策

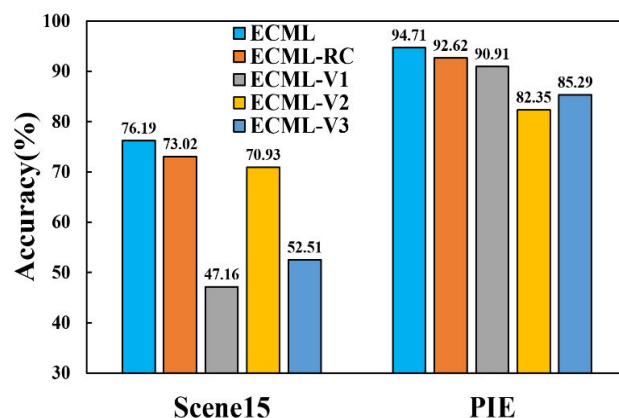


图6 一致性损失有效性分析

略对所有视角信息进行整合相比于单独的单视角信息，其所训练出的模型的性能更好。(2) 使用一致性损失函数能够增强模型的学习能力。综上所述，消融实验的结果进一步验证了本章提出的模型中各个组成部分的有效性。冲突观点的聚合策略和一致性损失函数在多视角学习任务中各自发挥了重要作用，而将它们集成在一起则能够更全面地捕捉多视角间的冲突信息，提高多视角任务的可靠性和安全性。

四、总结

针对可靠的多视角学习问题，本文提出了一种可信冲突多视角学习算法。该方法试图形成由信念质量向量和决策可靠性组成的特定视角的观点。它通过一个简单有效的平均池化层进一步聚合冲突观点。从理论上证明了该方法可以准确地模拟多视角公共可靠度和视角特定可靠度之间的关系。此外，本文对方法进行了扩展，通过最小化观点之间的冲突程度来保证不同观点之间结果的一致性。大量实验验证了所提出的方法在性能上的优越性以及其良好的稳定性和可靠性。

责任编辑 王金甲

参考文献

- [1] Cai Xu, Ziyu Guan, Wei Zhao, Hongchang Wu, Yunfei Niu and Beilei Ling. Adversarial incomplete multi-view clustering. In Proceedings of International Joint Conference on Artificial Intelligence(IJCAI), 2019.
- [2] Shudong Huang, Hongjie Wu, Yazhou Ren, Ivor Tsang, Zenglin Xu, Wentao Feng and Jiancheng Lv. Multi-view subspace clustering on topological manifold. In Neural Information Processing Systems(NeurIPS), 2022.
- [3] Jie Wen, Zheng Zhang, Lunke Fei, Bob Zhang, Yong Xu, Zhao Zhang and Jinxing Li. A survey on incomplete multiview clustering. In IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2022.
- [4] Nasrin Mostafazadeh, Chris Brockett, Bill Dolan, Michel Galley, Jianfeng Gao, Georgios P. Spithourakis and Lucy Vanderwende. Image-grounded conversations: Multimodal context for natural question and response generation. ArXiv preprint arXiv:1701.08251, 2017.
- [5] Ge Fan, Chaoyun Zhang, Kai Wang, and Junyang Chen. MV-HAN: A hybrid attentive networks based multi-view learning model for large-scale contents recommendation. In Proceedings of IEEE/ACM International Conference on Automated Software Engineering, 2022.
- [6] Yanchao Tan, Chengjun Kong, Leisheng Yu, Pan Li, Chaochao Chen, Xiaolin Zheng, Vicki S. Hertzberg, and Carl Yang. 4sdrug: Symptom-based set-to-set small and safe drug recommendation. In Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022.
- [7] Paul Pu Liang, Amir Zadeh and Louis-Philippe Morency. Foundations and Trends in Multimodal Machine Learning: Principles, Challenges, and Open Questions. ArXiv preprint arXiv:2209.03430, 2022.
- [8] Shuping Zhao, Jie Wen, Lunke Fei and Bob Zhang. Tensorized incomplete multi-view clustering with intrinsic graph completion. In Proceedings of the AAAI Conference on Artificial Intelligence, 2023.
- [9] Chaoyang Zhang, Zhengzheng Lou, Qinglei Zhou and Shizhe Hu. Multi-View Clustering via Triplex Information Maximization. In IEEE Transactions on Image Processing, 2023.
- [10] Jing Gao, Wei Fan, Deepak Turaga, Srinivasan Parthasarathy and Jiawei Han. A spectral framework for detecting inconsistency across multi-source object relationships. In IEEE International Conference on Data Mining, 2011.
- [11] Alejandro Marcos Alvarez, Makoto Yamada, Akisato Kimura, and Tomoharu Iwata. Clustering-based anomaly detection in multi-view data. In Proceedings of ACM international conference on Information & Knowledge Management, 2013.
- [12] Sheng Li, Ming Shao and Yun Fu. Multi-view low-rank analysis for outlier detection. In Proceedings of SIAM International Conference on Data Mining, 2015.
- [14] Handong Zhao, Hongfu Liu, Zhengming Ding and Yun Fu. Consensus regularized multi-view outlier detection. In IEEE Transactions on Image Processing, 2017. Kai Li, Sheng Li, Zhengming Ding, Weidong Zhang and Yun Fu. Latent discriminant subspace representations for multi-view outlier detection. In Proceedings of the AAAI Conference on Artificial Intelligence, 2018.
- [15] Dongdong Hou, Yang Cong, Gan Sun, Jiahua Dong and Jun Li; Kai Li. Fast multi-view outlier detection via deep encoder. In IEEE Transactions on Big Data, 2020.
- [16] Lampert C H, Krömer O. Weakly-paired maximum covariance analysis for multimodal dimensionality reduction and transfer learning. In ECCV 2010.
- [17] Xianchao Zhang, Linlin Zong, Xinyue Liu and Hong Yu. Constrained NMF-based multi-view clustering on unmapped data. In Proceedings of the AAAI Conference on Artificial Intelligence, 2015.

- [18] Zhenyu Huang, Peng Hu, Joey Tianyi Zhou, Jiancheng Lv and Xi Peng. Partially view-aligned clustering. In Neural Information Processing Systems(NeurIPS), 2020.
- [19] Mouxing Yang, Yunfan Li, Zhenyu Huang, Zitao Liu, Peng Hu and Xi Peng. Partially view-aligned representation learning with noise-robust contrastive loss. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition(CVPR), 2021.
- [20] Xianchao Zhang, Mengyan Chen, Jie Mu and Linlin Zong. Adaptive View-Aligned and Feature Augmentation Network for Partially View-Aligned Clustering. In Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2023.
- [21] Jakob Gawlikowski, Cedrique Rovile Njietcheu Tassi, Mohsin Ali, Jongseok Lee, Matthias Humt, Jianxiang Feng, Anna Kruspe, Rudolph Triebel, Peter Jung, Ribana Roscher, Muhammad Shahzad, Wen Yang, Richard Bamler and Xiao Xiang Zhu. A survey of uncertainty in deep neural networks. In Artificial Intelligence Review, 2023.
- [22] Murat Sensoy, Lance Kaplan and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In Neural Information Processing Systems(NeurIPS), 2018.
- [23] Yu Geng, Zongbo Han, Changqing Zhang and Qinghua Hu. Uncertainty-Aware Multi-View Representation Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, 2021.
- [24] Denker John and LeCun Yann. Transforming neural-net output levels to probability distributions. In Neural Information Processing Systems(NIPS), 1990.
- [25] Radford M. Neal. Bayesian training of backpropagation networks by the hybrid Monte Carlo method. Technical Report CRG-TR-92-1, Dept. of Computer Science, University of Toronto, 1992.
- [26] Hinton G E, Van Camp D. Keeping the neural networks simple by minimizing the description length of the weights. In Proceedings of annual conference on Computational learning theory, 1993.
- [27] Lakshminarayanan B, Pritzel A, Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. In Neural Information Processing Systems(NIPS), 2017
- [28] Shanmugam D, Blalock D, Balakrishnan G, et al. Better aggregation in test-time augmentation. In Proceedings of the IEEE/CVF international conference on computer vision, 2021.
- [29] Zongbo Han, Changqing Zhang, Huazhu Fu and Joey Tianyi Zhou. Trusted Multi-View Classification. In International Conference on Learning Representations, 2021.
- [30] Myong Chol Jung, He Zhao, Joanna Dipnall, Belinda Gabbe and Lan Du. Uncertainty estimation for multi-view data: The power of seeing the whole picture. In Neural Information Processing Systems(NeurIPS), 2022.
- [31] Wei Liu, Xiaodong Yue, Yufei Chen and Thierry Denoex. Trusted Multi-View Deep Learning with Opinion Aggregation. In Proceedings of the AAAI Conference on Artificial Intelligence, 2022.
- [32] Wei Liu, Yufei Chen, Xiaodong Yue, Changqing Zhang and Shaorong Xie. Safe Multi-View Deep Classification. In Proceedings of the AAAI Conference on Artificial Intelligence, 2023.
- [33] Qingyang Zhang, Haitao Wu, Changqing Zhang, Qinghua Hu, Huazhu Fu, Joey Tianyi Zhou and Xi Peng. Provable Dynamic Fusion for Low-Quality Multimodal Data. In Proceedings of International Conference on Machine Learning, PMLR, 2023.
- [34] Audun Jsang. Subjective Logic: A formalism for reasoning under uncertainty. Springer Publishing Company, Incorporated, 2018.

- [35] Perkins S, Theiler J. Online feature selection using grafting. In Proceedings of International Conference on Machine Learning(ICML), 2003.
- [36] He X, Peng Y. Fine-grained visual-textual representation learning. In IEEE Transactions on Circuits and Systems for Video Technology, 2019.
- [37] Wishart D S, Tzur D, Knox C, et al. HMDB: the human metabolome database. In Nucleic acids research, 2007.
- [38] Fei-Fei L, Perona P. A bayesian hierarchical model for learning natural scene categories. In IEEE computer society conference on computer vision and pattern recognition, 2005.
- [39] Fei-Fei L, Fergus R, Perona P. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In IEEE conference on computer vision and pattern recognition workshop, 2004.
- [40] Tulyakov S, Fitzgibbon A, Nowozin S. Hybrid vae: Improving deep generative models using partial observations. ArXiv preprint arXiv:1711.11566, 2017.
- [41] Sheng Wang, Jianfeng Lu, Xingjian Gu, Benjamin A. Weyori and Jing-yu Yang. Unsupervised discriminant canonical correlation analysis based on spectral clustering. In Neurocomputing, 2016.
- [42] Changqing Zhang, Zongbo Han, yajie cui, Huazhu Fu, Joey Tianyi Zhou and Qinghua Hu. CPM-Nets: Cross partial multi-view networks. In Neural Information Processing Systems(NIPS), 2019.



徐 偲

西安电子科技大学副教授，硕士生导师。主要研究方向为可信多模态深度学习。在 IEEE TPAMI、IEEE TNLS、IEEE TCyb、IEEE TII、NeurIPS、SIGKDD、AAAI、IJCAI、ACM MM 等中科院一区 IEEE 汇刊或 CCF A 类学术会议发表论文 22 篇，其中一作 9 篇，通讯 3 篇。主持国家自然科学基金青年基金项目等 5 项科研项目，作为核心骨干成员参与国家自然科学基金重点项目 2 项。担任 SCI 期刊 Array 编委，SCI 期刊 Sensors 和 Mathematics 的客座编委，IEEE TPAMI、IEEE TKDE、ICML、AAAI 等 20 余个中科院一区期刊或 CCF A 类会议的审稿人。

Email: cxu@xidian.edu.cn



司佳俊

西安电子科技大学计算机科学与技术学院 2021 级硕士研究生，导师为赵伟教授，主要研究方向为可信多模态深度学习。

Email: jiajungsi@stu.xidian.edu.cn



管子玉

西安电子科技大学教授，博士生导师，曾获国家级青年人才项目。主持基金委重点项目、面上项目等国家级项目。在数据挖掘、信息检索、数据管理等领域的顶级国际会议和期刊发表论文 70 余篇，包括 TKDE、TPAMI、TNNLS、TIP、VLDB、SIGMOD、SIGIR、ICDE、WWW、AAAI、IJCAI、SIGKDD、CVPR 等。担任 TKDE、Neurocomputing 和 International Journal of Machine Learning and Cybernetics 编委，担任多个高水平国际会议程序委员会委员/资深委员，如 SIGKDD、IJCAI、AAAI、NeurIPS、SIGIR、CIKM 等，担任领域内知名国际会议环太平洋多媒体会议 PCM 2016 组织主席。

Email: zyguan@xidian.edu.cn



赵伟

西安电子科技大学教授，博士生导师。主要从事智能媒体计算和机器学习等方面的研究工作。近五年，以第一作者/通讯作者在相关领域国际著名期刊及学术会议发表论文 20 余篇，如：IEEE TPAMI、IEEE TKDE、NeurIPS、AAAI、SIGKDD 等。担任 CCF A 类期刊 IEEE TKDE 和中科院一区期刊 IEEE TNNLS 编委，常年担任领域内顶级会议和期刊审稿人。研究成果获浙江省自然科学一等奖、陕西省自然科学优秀学术论文二等奖等。

Email: ywzhao@mail.xidian.edu.cn