

项目编号	
------	--

华东理工大学

大学生创新创业训练计划创新训练类 项目申报书

项目名称	针对异构数据缺陷的多模态联邦学习算法研究
项目所属一级学科	计算机科学与技术
项目负责人	朴贞霓
项目组成员	陶若愚、蔡鹏祥、刘若冰
指导教师	董文波（讲师）
所在学院	信息科学与工程学院
填报日期	2024.10.24

华东理工大学创新创业教育中心制表

填表说明

一、本表要求按顺序逐项填写，项目要实事求是，讲究诚信，不能有雷同，表达要明确、严谨。空缺项要填“无”。一律用 A4 纸打印，于左侧装订成册。

二、申请参加本项目团队人数不得超过 4 人（1 人为立项负责人，参与合作研究者 1-3 人）。

三、申请参加本项目的个人或团队必须聘请教师作为项目指导教师，并请指导教师在申请书上签名。

四、项目执行周期一般为 1 年半，允许项目开展 1 年后申请提前结题。

五、本表由指导教师所在院系初审，签署意见后，由学院统一保存。

一、项目基本情况

项目名称	针对异构数据缺陷的多模态联邦学习算法研究					
项目所属一级学科（可以是一种或多种跨学科）	计算机科学与技术					
项目来源	<input type="checkbox"/> 学生自选课题 <input checked="" type="checkbox"/> 导师科研项目					
	<input checked="" type="checkbox"/> 与专业实践相关 <input type="checkbox"/> 与课程设计相关 是否以 USRP 课题为基础 <input type="checkbox"/> 是 <input checked="" type="checkbox"/> 否 如果是，请填写 USRP 课题名称：					
成果导向	<input type="checkbox"/> 以参与科创竞赛为导向，请填写竞赛名称： <input checked="" type="checkbox"/> 以发表学术论文或专利为导向					
负责人	姓名	学号	所在院系	专业	手机	E-mail
	朴贞霓	23013150	信息科学与工程学院	软件工程	15844231749	2284029206@qq.com
参与组员	陶若愚	23013187	信息科学与工程学院	计算机科学与技术	18698712178	1932529304@qq.com
	蔡鹏祥	22010402	信息科学与工程学院	软件工程	18762362387	caipengxiang1230@163.com
	刘若冰	23013293	信息科学与工程学院	人工智能	13818493585	1542136761@qq.com
指导教师	姓名	董文波	所在院系	信息科学与工程学院		
	联系电话	15901920120	手机	15901920120	E-mail	wbdong@ecust.edu.cn
	职务/职称	讲师		主要研究方向	多模态学习和多目标优化	

二、项目申请理由

(选题的来源, 国内外同类研究工作的现状、方法、难点等, 前期准备工作)

2.1 选题来源

随着大数据时代的到来, 数据的获取和利用变得愈发重要。然而, 数据的隐私保护和安全性问题也随之而来, 尤其是在涉及个人隐私的领域, 如医疗、金融和社交网络等, 如何保护数据隐私, 加强数据安全性的问题迫在眉睫。联邦学习 (federated learning, FL) 作为一种新兴的分布式学习框架, 为解决这些问题提供了创新的解决方案。例如, 在医疗领域^[1-4], 各医院可以通过联邦学习共享经验, 而无需暴露患者的个人信息。在金融行业^[5-7], 银行可以通过协作学习提升反欺诈模型的准确性, 同时保护客户的隐私。联邦学习允许多个参与方在保护数据隐私的前提下, 共同训练机器学习模型。通过在本地设备上计算, 联邦学习避免了将敏感数据集中到一个服务器上, 不仅提高了模型的性能, 也加强了数据的安全性。

尽管联邦学习在隐私保护和数据安全方面具有显著优势, 但大多数现有的 FL 方法在训练时仍使用的是单模态数据。随着边缘计算的不断发展, 不同传感器和设备产生的数据具有不同模态 (如触觉、视觉、听觉等)。不同模态的数据可能具有不同的分布特征、维度和类型, 这种异构性使得模型训练变得复杂。因此, 对于多模态联邦学习 (multi-modal machine federated learning, MMFL) 的研究是有必要的。

然而, 目前针对 MMFL 的研究大多集中在提升模型性能、提高模型准确率等方面, 对数据异构的研究相对较少。市面上缺乏解决异构数据缺陷问题的针对性算法, 大多是基于图神经网络、整合不同模态、构建模态之间关联图的算法方式。由此可见, 在现有算法上进行改进与创新, 进一步加强 MMFL 对于异构数据缺陷的识别与应用能力, 是一项十分具有尝试价值的研究课题。

因此, 本研究将重点探讨“针对数据异构缺陷的多模态联邦学习”, 旨在有效整合不同模态的数据, 同时考虑数据的异构性, 用以改善 MMFL 在面对不同模态、异构数据、缺陷数据的作用结果, 并为实际应用提供更强的支持。本项目一方面, 可以推动联邦学习的进一步发展, 使其在更多实际场景中发挥作用。另一方面, 这项研究能提高模型的准确性和鲁棒性, 从而为用户提供更优质的服务, 最终实现数据价值的最大化和隐私保护的平衡, 从而为推动人工智能技术在各行业的应用做出积极贡献。

2.2 国内外同类研究现状

随着人工智能技术的迅速发展, MMFL 作为一种新兴的学习框架, 开始受到广泛关注。它允许不同的数据源在没有共享原始数据的情况下进行联合学习, 尤其适用于数据隐私至关重要的领域。然而, 数据异构性, 即不同客户端之间数据类型、分布和质量的差异, 给多模态联邦学习带来了诸多挑战。

项目背景和意义

2.2.1 研究现状

多模态联邦学习（MMFL）指的是在保证数据隐私的前提下，结合来自不同模态的数据（如图像、文本、语音等），通过分布式学习算法进行训练。目前，这一领域的研究主要集中在如何应对数据异构性、模态缺失、噪声干扰以及如何实现不同模态间的知识共享和整合等问题。

（1）异构性问题与解决方案

多模态数据的异构性是多模态联邦学习中的一个主要挑战。数据来自不同源且具有不同的表示形式，常见的如视觉数据、文本数据、语音数据等。各模态的分布差异、数据的质量差异以及模态之间的不平衡性，都会影响联邦学习模型的训练效果。因此，如何有效处理这些异构数据，尤其是在设备计算能力差异较大的情况下，成为了一个亟待解决的问题。

Lin 等^[8]在《Federated Learning on Multimodal Data: A Comprehensive Survey》这篇综述文章中做出了总结，为从事多模态联邦学习的研究人员提供了系统的指导。为了处理异构数据中的模态缺失问题，Ma 等^[9]提出的 SMIL 模型基于贝叶斯元学习，使用特征重建网络将利用可用的模态以高效的方式生成缺失模态特征的近似值，具有灵活性和高效性，但是该模型仅关注模态的整体缺失，缺乏对每组数据的确切评估与调整。在此之上，Ma 等^[10]进一步研究了异构数据环境下的多模态联邦学习，提出了一种自适应重加权机制。该机制能够动态调整各模态的影响力，从而有效提升模型性能。通过在训练过程中实时评估模态的重要性，该方法不仅增强了模型的学习能力，还改善了整体的学习效率。这一研究为多模态数据的有效整合提供了新的思路。类似的，Zheng 等^[11]构建了一种名为 AutoFed 的模型，通过引入伪标签模型和数据填补方法，解决了在异构感知数据下的联邦学习问题。通过增强模态间的信息共享和补充，AutoFed 能够更好地适应不同模态的数据缺失和不一致性问题，提高了系统的鲁棒性和学习效率。Chen 等^[12]提出了 FedDAT 框架，这是首个针对异构多模态（视觉-语言）FL 的基础模型进行分布式微调的方法。FedDAT 基于双适配器教师（Dual-Adapter Teacher, DAT）模块的框架，利用互知识蒸馏（Mutual Knowledge Distillation, MKD）技术来实现知识的有效转移。该方法通过正则化客户端本地更新，缓解了异构数据下模型训练的挑战，并实现了跨模态的知识共享。该框架不仅提高了多模态联邦学习的鲁棒性，还为基础模型的分布式微调提供了新的方法。

（2）模态融合与协作

模态间的有效融合与协作对于多模态联邦学习的成功至关重要。模态之间往往存在着相互补充的关系，因此如何通过合理的策略进行模态融合，进而实现信息的有效整合，仍是一个重要的研究方向。

Gao 等^[13]的研究指出，多模态数据融合仍处于早期阶段，难点在于如何建模不同模态间的相互关系和互补性。该综述进一步探讨了多模态数据融合的主要挑战，包括数据缺失、噪声干扰、模态不平衡等问题，并提出了一些应对措施，如利用对抗学习来增强模型对噪声的鲁棒性，以及通过注意力机制动态分配各模态的权重。作者指出未来的研究

方向，强调了可解释性、资源优化和在实际复杂环境中的应用前景。论文为多模态数据融合领域的研究者提供了丰富的理论参考和实践指导。Li 等^[14]提出了 Cross-Modal Meta Consensus (CMMC)，通过元学习 (Meta-Learning) 策略来协调模态内和模态间的特征空间，从而促进不同模态之间的知识共享。该方法能够有效解决模态分布差异带来的初始化点冲突问题和梯度冲突问题，实现了更为顺畅的模态融合。此外，Chen 等^[15]提出了 FedMSplit 方法，在多模态联邦学习中引入了分割网络和动态权重自适应机制，通过在不同任务和异构数据之间找到合适的平衡来优化模型的性能。该方法提升了多模态数据在复杂场景下的学习能力，但其高计算和通信负载要求也对实际应用提出了挑战。

(3) 应对慢节点与计算资源的挑战

在异构系统中，不同设备的计算能力和网络连接速度差异较大，部分设备可能成为“慢节点”，影响整个系统的训练效率。这一问题尤其在大规模联邦学习环境中更为突出。为了提高训练效率，研究者们提出了一些创新方法来缓解这一问题。Chen 等^[16]提出了一种基于知识蒸馏的加速方法，通过设计一个知识蒸馏框架，使得不同节点可以通过蒸馏技术从全局模型中提取重要信息，从而加速训练过程。该方法不仅能够有效应对设备异构性问题，还能在保证数据隐私和模型准确性的前提下，减少慢节点对全局训练进度的影响。

(4) 领域特定应用与挑战

多模态联邦学习在许多领域具有广泛的应用潜力，尤其是在医疗、自动驾驶等高需求场景中。随着医学领域中数据来源的多样性（如医学影像、电子病历、基因数据等），传统的联邦学习方法面临着数据模态间融合问题。在关于医疗领域的多模态联邦学习研究中，Thrasher^[17]重点探讨了多模态联邦学习在医疗数据中的应用，提出了一种改进的联邦学习框架，能够有效处理医疗数据中的异构模态。Yang 等^[18]针对医疗领域提出了一种新颖的模态融合策略，旨在结合医疗图像和文本数据，以提高疾病预测的准确性。在实际医疗场景中，不同模态的医疗数据可以提供互补信息，通过多模态学习能显著提高模型对复杂疾病的诊断能力。

自动驾驶是另一个典型的多模态应用场景。在自动驾驶领域的研究中，AutoFed^[11]使用了面向异构感知数据的多模态联邦学习方法，通过利用伪标签和数据填补来解决不同感知模态（如视觉、雷达、激光雷达等）之间的差异，从而提高系统的感知能力和鲁棒性。

2.2.2 研究现状总结

综上所述，现有研究在模态缺失、模态融合、快慢节点等方面作出了许多贡献，但在“数据异构的多模态联邦学习”领域，依旧面临多重挑战。首先，数据缺失是一个突出问题。由于设备异构性和环境影响，某些模态的数据可能不完整，这使得模型在训练时难以获得全面的信息，从而影响准确性和鲁棒性。其次，计算成本也是一个重要挑战。在异构数据环境下，缺失模态的客户端往往需要更多计算资源来弥补数据的不足，这与联邦学习旨在降低计算负担的目标相悖。最后，数据异构性使得不同客户端的数据难以有效整合，导致模态间的融合差异影响训练效果，增加了模型处理多模态数据的复杂性。

针对上述问题与缺陷，本项目在关于模态缺失的多模态学习的算法上作出优化，提出

了 Data Loss Processing Module (DLPM) 的模块, 通过利用新的数据补全机制, 化解了不同模态中的数据缺失问题; 通过建立数据矫正模块, 增强了模型的准确性和鲁棒性; 通过采取对抗性学习机制, 实现了不同模态间的数据融合。在以上机制与模块的作用下, DLPM 于“具有数据缺陷的异构数据的多模态联邦学习”领域效果显著, 并减少了模型的计算成本。

2.3 前期准备工作

(1) 多模态学习与联邦学习的知识积累

1) 多模态学习基础: 深入学习多模态学习的概念、技术框架以及其在计算机视觉、自然语言处理、智能医疗等领域的典型应用。重点理解如何在模型中对多模态信息进行特征提取、融合以及联合学习。熟悉如图像、音频、文本等模态特征提取的基础模型和方法, 包括卷积神经网络 (Convolutional Neural Network, CNN)、长短期记忆网络 (Long Short-Term Memory network, LSTM) 等, 并重点学习如何利用 LSTM 在时间序列分析中处理动态数据。

2) 联邦学习的概念与实现: 了解联邦学习的基本架构与工作原理, 尤其是如何在分布式数据环境下实现模型训练, 保护用户数据隐私。研究联邦平均算法 (Federated Averaging, FedAvg) 等常用的联邦学习方法, 分析联邦学习在数据隐私保护、多方协作中的优缺点, 熟悉不同模态数据在多方环境下的共享与处理方法。

3) 异构数据的处理: 学习在多模态学习中如何处理数据缺陷问题, 尤其是缺失模态、噪声模态的相关处理技术。查阅并理解有关不完全模态数据处理的现有模型, 如缺失模态补全模型、权重调整模型等, 重点掌握如何通过动态权重调整、数据填充等方式解决数据不完整的影响。

(2) 文献查阅

1) 不完全模态多模态学习现状: 系统检索国内外关于不完全模态学习的最新文献, 分析常见的解决方法, 如模态补全、动态权重调整、特征变换等, 归纳出当前领域中解决数据缺陷问题的有效方法。探讨现有方法在不同应用场景 (如多模态情感分析、医学影像分析、环境监控) 中的优缺点, 为本项目的模型设计提供借鉴。

2) 基于 LSTM 的时间序列特征提取方法: 查阅时间序列数据处理的相关文献, 分析 LSTM 在时间序列数据中的应用情况, 尤其是如何在数据不完整的条件下提取关键特征。同时研究动态特征处理方法, 关注不同模态中对特征权重的自适应调整机制, 为后续研究动态学习权重提供技术支持。

3) 动态权重学习的前沿研究: 调研国内外在多模态数据中的动态权重调整策略, 分析现有研究在处理多模态数据缺失、模态异构性问题时采用的权重调整方案。参考权重图的生成方法与模型设计, 探索动态权重图生成方法的技术细节与实际应用效果, 为项目中各模态特征的加权组合提供理论支撑。

(3) 数据库的选择

从各种边缘设备、传感器和可穿戴设备中收集不同模态的本地数据。例如，在医疗领域可能包括 X 光图像、CT 扫描图像、MRI 图像等；在智能家居场景中，可能涉及 RGB 摄像头拍摄的视频、身体传感器记录的数据等。这些数据的收集要确保质量和多样性，以更好地反映真实世界的情况。下面介绍几个常用的数据集。

1) The Multi-Temporal Urban Development SpaceNet Dataset^[19]: 该数据集是一个新的 SpaceNet 数据集，包含每个月拍摄的建筑区域的卫星图像。目标是在空间时间序列的帮助下在全球范围内跟踪这种建筑活动。

2) Spoken Moments^[20]: 这是一个包含 50 万个描述各种不同事件的短视频音频描述的语料库。该数据集构建了一个结合视听信息的提议架构，并提供了一个优雅的解决方案来使用自适应平均边距 (Adaptive Margin Metric, AMM) 方法来解决视频/字幕检索问题。

3) Conceptual 12M^[21]: 这是谷歌研究团队通过放宽数据抓取的限制，将 CC-3M 数据集扩展到 1200 万个图像字幕对的数据集。该数据集对于学习视觉概念的更广义的文本表示非常有帮助。在概念 12M 数据集上预训练的神经图像标题模型能够学习长尾概念，即数据集中非常具体且罕见的概念。

4) Euro-PVI^[22]: 这是旨在通过在行人和骑自行车者轨迹的标记数据集上训练模型来预测行人将采取什么样的轨迹来响应接近的车辆的数据集。Euro-PVI 数据集包含有关行人车辆交互的丰富信息，例如场景中所有参与者的视觉场景、速度和加速度。

(4) 损失函数设计

针对多模态数据的异构数据缺陷的特点，设计合适的损失函数来衡量模型的性能。例如，可以使用交叉熵损失函数来分类任务，或者使用均方误差损失函数来回归任务。同时，还可以考虑引入对比学习、自监督学习等方法来提高模型的泛化能力和鲁棒性。

Yang 等^[23]提出蒸馏损失函数，结合传统交叉熵损失和基于教师模型输出概率分布的蒸馏损失的损失函数，通过最小化学生模型输出与教师模型输出之间的差异，使得学生模型能够学习到教师模型的知识，从而提高学生模型的性能。Yang 等^[24]提出 ConFEDE 框架，通过统一损失函数将多种学习机制整合在一起，使得模型能够通过多任务预测损失同时学习多模态预测和单模态预测。这种学习策略有助于模型更好地捕捉情感的细微差别，尤其是在不同模态之间提供互补信息的情况下。

2.4 项目意义

(1) 提升数据采集系统的鲁棒性与精确性

在实际应用中，数据采集设备常常面临诸如电源波动、噪声干扰、光线变化和磁场影响等多种环境因素的干扰。这些干扰会导致多模态数据的缺失或偏差，从而影响设备对环境的监测与反馈。通过设计一种鲁棒的多模态联邦学习方法，本项目旨在确保设备即使在数据不完整的情况下，依然能够基于有限的多模态数据做出准确决策，进而提升数据采集系统的可靠性和准确性。例如，在智能交通系统中，项目成果可以帮助交通监控设备在夜

间光线不足或存在干扰的环境中，持续提供交通状况的准确反馈。

(2) 提升智能设备的适用性与用户体验

随着智能设备在日常生活中的普及，对其在多样化环境中稳定运行的需求日益增加。本项目能够确保智能设备在面对复杂环境或数据缺陷时，依然保持有效性和稳定性。以便携式健康监测设备为例，许多此类设备在长期使用过程中可能会遭遇信号衰减或数据丢失的问题。通过本项目的技术，即便在数据不全的情况下，设备仍能提供稳定、可靠的健康状态预测，从而提升用户体验，增强用户对智能设备的信任度。

(3) 为多模态学习与联邦学习技术研究提供新方向

本项目的技术路线包括多模态数据处理、动态权重学习机制及不完全模态数据的自适应融合。这些技术思路不仅为多模态联邦学习中的数据缺陷问题提供了解决方案，也为学术界探索如何处理不完整模态提供了新的研究视角。项目的研究成果将为多模态学习和联邦学习领域的进一步研究提供理论基础和实践经验，促进相关领域的技术进步。

参考文献：

- [1] Pallavi Dhade and Prajakta Shirke, Federated Learning for Healthcare: A Comprehensive Review, Engineering Proceedings, 2023, vol. 59, pp. 230-239.
- [2] Sunghwan Moon and Wonhee Lee, Privacy-Preserving Federated Learning in Healthcare, 2023 International Conference on Electronics Information and Communication (ICEIC), 2023, pp. 1 - 4.
- [3] Guodong Long, Tao Shen, Yue Tan, Leah Gerrard, Allison Clarke and Jing Jiang, Federated Learning for Privacy-Preserving Open Innovation Future on Digital Health, Humanity Driven AI, 2021, pp. 113 - 133.
- [4] Maryamossadat Aghili, Solale Tabarestani and Malek Adjouadi, Addressing the missing data challenge in multi-modal datasets for the diagnosis of Alzheimer's disease, Journal of Neuroscience Methods, 2022, vol. 375, pp. 109582 - 109591.
- [5] Aydin Abadi, Bradley Doyle, Francesco Gini, Kieron Guinamard, Sasi Kumar Murakonda, Jack Liddell, Paul Mellor, Steven J. Murdoch, Mohammad Naseri, Hector Page, George Theodorakopoulos and Suzanne Weller, Starlit: Privacy-Preserving Federated Learning to Enhance Financial Fraud Detection, IACR Cryptology ePrint Archive, 2024, vol. 2024, pp. 90 - 123.
- [6] Tao Liu, Zhi Wang, Hui He, Wei Shi, Liangliang Lin, Ran An, and Chenhao Li, Efficient and Secure Federated Learning for Financial Applications, Applied Sciences, 2023, vol. 13, pp. 5877 - 5894.
- [7] Theodora Nevrataki, Anastasia Iliadou, George Ntolkeras, Ioannis Sfakianakis, Lazaros Lazaridis, Georgios Maraslidis, Nikos Asimopoulos, and George F. Fragulis, A Survey on Federated Learning Applications in Healthcare, Finance, and Data Privacy/Data Security, AIP Conference Proceedings, 2023, pp. 120015 - 120021.

- [8] Yiming Lin, Yuan Gao, Maoguo Gong, Sijia Zhang, Zhiyuan Li and Yuanqiao Zhang, Federated Learning on Multimodal Data: A Comprehensive Survey, *Machine Intelligence Research*, 2023, Res. 20, pp. 539 - 553.
- [9] Mengmeng Ma, Jian Ren, Long Zhao, Sergey Tulyakov, Cathy Wu and Xi Peng, SMIL: Multimodal Learning with Severely Missing Modality, *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, pp. 2302 - 2310.
- [10] Benteng Ma, Yu Feng, Geng Chen, Changyang Li and Yong Xia, Federated adaptive reweighting for medical image classification, *Pattern Recognition*, 2023, vol. 144, pp. 109880 - 109899.
- [11] Tianyue Zheng, Ang Li, Zhe Chen, Hongbo Wang and Jun Luo, AutoFed: Heterogeneity-Aware Federated Multimodal Learning for Robust Autonomous Driving, *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, 2023, pp. 1 - 14.
- [12] Haokun Chen, Yao Zhang, Denis Krompass, Jindong Gu and Volker Tresp, FedDAT: An Approach for Foundation Model Finetuning in Multi-Modal Heterogeneous Federated Learning, *AAAI Conference on Artificial Intelligence*, 2023, pp. 11285 - 11293.
- [13] Jing Gao, Peng Li, Zhikui Chen and Jianing Zhang, A Survey on Deep Learning for Multimodal Data Fusion, *Neural Computation*, 2020, vol. 32, pp. 829 - 864.
- [14] Shuai Li, Fan Qi, Zixin Zhang and Changsheng Xu, Cross-Modal Meta Consensus for Heterogeneous Federated Learning, In *Proceedings of the 32nd ACM International Conference on Multimedia*, 2024, pp. 975 - 984.
- [15] Jiayi Chen and Aidong Zhang, FedMSplit: Correlation-Adaptive Federated Multi-Task Learning across Multimodal Split Networks, In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2022, pp. 87 - 96.
- [16] Jinqian Chen, Haoyu Tang, Junhao Cheng, Ming Yan, Ji Zhang, Mingzhu Xu, Yupeng Hu and Liqiang Nie, Breaking Barriers of System Heterogeneity: Straggler-Tolerant Multimodal Federated Learning via Knowledge Distillation, *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 2024, pp. 3789 - 3797.
- [17] Jacob Thrasher, Alina Devkota, Prasiddha Siwakoti, Rohit Chivukula, Pranav Poudel, Chaunbo Hu, Binod Bhattarai and Prashna Gyawali, Multimodal Federated Learning in Healthcare: A Review, *arXiv*, 2023, abs/2310.09650, pp. 1 - 28.
- [18] Yujun Yang, Haibo Zhang and Ming Chen, Multi-modal Federated Learning for Medical Data with Heterogeneous Distributions, *IEEE Journal of Biomedical and Health Informatics*, 2023, vol. 27, no. 4, pp. 1080 - 1090.
- [19] Adam Van Etten, Daniel Hogan, Javier Martinez-Manso and Ryan Lewis, The Multi-Temporal Urban Development SpaceNet Dataset, *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 6394 - 6403.

	<p>[20] Mathew Monfort, SouYoung Jin, Alexander Liu, David Harwath, Rogerio Feris, James Glass and Aude Oliva, Spoken Moments: Learning Joint Audio-Visual Representations from Video Descriptions, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14866 – 14876.</p> <p>[21] Soravit Changpinyo, Piyush Sharma, Nan Ding and Radu Soricut, Conceptual 12M: Pushing Web-Scale Image-Text Pre-Training to Recognize Long-Tail Visual Concepts, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3357 – 3567.</p> <p>[22] Apratim Bhattacharyya, Daniel Olmeda Reino, Mario Fritz and Bernt Schiele, Euro-PVI: Pedestrian Vehicle Interactions in Dense Urban Centers, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 6404 – 6413.</p> <p>[23] Zhiqin Yang, Yonggang Zhang, Yu Zheng, Xinmei Tian, Hao Peng, Tongliang Liu and Bo Han, FedFed: Feature Distillation Against Data Heterogeneity in Federated Learning, ArXiv, 2023, abs/2310.05077, pp. 1 - 32.</p> <p>[24] Jiuding Yang, Yakun Yu, Di Niu, Weidong Guo and Yu Xu, ConFEDE: Contrastive Feature Decomposition for Multimodal Sentiment Analysis, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, 2023, pp. 7617 – 7630.</p>
项目申请理由	<p>(项目团队成员具备的知识基础、能力素养、兴趣特长和已参加的科研等条件)</p> <p>在当今数据驱动的社会中，数据的准确性和完整性至关重要。本项目针对数据缺失问题，尤其是在多模态数据收集过程中，旨在提出有效的解决方案。这不仅有助于提升数据采集系统的鲁棒性与准确性，还能改善智能设备在复杂环境中的性能，为日常生活提供更可靠的技术支持。</p> <p>本项目在多模态学习与联邦学习领域的创新研究，将为学术界提供新的理论支持与实践方法。通过针对数据缺失的动态加权融合机制和自适应填补方法，项目将丰富现有文献，推动相关领域的进一步探索。我们的研究成果将为学术界如何处理不完整模态提供新的视角，并为未来的研究指明方向。</p> <p>团队成员系统学习了数据结构、程序设计、人工智能等核心课程，对多模态学习、深度学习的基本原理有深入理解。这种知识基础使我们能够理解和应用当前的多模态学习和联邦学习模型。</p> <p>团队成员熟悉常用的编程语言如 Python 和 Matlab，能够利用 Pytorch 等深度学习框架构建、训练和优化模型。此外，团队成员对时间序列数据、动态权重调整等技术有初步掌握，能够较好地应对项目中设计和实现 LSTM 模块以及动态权重学习的需求，为模型的设计与优化提供技术支持。团队成员对人工智能、多模态学习等领域有浓厚兴趣，尤其关注在数据不完整的情况下如何保持模型的鲁棒性。部分成员专注于深度学习模型优化、算法调整，尤其对动态权重调整、数据特征提取方面有深入研究的热情。这种兴趣促使团队在课外积极学习前沿技术、探索新算法，具备一定的技术储备。此外，成员在多模态数据融合与处理、隐私保护方面也有实践经验，能够将兴趣转化为行动力，为项目的创新性研究奠定坚实基础。</p>

三、项目研究内容

(主要研究内容, 重点和难点, 研究思路和方法等)

3.1 主要研究内容

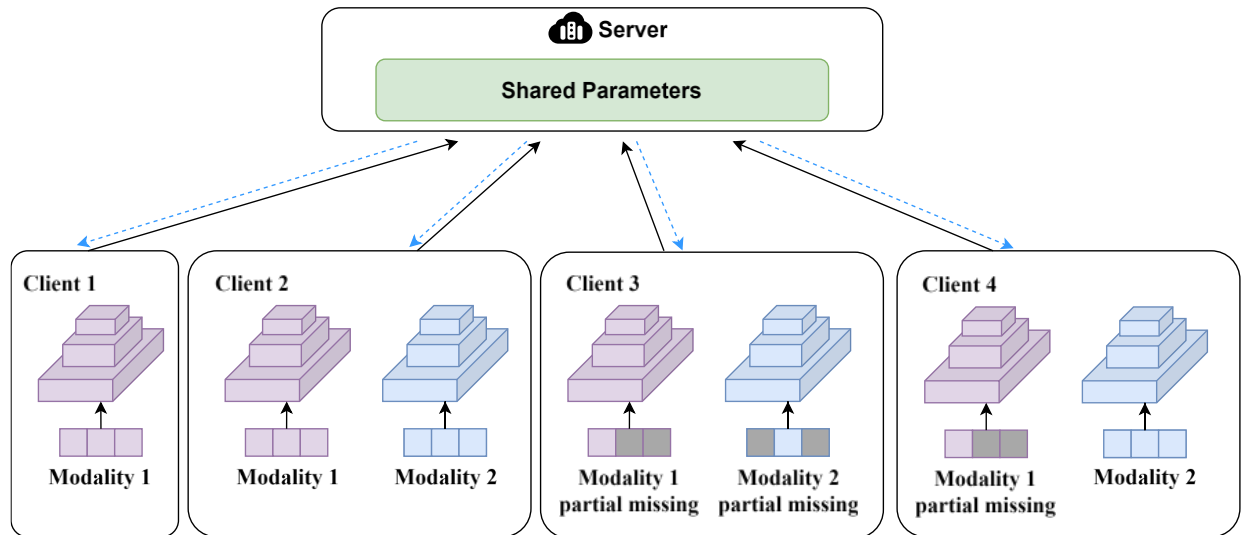


图 1 多模态联邦学习实际应用时面临的四种普遍场景——模态缺失(client 1)、模态完整(client 2)、完全数据缺陷(client 3)、部分数据缺陷(client 4)。本研究提出的模型期望能应对数据异构导致的数据不完整和模态缺失等场景

本研究的核心内容集中在多模态联邦学习 (Multi-modal Federated Learning, MMFL) 领域, 旨在解决多模态数据的异构性和不完整性所带来的挑战。研究将深入探讨如何通过先进的机器学习技术, 特别是长短期记忆网络 (Long Short-Term Memory, LSTM) 和注意力机制, 来提升模型对不完整多模态数据的处理能力。

本研究提出了一种新颖的数据缺陷处理模块 (Data Loss Processing Module, DLPM), 旨在对数据集中的缺失部分进行有效的补全。该模块首先利用 LSTM 对含有缺陷的原始数据进行处理, 以捕捉数据的时间序列特性。随后, 通过补全器 (Complementor) 模块, 对缺失值进行预测并生成补齐后的原始数据。在此过程中, 模型将预测的补齐数据与真实值进行比较, 通过 L2 范数 (即欧几里得范数) 来衡量两者之间的差异, 以确保补齐数据在特征上与原始数据保持一致性。

鉴于补全过程中可能产生的误差以及对模型鲁棒性的挑战, 本研究进一步引入了模态数据校正模块 (Modality Data Correction)。该模块综合利用其他模态的信息, 生成一个综合特征值, 并与对齐后的数据进行加权融合。加权融合过程中的权重系数由模型预测的数据缺陷率决定, 从而实现了对数据缺陷的自适应调整。

最终, 通过对抗性学习机制, 模型能够生成与原始特征分布相匹配的特征数据, 同时能够有效地利用其他模态信息, 补充当前数据可能存在的信息缺失。这一方法不仅提高了数据补全的准确性, 也增强了模型对异常值和噪声的鲁棒性。

项目研究的主要内容及目标

3.2 研究重点和难点

在多模态联邦学习领域，研究者们面临着由数据异构性和不完整性所带来的挑战。多模态学习通过整合不同模态的信息，能够显著提升模型的泛化能力。然而，现实世界中的多模态数据常常存在不完整性，这限制了全局模型的准确性。现有的数据融合策略，例如渐进式融合和正交序列融合，以及数据对齐方法，如 ALBEF 中采用的动量蒸馏和动态时间规整，虽然能够提升模型训练的效果，但在处理存在缺陷的多模态数据时，可能会引起对齐紊乱和融合异常，从而影响模型的学习效果。

特别地，当客户端数据模态存在缺失时，模型可能会出现不平衡学习，导致模型在某些方面的性能存在显著缺陷。在这种情况下，传统的数据对齐和融合技术将不再适用。如果利用这种模态不平衡的数据进行训练，将使得模型性能偏向大数据量的模态，导致模型的鲁棒性降低。

目前常用的方法是人工数据清洗，例如通过 K 最近邻 (K-Nearest Neighbors, KNN)、长短期记忆网络 (LSTM) 补全或使用生成对抗网络 (Generative Adversarial Network, GAN) 生成缺失数据。这些方法虽然能够缓解数据不完整性的问题，但同时也增加了额外的成本，并且可能不利于模型的学习。因此，在数据缺陷和异构性并存的条件下，如何有效地融合和分析不完整的多模态数据，以提升多模态联邦学习模型的准确性和鲁棒性，成为一个亟待解决的学术难题。

3.3 研究思路和方法

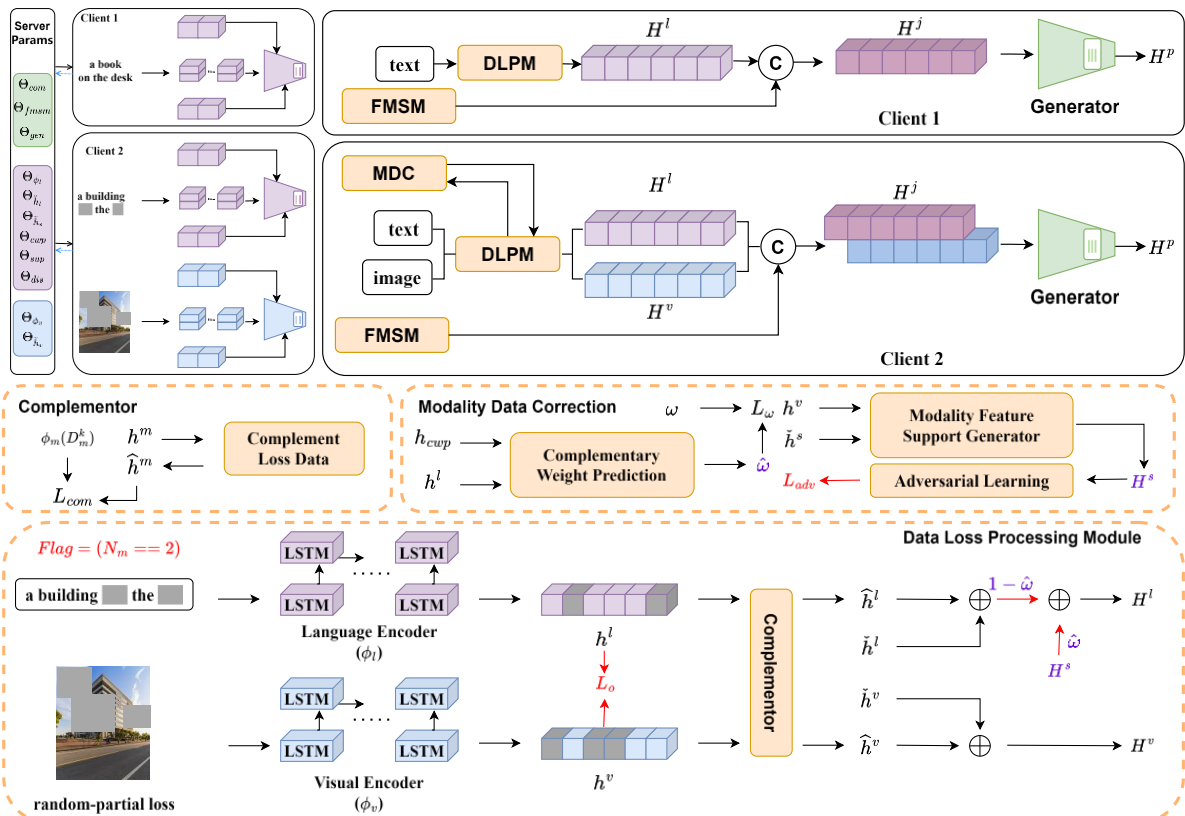


图2 (a) 客户端模型主体结构, 本地通过输出数据, 计算梯度后, 将与服务器上进行参数共享, 服务器返回训练后的模型参数, 辅助本地模型进行优化。(b) DLPM 模块结构图, 该模块通过数据补全以及数据协同纠正, 很大程度上降低了数据缺陷带来的负面影响。为了简化流程, 我们假设模态数据为文本和图像, 并仅对文本进行模态数据纠正(Modality Data Correction, MDC)优化。此情景下, 仅当本地模态数量等于2时, 会调用 MDC 模块

在本研究中, 我们的目标是通过改进 Multimodal Federated Learning Framework for Incomplete Modalities (FedInMM), 提升模型在数据缺陷时的性能。本研究提出了一种新颖的数据缺陷处理模块 (DLPM), 旨在对数据集中的缺失部分进行有效的补全。

(1) 问题重新和变量定义

我们定义每个客户端的输入数据为 $D = \{(D_i^k, i)_{i \in M}\}_{k=1}^K$, 其中 $D_i^k \in \mathbb{R}^{1 \times d_k}$, 表示 k 号客户端的输入模态 i 的数据, 每个数据的维度为 d_k 。我们定义 N_M 表示模态种类数。为了简化问题, 本文假设客户端最多只输入两种模态的数据(即 $M = \{l, v\}$, $N_M = 2$), 分别为用 l 表示语言模态, v 表示视觉模态。

(2) 数据随机掩码

为了训练模型对数据缺陷的处理能力, 我们将在训练数据上随机进行掩码, 人为制造缺陷率 ω 为 0.1 ~ 0.4 的仿真数据, 模拟真实场景下的数据缺陷问题。具体来说, 我们生成满足均匀分布的随机数矩阵 $Mask_i^k$, 尺寸与 D_m^K 相等。然后根据 ω 将 $Mask_i^k$ 数值映射到 $\{0,1\}$, 然后按位与 D_m^K 相乘, 计算公式如下:

$$\tilde{D}_m^K = Mask_i^k * D_m^K \quad (1)$$

其中 \tilde{D}_m^K 表示 k 号客户端的模态 m 数据经过随机掩码后的数据

(3) 正交损失优化

我们在 FedInMM 上进行改进, 争对每种模态都设计了一个编码器(Encoder), 表示为 ϕ_m , 模型根据数据中的模态标签值 i 选择合适的编码器进行特征提取, 得到不同模态的特征值 h^m 。我们期望不同的编码器可以提取出正交互补的特征信息, 因此我们设计了正交损失(orthogonality loss), 计算公式如下:

$$h^m = \phi_m(\tilde{D}_m^K) \quad (2)$$

$$L_o = \frac{1}{N_M} \left\| \prod_{i \in M} h^i \right\|_F^2 \quad (3)$$

其中 $\|\cdot\|_F$ 表示 Frobenius norm。通过优化 L_o 损失, 各个模态间特征数据将互补信息, 有效减少了信息冗余问题。

(4) 数据补全模块

我们参考了 LNLN 模型，设计了数据补全模块(Complementor)，用于补全丢失的部分数据，一定程度上可以弥补丢失的特征信息。Complementor 由两个 Transformer Layer 构成，通过注意力机制，期望通过同模态其他部分的数据补齐缺失的信息，计算公式如下：

$$\hat{h}^m = \text{Com}(h^m) \quad (4)$$

其中 \hat{h}^m 为补全特征后的特征数据，Com表示数据补全模块

我们希望补全后的数据所携带的信息，能尽可能与原始数据保持一致，因此我们将使用 L2 范数，计算 \hat{h}^m 与原始特征之间的差异

$$L_{\text{com}} = \frac{1}{N_M} \sum_{i \in M} \sum_{k=1}^{d_k} \|\hat{h}^i - \phi_i(D_i^k)\|_2^2 \quad (5)$$

其中 $\phi_i(D_i^k)$ 表示模态 i 提取的原始特征。该损失函数有助于最小化原始特征和补全特征之间的差异，从而提高其他组件的准确性。

(5) 模态数据纠正

我们认为，仅仅通过数据补全模块，并不能充分的完善缺失信息。同时，补全时产生的特征数据有概率是错误信息，会给模型造成一定的负影响。因此我们在 Complementor 后设计了模态数据纠正模块(Modality Data Correction)，该模块可以综合其他模态的特征，对当前模态的信息做进一步完善补充。并且我们考虑到当前模态原本携带的信息更加也很关键，我们希望其他模态的信息是用来补充缺陷，而不是取代模态特征。具体来说，首先将使用补充权重预测模块(Complementary Weight Prediction)，该模块将根据经 Complementor 完善前的特征，预测模态数据的缺陷率，并以此为权重指导后续补充信息与原始信息的加权融合。其计算公式如下：

$$\hat{\omega}^m = \text{Pred}(h^m; h_{\text{cwp}}) \quad (6)$$

其中 h_{cwp} 是随机初始化的可学习参数，用于辅助预测模块的学习。该模块使用 L2 范数进行优化：

$$L_{\omega} = \frac{1}{N_M} \sum_{i \in M} \|\omega^i - \hat{\omega}^i\|_2^2 \quad (7)$$

接着模态特征辅助生成器Sup(由两层 Transformers Encoder 组成)将结合其他模态的信息，以及随机初始化的可学习参数 \check{h}^s ，生成辅助特征 H^s ，用于补全和纠正主导模块。为了让特征值具有更高的可学习性和自由度，我们在此基础特征上加上了可学习参数 \check{h}^m 。假设当前考虑模态 m，则其计算方式如下：

$$H^s = \text{Sup}\left(\text{Concat}\left(\left[h^i\right]_{i \in M, i \neq m}^M, \check{h}^s\right)\right) \quad (8)$$

$$H^m = \omega H^s + (1 - \omega)(\hat{h}^m + \check{h}^m) \quad (9)$$

为了确保辅助特征提供了与当前模态特征不同的视角，我们使用了对抗学习。我们设计了二分类鉴别器Dis，用于判断数据是否来自当前模块。

$$\hat{y}_i = \text{Dis}(H^s/h^m) \quad (10)$$

通过鉴别器来训练生成器，使得生成器生成的特征值不是当前模态特征的简单复制，而是包含了其余模态的综合特征。这种对抗性的优化损失函数如下：

$$\min_{\theta_{\text{Dis}}} \max_{\theta_{\text{Sup}}} L_{\text{adv}} = -\frac{1}{N_M} \sum_{i \in M}^{N_M} y_i \cdot \log \hat{y}_i \quad (11)$$

其中 θ_{Dis} 和 θ_{Sup} 表示Dis和Sup的参数

经过上述操作，模型将很大程度上解决数据缺陷问题，有利于后续数据融合，提高模型的性能。

1. 基于不完全模态的动态权重学习机制

传统的多模态学习方法通常假设所有模态数据完备，在数据缺失的情况下难以保证模型的有效性。本项目提出一种基于不完全模态的动态权重学习机制，能够根据数据的完整性和质量实时调整各模态的权重，使模型在不同模态缺失的情况下依旧保持良好的准确性。通过这种机制，模型可以在数据缺陷的条件下进行有效的特征融合，增强了多模态学习的鲁棒性和适应性。

2. 利用长短期记忆网络（LSTM）进行时间序列信息提取

在多模态数据中，时间序列信息的提取是提升模型性能的重要因素。本项目设计了一个基于 LSTM 的模块，专注于从时间序列中提取特征，使得模型不仅可以处理静态多模态数据，还能充分捕捉动态时间信息。相比于常规的静态特征提取方法，LSTM 模块能够更好地适应数据中时间相关的模式和趋势，在多模态学习中进一步提升了动态环境的适应性。

3. 自适应多模态数据融合方法

现有的多模态数据融合方法通常采用固定的特征组合方式，难以在模态数据缺失时有效应对。本项目提出了一种自适应的多模态数据融合方法，使模型能够在数据缺失条件下，重新构建有效的统一表示。通过动态调整不同模态特征的融合策略，本方法在融合缺失模态时能够有效保留各模态的关键特征，为多模态数据不完整条件下的统一表示提供了更高的灵活性和准确性。

项目
创新
点

四、项目进度安排

(查阅资料、自主设计项目研究方案、实验研究、数据统计、处理与分析、研制开发、中期检查、填写结题表、撰写研究论文和总结报告、参加结题答辩和成果推广等)

第一阶段：查阅资料与文献研究 (2024. 9)

阶段目标：全面了解多模态学习、联邦学习和数据缺陷处理的最新研究进展。

- 1) 查阅与该领域相关的论文、专利和技术报告。
- 2) 汇总国内外相关研究的现状与挑战，了解现有技术的不足之处。
- 3) 为导师的研究提供文献支持，提出初步的研究思路。

第二阶段：参与项目研究方案的设计 (2024. 10)

阶段目标：协助导师制定详细的项目研究方案。

- 1) 参与讨论项目的具体研究内容与创新点。
- 2) 协助设计实验方案，包括数据采集方法、模型结构与训练流程。
- 3) 明确项目实施计划，记录各阶段的时间节点和任务分配。

第三阶段：实验研究与数据采集 (2024. 11 - 2025. 2)

阶段目标：参与实际实验，验证研究方案的可行性。

- 1) 协助数据采集，获取多模态数据，并处理不完整模态的数据问题。
- 2) 参与模型训练，协助导师进行模型调优。
- 3) 对模型进行初步测试，评估其在缺失数据情况下的性能。

第四阶段：数据统计、处理与分析 (2025. 3 - 2025. 4)

阶段目标：对实验数据进行统计分析，提取有效结论。

- 1) 协助收集模型输出结果，并进行数据统计分析。
- 2) 比较不同实验设置下模型的表现，记录和分析结果。
- 3) 整理分析结果，为后续的研制开发提供数据支持。

第五阶段：模型优化与结果汇报 (2025. 5 - 2025. 6)

阶段目标：参与模型的优化与改进。

- 1) 根据实验结果调整模型参数和结构，提升模型性能。
- 2) 协助开发项目成果的应用原型，并进行现场测试。
- 3) 准备中期汇报，向导师和团队展示研究进展。

第六阶段：中期检查 (2025. 7)

阶段目标：对项目进展进行全面评估，确保各项任务按计划进行。

- 1) 汇报项目进展与成果，接受导师或专家的反馈意见。
- 2) 根据反馈调整项目计划，确保后续工作的有效性。

第七阶段：填写结题表与总结报告 (2025. 8 - 2025. 9)

阶段目标：总结项目成果，准备结题材料。

- 1) 整理项目文档、实验数据和相关材料，准备结题表格。
- 2) 撰写总结报告，回顾项目实施过程与收获。

项目
进度
安排

五、项目实施条件

项目实施的条件和资源	<p>(依托的科研或教学实验基地, 实验仪器设备的配置, 图书资料, 实验或实践场地等项目实施的条件)</p>
	<p>本项目依托于导师在多模态学习和联邦学习等领域的丰富研究经验, 为我们的研究提供了必要的指导与支持。这不仅使我们能够有效地规划研究方向, 还能在遇到挑战时获得专业建议。</p>
	<p>团队成员均为信息学院的学生, 系统学习了数据结构、程序设计和人工智能等核心课程, 深入理解多模态学习和深度学习的基本原理。这种坚实的知识基础, 使我们能够更好地理解和应用当前的多模态学习和联邦学习模型, 为项目的顺利推进奠定了基础。</p>
	<p>在实验仪器设备方面, 项目将依托于学校的计算机科学与工程实验室。实验室配备高性能计算机, 能够满足大规模模型训练和数据处理的需求。这为我们的实验提供了强大的计算能力, 确保我们能高效地进行数据分析与模型训练。同时, 我们将使用深度学习框架(如 TensorFlow 和 PyTorch) 以及相关的科学计算库(如 NumPy 和 Pandas) 进行模型开发和数据分析, 提升研究的专业性和准确性。</p> <p>在图书资料方面, 我们可以充分利用学校的图书馆, 获取相关领域的书籍、期刊和电子资源, 以支持文献调研和参考。同时, 团队还将利用学校的科研资源, 访问各种在线数据库(如 IEEE Xplore、SpringerLink 和 ACM Digital Library), 获取最新的研究成果和技术资料。这将为我们的研究提供必要的理论支持和数据参考, 确保项目的科学性和前沿性。</p>

六、项目预期效果及成果形式

项目预期成果及其形式	<p>(学术论文、专利申请、竞赛获奖、开发软件、研制产品、项目鉴定, 推广应用等, 其中前三项至少任选其一)</p>
	<ol style="list-style-type: none">1. 在中文学术期刊上发表 1 篇学术论文2. 参加上海市计算机应用能力大赛、中国国际大学生创新创业大赛等

结合专业学习的计划	<p>(是否有进一步深造的想法? 如何通过参与项目研究加强专业学习, 培养自身的实践能力、科研能力。)</p>	
	<p>1. 理论知识的深化</p> <p>参与该项目将使我和团队成员深入理解多模态学习和联邦学习的核心理论与技术。这些领域在人工智能、机器学习和数据科学等专业中占据重要地位。通过文献调研和实验研究, 我们不仅能掌握前沿的研究动态, 还能将理论知识与实践相结合, 增强对所学内容的理解和应用能力。</p>	
	<p>2. 科研能力的培养</p> <p>实际项目的参与将帮助我们培养科研能力。通过实验设计、数据采集和模型训练, 我们将接触到真实的科研流程, 锻炼数据分析与解决问题的能力。同时, 与导师和团队成员的紧密合作将提升我们的沟通和协作能力, 帮助我们更好地适应未来的研究或工作环境。</p>	
	<p>3. 实践能力的提升</p> <p>在实践能力方面, 项目中的多模态数据处理和模型优化将提升我们的编程技能与数据分析能力。我们将使用深度学习框架和数据处理工具进行实际操作, 增强对相关技术的熟悉度。此外, 项目实施过程中的问题解决和创新思维将促使我们不断反思和调整自己的学习方法, 提高自主学习和实践的能力。</p>	
	<p>4. 学术与职业发展</p> <p>参与该项目将为我和团队成员提供宝贵的科研经验, 这不仅有助于我们在专业领域内的深造, 还能增强我们在求职市场上的竞争力。我们计划在完成项目后撰写研究论文并参加学术会议, 进一步提升我们的学术素养和影响力。</p>	

七、项目经费预算

项目经费使用计划	(包括大概支出科目(含配套经费)、金额、计算根据及理由, 如果别的经费支持请说明)		
	支出科目	金额(元)	计算根据及理由
	版面费	5000 元	用于支付项目研究成果发表时的期刊版面费
	差旅费	2000 元	用于团队成员参加学术会议、调研或实验数据采集所需的交通和住宿费用
	材料费	3000 元	包括实验所需的硬件设备、传感器、存储设备和小型配件

八、指导老师意见

指导教师意见及指导计划	<p>(指导教师指导项目实施的机会和安排, 并从项目学科性、前沿性、可行性、研究性、可操作性和成效性加以评价)</p> <p style="text-align: right;">指导教师签字:</p> <p style="text-align: right;">年 月 日</p>
-------------	---

九、审查意见

答辩专家组意见	<p style="text-align: right;">答辩专家组签字：</p> <p style="text-align: right;">年 月 日</p>
学院意见	<p style="text-align: right;">学院负责人签字：</p> <p style="text-align: right;">年 月 日 (学院盖章)</p>
学校意见	<p style="text-align: right;">负责人签字：</p> <p style="text-align: right;">年 月 日 (盖章)</p>

附：华东理工大学“大学生创新创业训练计划”项目承诺书

为确保“大学生创新创业训练计划”项目的顺利实施，本项目承诺遵守以下条款要求：

- 1 在项目实施的过程中遵守学校有关规定，恪守学术规范，真正做到诚实守信，实事求是；
- 2 项目经费严格按学校“大学生创新创业训练计划”项目管理的有关规定执行，按照专款专用原则，保证项目经费的有效合理使用；
- 3 接受学校及学院对本项目的检查与监督；
4. 保证研究项目的有序进行，并按华东理工大学“大学生创新创业训练计划”项目管理办法的要求，接受学校的中期检查。
- 5 按期完成研究项目并由负责人提交结题报告及有关研究成果、项目经费决算（含使用明细）、等材料，并且归还仪器设备、工具、资料等；
- 6 因主观原因致使项目无法执行的，将根据实际情况退回全部或部分资助经费；
- 7 研究成果如果需要按技术秘密进行保密的，项目负责人及指导教师应制定具体方案并报学校批准；
- 8 凡研究工作有发明创造，经学校科研院审查符合条件的，可由科研院协助申报专利，以保护知识产权，维护学校的合法权益；
- 9 本承诺书由本项目全体成员及指导老师签字并经创新创业教育中心确认后生效，如有违反，愿承担相应责任。

项目成员（签字）：

指导教师（签字）：

年 月 日

说明： 本承诺书是华东理工大学“大学生创新创业训练计划”项目管理和经费划拨的依据。