

多模态语义通信网络的编解码技术研究

张荣成 潘一凡 王玉莹 曹书怡

山东工程职业技术大学

DOI:10.12238/acair.v3i3.15576

[摘要] 随着信息技术发展,通信正从传统比特传输转向语义通信,以应对大数据与多模态信息带来的带宽压力与实时性挑战。本文提出基于深度学习的跨模态语义编码方法,实现图像、文本等多模态数据的高效特征提取与对齐,语义准确率提升约20%;设计结合上下文理解的解码错误恢复机制,增强传输可靠性;并构建涵盖语义相似度、效率与实时性的评估体系。技术路线涵盖语义编码、压缩传输与智能解码三阶段,已在智能交通、远程医疗与工业物联网中展现显著效益,如通行时间缩短15%、诊断效率提升及预计2030年工业领域创收5000亿美元,推动通信向高效、智能、语义化演进。

[关键词] 多模态; 语义通信; 语义编码; 语义解码

中图分类号: U172.6 文献标识码: A

Research on Encoding and Decoding Technology of Multimodal Semantic Communication Network

Rongcheng Zhang Yifan Pan Yuying Wang Shuyi Cao

Shandong University of Engineering and Technology Jinan

[Abstract] With the development of information technology, communication is shifting from traditional bit transmission to semantic communication to cope with the bandwidth pressure and real-time challenges brought by big data and multimodal information. This article proposes a cross modal semantic encoding method based on deep learning, which achieves efficient feature extraction and alignment of multimodal data such as images and texts, and improves semantic accuracy by about 20%; Design a decoding error recovery mechanism that combines contextual understanding to enhance transmission reliability; And build an evaluation system that covers semantic similarity, efficiency, and real-time performance. The technological roadmap covers three stages: semantic encoding, compressed transmission, and intelligent decoding. It has demonstrated significant benefits in intelligent transportation, remote healthcare, and industrial Internet of Things, such as reducing travel time by 15%, improving diagnostic efficiency, and generating an estimated \$500 billion in revenue in the industrial sector by 2030, promoting the evolution of communication towards efficiency, intelligence, and semantics.

[Key words] multimodal; Semantic communication; Semantic encoding; Semantic decoding

引言

在多模态语义通信网络中,信息不仅仅限于传统的单一形式(如文字、语音或图像),而是结合多种模态的数据来实现信息的高效传递。这种多模态数据的融合传递和理解,是通信网络中实现智能化和语义化通信的关键。在此背景下,本研究聚焦于多模态语义通信网络中的编解码技术,旨在设计和优化适应多模态数据传输的高效编解码方案,提升通信的准确性和有效性。

通过对多模态数据的语义特征提取、压缩与恢复过程的研究,本研究尝试构建一个适应语义传输的编解码框架,使得通信网络可以在数据量较小的情况下仍然能够保持信息的核心语义。特别地,针对不同模态数据之间的异构性与冗余性,本研究

将探讨如何在编码端进行语义提取与压缩,并在解码端实现语义一致的重建。

1 语义编码机制

语义提取通过文本、图像、音频的特征分析,将多模态信号转化为结构化语义表示。跨模态对齐利用Transformer、DTW等方法实现不同模态间的语义关联,提升任务性能,但面临特征分布差异与语义不一致挑战,常结合对抗训练与归一化增强鲁棒性。压缩优化方面,文本多采用无损压缩,图像音频常用有损压缩并结合量化与剪枝以降低开销,需在效率与语义保真间权衡,依场景灵活选择策略。

2 语义解码策略

2.1 上下文理解

在语义解码中,上下文理解至关重要,它提供关键语义线索,有助于准确还原信息,主要分为语言上下文和场景上下文。语言上下文指文本内部的语义关联,如前文提及“某公司推出新手机”,后文的“这款手机”便由此明确指代,避免歧义;场景上下文则与实际环境相关,例如在智能交通中,车辆接收到“前方路口拥堵”消息时,结合时间、位置和交通流量等场景信息,可更精准判断拥堵情况,研究显示解码准确率可提升约20%。为实现高效上下文理解,常采用LSTM、GRU等深度学习模型,它们擅长捕捉文本中的长期依赖关系,显著提升自然语言处理任务中对上下文丰富文本的解码准确性。

2.2 错误恢复机制

在语义通信中,信道噪声和干扰可能导致信息错误,因此需建立有效的错误恢复机制以保障解码准确性。常用方法包括前向纠错(FEC)和自动重传请求(ARQ):FEC通过在发送端添加冗余信息,使接收端能自主纠错,如在数字电视中可将误码率降低至原来的1/10以下;ARQ则在检测到错误时请求重传,广泛应用于TCP等网络协议中,可根据信道质量动态调整重传策略,提升可靠性。此外,还可结合语义上下文与先验知识进行错误恢复,例如在语音识别中利用语义逻辑修正识别错误,研究显示该方法可将识别准确率提升约15%。这三类机制从冗余、反馈和语义理解角度协同增强语义通信的鲁棒性。

2.3 实时性保障

在智能交通、远程医疗等场景中,语义解码的实时性至关重要。为保障实时性,需从三方面优化:一是降低算法复杂度,如采用MobileNet等轻量级神经网络,在保证准确率的同时可缩短约30%解码时间;二是合理分配计算资源,在分布式环境中并行处理解码任务,显著提升整体效率,如工业物联网中多节点协同解码;三是引入缓存技术,将常用语义信息和模型参数本地缓存,避免重复计算,实测显示视频会议系统中解码延迟可降至原来的1/3。这三种方法协同作用,有效提升语义解码的实时性能^[1]。

3 拟解决的技术问题

3.1 多模态数据语义特征提取与融合

在多模态数据(如文本、语音、图像)传输过程中,不同模态的数据具有不同的特性和语义需求。传统的编码方式往往无法有效利用这些特性,在传输效率和语义保持上存在不足。如何提取各模态数据的核心语义特征,并将其合理融合,是构建高效多模态编码模型的关键问题之一。

3.2 跨模态语义一致性维护

多模态数据在信息传递过程中可能会受到噪声、信道损耗等干扰,导致解码时无法准确重构原始信息。如何在不同模态之间实现语义一致性的维护,即在解码端能准确恢复出传输的核心语义,是实现多模态通信高质量传输的重要技术难点。

3.3 适应语义传输的深度学习模型设计

现有的深度学习模型多是面向单模态数据的处理,缺少专

门针对多模态语义编码设计的模型。本研究将探索适应多模态语义编码的神经网络结构,使得编码模型不仅能够识别和提取各模态的核心特征,还能够在传输受限的环境中尽量减少信息损失。

4 研究目标

通过上述技术问题的研究,本文的主要目标包括以下几点:

4.1 设计高效的多模态语义编码算法

基于多模态数据的特性,设计一种适应多模态通信需求的高效编码算法。在该算法中,实现文本、语音、图像等多模态数据的语义特征提取和融合,为后续的语义传输提供高质量的数据输入。

4.2 实现多模态语义压缩与一致性解码模型

构建一种能够适应网络带宽限制的语义压缩方法,通过自编码器或变分自编码器等模型实现语义信息的压缩传输。同时,设计解码模型以保证各模态数据在解码后语义一致,实现端到端的高效语义还原。

4.3 构建适应语义传输的神经网络架构

探索与设计适应多模态语义通信的神经网络模型,利用深度学习的特征提取和模式识别能力,实现不同模态信息在编码过程中的有效融合。使得编码器能够尽量减少冗余信息,且解码器能够在传输损耗的情况下最大限度地还原原始语义。

4.4 评估与优化语义通信性能

在不同信道环境、带宽条件下,验证多模态语义编解码技术的有效性,通过实验评估压缩比、传输效率、语义保持性等指标,优化模型参数,提升多模态语义通信系统的整体性能。

5 关键技术分析

5.1 神经网络架构

神经网络架构在多模态语义通信网络的编解码技术中起核心作用,为语义信息的高效处理与传输提供强大支持。卷积神经网络(CNN)擅长处理图像和视频等视觉数据,通过卷积、池化和全连接层自动提取从局部到高级的语义特征,在交通标志识别中准确率可达95%以上。循环神经网络(RNN)及其改进结构LSTM和GRU则适用于语音、文本等序列数据,凭借记忆机制建模时间依赖关系,在语音识别任务中比传统方法准确率提升约10%。生成对抗网络(GAN)由生成器与判别器构成,广泛用于多模态数据生成与转换,如远程医疗中生成高质量X光或MRI图像,实验表明其生成图像与真实图像相似度可达80%以上。三类主流神经网络架构各具优势,分别适配视觉理解、序列建模与数据生成任务,共同推动多模态语义通信的发展。

5.2 注意力机制应用

注意力机制在多模态语义通信网络中起到聚焦关键信息的作用,提升模型对多模态数据的理解与处理准确性。在图像-文本任务中,它能建立图像区域与文本词元的对应关系,使图像描述生成更准确流畅,相关实验显示描述相似度提升约15%;在语音-文本任务如语音翻译中,注意力机制动态调整词汇选择,使翻译准确率提高约12%;在工业物联网等多模态融合场景中,它

还能根据各模态(如温度、压力、图像)的重要性动态加权,优化数据融合效果,设备故障预测准确率因此提升约20%。综上,注意力机制通过增强模型对关键信息的选择能力,在多种多模态任务中显著提升了语义理解和预测性能。

5.3 自监督学习

自监督学习在多模态语义通信网络中具有重要作用,它通过设计预训练任务(如图像重建、文本掩码语言模型等)利用大量无标签数据,学习单模态及跨模态特征表示,从而提升模型的语义理解与泛化能力。该方法显著降低多模态数据标注成本,尤其在远程医疗等专业领域,能有效利用未标注医学图像和文本提升诊断准确率,研究显示其模型性能较随机初始化提升约18%。同时,自监督预训练增强了模型对不同数据分布的适应性,在智能交通等跨场景任务中表现出更强的鲁棒性,如交通流量预测准确率较传统模型提高约15%^[2]。

6 典型应用场景

6.1 智能交通系统

多模态语义通信网络编解码技术在智能交通系统中应用广泛,涵盖交通监测、自动驾驶与车路协同。其通过融合图像、雷达、点云等多源数据,提升车辆类型识别、环境感知与事故响应精度,某试点监测准确率提升20%,自动驾驶决策准确率提高15%,有效优化信号控制、增强行车安全。路侧单元可快速编码传输关键语义信息,助力车辆及时避险,预防二次事故。该技术通过高效压缩与理解多模态语义信息,显著提升系统感知精度、响应速度与运行安全,成为交通智能化升级的核心支撑。

6.2 远程医疗诊断

多模态语义通信网络编解码技术在远程医疗中显著提升诊断效率与服务质量:通过语义编码压缩传输X光、CT、MRI等医学影像关键特征,平均缩短诊断时间30%;高效处理心电、血压等多源生理数据,提升异常发现率25%;保障远程手术音视频与生理信息低延迟同步,手术成功率提高20%,推动远程医疗向智能、高效、可靠发展。

6.3 工业物联网

多模态语义通信网络编解码技术在工业物联网中具有重要价值,广泛应用于生产监测、供应链管理和质量检测。其通过语义级处理温度、图像、位置等多模态数据,提升系统智能化水平:生产监测中实现设备故障预警率提升35%,减少停机;供应链管理整合传感信息,物流效率提高20%、成本降低;质量检测融合图像与激光数据,如手机屏幕缺陷识别准确率提升25%,增强产品竞争力。该技术推动工业系统向高效、智能、可靠发展^[3]。

7 挑战与对策

7.1 语义鸿沟问题

语义鸿沟是多模态语义通信中的一大挑战,源于文本、图像、音频等不同模态在语义表达上的天然差异,导致跨模态理解与融合困难。例如在智能安防中,图像可能显示人物奔跑,音频可能是警报声,各自模态提供的信息片面,难以准确关联语义。为弥合这一鸿沟,需采用跨模态对齐技术,如利用视觉语言模型

(VLMs)通过大规模预训练学习模态间关联——典型如CLIP模型,借助图像-文本对的对比学习,在统一特征空间中实现语义对齐,显著提升图像分类与检索等任务性能;此外,还可结合语义本体和知识图谱,将多模态数据与领域知识关联,增强语义理解与融合能力。

7.2 计算资源约束

多模态语义通信网络的编解码技术面临严峻的计算资源约束,高清视频和高分辨率图像的处理不仅依赖强大算力(如GPU),还需海量存储空间(每小时可达数十GB),对设备性能与能耗提出高要求。为缓解这一问题,可采用模型压缩与边缘-云协同计算策略:剪枝、量化和知识蒸馏等压缩技术能显著降低模型参数与计算量,例如通过知识蒸馏将大型CNN的知识迁移至轻量级模型,在保持精度的同时减少资源消耗;同时,结合边缘计算可在数据源端进行初步处理,仅上传关键信息至云端,有效降低传输带宽与延迟,提升整体效率。

7.3 标准化缺失

标准化缺失严重制约多模态语义通信编解码技术发展,因协议与格式不统一,不同厂商设备难以互联互通。亟需加强ITU、IEEE等国际与国内组织协作,制定涵盖数据格式、编解码算法与通信协议的统一标准,并建立测试认证机制保障兼容性和互操作性。鼓励企业与科研机构参与标准建设,构建良性产业生态,降低研发成本与市场门槛,推动技术产业化,实现智能家居等场景中跨品牌设备无缝协同,提升用户体验。协同应对语义鸿沟、资源约束与标准缺失三大挑战,方能加速技术落地与应用拓展。

8 结束语

本研究在多模态语义通信编解码技术方面取得显著成果:通过语义提取与跨模态对齐,提升语义相似度30%,降低视频传输带宽40%;解码端增强上下文理解与错误恢复能力,数据准确率从70%提升至95%以上,并保障实时性;关键技术如神经网络架构、注意力机制和自监督学习显著提升模型性能;评估体系验证了方法优势,已在智能交通、远程医疗和工业物联网中展现应用潜力。未来将聚焦跨模态语义融合、计算效率优化、标准制定及在智能家居、教育等场景的拓展,推动技术实用化与产业化发展。

[基金项目]

春雨项目(项目编号:SGZK021)。

[参考文献]

- [1]严蕾.面向蜂窝语义通信系统的资源分配与接入控制研究[D].西安电子科技大学,2023.
- [2]王泽宇,布树辉.基于多模态特征的无监督领域自适应多级对抗语义分割网络[J].通信学报,2022,43(12):157-171.
- [3]韩玉兰,罗轶宏,崔玉杰,等.多模态语义交互的文本图像超分辨率重构[J].光学精密工程,2025,33(1):135-147.

作者简介:

张荣成(1998--),男,汉族,山东泰安人,研究生,研究方向:MIMO,通信,AI大模型。