**报告正文**

参照以下提纲撰写，要求内容翔实、清晰，层次分明，标题突出。**请勿删除或改动下述提纲标题及括号中的文字。**

**（一）立项依据与研究内容**（**建议8000字以下**）：

1．**项目的立项依据**（研究意义、国内外研究现状及发展动态分析，需结合科学研究发展趋势来论述科学意义；或结合国民经济和社会发展中迫切需要解决的关键科技问题来论述其应用前景。附主要参考文献目录）；

# 1.1研究意义

多模态学习旨在通过联合不同类型的数据模态，并利用其间的相互关系和信息交互来提高学习任务的效果。例如，在医疗辅助诊断中，系统通过化验指标、影像等患者的多模态数据来诊断病情；在自动驾驶中，车辆通过摄像头、激光雷达等多个传感器收集到的多模态数据感知周围环境。**在2017年国务院印发的《新一代人工智能发展规划》中将“多模态（媒体）学习”的相关研究拔高到国家战略层面**，要求“重点突破跨媒体统一表征、关联理解与知识挖掘、知识图谱构建与学习、知识演化与推理、智能描述与生成等技术，实现跨媒体知识表征、分析、挖掘、推理、演化和利用，构建分析推理引擎。”在**《国家自然科学基金“十四五”发展规划》**的“大数据与交互计算技术”、“精准探测与信息融合处理”、“新型网络及网络安全”、“跨时空、跨尺度生物分子事件探测与解析”、“多学科交叉新型诊疗技术”等多个优先发展领域中，**明确指出要研究**“大数据融合”、“多源融合探测成像”、“基于多模态影像的个体化手术规划”等**多模态学习相关的技术**。因此，多模态学习是迈向强人工智能的必经之路，对计算机学科乃至整个科学领域都具有重大意义。

传统的多模态学习大都只考虑“实验室”数据环境下，即假设数据是高质量的且具有完善的标注，其目标主要在于提升决策结果的精确度。然而，随着多模态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[1,2]。相较于 “实验室”数据环境，复杂开放环境中具有更多的未知性，其中重要一点就是数据的质量难以保证，从而导致传统面向实验室环境的研究的假设难以满足[3]。因而，如何使多模态学习适应开放环境成为当前重要的研究课题。另外，复杂开放环境应用大都要求模型提供的结果是可靠的，以便用户和利益相关者能够理解和信任这些结果，并根据它们做出决策[4,5]。尽管相关领域学者已经面向多种复杂开放环境开展相关研究，但仍未形成系统的理论方法。面向复杂开放环境的多模态学习相比于传统面向“实验室”数据环境的多模态学习有如下新的挑战：

**（1）数据质量难以保证**。在复杂开放环境中，由于数据采集不规范、不可控等问题，导致数据质量低。由于每个模态的数据资源都可能存在问题，因此多模态数据加剧了这一问题。因此，面向复杂开放环境的多模态应用时，数据质量问题亟需解决，具体包括三个方面：1）模态随机缺失。例如在多模态医疗辅助诊断系统中，并非每个患者都会进行化验、影像、病理等完整的检查。此时现有针对完整多模态数据的方法性能会急剧下降[6]；2）模态维度非平衡。例如影像的维度通常接近107，而化验指标数据的维度只有10维左右。现有方法在融合非平衡数据时容易忽略低维模态信息、并学习到高维模态的冗余信息[7]；3）模态内容有冲突。例如由于设备噪声、病情差异等原因，化验和病理结果可能会诊断出不同的病情。现有方法融合冲突模态时将会得到模糊的信息，这一信息与所有模态体现的内容都存在差异。此时，多模态模型性能反而弱于单模态模型，严重影响决策结果的鲁棒性[8]。

**（2）决策风险需要评估**。复杂开放环境的决策任务通常在决策错误时会产生较大代价，如生命、财产损失等。因此，复杂开放环境往往要求人工智能模型必须安全可信。考虑到复杂开放环境的数据质量难以保证，这进一步增加了评估决策风险、降低决策不确定度的重要性。观察到人在面对高风险场景时，会综合多个器官感知的信息以提升决策置信度，而多模态学习人的决策过程类似，天然适合提升决策的置信度，有助于缓解决策风险高这一难题。但现有可信多模态学习仍处于起步阶段，缺乏针对不同类型低质量多模态数据的可信学习研究。特别地，现有方法一般都依赖高质量标签，而实际开放场景难以获取足够多的高质量标注数据。此时，标签噪声将导致现有可信多模态学习方法的决策置信度的失准[9]。

**针对上述挑战，本项目拟研究复杂开放环境下可信多模态学习方法，重点解决“模态随机缺失”、“模态维度非平衡”、“模态内容有冲突”和“数据标签错误”四大关键问题，旨在提升复杂开放环境下多模态学习方法的泛化性、鲁棒性、安全性、自适应性。本项目还计划在团队现有平台积累的基础上**，将相关研究成果**应用到上海交通大学医学院附属瑞金医院的医疗辅助诊断系统**，提升诊断结果的可信度；以及应用到**阿里巴巴电商推荐平台**，增加平台效益，具有较强的社会意义和经济意义。此外，本项目**拟构建并开源2-3个低质量多模态数据集**，有助于解决其他基于多模态学习相关的应用问题（如自动驾驶[10]等），将对多模态学习乃至可信人工智能领域的发展起到积极推动作用。

# 1.2国内外研究现状及发展动态分析

## 1.2.1 缺失多模态学习

现实场景中收集到的多模态数据经常存在模态缺失现象。例如，在社交媒体平台上，用户发布的内容可以包含文字、图片、视频等多种模态，但并非每条内容都会包含所有模态。近年来，缺失多模态学习问题引起了广泛关注[11, 12]。早期的解决思路基于矩阵分解、图学习等方法，借助非缺失模态来捕获模态间或样本间的关联，进而缓解缺失模态的影响，其中的代表性方法包括：（1）基于矩阵分解的缺失多模态学习方法。该类方法旨在利用非负矩阵分解来挖掘缺失多模态数据的协同表征，以此恢复缺失模态数据。Hu等人[13]提出基于加权半非负矩阵分解的方法，通过强制多个模态的基矩阵进行相互对齐，来减少模态缺失对聚类的影响；（2）基于多核学习的缺失多模态学习方法。该类方法首先构造模态私有核矩阵，接着借助核矩阵学习模态协同表征。Liu等人[14]通过增强最优核的可表征性，提升了核学习与聚类之间的协调性；（3）基于图学习的缺失多模态学习方法。这类方法旨在从缺失数据中自适应地学习多个图，接着借助多图融合来缓解数据缺失的影响。Zhao等人[15]考虑协同表征中全局结构紧凑性，利用完整数据样本上的拉普拉斯图，来桥接完整模态的样本和部分模态的样本。这些方法在探索高维复杂数据的特征表示方面存在局限性，并且严重依赖于原始特征的质量。

近年来，由于深度神经网络强大的表示能力和非线性建模能力，有助于捕获模态间的多层级复杂关联，因而被广泛应用于缺失多模态学习任务中。现有的基于深度学习的缺失多模态学习方法可分为以下四类: （1）基于图卷积神经网络的缺失多模态学习方法。该类方法旨在深度融合多个模态的不完全图以获得一个共识图或协同表征，并基于此进行下游任务。Gerome等人[16]使用多个递归图卷积网络分析每个模态的关联图，并基于自注意力机制进行多图融合；（2）基于自编码器的缺失多模态学习方法。这类方法通过提取已有模态数据的特征，借助深度自编码器学习协同表征来填补缺失数据，并获得了良好的性能。Hamghalam等人[17]利用变分自编码器来计算可用模态和所有样本模态之间的相关性，进而学习模态生成模型。Wen等人[18]学习了多个模态私有卷积编码器，以得到数据的高层特征和高阶几何结构信息，籍此构建最优聚类指标矩阵；（3）基于生成对抗网络的缺失多模态学习方法。这类方法通过生成器补全缺失数据，并通过判别器来促使补全的数据和原始的数据尽可能接近，从而提升补全效果。Xu等人[19]通过结合元素重建目标和语义重建目标来推断缺失的数据。Nikhilanand等人[20]利用不同模态之间的互补知识来学习模态协同表征，并构建多个模态私有生成对抗网络生成缺失的模态数据；（4）基于对比学习的缺失多模态学习方法。该类方法通过多个模态之间的自监督对比来提升补全效果。例如，Lin等人[21]通过对比学习最大化不同模态的互信息，并通过对偶预测最小化条件熵来恢复缺失的模态。

虽然现有的多模态补全方法已经取得了良好的效果，但其补全内容的可靠性仍难以保证，并且这些方法通常建立在模态无冲突的基础之上。但是，**真实多模态数据往往存在模态冲突现象，导致现有方法进行了错误的补全，进而产生整体样本的信息偏差，从而严重影响下游任务的效果**。因此，如何可信地补全缺失多模态数据值得进一步研究。本项目拟结合原型学习和证据神经网络，研究可信的多模态补全方法，进而提高下游多模态学习任务的鲁棒性和可靠性。

## 1.2.2 非平衡多模态学习

非平衡多模态数据广泛存在现实生活中，例如在医学领域，医生通常利用电子计算机断层扫描图（Computed Tomography，CT）和血液检测数据来诊断疾病，CT的维度通常接近107，而血液检测指标数据的维度只有10维左右；互联网中的多模态UGC数据，也包含大量的不平衡的文本、图片数据。这种模态数据维度不平衡的现象给多模态学习带来了新的挑战。以往多模态表征学习工作都假定不同模态数据间的维度是平衡的[22,23,24]，这种不平衡的问题使得这些方法在模态融合过程中不仅容易忽略低维模态信息，而且还容易受到高维模态的冗余信息的影响。事实上，虽然低维模态维度低，但其却蕴含着重要的互补信息，例如血液检测数据包含着电子计算机断层扫描图没有的细胞含量等信息。

当前，围绕非平衡多模态学习主要有三类方法：（1）基于深度神经网络（Deep Neural Networks，DNNs）的平衡模态表征方法。该方法通常构造编码器来学习各模态维度平衡的高层特征，实现模态间高层特征的对齐[25，26]。然而这种直接使用DNNs对高维模态进行降维、对低维模态进行升维的方法会损失高维模态的信息，同时也造成了低维模态高层特征信息冗余。当维度相差很大时，低维模态的编码器会产生大量的冗余参数，从而导致网络过拟合。Xu[27]等人通过字典学习升维低维的模态维度，再使用DNNs对齐多模态，很好地避免了低维模态的DNNs过拟合问题。（2）多模态决策级融合方法。为了避免不平衡问题带来的影响，一些研究人员直接学习融合多模态的决策信息，例如不同模态的分类、聚类结果。Han[28]等人通过模态的可信决策，获得对应模态的所有类别的证据，然后在融合阶段重用具有高证据的模态。然而，基于决策的方法不能很好地学习各模态间相关性，导致不同模态之间的交互仅仅局限于类别空间。此外，决策级融合方法侧重于与类别相关的任务[29]，如分类和聚类，并不能很好地应用于分割等下游任务。（3）多模态样本相似度融合方法。这类方法通过计算模态内样本间的相似度[30,31]，得到平衡的多模态的相似图，利用自监督的学习方法得到多模态的综合表征，同时也保留了样本间的相似性。Zhang等人[32]构建平衡的特定模态的相似图，并执行图嵌入算法来学习特定模态的表征，然后拼接特定模态的表征得到综合表征。这类方法同样面临着和决策级融合方法一样的问题，不能充分挖掘模态间的信息，且适用的下游任务有限。此外，当模型遇到新的样本时，这种自监督模型需要重新训练。

虽然上述方法可以一定程度上避免不平衡的问题，但是仍存在以下限制：**（1）决策级融合方法和相似图融合方法局限于分类、聚类任务，方法不能应用到分割任务等；（2）直接应用DNNs容易造成低维模态的编码器网络出现过拟合的问题；（3）不能充分地挖掘模态间的一致性和互补性的信息，且模态间交互浅，有的工作只局限于类别空间和相似度空间的对齐**。因此如何实现特征级的多模态对齐融合、避免低维模态的编码器网络出现过拟合问题值得进一步的研究。本项目拟研究特征级的不平衡多模态融合对齐方法，并在融合过程中充分挖掘模态高级特征间的一致性和互补性的关系。

## 1.2.3可信多模态学习

多模态学习虽然通过综合多个模态信息提升了模型的性能表现，然而在面对某些模态数据质量较差的情况时，模型往往会产生不可靠的决策结果，这导致多模态学习无法部署到医疗诊断、自动驾驶等高风险应用中。

为了提高模型决策结果的可信性，一种较为有效的方法是对决策结果的不确定性进行量化[33,34]。对不确定性进行的方法主要分为两类，一类是基于贝叶斯方法的不确定性量化方法，一类是基于非贝叶斯的不确定性量化方法。基于贝叶斯方法通过学习模型参数在分布表征下的不确定性来衡量模型决策的可靠程度。其中较为经典的贝叶斯方法包括拉普拉斯近似[35]，马尔可夫链蒙特卡罗算法[36]和变分推理[37]等。然而，在模型中设定权值（模型参数）的分布以及模型训练的收敛性方面存在一些难题，贝叶斯方法的计算成本非常高昂。为了避免显式学习参数的分布时成本过高，研究者提出多种非贝叶斯方法，包括深度集成训练多个深度网络[38]等，并取得了良好的性能。为了进一步减少模型的参数，证据深度学习方法[39]逐渐得到学术界的关注。该方法没有通过网络权值来进行不确定性建模，而是引入主观逻辑理论进行直接建模，因此模型也更加地轻量高效。

单一模态的可信学习无法有效利用现实场景中充分存在的多模态数据的一致性和互补性信息，因此相关学者将证据深度学习方法引入了多模态学习框架中[40]，可信多模态学习这一领域开始逐渐受到关注。可信多模态学习旨在聚合多个模态信息的同时能够量化结果的可信性，使模型做出可信决策。现有的可信多模态学习方法[41]大多都是通过证据深度学习（EDL, Evidence Deep Learning）获得各个模态的证据，借助主观逻辑框架将特定模态的证据转化为意见，最终利用Dempster-Shafer证据理论对不同模态的意见进行融合。需要注意的是，这类方法大都基于一个严格的前提：多个模态数据之间是无冲突的。然而，从现实世界中获得的多模态数据之间往往会存在冲突的情况。在这种情况下，冲突模态会导致模型生成模糊意见，进而导致不确定性量化的失效[8]。针对这类冲突数据，现有的方法主要分为两类：（1）冲突消除方法[42,43]。该方法主要致力于消除冲突的模态数据，包括将冲突数据视为异常进行消除，以及为原始冲突数据学习新的对齐关系，并据此构建新的模态数据来消除冲突。然而，在开放环境的可信应用是需要对冲突的模态数据也做出决策，甚至从这些冲突模态数据中学到可用信息，简单地消除冲突数据会导致部分信息的缺失。（2）一致性最大化方法[44]。这种方法通过最大化模态间的一致性信息来实现可信融合，从而消除互补性信息带来的冲突影响。然而，这种方法忽略了意见间的互补关系，在一致性信息不足时无法做出可信决策。

虽然存在上述关于冲突模态融合方法的研究，但这些方法都**过于依赖模态一致性，而忽略了模态间高度互补的信息，导致模型无法做出准确且全面的不确定性量化，无法满足“可信”的多模态学习需求**。为此，本项目拟对决策级的一致性证据和互补性证据进行解耦，明确区分证据层的一致性和互补性内容，通过充分利用模态间的一致性和互补性信息，消解冲突的多模态数据，并分别自适应地感知并融合一致性和互补性证据，实现决策级的冲突模态数据可信融合，进而提高多模态学习在模态内容冲突情况下的可靠性。

## 1.2.4噪声标签学习

现有的可信多模态学习方法[45][46]大都依赖于高质量标签。实际开放场景由于不断有大规模数据产生，且学习任务繁多，因此难以获取足够多的包含高质量标签的数据。因此，噪声标签学习引起了学术界的广泛关注[47]。现有工作大多将标签噪声分为类别相关噪声和样本相关噪声[48][54][55]：类别相关噪声表示一个类中的每个样本被错误标注为另一个类的概率是固定的，而样本相关噪声则表示一个样本被错误标注的概率不仅取决于它所属的类，还与其自身特征有关。

现有的噪声标签方法主要分为三类：（1）基于结构优化的噪声适应方法[49]。这类方法通过建模针对噪声数据的标签转移概率矩阵，设计专用的噪声适应层，借助该标签转移概率修改神经网络输出。但该类方法对所有样本的预测标签都进行了修改，缺乏识别标签错误的样本的能力，会对标签正确的样本产生负面影响。（2）基于正则化的噪声适应方法[50，51]。通常来说，显示和隐式正则化方法能提高各类学习模型的鲁棒性，同时避免模型训练中的过拟合，提高模型对噪声的耐受性。同时由于其灵活性与易用性，在各类场景中得到了广泛的使用。然而，显式正则化方法会引入敏感的超参数，需要进行细致的调整；而隐式正则化方法扩展的特征或标签空间会减缓训练的收敛速度。（3）基于损失调整的噪声识别方法。该类方法使用估计的噪声转移矩阵来校正预测损失[52,53]，可以有效地减少噪声标签对模型训练产生的负面影响。与第（1）类方法的区别在于，其转移概率的学习与网络模型的学习是异步的。该类方法高度依赖于转移矩阵估计的精确程度。通常需要先验知识（如无噪锚点集）来增加估计精确度。

尽管上述方法可以缓解噪声标签造成的预测性能下降，但噪声标签在模型训练过程带来的模型不确定度仍然不可避免且难以度量。**标签噪声导致的模型不确定度使得现有可信多模态学习方法对总体决策置信度的度量失准，无法正确反映模态数据的质量，从而影响各模态间意见融合过程，难以确保聚合意见的正确性**。因此，需要对总体决策不确定度进行解耦，准确量化由标签噪声带来的模型不确定度与样本自身存在的数据不确定度。为此，本项目拟研究真实噪声标签环境下的可信多模态学习方法，旨在使该环境下训练的模型具备准确且可靠的决策结果。

综上所述，尽管多模态学习已经得到了广泛的关注，并在众多相关问题上都表现出了优秀的效果。但是现在多模态学习方法大都面向实验室场景，应用到真实开放场景时仍面临着很大的挑战。**具体来说，现实场景的开放性与多样性使得多模态数据质量难以保证，在实际应用中，可能会面临模态数据随机缺失、模态维度非平衡以及模态内容有冲突等问题，而现有的多模态学习算法难以解决这些问题。除此之外，在上述的低质量训练数据的背景下，模型决策的风险和可靠性至关重要，尽管当前在不确定性估计上已有相关工作进行了初步探索，但大多都只聚焦于单个模态数据，因此仍须对多个模态情况下的决策可靠性进行探究。**因此，本项目将在上述两个方面开展研究工作，针对复杂开放场景的多模态学习提出创新的解决方案，促进多模态学习领域的发展。

# 参考文献

[1] Fei Wu, Cewu Lu, Mingjie Zhu, Hao Chen, Jun Zhu, Kai Yu, Lei Li, Ming Li, Qianfeng Chen, Xi Li, Xudong Cao, Zhongyuan Wang, Zhengjun Zha, Yueting Zhuang, and Yunhe Pan. Towards a new generation of artificial intelligence in China[J]. Nature Machine Intelligence, vol. 2, pp. 312-316, 2020.

[2] Michael Moor, Oishi Banerjee, Zahra Shakeri Hossein Abad, Harlan M. Krumholz, Jure Leskovec, Eric J. Topol, and Pranav Rajpurkar. Foundation models for generalist medical artificial intelligence[J]. Nature, vol. 616, pp. 259-265, 2023.

[3] 袁晓彤,张煦尧,刘希,刘成林.面向开放环境的机器学习理论研究进展[J]. 模式识别与人工智能, vol.36, pp. 1059-1071, 2023.

[4] Jakob Gawlikowski, Cedrique Rovile Njieutcheu 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[J]. Artificial Intelligence Review, vol. 56, pp. 1513-1589, 2023.

[5] Davinder Kaur, Suleyman Uslu, Kaley J. Rittichier, and Arjan Durresi. Trustworthy artificial intelligence: a review[J]. ACM Computing Surveys (CSUR), vol. 55, pp. 1-38, 2022.

[6] Reuben Dorent, Nazim Haouchine, Fryderyk Kogl, Samuel Joutard, Parikshit Juvekar, Erickson Torio, Alexandra J. Golby, Sebastien Ourselin, Sarah Frisken, Tom Vercauteren, Tina Kapur, and William M. Wells III. Unified brain MR-ultrasound synthesis using multi-modal hierarchical representations[C]//International conference on medical image computing and computer-assisted intervention. 2023: 448-458.

[7] Arsha Nagrani, Shan Yang, Anurag Arnab, Aren Jansen, Cordelia Schmid, and Chen Sun. Attention bottlenecks for multimodal fusion[C]//Advances in Neural Information Processing Systems. 2021: 14200–14213.

**[8] Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao, Yue Wu, and Xiyue Gao. Reliable Conflictive Multi-View Learning[C]// In Proceedings of the AAAI Conference on Artificial Intelligence. 2024.**

[9] Lucas Fidon, Michael Aertsen, Florian Kofler, Andrea Bink, Anna L. David, Thomas Deprest, Doaa Emam, Fred eric Guffens, Andras Jakab, Gregor Kasprian, Patric Kienast, Andrew Melbourne, Bjoern Menze, Nada Mufti, Ivana Pogledic, Daniela Prayer, Marlene Stuempflen, Esther Van Elslander, Sebastien Ourselin, Jan Deprest, and Tom Vercauteren. A Dempster-Shafer Approach to Trustworthy AI with Application to Fetal Brain MRI Segmentation[J]. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 11, pp. 12878-12895, 2023.

[10] Kashyap Chitta, Aditya Prakash, Bernhard Jaeger, Zehao Yu, Katrin Renz, and Andreas Geiger. Transfuser: Imitation with transformer-based sensor fusion for autonomous driving[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 11, pp. 12878-12895, 2023

[11] 李学龙.多模态认知计算[J].中国科学:信息科学, vol. 53, no.01, pp. 1-32, 2023.

[12] 吴友政,李浩然,姚霆,何晓冬.多模态信息处理前沿综述：应用、融合和预训练[J].中文信息学报, vol.36, no.05, pp.1-20, 2022.

[13] Menglei Hu and Songcan Chen. Doubly aligned incomplete multi-view clustering[C]//In Proceedings of the International Joint Conference on Artificial Intelligence. 2018: 2262-2268.

[14] Xinwang Liu, Sihang Zhou, Yueqing Wang, Miaomiao Li, Yong Dou, En Zhu, and Jianping Yin. Optimal neighborhood kernel clustering with multiple kernels[C]//In Proceedings of the AAAI Conference on Artificial Intelligence. 2017: 2266-2272.

[15] Handong Zhao, Hongfu Liu, and Yun Fu. Incomplete multi-modal visual data grouping[C]//In Proceedings of the International Joint Conference on Artificial Intelligence. 2016: 2392-2398.

[16] Gerome Vivar, Anees Kazi, Hendrik Burwinkel, Andreas Zwergal, Nassir Navab, Seyed-Ahmad Ahmadi. Simultaneous imputation and classification using Multigraph Geometric Matrix Completion (MGMC): Application to neurodegenerative disease classification[J]. In Artificial Intelligence in Medicine, vol.117, pp. 102097, 2021.

[17] Hamghalam, M., Frangi, A.F., Lei, B., Simpson, A.L. Modality Completion via Gaussian Process Prior Variational Autoencoders for Multi-modal Glioma Segmentation[J]. In Medical Image Computing and Computer Assisted Intervention, vol 12907, pp. 442-452, 2021.

[18] Jie Wen, Zhihao Wu, Zheng Zhang, Lunke Fei, Bob Zhang, and Yong Xu. Structural Deep Incomplete Multi-view Clustering Network[C]//In Proceedings of the ACM International Conference on Information & Knowledge Management. 2021: 3538-3542.

**[19] Cai Xu, Ziyu Guan, Wei Zhao, Hongchang Wu, Yunfei Niu, and Beilei Ling. Adversarial Incomplete Multiview Clustering[C]//In Proceedings of the International Joint Conference on Artificial Intelligence. 2019: 3933-3939.**

[20] Nikhilanand Arya and Sriparna Saha. Generative Incomplete Multi-View Prognosis Predictor for Breast Cancer: GIMPP[J]. In IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 19, no. 4, pp. 2252-2263, 2021.

[21] Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv and Xi Peng. Dual Contrastive Prediction for Incomplete Multi-View Representation Learning[J]. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 4, pp. 4447-4461, 2023.

[22] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, [Ilya Sutskever](https://dblp.uni-trier.de/pid/60/5276.html). Learning transferable visual models from natural language supervision[C]//In International Conference on Machine Learning. 2021: 8748-8763.

[23] Galen Andrew, Raman Arora, Jeff A. Bilmes, Karen Livescu. Deep canonical correlation analysis[C]//In International Conference on Machine Learning. 2013: 1247-1255.

[24] Arsha Nagrani, Shan Yang, Anurag Arnab, Aren Jansen, Cordelia Schmid, Chen Sun. Attention bottlenecks for multimodal fusion[C]//Advances in Neural Information Processing Systems. 2021: 14200-14213.

**[25] Cai Xu, Wei Zhao, Jinglong Zhao, Ziyu Guan, Yaming Yang, Long Chen, Xiangyu Song. Progressive deep multi-view comprehensive representation learning[C]//In Proceedings of the AAAI Conference on Artificial Intelligence. 2023: 10557-10565.**

**[26] 徐偲. 基于深度学习的多视角表达学习技术研究[D]. 西安电子科技大学, 2022.**

**[27] Cai Xu, Zehui Li, Ziyu Guan, Wei Zhao, Xiangyu Song, Yue Wu, Jianxin Li. Unbalanced Multi-view Deep Learning[C]//In Proceedings of the 31st ACM International Conference on Multimedia. 2023: 3051-3059.**

[28] Zongbo Han, Changqing Zhang, Huazhu Fu, Joey Tianyi Zhou. Trusted Multi-View Classification With Dynamic Evidential Fusion[J]. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 2, pp. 2551-2566, 2023.

[29] Xinwang Liu, Xinzhong Zhu, Miaomiao Li, Lei Wang, Chang Tang, Jianping Yin, Dinggang Shen, Huaimin Wang, Wen Gao. Late Fusion Incomplete Multi-View Clustering[J]. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 10, pp. 2410-2423, 2019.

[30] Ruihuang Li, Changqing Zhang, Qinghua Hu, Pengfei Zhu, and Zheng Wang. Flexible Multi-View Representation Learning for Subspace Clustering[C]//In International Joint Conference on Artificial Intelligence.2019:2916–2922.

[31] Bin Zhang, Qianyao Qiang, Fei Wang, Feiping Nie. Flexible Multi-View Unsupervised Graph Embedding[J]. In IEEE Transactions on Image Processing, vol. 30, pp. 4143-4156, 2021.

[32] Changqing Zhang, Huazhu Fu, Qinghua Hu, Pengfei Zhu, Xiaochun Cao. Flexible Multi-View Dimensionality Co-Reduction[J]. In IEEE Transactions on Image Processing, vol. 26, no. 2, pp. 648-659,2017.

[33] Charpentier B, Zügner D, Günnemann S. Posterior network: Uncertainty estimation without ood samples via density-based pseudo-counts[C]//In Advances in Neural Information Processing Systems.2020: 1356-1367.

[34] Kendall A, Gal Y. What uncertainties do we need in bayesian deep learning for computer vision[C]//In Advances in neural information processing systems. 2017:5574-5584.

[35] MacKay D J C. A practical Bayesian framework for backpropagation networks[J]. In Neural computation, vol 4, pp. 448-472, 1992

[36] Neal R M. Bayesian learning for neural networks[J]. In Springer Science & Business Media, vol. 56, pp. 11773-11823, 2012.

[37] Alex Graves. Practical variational inference for neural networks[C]//In Advances in neural information processing systems. 2011:2348--2356.

[38] Balaji Lakshminarayanan，Alexander Pritzel，Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles[C]//In Advances in neural information processing systems. 2017: 6402-6413.

[39] Murat Sensoy，Lance M. Kaplan，Melih Kandemir. Evidential deep learning to quantify classification uncertainty[C]//In Advances in neural information processing systems.2018:3183-3193.

[40] 张北辰, 李亮, 查正军, 黄庆明. 基于跨模态对比学习的视觉问答主动学习方法[J]. 计算机学报, vol. 45, pp. 1730-1745, 2022

**[41] Cai Xu, Wei Zhao Wei, Jinglong Zhao, Ziyu Guan, Xiangyu Song, Jianxin Li, Uncertainty-Aware Multiview Deep Learning for Internet of Things Applications[J]. In IEEE Transactions on Industrial Informatics, vol. 19, no. 2, pp. 1456-1466, 2023.**

[42] Marcos Alvarez, A.; Yamada, M.; Kimura, A.; and Iwata, T. Clustering-Based Anomaly Detection in Multi-View Data[C]//In Proceedings of the ACM International Conference on Information & Knowledge Management, pp. 1545-1548, 2013

[43] Hou Dongdong, Cong Yang, Sun Gan, Dong Jiahua, Li Jun, Li Kai. Fast Multi-View Outlier Detection via Deep Encoder[J]. In IEEE Transactions on Big Data, vol.8, pp. 1047-1058, 2022.

[44] Wei Liu, Xiaodong Yue, Yufei Chen, and Thierry Denoeux. Trusted multiview deep learning with opinion aggregation[C]// In Proceedings of the AAAI Conference on Artificial Intelligence. 2022:7585-7593.

[45] Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification[C]// In International Conference on Learning Representations, 2020.

[46] Wei Liu, Xiaodong Yue, Yufei Chen, and Thierry Denoeux. Trusted multi-view deep learning with opinion aggregation[C]//In Proceedings of the AAAI Conference on Artificial Intelligence. 2022: 7585– 7593.

**[47] Wei Zhao, Cai Xu, Ziyu Guan, Xunlian Wu, Wanqing Zhao, Qiguang Miao, Xiaofei He and Quan Wang, TelecomNet: Tag-Based Weakly-Supervised Modally Cooperative Hashing Network for Image Retrieval[J]. In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 11, pp. 7940-7954, 2022.**

[48] Hwanjun Song, Minseok Kim, Dongmin Park, Yooju Shin and Jae-Gil Lee. Learning from noisy labels with deep neural networks: A survey[J]. IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 11, pp. 8135-8153, 2023.

[49] Alan Joseph Bekker and Jacob Goldberger. Training deep neural-networks based on unreliable labels[C]//2016 IEEE International Conference on Acoustics, Speech and Signal Processing. 2016: 2682-2686.

[50] Simon Jenni and Paolo Favaro. Deep bilevel learning[C]//Proceedings of the European conference on computer vision. 2018: 632-648.

[51] Ryutaro Tanno, Ardavan Saeedi, Swami Sankaranarayanan, Daniel C. Alexander and Nathan Silberman. Learning from noisy labels by regularized estimation of annotator confusion[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 11244-11253.

[52] Xiaobo Xia, Tongliang Liu, Nannan Wang, Bo Han, Chen Gong, Gang Niu and Masashi Sugiyama. Are anchor points really indispensable in label-noise learning?[C]//Advances in neural information processing systems, 2019: 6835-6846.

[53] 魏琦,孙皓亮,马玉玲,尹义龙.面向标签噪声的联合训练框架[J].中国科学:信息科学, vol. 54, no. 1, pp. 144-158, 2024.

[54] De Cheng, Tongliang Liu, Yixiong Ning, Nannan Wang, Bo Han, Gang Niu, Xinbo Gao and Masashi Sugiyama. Instance-dependent label-noise learning with manifold-regularized transition matrix estimation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 16630-16639.

[55] Antonin Berthon, Bo Han, Gang Niu, Tongliang Liu and Masashi Sugiyama. Confidence scores make instance-dependent label-noise learning possible[C]//Proceedings of the 38th International Conference on Machine Learning, 2021: 825-836.

2．**项目的研究内容、研究目标，以及拟解决的关键科学问题**（此部分为重点阐述内容）**；**

# 2.1研究内容

在充分考虑复杂开放环境中多模态数据存在的不完整、不平衡、有冲突等特性的基础上，本项目拟从四个方面开展可信多模态学习方法研究工作：首先，研究多模态可信补全方法，为后续任务提供完整且可靠的多模态数据；接着，研究特征级的多模态数据自适应平衡对齐方法，以挖掘模态间高层特征的语义级关联；随后，研究决策级多模态可信融合旨在提高决策性能的同时，可靠地评估冲突多模态数据的决策置信度；最后，研究噪声标签环境下的可信多模态学习方法，进一步提升多模态模型在真实环境中的决策可信度。研究内容框架如图 1 所示。

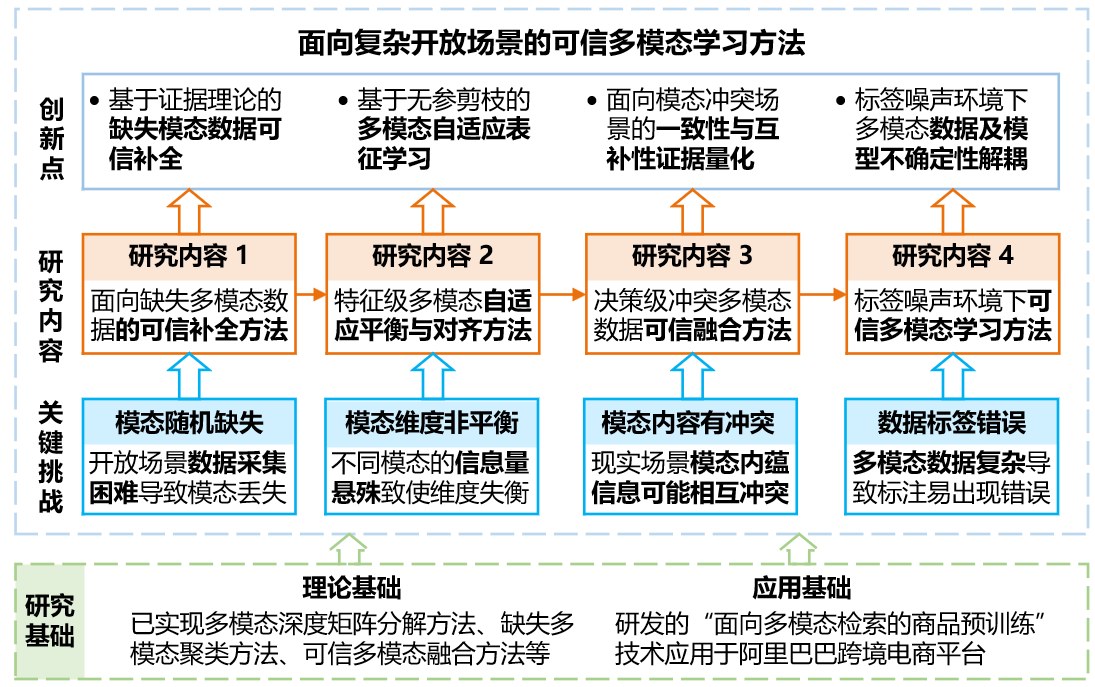


图 1面向复杂开放场景的可信多模态学习方法研究内容框架图

## 2.1.1多模态可信补全

在现实世界中，多模态数据普遍存在数据缺失问题，导致现有针对完整多模态数据的算法失效。面对模态随机缺失这一挑战性问题，其关键在于挖掘模态间的一致性内容，进而精准补全缺失的数据内容。然而，现有补全方法所生产的缺失模态内容并非总是可靠的，尤其当不同模态间存在冲突时，补全过程可能由于参考了误导性的信息而产生错误，进而对数据质量产生了严重的负面影响。

因此，本项目拟解决可信的缺失多模态学习问题，要求模型能够可信地生成缺失模态内容。为此，本项目拟研究一种基于证据理论的可信缺失多模态补全方法。该方法拟结合信念质量向量和决策可靠性，形成多模态类原型意见，并基于此衡量和提升补全内容的可信度，实现可信补全，为后续任务提供完整可靠的多模态数据。

## 2.1.2特征级多模态数据自适应平衡及对齐

随着深度学习技术的快速发展，多模态表征学习取得了很大的进步。其中，特征级多模态的对齐表征学习旨在利用深度神经网络挖掘特定模态的高级特征，再通过加权融合的方法计算其综合表征。当数据维度不平衡时，大多数现有的对齐表征方法都使用DNNs来直接对齐低维模态和高维模态，这会损失高维模态的信息以及造成低维模态DNNs过拟合。由于低维模态的信息少，现有方法都忽略了从低维模态特征到多模态综合表征的映射是稀疏的。此外，现有解决不平衡的方法，需要人为指定低维模态和高维模态，方法不灵活。

因此，本项目拟研究一种特征级的多模态融合对齐方法。由于对齐后的特征应该包含完整的信息，而特定模态的高级特征包含部分信息，故从部分信息到完整信息的映射应该是稀疏的。为此，本项目拟建立一个多模态稀疏批处理归一化层来灵活地来捕获多模态数据的互补性和一致性信息。对于网络过拟合问题和参数冗余问题，本项目拟提出一种无参数剪枝算法，来自动确定模型剪枝率。算法能自适应地剪枝多模态编码器中的冗余神经元，实现多模态数据自适应对齐。

## 2.1.3决策级冲突多模态数据可信融合

多模态可信决策旨在将不确定性量化整合进多个模态特征的结合过程，整合多个模态一致性和互补性信息的同时，在不确定性感知框架下融合不同模态的特定意见来进行更可信的决策。这种可信的前提是各模态意见能相互独立不存在冲突，而在开放环境中不同模态数据间往往存在冲突信息，感知到的意见也并不独立。低维度的数据表征上，不同模态数据由于来源不同，可能描述存在相互冲突的场景，这种冲突会导致以往的可信模态融合方法不可避免地退化甚至失效；高维度的模态特征上，模态冲突起源于不同模态提供的高度互补的信息，融合过程中不同模态的一致和互补信息无法被模型消解且互相耦合，这导致模型既无法通过一致信息进行整体上的判别，也无法通过互补信息学习到可区分的决策边界，这样直接融合上述模糊意见进一步导致模型做出不可信的决策。现有冲突模态融合方法大多着眼于消除冲突数据或增大非冲突数据的占比来减弱冲突的干扰，而这些方法忽视了冲突模态中存在的高度互补信息，这些信息对于区分模糊意见、量化不确定性并做出可信决策具有重要意义。

为此，本项目拟研究冲突多模态数据融合方法，对决策级的一致性证据和互补性证据进行解耦，旨在通过轻量高效的算法，在不对融合引入额外干扰的前提下，衡量决策阶段不同模态间的一致性和互补性信息，明确区分证据层的一致性和互补性内容，通过充分利用模态间的一致性和互补性信息，对一致性证据进行还原来修正模态间证据冲突导致的失调性，消解冲突的多模态数据，并分别自适应地感知并融合一致性和互补性证据，实现决策级的冲突模态数据可信融合。

## 2.1.4噪声标签环境下的可信多模态模型学习

随着可信多模态学习的发展，深度学习模型在“实验室”数据环境下的准确性与可靠性得到进一步的提升。然而，现有可信多模态学习方法依赖高质量标签，而真实场景难以得到与数据量和学习任务规模相匹配的高质量标签。使用噪声标签训练可信多模态模型会极大地影响决策不确定度的量化，进而破坏多模态融合过程并降低最终的决策精度与可信度。而现有方法未考虑噪声标签对模型不确定度量化带啊来的影响。

为此，本项目拟研究一种可信的多模态标签噪声矫正方法。通过设计每个样本的模态私有噪声关联矩阵，并利用多模态一致性信息，识别每个样本的噪声产生模式，以实现对样本级标签噪声的矫正。同时，通过证据理论量化由于使用噪声标签进行训练而产生的模型不确定性，增强模型决策结果的可信程度和可解释性，使得在含噪声标签的大规模真实世界数据上仍能得到可信的多模态模型。

# 2.2研究目标

本项目旨在解决复杂开放场景中传统多模态学习方法的局限性，针对多模态系统如何适应开放环境下数据质量动态变化的实际需求进行研究，重点解决在真实世界场景中多模态学习在决策过程中的可信问题。首先从数据质量角度，研究一种基于证据理论的可信的缺失多模态学习方法，解决现实环境下多模态数据不完整的问题；其次从数据自适应的角度，研究适合不平衡模态数据的模态融合对齐方法，达到数据间自适应平衡及对齐；另外从数据内容角度，针对冲突模态数据，研究证据层面的多模态一致性和互补性证据解耦理论，达到消解冲突的多模态数据的目的。最后从数据标签角度，研究一种基于可信理论的标签噪声矫正方法，达到识别并消除样本级标签噪声的目的。通过本项目的研究，希望能够解决多模态学习方法在真实开放场景中的应用难点，有望突破现有学习方法依赖高质量训练数据的局限，推动开放环境场景下的多模态学习理论和方法的发展。

# 2.3拟解决的关键科学问题

本项目拟解决的关键科学问题如下：

**（1）如何自适应对齐非平衡多模态数据的高层表征**

维度不平衡的多模态数据给多模态表征学习带来了新的挑战。主要体现为在模态融合过程中模型容易忽略低维模态的信息。虽然现有工作可以缓解此问题。但是现有方法仍存在以下缺点：首先，低维模态和高维模态需要先进行人为判断，再制定避免低维模态DNNs过拟合的方法，因此方法不够灵活。其次，现有通过剪枝避免网络过拟合的方法，都是人为设定剪枝率，方法自适应性差。如何实现非平衡多模态表征自适应对齐是本项目拟解决的核心问题之一。

**（2）如何可靠度量模态冲突场景下的决策置信度**

在开放环境的现实世界中，不同模态的数据间往往存在着冲突信息。现有方法一般致力于增大非冲突内容（即一致性信息）的重要性，当不同模态间的一致性信息不足以支撑模型做出正确决策时，由于忽略了包含高度互补的信息的冲突内容，将无法区分模糊意见以及进行不确定性量化的过程，这将导致结果的可信性大幅下降。面对这种模态冲突数据，如何在消解冲突的同时，挖掘冲突内容中的可信部分，进而准确度量决策置信度可信度是本项目拟解决的关键科学问题之一。

**（3）如何解耦噪声标签场景下的多模态数据与模型不确定度**

现有可信多模态学习方法所度量的总体决策不确定度，是由数据不确定度和模型不确定度共同组成的。理想情况（标签无噪声）下一般可通过增加训练数据量或改善模型结构来减小模型不确定度，使得总体不确定度能够反映出每个样本特异性的数据不确定度，据此提升模型决策的可靠性。然而标签噪声会导致模型不确定度以未知模式不规律地增长，进而难以通过总体不确定度来寻找决策风险的来源。因此，如何解耦噪声标签场景下的多模态数据与模型不确定度，并据此设计对高不确定性决策的解决方案，是本项目拟解决的前沿科学问题之一。

3．**拟采取的研究方案及可行性分析**（包括研究方法、技术路线、实验手段、关键技术等说明）；

# 3.1研究方案与技术路线

为实现项目研究目标，本项目的总体研究技术路线如下图所示：

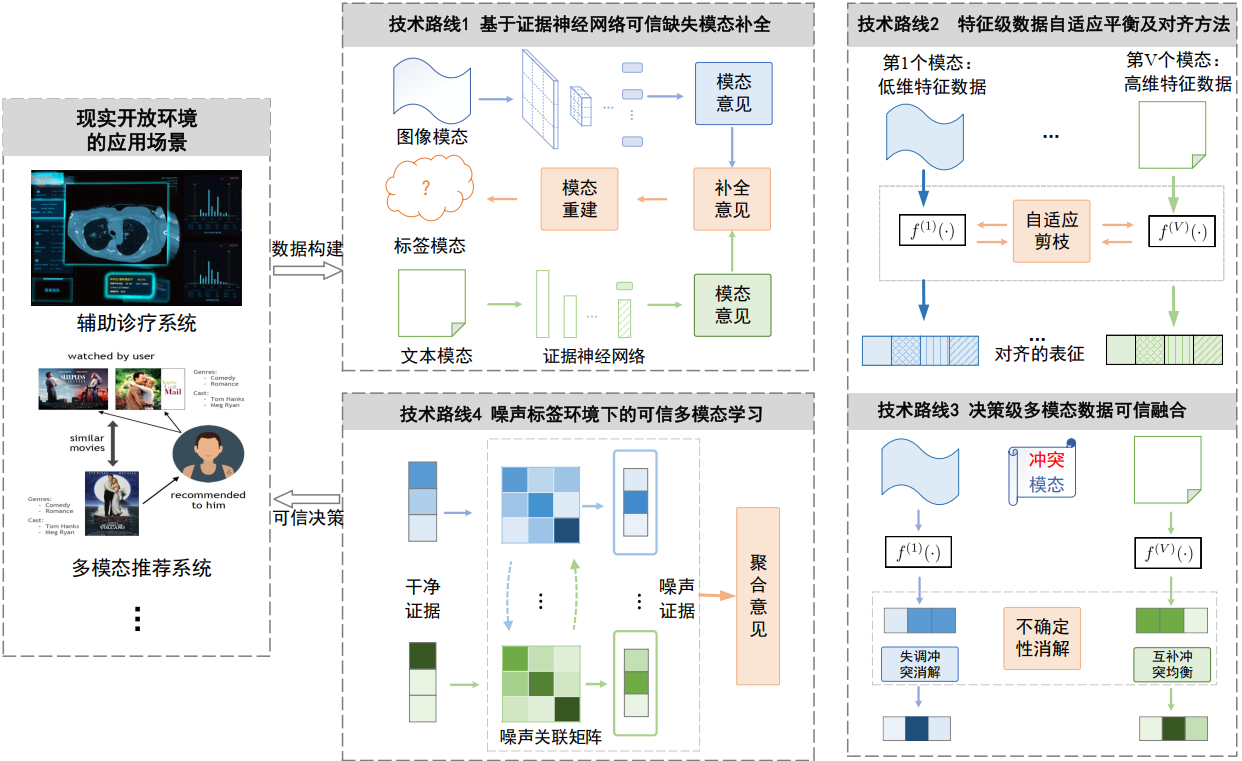


图2 本项目技术路线图

## 3.1.1 多模态可信补全

在现实场景中，由于数据损坏或传感器故障等诸多因素的影响，所收集的多模态数据往往会出现不同程度的缺失导致多模态数据不完整。不仅如此，这些因素还可能导致多模态数据中出现大量的模态冲突，使得数据的完整性和准确性受到严重影响。但是，目前却很少有工作对这类模态缺失并且存在模态冲突的情况开展研究工作。该问题包含了两大挑战：1）多模态数据存在模态缺失的情况。传统的多模态学习方法通常假设每个样本的所有模态都是可用的。然而，在实际应用中，多模态数据通常存在不完整的情况。例如，在多模态情感分析中，有些数据可能包含文本、图片和视频，而其他数据可能只包含一种或两种类型。当样本存在某些模态数据缺失时，多个模态之间的对齐属性就会被破坏，可能会导致对模态间的互补和一致性信息的挖掘不足。2）多模态数据存在模态冲突的情况。以往关于多模态学习的研究大多假设不同模态的数据是严格对齐的。例如，在分类任务中，一个样本的不同的模态始终属于同一类别。然而，在现实环境中，这种假设并不总是能得到保证。例如，在用户评价中，由于用户的多样化评价导致图片和文本使用了不同的物体。因此，这种模态信息的冲突使得大多数多模态学习方法不可避免地退化甚至失效。

为了更好地从缺失多模态数据中获得安全可靠的信息，本项目提出一种新的冲突的缺失多模态学习问题，并针对该问题提出一个新的基于原型学习和证据理论的可信的缺失多模态学习方法，整个网络框架如图3所示。提出的方法采用四个步骤来识别、补全和整合多模态信息，其中第一阶段以特定的模态信息作为输入，采用证据神经网络来提取证据，生成特定模态的Dirichlet分布意见，并通过各自模态的编码器获得各个类别的标准意见原型；第二阶段是计算样本各模态的意见与各个类别的各个模态的标准意见原型之间的差异，选择出合适的类原型意见，并结合各模态的不确定性对缺失模态进行意见补全；第三阶段是利用补全后的多模态意见信息指导缺失模态的特征重构；第四阶段是对缺失模态进行模态重建。

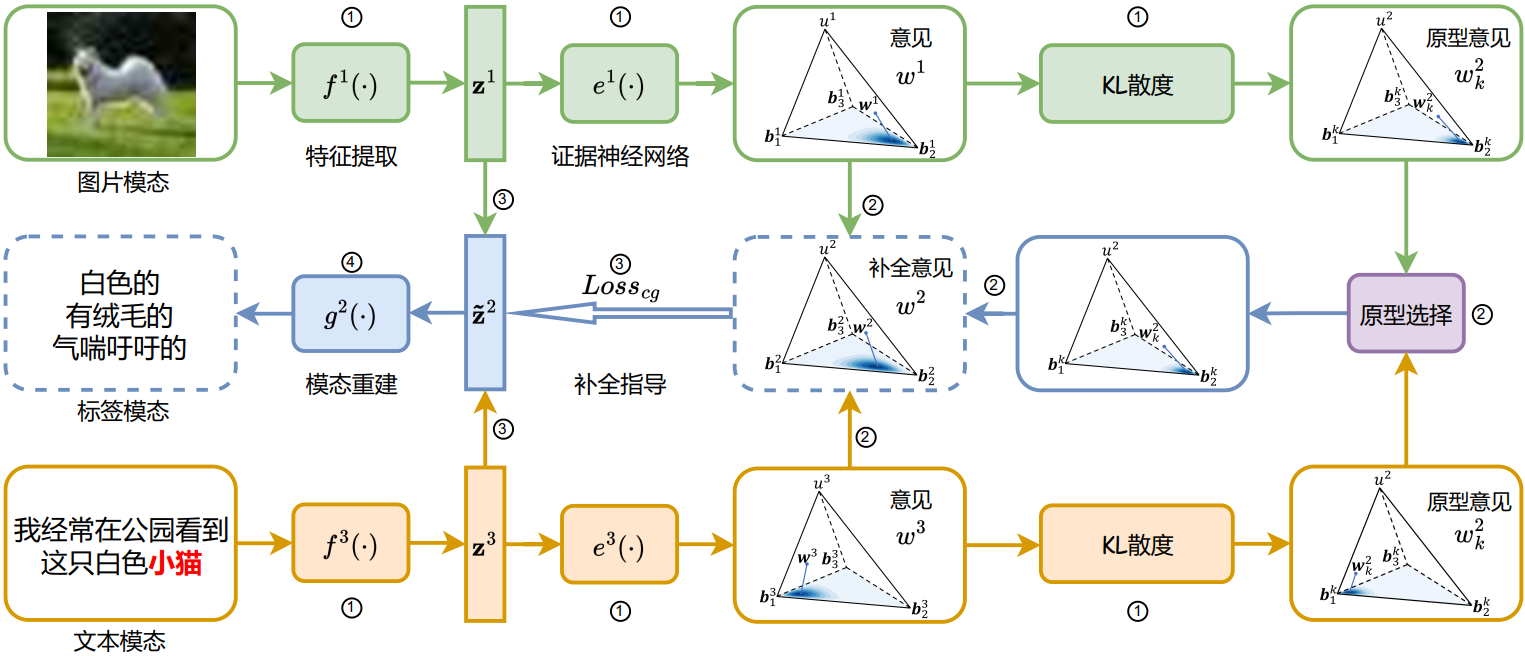


图 3 面向冲突场景的可信缺失多模态学习方法示意图

首先，为了能够量化多模态数据中每个模态的不确定性，我们通过利用证据神经网络和主观逻辑将各模态数据转化为对应的Dirichlet分布意见。在多模态学习任务中，深度自编码器已经被广泛用于提取原始特征的高层表征信息。因此，本项目采用自编码器来生成对应的各模态各类别的类意见原型。具体来说，对于特定模态，使用非缺失模态数据来训练得到该模态对应的证据神经网络。接下来，我们通过证据神经网络获取每个包含完整模态信息样本在模态对应的证据，进而获得对应模态的Dirichlet分布的参数和意见信息。假设是模态的第个类别所对应的类原型意见的Dirichlet分布的参数，则可根据自编码器方法通过优化以下目标函数来获取该原型：

|  |  |  |
| --- | --- | --- |
|  |  | （1） |

如上所述，利用自编码器和主观逻辑，可以得到每个模态的每个类别的类原型意见。

其次，任何物体都拥有着属于该类物体的原型特征，每一个物体都是从该类的原型特征加上一些独有的特性所构成的。因此，该方法基于这些原型意见对缺失模态进行补全可以更好的利用多模态数据之间的一致性和互补性。并且，考虑到多模态数据往往存在模态信息冲突的问题，采用以前的补全方法可能存在选择了不可靠的模态信息参考这一问题。因此，我们设计了如下方案：具体的流程如下：对于某一个样本，假设是非缺失模态集合。首先，根据证据神经网络获得非缺失模态的Dirichlet分布的参数和对应的意见。然后，通过比较样本各模态意见与类原型意见之间的KL散度度量其相似性，将相似性大的作为每个模态所对应的类原型信息。然后，结合每个模态的不确定性，将属于同一类的模态的相似性进行加权相加，选择出可靠的缺失模态对应的类原型意见。最后，根据类原型意见，利用多模态数据的一致性和互补性对缺失模态进行意见层面上的补全。

|  |  |  |
| --- | --- | --- |
|  |  | （2） |
|  |  | （3） |
|  |  | （4） |
|  |  | （5） |

其中，代表了该样本的第个模态对应的第类的原型意见，是该样本的第个模态的类原型意见对应的Dirichlet分布的参数,是最终选择第类的原型意见作为该样本补全时的参考意见，是该样本中的缺失模态，是模态转变成模态时的权重。最后，根据上述规则，我们可以获得一个完整可靠的多模态信息。

根据均方差损失函数微调可以得到该部分的损失函数，该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。但是，该损失函数不能保证错误的标签产生较少的证据。因此，在该损失函数的基础上，引入了可以去除非误导性证据后的狄利克雷分布与标准Dirichlet分布的KL散度。另外，通过逐渐增加KL散度对损失的影响，可以防止网络在训练初始阶段过多关注KL散度，避免过早收敛到错误的均匀分布。

|  |  |  |
| --- | --- | --- |
|  |  | （6） |
|  |  | （7） |
|  |  | （8） |
|  |  | （9） |

其中，，是平衡因子。最终的损失函数如下所示：

|  |  |  |
| --- | --- | --- |
|  |  | （10） |

最后，根据获得的缺失模态的补全意见，通过优化以下目标函数指导来获取该模态的特征信息：

|  |  |  |
| --- | --- | --- |
|  |  | （11） |
|  |  | （12） |

其中，是模态的特征向量，是特征重构网络，是第个模态的证据神经网络。然后再根据已有的模态重建方法对缺失模态进行重建。

## 3.1.2 特征级多模态数据自适应平衡及对齐

由于不同模态数据的来源不同，它们往往在特征表示上存在着巨大的维度差异。例如，在疾病诊断的过程中，医生参考的病理特征通常会涉及到维度差异非常大的数据，比如需要高维表示的CT图像和仅仅需要几个维度就可以表示的血液检测指标。不平衡的模态数据广泛存在于现实生活中，而现有的大部分多模态对齐表征方法都忽略了模态维度不平衡这一现象，并不能充分利用模态间的互补信息，特别是当低维模态的信息特别少时，其对应的特征提取网络极易出现参数冗余和过拟合问题。为了解决上述问题，本项目提出一种用于实现不平衡模态数据对齐融合的方法，同时为了解决不平衡带来的模型参数冗余问题，提出一种无参数的自动化剪枝方法，降低模型由于参数冗余而造成的过拟合风险。

针对多模态表征学习问题，为了挖掘模态间的互补性信息和一致性信息，本项目提出一种稀疏多模态对齐表征学习方法。如图4所示，本项目通过建立特定模态的编码器，将特定模态的数据映射为对齐的特定模态的表征。之后，我们在每个模态的编码器的每一层中添加批处理归一化层（Batch Normalization，BN）对不同模态的特征进行强制对齐。这样做不仅可以加快网络的训练和收敛同时还能防止网络产生过拟合问题。我们把编码器第层的第c维定义为，其中。BN公式如下：

|  |  |  |
| --- | --- | --- |
|  |  | （13） |

其中，和分别为可训练的比例因子和偏移量。ϵ是一个小的避免零值的常数。和分别表示第层当前小批量数据的所有维度的平均值和标准差：

|  |  |  |
| --- | --- | --- |
|  |  | （14） |
|  |  | （15） |

由于多模态数据之间存在着固有的一致性和互补性，所以单模态的信息往往是不完整的，而综合表征包含了完整的信息，我们认为单个模态表征的某些维度与综合表征无关。因此，我们在对齐的特定模态的表征上设计了一个多模态稀疏批归一化层，以此来实现稀疏融合。我们对该层的比例因子施加稀疏约束。当时，表示的第c维与的第c维无关。这表明该模态对第c维的生成没有贡献。我们通过最小化以实现稀疏连接。最后，在融合阶段，我们对对齐的特定模态的表征进行平均，得到多模态综合表征：

|  |  |  |
| --- | --- | --- |
|  |  | （16） |

中的一个特征至少被一个模态激活，激活的强度取决于被激活模态的数量，即一致的特征的激活程度往往高于互补特征的激活程度。

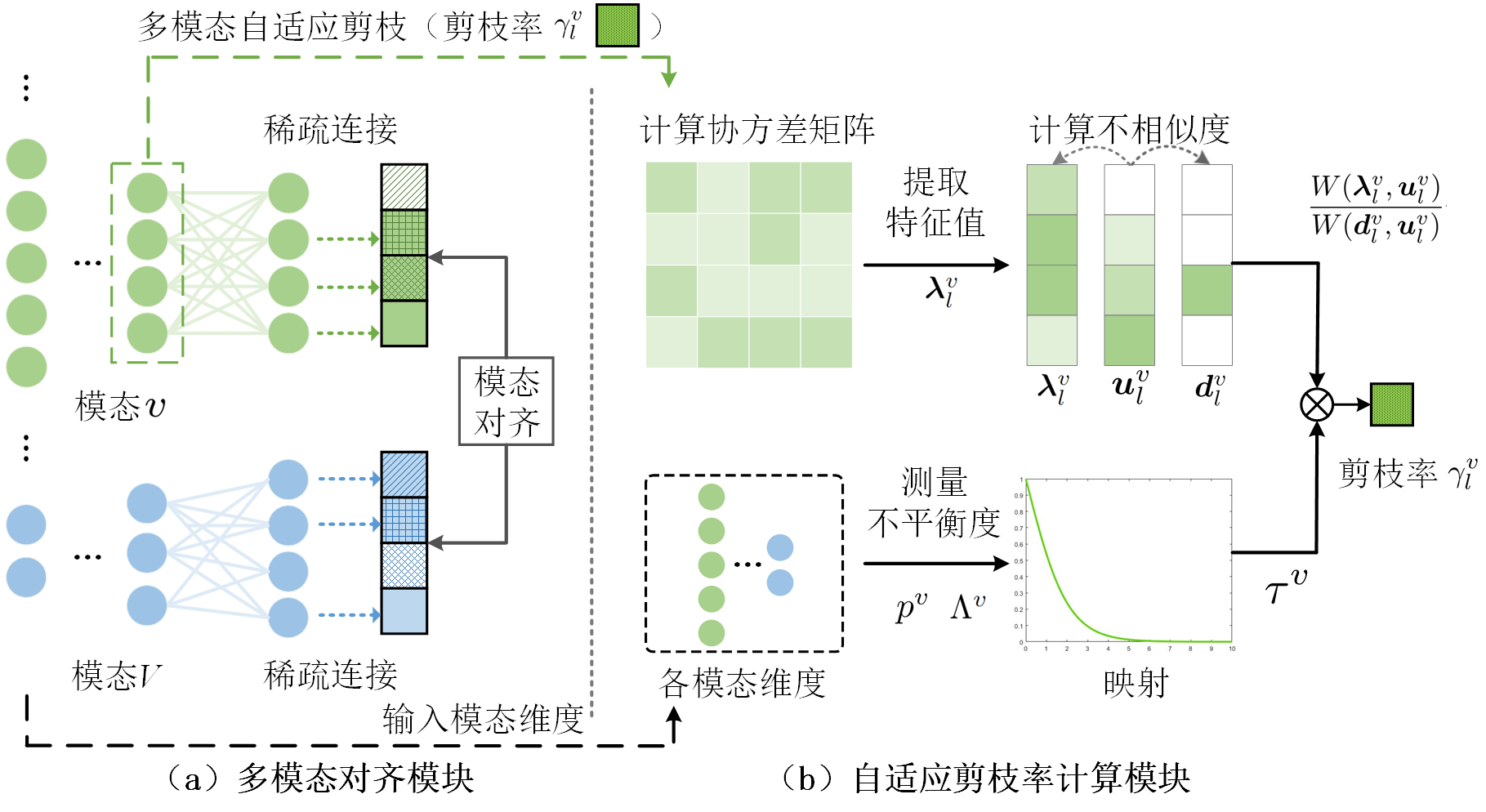


图4 自适应剪枝方法流程图

针对模态维度不平衡带来的挑战，常见的做法通常直接使用编码器进行对齐。然而，在模态维度非常不平衡时，低维模态的网络存在参数冗余和过拟合的风险。最近的一些研究认为，网络剪枝有利于网络的泛化能力，可以缓解模型过拟合的问题。多模态中常用的剪枝方法通常根据神经元输出的绝对值、对网络激活的贡献或梯度来对其进行评分，然后对每一层得分最低的神经元进行修剪。每一层的剪枝率通常被认为是一个超参数，并根据模型的性能进行调整。然而，随着多模态编码器的增加，手动调整剪枝率的方法将变得低效和不方便。因此，本项目提出了一种自适应的网络剪枝方法，即主神经元分析法（Principal Neuron Analysis, PNA)。该剪枝方法能够自动地根据网络的冗余度计算剪枝率并神经元进行剪枝，以缓解模型的过拟合问题。具体来说，该方法根据网络每层神经元输出的相关性来决定每层的剪枝率。我们发现，通过移除不重要的神经元或连接来减少模型参数的方法等同于对数据进行降维，经典的数据降维算法包括主成分分析法（Principal Component Analysis，PCA），该算法首先提取原始数据的协方差矩阵的特征向量和特征值，然后选取前项最大特征值对应的特征向量来重建原始数据。在特殊情况下，如果特征值大小服从均匀分布，则意味着原始数据的特征不相关，这种情况下就没有必要进行降维。相反，如果特征值大小遵循Dirac δ分布，则说明原始数据的特征具有高度相关性。

受此启发，我们提出基于PNA的剪枝算法，如图3所示。我们首先计算的协方差矩阵，并提取了其特征值。之后，我们通过的分布推断每一层神经元输出的相关性，也就是特征的冗余度。当越接近均匀分布时，神经元输出的相关性越小，大多数神经元将被保存下来。相反，越接近Dirac δ分布，神经元的输出的相关性越大，大量的神经元将被修剪。最后，我们利用瓦瑟斯坦度量来度量和均匀分布之间的差异计算剪枝率。剪枝率计算公式如下：

|  |  |  |
| --- | --- | --- |
|  |  | （17） |

同时为了防止高剪枝率导致网络表达能力降低，我们用来适度控制。我们假设每个模态的不平衡度与成正比。计算公式如下：

|  |  |  |
| --- | --- | --- |
|  |  | （18） |
|  |  | （19） |

其中，为每个模态的维度，其中，为的维数与之间的比例关系。和是期望和方差。之后，我们根据剪枝率计算网络每一层待修剪的神经元数量，并计算每个神经元与其他神经元的皮尔逊相关系数的L1范数。最后，我们对范数最大的前个神经元进行剪枝。

## 3 1.3 决策级冲突多模态数据可信融合

为解决现实场景下多模态数据中广泛存在模态冲突的问题，关键在于如何解决模态间的一致性和互补性信息的相互纠缠和在模态冲突的条件下实现可信融合的挑战，以及如何在可信融合过程中进一步消解冲突。本项目针对如上三项问题提出了相应的技术路线。

针对模态间的一致性和互补性信息相互纠缠，对可信决策造成影响的问题，本项目提出了证据层面的多模态一致性和互补性证据解耦理论，旨在通过轻量高效的算法，在不对融合引入额外干扰的前提下，衡量决策阶段不同模态间的一致性和互补性信息。具体来说，我们需要对从特定模态中提取到的证据进行解耦，以此得到一致性证据和互补性证据。我们使用来表示来自特定模态的证据，其中，表示为类别，将每个模态所支持的每种类别的最小证据定义为模态间的一致性证据：

|  |  |  |
| --- | --- | --- |
|  |  | （20） |
|  |  | （21） |

我们将互补性证据定义为：,其中指的是高度互补的证据。通过这种方法将多模态数据的一致性信息和互补性信息在证据层面解耦，在下一阶段中可以对模态间的一致性和互补性证据采取不同的处理方式，来最大化它们在最终融合结果中的作用，进而消除模态间冲突。

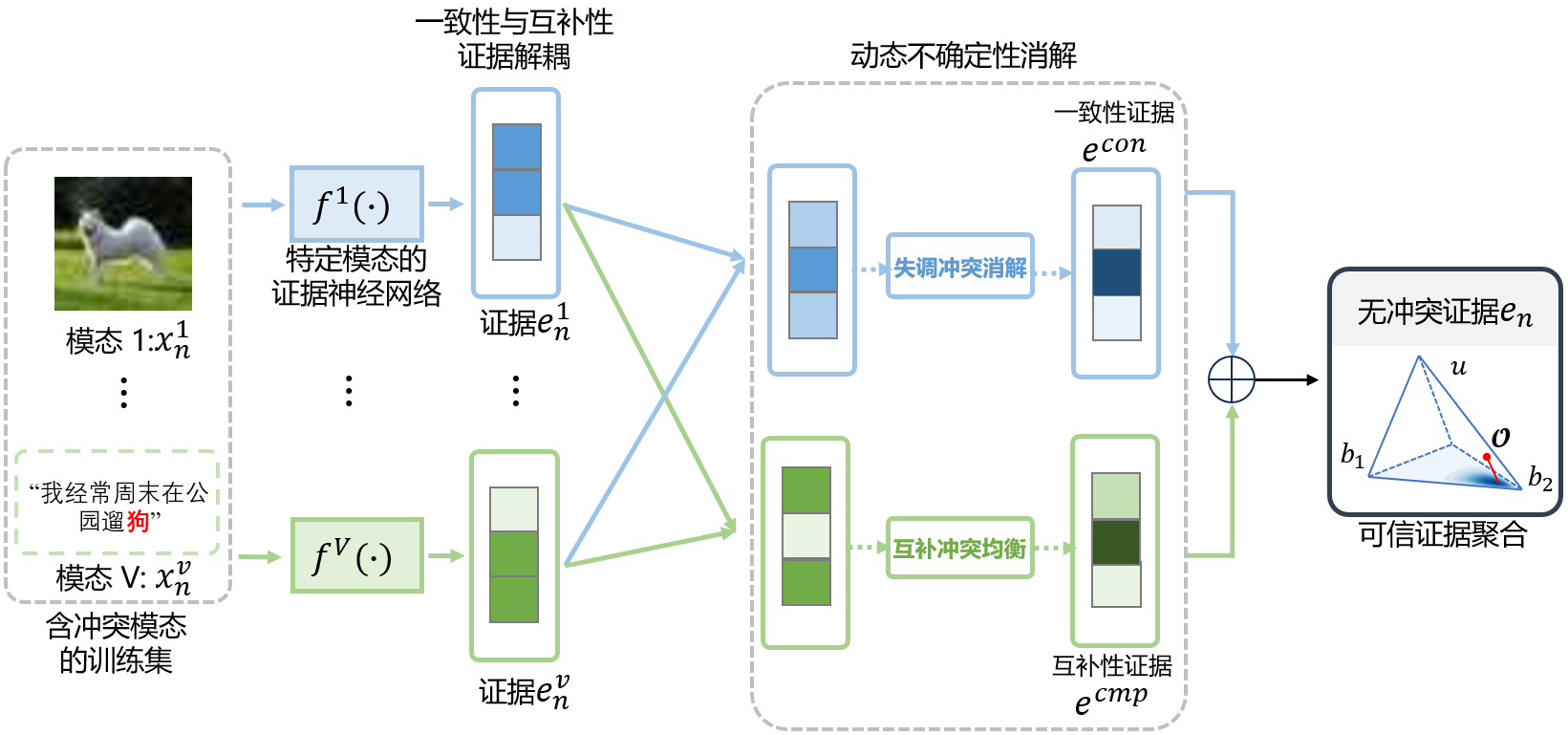


图 决策级证据多模态可信融合框架图

针对在模态冲突的条件下实现可信融合的挑战，本项目提出了通过一致性证据还原模块来修正模态间证据冲突导致的失调性。首先需要量化模态间的冲突程度来减少融合过程中的不确定性，本项目提出将由模态间冲突导致的不确定性解耦为空洞不确定性和失调不确定性，通过结合一致性和互补性证据来动态减少融合过程中的整体不确定性。其中空洞不确定性用来衡量由于模态间冲突所导致的证据的缺乏程度：即证据越缺乏，模态间存在的空洞不确定性就越多。另一方面，失调不确定性代表在解耦了模态间冲突后，剩余不同类别的证据间缺乏变化的程度，这意味着模型将更难依赖一致性信息做出可信决策：这是因为不同类别间的证据差异越大，对分类决策的贡献就越大。在主观逻辑框架下，我们将意见的失调不确定性定义为：

|  |  |  |
| --- | --- | --- |
|  |  | （22） |

其中为相对质量的平衡因子，定义为：

|  |  |  |
| --- | --- | --- |
|  |  | （23） |

接下来，通过一致性证据还原模块对不确定性进行调整来消解模态间冲突，保证调整前后一致互补证据的数量不变的同时，实现对一致性证据的失调不确定性和整体证据的空洞不确定性的调整。具体来说，将一致性意见通过以下过程获得减少失调性后的最终一致性意见:

|  |  |  |
| --- | --- | --- |
|  | ; | （24） |
|  | =(,); | （25） |
|  |  | （26） |

其中表示支持类别c的信任质量，(,)是幂运算函数，是一个大于1的超参数。

针对冲突模态的可信融合过程，本项目提出一种可信的冲突融合方法。考虑到证据的一致性部分被所有模态支持，所以具有更高的置信度，因此它们应该在最终的证据聚合过程中产生更多的证据。而不同模态间具有差异的部分是不同模态间具有区别性或互补性的信息，应该被选择地保留在融合结果中。通过对一致性和互补性的证据进行可信融合，来获得最终的无冲突聚合证据和狄利克雷分布参数:

|  |  |  |
| --- | --- | --- |
|  | *=* | （27） |
|  |  | （28） |

在最终的证据融合中，一致性证据部分()的证据量等于所有模态一致部分的总和，而互补证据部分()的证据量等于所有模态互补部分的平均值。这是因为对于多分类任务，所有模态之间的不同部分代表支持每个类别的证据之间的差异，而这部分区分可能会因为噪声或互补信息等因素而增加。因此，所有这些证据不应该全部保留，而应该通过取平均值的方式适当地保留。

## 3.1.4 噪声标签环境下的可信多模态学习

在海量数据的帮助下，深度学习在许多领域都取得了显著的成功。然而在大规模数据集上训练标签的质量问题日益凸显，因为取得大量数据的高质量标签成本是非常高昂的，低质量标注信息中的噪声标签会严重降低可信模型预测的可靠性。针对这一挑战，我们考虑对现实世界数据集中的标签噪声进行处理，图6给出了来自两个方面的标签噪声示意图：易混淆类别（如“狗”和“狼”类别）和低质量数据特征（如模糊和不完整特征）。因此，如何在解耦模型与数据不确定性的同时，发掘类别间可能的破坏关系与识别样本自身的质量是该问题亟待解决的目标之一。同时多模态学习的发展已经证明，结合多个模态的信息能够提高模型在不同领域任务上的性能，而现有方法都只尝试在单模态场景中消除标签噪声的影响。因此，本项目拟利用多模态一致性信息以更好地识别与消除标签噪声。

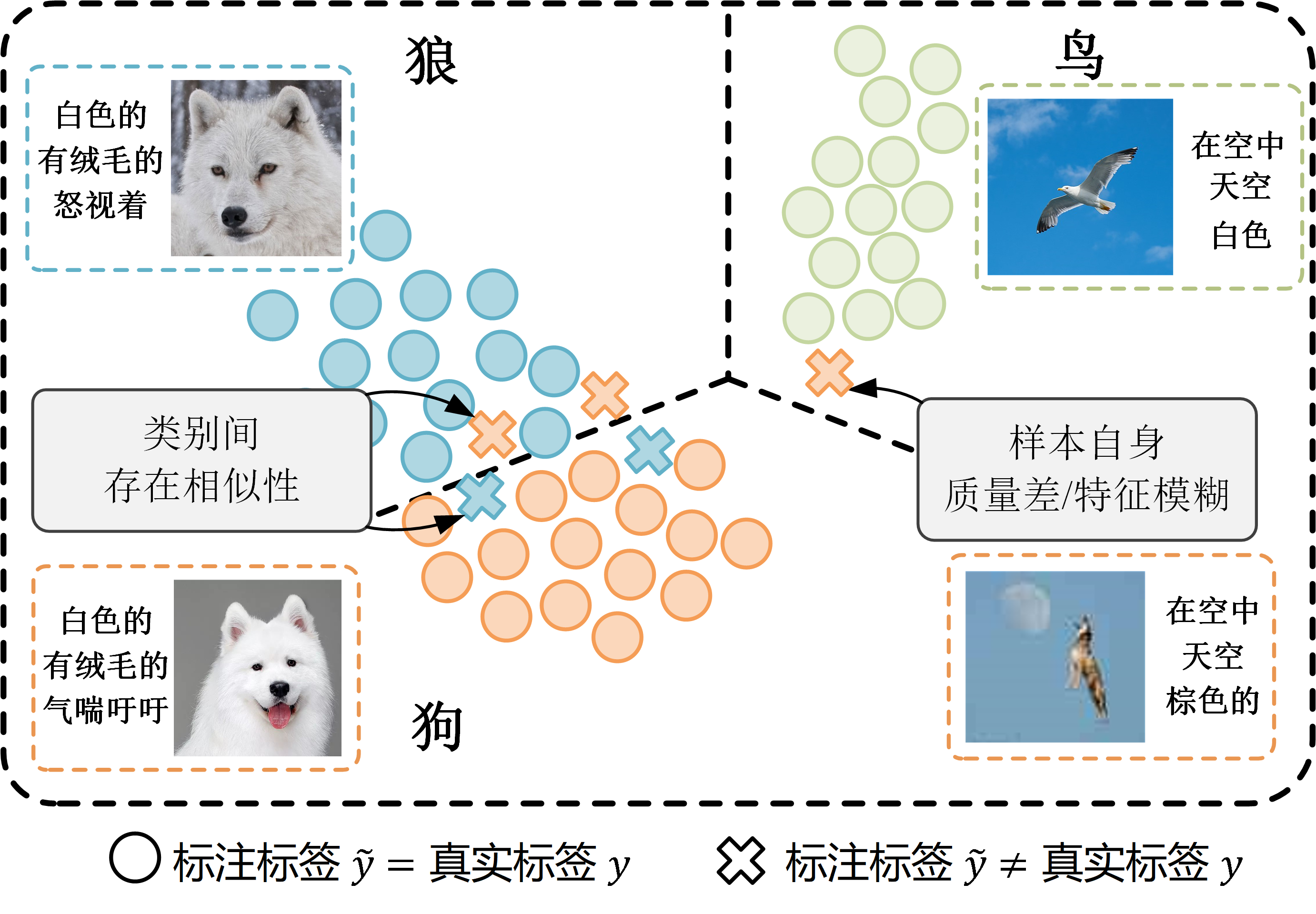


图6 标签噪声的产生来源示意图

为在多模态场景下有效减轻噪声标签对模型噪声的产生的影响，本项目提出一种新的基于样本模态特定噪声关联矩阵的可信模型学习方法。此方法可以作为一种即插即用的通用方法应用在各类可信学习模型中，使得模型使用含大量标签噪声的数据集进行训练也能做出准确而可靠的决策。如图7所示，我们所提出的模型包含模态特定的证据神经网络 和仅在训练阶段使用的样本模态特定的噪声关联矩阵。前者基于证据理论学习提取每个模态各类别的证据,并以此为参数使用狄利克雷分布对每个模态的类别概率分布进行建模。后者则是建立类别概率干净分布与噪声分布之间的联系，显式的发掘标注信息被破坏的规律，从而对模型损失进行调整以实现对标签噪声的鲁棒性。其中表示模态数量，表示训练集中样本数量。

首先对噪声标签问题进行定义，样本真实所属的类别由表示，表示对该样本进行标注后可能包含噪声的标签。我们的目标是根据含样本相关标签噪声的多模态数据集来训练一个可信的分类模型。对于测试样本，模型可以预测出其所属类别并给出关于此决策的不确定性。在训练模型的证据神经网络以提取证据时，我们希望使该样本标签对应类别的证据量增加，而其他类别对应的证据量应该尽可能的低。由此将所提模型的多模态证据学习优化目标函数定义如下：

|  |  |  |
| --- | --- | --- |
|  |  | （29） |
|  |  | （30） |

其中表示函数，为平衡因子，我们在训练中逐渐增加的值以保证网络具有良好的参数探索空间，不会输出一个平面均匀的分布。是狄利克雷分布的调整参数，避免将真实类别的证据惩罚为0。基于此优化函数对神经网络进行训练后，我们可以从输入样本中经过推理得到由信念质量向量和不确定性估计值组成的模态特定意见，这被称为模型预测得到的干净意见。

为了明确意见中不确定度的组成，以指导对高不确定性决策的解决方案，我们需要对数据与模型的不确定度进行解耦。为了获得数据不确定性的度量，我们可以评估数据分布的期望熵。由于熵捕获了输出分布的“峰值”，较低的熵表明模型将大概率质量集中在单一类上，而高熵则描述了更均匀的分布即模型尚未确定正确的预测。对于狄利克雷证据网络，数据不确定度有封闭形式的解:

|  |  |  |
| --- | --- | --- |
|  |  | （31） |

在计算出数据不确定度后，总体不确定度中的其余部分即为模型不确定度，表示为。在实现对数据与模型不确定度的解耦后，高数据不确定度的决策即表明其自身特征难以被识别，需要通过人工介入或其他方法做出决策；高模型不确定度的决策则可能表示模型受到噪声标签影响或对于该类别样本学习不充分，需要增加高质量的标注信息进一步训练模型。

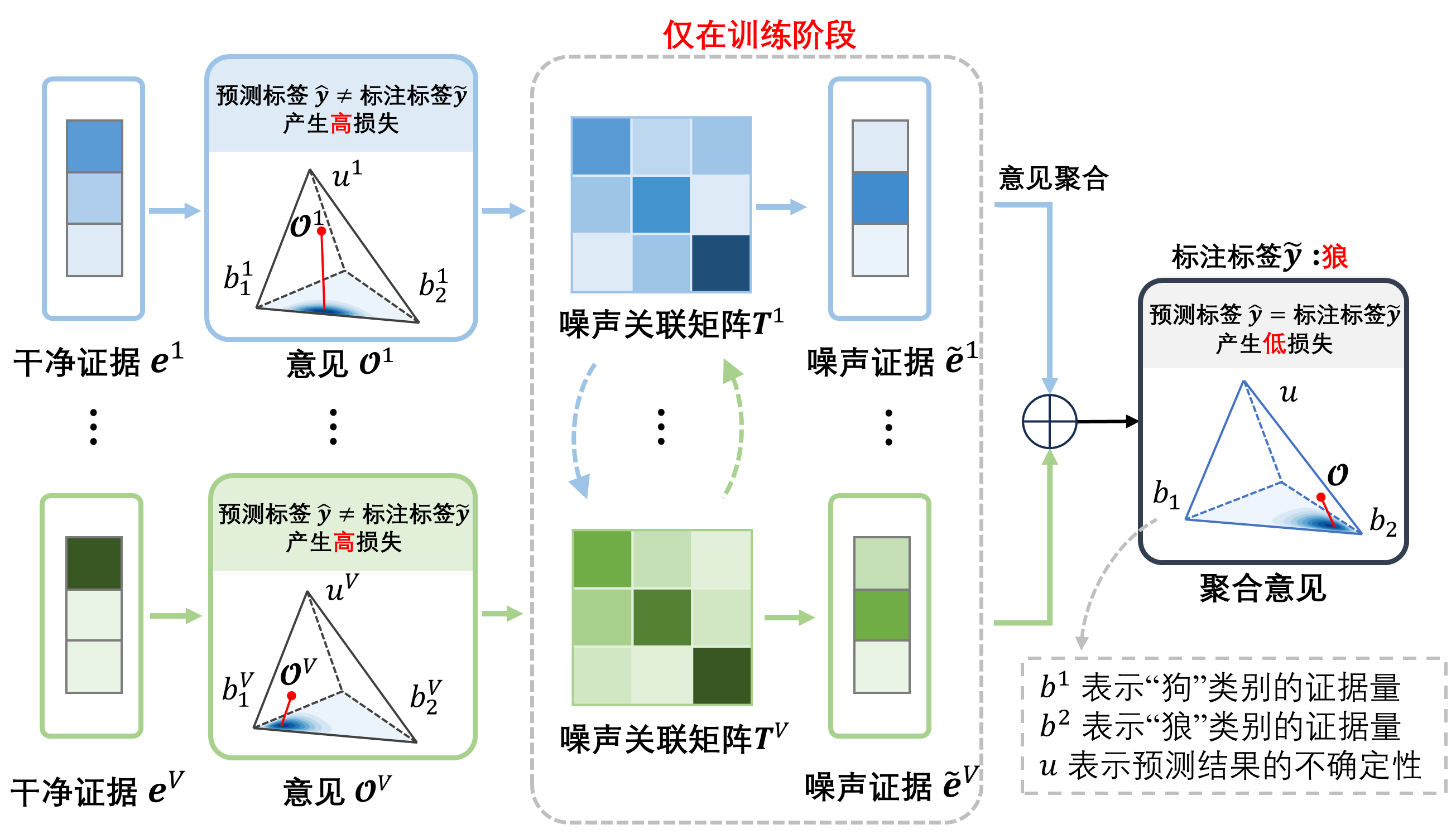


图7 可信的样本级多模态噪声标签修正模型示意图

噪声关联矩阵是通过调整模型损失以避免标签噪声的影响，它期望将干净意见转换为噪声意见，并使得其与噪声标签表达的信息保持一致。因此我们为每个样本设计了针对特定模态的噪声关联矩阵。它是一个大小为的矩阵，其中每个元素保证在区间内，表示该样本真实类别为时被错误标记为类的概率。由于我们所处理的样本相关噪声在不加约束的情况下是非常高维的函数，难以对样本模态特定的噪声关联矩阵进行训练。为此我们考虑到低质量数据特征更容易被错误地标注，我们要求噪声关联矩阵的对角元素与不确定性成反比。而对于非对角线元素，考虑到现实世界中特征相似的样本被错误标记为同一类的概率更大，我们基于样本间相似性设计了一个正则化项来确保相似样本间的噪声关联矩阵接近。在不同模态间，我们则利用多模态信息的一致性原则来确保所有模态关联矩阵的整体相似性，一致性损失定义为：

|  |  |  |
| --- | --- | --- |
|  |  | （32） |

其中为不同模态真实类别为时被错误标记为类的概率均值。在训练时通过各模态特定的噪声关联矩阵，将各模态干净意见转换为噪声意见并根据合适的多模态意见聚合方法得到聚合意见，以此来拟合含噪声的标签并更新模型参数。最终使用训练完成的证据神经网络模型预测新样本时，模型的输出即为干净的决策意见。

# 3.2可行性分析

本项目的选题建立在对国内外多模态学习领域研究现状和应用成果的深入调研和分析之上，紧密围绕复杂开放环境下可信多模态学习关键算法展开研究。本项目拟采用的方法和技术路线均是在项目申请人现有研究基础上发展而来，也充分借鉴了弱监督学习、可信深度学习等相关领域的最新研究成果。除此之外，申请人团队承担过多项国家和省部级重点科研项目，为本项目的顺利开展提供了坚实的资源平台支撑。本项目的可行性主要体现在以下三个方面：

**（1）在研究问题选择和研究思路方面：**开放环境下的多模态学习方法是当前国际研究的热点，有着重要的现实意义和广泛应用前景。目前开放环境下的多模态学习方法在缺失多模态学习、多模态对齐融合以及噪声多模态学习方面已经取得了一定的进展。然而这些研究工作注重于提升多模态算法的性能，对算法可靠性方法考虑较少。如果能够有效度量开放环境下缺失多模态学习中补全数据/表征的可靠性，可以减少模态信息在融合时产生的信息混淆。与此同时，经过调研发现，多模态的一致性信息与互补性信息能够很好地建立多个模态之间的联系。如果可以将上述两种关联的信息解耦，就能够有效缓解多模态信息融合过程中的冲突问题，提升融合结果的可靠性。而从信息论方面相关研究来看，通过信息解耦来实现冲突消解是可行的。再从噪声标签学习的相关研究来看，噪声关联矩阵已被证明能很好地识别正确标签分布与噪声标签分布之间的转换模式，而多模态数据的一致性关联有助于进一步检测并修正噪声标签，这为处理开放世界的多模态数据的标签噪声提供了可能的依据。

因此，本项目拟设计一个面向开放复杂场景的可信多模态学习框架，该框架由四个研究内容组成，分别是缺失模态可信补全、多模态数据自适应平衡及对齐、决策级冲突多模态数据融合、噪声环境下可信多模态学习。研究内容一是后三个研究内容的基础，研究内容二和研究内容三是对开放环境下多模态学习出现的普遍问题的进一步探索。在此基础上，研究内容四考虑了标签不准确时，如何学习可靠的多模态模型。这四个研究内容逐次递进，面向开放复杂环境中多模态学习出现的重点问题和难点问题提出了一个系统的、全面的解决方案，在多媒体/多传感器情景感知、疾病辅助诊断等领域具有广泛的应用前景。

**（2）在技术积累方面：**项目申请人长期从事多模态聚类、多模态推荐、跨模态检索等方面的研究工作，拥有丰富的研究经验和实践积累，熟悉这些研究方向的专业知识和国内外发展现状。近5来，在本项目涉及的缺失多模态补全、特征级多模态对齐、决策级多模态融合和噪声标签多模态学习四个领域取得了一定的研究成果：**在IEEE TPAMI**、IEEE TNNLS、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21篇，其中一作/通讯11篇，获得AAAI-24最佳论文奖**（**当年国内唯一、唯一第一作者**、见附件6），博士毕业论文《基于深度学习的多视角表达学习技术研究》获评**“陕西省优秀博士学位论文”**（**当年全省“计算机科学与技术学科”共3人**、见附件6）。这些前期研究工作都为本项目的顺利开展打下了坚实的研究基础。

**（3）在资源平台积累方面：**在项目研究工作开展过程中，不可避免要面临数据资源和平台方面的挑战。在公开可获得数据资源方面，申请人团队在以往的研究工作中已通过购买、下载、抓取等方式获得了一定规模的医疗辅助诊断和社会媒体应用中的多模态数据集，如上海交通大学医学院附属瑞金医院的患者多模态数据集，新浪微博、亚马逊、Yelp的用户多模态评论数据集等，能够很好地支持理论研究的开展。

本项目的应用验证平台有两个：

* 医疗辅助诊断平台：申请人所在团队已与上海交通大学医学院附属瑞金医院建立了长期合作，研发的消化道疾病辅助诊断平台已经应用于该医院，降低了医生工作量，取得了较为显著的社会效益。在此合作基础上，**申请人作为骨干成员（4/10，校内第2），联合上海交通大学医学院附属瑞金医院等申请2021年度国家自然科学基金重点项目**“基于医学影像的消化道肿瘤和炎症类疾病辅助诊断方法研究”，**并获批资助**（见附件7）。这些合作基础能够支持本项目在可信多模态疾病辅助方面的实际验证。
* 电商平台：申请人与阿里巴巴已开展数年的深入合作，与一些具体部门如阿里巴巴跨境电商平台建立了良好的合作关系，并已与阿里巴巴联合发表高水平研究成果3篇。特别地，申请人研发的“**面向多模态检索的商品预训练**”技术“**应用于阿里巴巴跨境电商平台速卖通（AliExpress）的商品检索系统中，服务于全球220个国家和地区的数亿消费者，显著提升了用户检索的效率和准确度，同时提升了平台广告的营收，2022 年，为平台带来了近1.5%的广告收入的相对提升**”（见附件8）。这些合作基础能够支持本项目在可信多模态推荐方面的实际验证。

# 4.1本项目的特色与创新之处；

围绕本项目的总体研究目标，本项目的主要创新之处有四点：

**（1）基于证据理论的缺失模态数据可信补全**

现有缺失多模态学习大都局限于模态数据是严格对齐的场景下，在模态冲突的数据时大多数缺失多模态学习方法不可避免地退化甚至失败。本项目首先利用证据神经网络获取模态的高级观点信息，结合原型学习理论，学习每个类别每个模态的类原型观点。然后，根据类原型观点对缺失模态进行观点层面的补全，之后再结合多模态数据的一致性和互补性对缺失模态数据进行重建。

**（2）基于无参剪枝的多模态自适应表征学习**

现有多模态表征学习大都受限于平衡多模态数据场景下，在处理不平衡数据时容易忽略低维模态的信息。本项目利用深度神经网络提取模态的高级表征，通过添加批处理归一化层来挖掘模态间一致性和互补性信息。项目提出无参剪枝算法，来自动地确定每个模态编码器的剪枝率，自适应的剪枝冗余的参数。

**（3）面向模态冲突场景的一致性与互补性证据量化**

现有的可信多模态学习方法没有考虑到现实场景下模态间普遍存在的冲突关系对可信融合过程造成的影响。本项目对决策级的一致性证据和互补性证据进行解耦，明确区分证据层的一致性和互补性内容，通过充分利用模态间的一致性和互补性信息，消解冲突的多模态数据，并分别自适应地感知并融合一致性和互补性证据，实现决策级的冲突模态数据可信融合,能够在充满冲突模态的真实世界大规模数据集上生成精准的模态意见的同时取得更可信的融合结果。

**（4）标签噪声环境下多模态数据及模型不确定性解耦**

现有的可信学习方法所度量的不确定度包含数据与模型不确定度，但却难以对其进行解耦分析处理，这使得模型在标签噪声场景下对总体决策置信度的度量失准。本项目基于证据理论并利用多模态信息间一致性，准确量化由标签噪声带来的模型不确定度与样本自身存在的数据不确定度。在减小标签噪声影响的同时保证所作决策的可靠性。

5．**年度研究计划及预期研究结果**（包括拟组织的重要学术交流活动、国际合作与交流计划等）。

# 5.1年度研究计划

本项目计划四年完成，各年度的具体研究计划如下：

**第一年度（2025年1月至2025年12月）：**

1. 以抓取、购买、与相关平台合作等方式收集真实世界的多模态数据；
2. 预处理多模态数据，初步构建2-3个低质量多模态数据集；
3. 设计基于主观逻辑的证据神经网络，提取特定模态意见并识别缺失模态；
4. 研究用于补全和整合缺失多模态信息的可信多模态补全方法；
5. 撰写和发表论文1-2篇，参加学术会议/研讨会2次。

**第二年度（2026年1月至2026年12月）：**

1. 在和阿里巴巴电商推荐平台已有合作基础上，了解当前推荐系统在多模态数据低质量环境下的应用难点；
2. 与上海交通大学医学院附属瑞金医院继续开展合作，熟悉当前医疗辅助诊断平台的流程与实际应用效果；
3. 设计面向已有模态内容冲突场景下可信多模态补全方法；
4. 研究模态维度非平衡条件下的特征级多模态自适应平衡及对齐方法；
5. 在之前研究的基础上，设计基于无参剪枝的多模态自适应表征学习方法；
6. 撰写和发表论文2-3篇，参加学术会议/研讨会2次。

**第三年度（2027年1月至2027年12月）：**

1. 设计模态冲突场景下的多模态数据可信融合方法，量化模态间的一致性和互补性证据；
2. 与阿里巴巴电商平台和上海交通大学医学院附属瑞金医院合作采集带有标签噪声的多模态数据，构建1-2个带有标签噪声的多模态数据集；
3. 在此基础上，研究用于标签噪声环境下的可信多模态学习方法；
4. 将研究成果应用于阿里巴巴电商平台与瑞金医院的医疗辅助诊断平台；
5. 撰写和发表论文2-3篇，参加学术会议/研讨会2次；

**第四年度（2028年1月至2028年12月）：**

1. 实现并完善阿里巴巴电商推荐系统平台决策系统的主要功能并测试上线；
2. 实现并完善医疗辅助诊断平台系统的主要功能并测试上线；
3. 撰写和发表论文2-3篇，参加学术会议/研讨会2次；
4. 研究成果总结、整理、完善，撰写项目结题验收报告。

# 5.2预期研究成果

* 在**学术成果方面**，预计将在人工智能、多媒体、数据挖掘等本领域主流期刊和学术会议上发表学术论文 8 篇以上，申请（或授权）发明专利 3-4 项。
* 在**模型算法方面**，实现多模态可信补全方法、特征级多模态数据自适应平衡及对齐方法、决策级冲突多模态数据可信融合方法、噪声标签环境下的可信多模态学习等，公开共享相关源代码，构建并开源2-3个低质量多模态数据集。
* 在**应用验证方面**，在医疗辅助诊断、电商推荐等2-3个场景实现应用示范。
* 在**人才培养方面**，通过项目研究工作的开展，培养博士研究生 2-3 名，硕士研究生 6-8 名，本科生3-4名。

**（二）研究基础与工作条件**

1．**研究基础**（与本项目相关的研究工作积累和已取得的研究工作成绩）；

申请人团队一直从事机器学习、多媒体计算、多模态/视角学习方面的研究工作。近5年，申请人团队在相关领域主流期刊和学术会议发表论文46篇，其中中科院一区期刊论文16篇，CCF A类学术会议论文11篇。中科院一区期刊包括：IEEE TPAMI、IEEE TC、Research、IEEE Tcyb、IEEE TNNLS、IEEE TIP、IEEE TGRS等，CCF A类会议包括：NeurIPS、CVPR、SIGKDD、AAAI、IJCAI、CVPR、ACM MM等。研究成果获得**AAAI-24最佳论文奖**(**当年国内唯一**、**唯一第一作者**)。此外，申请人和团队成员也积极参与相关领域学术活动和学术任职，包括**任职IEEE TITS、TNSM、TNSE、TCSS、ACM TOMM等领域内高水平期刊的编委**、**多次担任知名国际会议ICML、AAAI、IJCAI、ACM MM的Workshop主席、领域主席**、资深程序委员会委员、程序委员会委员等，具有一定国际影响。近年来的相关工作积累详述如下：

**（1）在缺失****多模态补全方面**，申请人在国际上较早提出“重构+对抗”式多模态补全方法来学习缺失多模态数据的高层表征，并藉此重构原始数据及推测缺失数据。该框架通过结合元素级重构损失捕获了数据的总体结构，并且通过对抗式训练促使原始数据和重构数据蕴含的高层知识尽量接近，进而获得更逼真的补全数据。相关研究成果获得加拿大皇家科学院院士、IEEE Life Fellow、阿尔伯塔大学的Witold Pedrycz教授团队，英国皇家工程院院士、IEEE Life Fellow、伦敦布鲁内尔大学的Asoke K. Nandi教授团队等知名学者团队的正面评价。代表性研究成果包括：

[1] **Cai Xu**, Hongmin Liu∗, Ziyu Guan, et. al. Adversarial Incomplete Multiview Subspace Clustering Networks. *IEEE Transactions on Cybernetics* **(IEEE Tcyb，中科院一区，影响因子：19.118)**, 10, no. 52 (2022): 10490-10503. （见附件3）

[2] **Cai Xu**, Ziyu Guan∗, Wei Zhao, et. al. Adversarial Incomplete Multi-view Clustering. In *IJCAI* **(CCF A类会议，oral，Google学术被引127次)**, pp. 3933-3939. 2019.

**（2）在特征级多模态对齐方面**，申请人提出多模态深度矩阵分解方法，通过组稀疏约束，允许融合多模态信息的高层表征空间的特征只与部分模态相关联并自动优化这一关联。为了提升场景适用性，申请人将其推广到深度神经网络框架中，提出多模态稀疏批归一化方法，通过批归一化促使不同模态的高层表征服从相似的分布，并对批归一化的缩放参数施加稀疏约束，藉此对齐模态独有高层表征和多模态共同表征。相关研究成果获得加拿大工程院院士、IEEE Fellow、英属哥伦比亚大学的Z. Jane Wang教授团队，IEEE Fellow、诺桑比亚大学的Wai Lok Woo教授团队等知名学者团队的正面评价。代表性研究成果包括：

[3] **Cai Xu**, Wei Zhao∗, Jinglong Zhao, et. al. Progressive Deep Multi-view Comprehensive Representation Learning. In *AAAI* **(CCF A类会议，oral)**, 2023.

[4] **Cai Xu**, Ziyu Guan∗, Wei Zhao, et. al. Deep Multi-View Concept Learning. In *IJCAI* **(CCF A类会议，oral)**, pp. 2898-2904. 2018.

[5] **Cai Xu**, Zehui Li, Ziyu Guan, et. al. Unbalanced Multi-view Deep Learning. In ACM MM **(CCF A类会议)**, 2023.

[6] Wei Zhao, **Cai Xu**∗, Ziyu Guan∗, et. al. Multiview concept learning via deep matrix factorization. *IEEE Transactions on Neural Networks and Learning Systems* **(IEEE TNNLS，中科院一区，影响因子：14.255)**, 32, no. 2 (2021): 814-825. （见附件4）

[7] Xiyue Gao, Zhuoqi Ma, Jiangtao Cui∗, Xiaofang Xia and **Cai Xu∗**. Hierarchical Category-Enhanced Prototype Learning for lmbalanced Temporal Recommendation. In ACM MM **(CCF A类会议)**, 2023.

[8] Linhao Zhang, Li Jin, Guangluan Xu, Xiaoyu Li, **Cai Xu**, Kaiwen Wei, Nayu Liu and Haonan Liu. CAMEL: Capturing Metaphorical Alignment with Context Disentangling for Multimodal Emotion Recognition. In AAAI (CCF A类会议), 2024.

**（3）在决策级多模态融合方面**，申请人针对质量不可控的多模态数据，设计了基于证据理论的多模态数据分析方法，通过直接建模每个源类别级的可信度，以及基于可信度的信息去伪存真，实现了决策级可信多模态融合。所提方法在提供高精度决策的同时，赋以对应的决策不确定性，实现了多模态数据可信分析。相关研究成果获得欧洲科学院院士、纽约州立大学的Keqin Li教授团队，IEEE Life Fellow、早稻田大学的Takuro Sato教授团队等知名学者团队的正面评价。代表性研究成果包括：

[9] **Cai Xu**, Jiajun Si, Ziyu Guan, et. al. Reliable Conflictive Multi-view Learning. In AAAI (CCF A类会议，最佳论文奖，当年国内唯一), 2024. （见附件6）

[10] **Cai Xu**, Wei Zhao∗, Jinglong Zhao, et. al. Uncertainty-aware Multi-view Deep Learning for Internet of Things Applications. *IEEE Transactions on Industrial Informatics*  **(IEEE TII，中科院一区，影响因子：11.648)**, 19, no. 2 (2023): 1456 - 1466. （见附件4）

[11]Jie Zhao, Ziyu Guan∗, **Cai Xu**, et. al. Charge Prediction by Constitutive Elements Matching of Crimes. In *IJCAI* **(CCF A类会议，oral)**, pp. 1-8. 2022.

**（4）在弱监督跨模态学习方面**，申请人初步建立了基于“大规模弱标签抗噪预训练+小规模精确标签精调”的弱监督深度表征学习框架。申请人在带不同类型弱标签的图像等模态数据上，提出适应多种弱标签场景（如单标签、多标签等）的弱监督深度表征学习方法，显著提升了深度学习算法在弱标签条件下的表征学习效果，验证了所提框架的有效性。相关研究成果获得英国皇家工程院国际院士、IEEE Fellow、赫尔辛基大学的Pan Hui教授团队，IEEE Fellow、A\*STAR的Ivor W. Tsang教授团队等知名学者团队的正面评价。代表性研究成果包括：

[12]Wei Zhao, **Cai Xu**, Ziyu Guan∗, et. al. TelecomNet: Tag-based Weakly-supervised Modally Cooperative Hashing Network for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **(IEEE TPAMI，中科院一区，CCF A类，影响因子：24.314)**, 11, no. 44 (2022): 7940 - 7954. （见附件2）

[13] **Cai Xu**, Ziyu Guan∗, Wei Zhao, et. al. Recommendation by users’ multimodal preferences for smart city applications. *IEEE Transactions on Industrial Informatics*  **(IEEE TII，中科院一区，影响因子：11.648)**, 17, no. 6 (2021): 4197-4205.

[14] Jiaming Liu, Yue Wu**∗**, Maoguo Gong, Qiguang Miao, Wenping Ma, **Cai Xu∗**. Exploring Dual Representations in Large-Scale Point Clouds: A Simple Weakly Supervised Semantic Segmentation Framework. In ACM MM (CCF A类会议), 2023.

综上，申请人团队在多模态学习领域具有较坚实的研究积累。通过本项目工作开展，期望进一步推动多模态学习乃至可信人工智能领域技术的进步。

2．**工作条件**（包括已具备的实验条件，尚缺少的实验条件和拟解决的途径，包括利用国家实验室、国家重点实验室和部门重点实验室等研究基地的计划与落实情况）；

申请人所在的西安电子科技大学计算机科学与技术学科，入选国家“双一流”重点建设学科，第四轮学科评估中获评“A”。据科睿唯安（Clarivate Analytics）发布的最新一期ESI学科排名数据，西安电子科技大学计算机科学学科稳居全球排名前1‰，在全球进入计算机科学学科前1%的765所科研机构中，西电位列全球第10，全国排名第3。申请人将依托信息感知技术“国家2011协同创新中心”和空天地一地化综合业务网全国重点实验室开展研究工作。“信息感知技术协同创新中心”于2014年以行业产业类第一名的优异成绩通过国家“2011计划”认定。空天地一地化综合业务网全国重点实验室原为“综合业务网理论及关键技术国家重点实验室”，于1991年开始建设，1995年10月通过国家验收并正式投入开放运行，2017年实验室评估成绩为“优秀”，2022年实验室通过重组进入全国重点实验室建设名单，并更为现名。实验室作为优秀人才的培养基地和科技创新的重要基地，汇聚了一批基础扎实、治学严谨的中青年学术骨干，形成了团结协作、勇于创新的科研群体。近年来，实验室获得国家科技奖励二等奖4项，省部级科技奖励10余项。拥有中国科学院/工程院院士、国家杰出青年科学基金获得者、教育部“长江学者”和国家万人计划领军等国家级高层次人才40余人，并建设有国家基金委创新研究群体、科技部重点领域创新团队、教育部创新团队和国防科技创新团队等国家级科研团队9个。同时，实验室建设有“高性能仿真与云计算平台”、“高性能视频采集与分析测量平台与实验环境”等仪器平台和高性能计算仿真系统等大型仪器设备，为项目的顺利完成提供了坚实保障。

申请人所在的“智能媒体计算”陕西省科技创新团队主要从事多媒体计算、机器学习和医疗辅助诊断等方面的研究工作，重点研究了社会媒体数据的表达、分析理解以及构建于其上的决策支持。团队现有教授2人，副教授1人，讲师3人，在读博士研究生10余人、硕士研究生20余人。团队成员主持了国家自然科学基金重点项目，以及陕西省重大基础研究计划项目、陕西省重点研发计划项目等20余项科研项目。相关研究成果获AAAI-24最佳论文奖、浙江省自然科学一等奖和陕西省自然科学优秀学术论文二等奖等科研奖励。

3．**正在承担的与本项目相关的科研项目情况**（申请人和主要参与者正在承担的与本项目相关的科研项目情况，包括国家自然科学基金的项目和国家其他科技计划项目，要注明项目的资助机构、项目类别、批准号、项目名称、获资助金额、起止年月、与本项目的关系及负责的内容等）；

（1）国家自然科学基金重点项目，基于医学影像的消化道肿瘤和炎症类疾病辅助诊断方法研究（项目编号：62133012），经费来源：国家自然科学基金委，经费额度 297 万元，起止日期：2022.01-2026.12。

项目负责人：赵伟

与本项目的关系：该项目建立了较全面的消化道疾病多模态医学数据库，包括多源影像数据生理指标以及病情描述等，接着研究了基于医学影像的消化道疾病辅助诊断方法，以及基于多模态医学数据的疾病识别和分期评估模型，最终搭建消化道疾病辅助诊断系统。该项目的研究成果为本项目的研究工作提供数据、技术以及应用平台方面的基础。本项目与之不同之处在于：1）该项目只局限于医疗辅助诊断场景，而本项目考虑更广的应用场景，包括医疗辅助诊断、推荐系统、自动驾驶等；2）该项目的研究目标在于提升辅助诊断算法的准确度，而本项目还力求提升算法的可信度，进一步促进算法的落地应用。

负责内容：申请人是项目组的骨干成员（4/10，校内第2），负责基于多模态数据特征的消化道疾病识别与评估方法。

（2）国家自然科学基金青年科学基金项目，基于弱标注非平衡多模态数据的可解释推荐方法研究（项目编号：62103314），经费来源：国家自然科学基金委，经费额度 30 万元，起止日期：2022.01-2024.12。

项目负责人：徐偲

与本项目的关系：该项目主要研究多模态异常数据检测方法，非平衡不完整多模态数据融合补全方法以及可解释推荐方法，进而实现高精度的个性化推荐。该项目的研究成果能够为本项目的研究工作提供基础和支持。本项目与之不同之处在于：1）该项目只局限于推荐场景，而本项目考虑通用的多模态应用场景，是该项目研究目标的升华与发展；2）该项目提出的异常检测方法，目标在于删除异常点，而本项目拟衡量针对异常数据的决策可信度，适用范围更广；3）该项目提出了不完整多模态数据的补全方法，本项目进一步考虑了补全内容的置信度，实现了可信补全。

负责内容：申请人主要负责方案总体设计、算法设计与实施、结果论证等。

4．**完成国家自然科学基金项目情况**（对申请人负责的前一个已资助期满的科学基金项目（项目名称及批准号）完成情况、后续研究进展及与本申请项目的关系加以详细说明。另附该项目的研究工作总结摘要（限500字）和相关成果详细目录）。

无。

**（三）其他需要说明的情况**

1. 申请人同年申请不同类型的国家自然科学基金项目情况（列明同年申请的其他项目的项目类型、项目名称信息，并说明与本项目之间的区别与联系）。

无。

2. 具有高级专业技术职务（职称）的申请人或者主要参与者是否存在同年申请或者参与申请国家自然科学基金项目的单位不一致的情况；如存在上述情况，列明所涉及人员的姓名，申请或参与申请的其他项目的项目类型、项目名称、单位名称、上述人员在该项目中是申请人还是参与者，并说明单位不一致原因。

无。

3. 具有高级专业技术职务（职称）的申请人或者主要参与者是否存在与正在承担的国家自然科学基金项目的单位不一致的情况；如存在上述情况，列明所涉及人员的姓名，正在承担项目的批准号、项目类型、项目名称、单位名称、起止年月，并说明单位不一致原因。

无。

4. 其他。

无。