

Application Research and Future Development of Multimodal Artificial Intelligence

Weiya Yuan

Gansu Suyixiu Electronic Technology Co., Ltd., Lanzhou, Gansu, 730030, China

Abstract

Multimodal artificial intelligence, as a strategic theoretical technology leading the new round of scientific and technological revolution and industrial transformation, has attracted much attention at home and abroad. AIGC multimodal artificial intelligence is an important development direction. It has cross-domain understanding capabilities and has already played a role in multiple industries, adding impetus to the intelligent industry and gradually penetrating society. Technological progress will make it smarter and the content created more realistic. This article sorts out the relevant content and applications, and looks forward to its future innovative development in major fields such as intelligent water conservancy. However, its application research in these fields is a multi-disciplinary cross-system engineering project. It requires the state to introduce policies, organize multi-disciplinary experts and strong teams, and steadily advance scientific and technological research.

Keywords

Multimodal AIGC; Research and application; Future development; Action Plan

多模态人工智能的应用研究与未来发展

袁伟亚

甘肃速易修电子科技有限公司, 中国·甘肃 兰州 730030

摘要

多模态人工智能作为引领新一轮科技革命与产业变革的战略性理论技术, 备受国内外关注。AIGC多模态人工智能是重要发展方向, 有跨领域理解能力, 已在多行业发挥作用, 为智能产业添动力, 正逐步渗透社会。技术进步将使其更智能, 创作内容更逼真。本文梳理了相关内容与应用, 展望其未来应向智能水利等重大领域创新发展。不过, 其在这些领域的应用研究是多学科交叉系统工程, 需国家出台政策, 组织多学科专家和强大团队, 扎实推进科技攻关。

关键词

多模态AIGC; 研究应用; 未来发展; 行动纲要

1 引言

多模态 (Multimodality) 是涉及多个领域和层面的数据或者信息的传递、交流和表现形式, 是指在信息传递和交流中, 使用视觉、听觉、触觉、嗅觉等多种不同的感官模式, 或语言、图像、声音、动作、文字等多种符号系统的组合, 达到对多个领域和层面的信息进行表达和理解的一种理论。多模态理论强调不同模态之间的互动和整合, 通过多种模态的协同作用, 整合和处理多种数据类型, 以提高智能系统对复杂环境的响应能力, 高效准确地理解传达和复杂信息, 在提高效率、节约成本和优化资源、优化决策等方面具有潜力和优势。在某些领域的应用取得了显著的效果, 展示了信息

智能化的巨大潜力。从而提高决策效果, 提高工作和生产效率, 减少人为错误, 应用潜力巨大。

现实世界的信息呈现出高度复杂性和多样性。传统的人工智能系统专注于单一的数据模态, 难以完整描述信息的复杂性和多样性。在处理复杂的多元信息时存在局限性, 无法有效整合来自不同类型数据的信息。无法满足人们对于复杂性和高质量内容的需求。

人工智能生成 (Artificial Intelligence), 是研究如何使计算机具备像人类一样智能行为的学科。以科学交叉, 涵盖了多个技术领域, 如机器学习、深度学习、自然语言处理、计算机视觉等, 旨在让计算机系统具备感知、理解、学习、推理等能力, 通过算法和模型, 自动创造出新的内容、产品或解决方案的过程。目标是实现智能化的任务处理和决策, 以提高效率、降低成本、改善人类生产生活质量等。

【作者简介】袁伟亚 (1983-), 男, 中国甘肃兰州人, 本科, 计算机科学与技术中级职称, 从事计算机科学与技术研究。

2 AIGC 技术的优势

2.1 AIGC 的基本概念

AIGC, 全称为 Artificial Intelligence Generated Content, 即利用人工智能技术来生成图像、视频、音频、模型等多种类型的内容的一种新型创作方式。它代表了人工智能领域创作的全新变革。

AIGC 的兴起源于深度学习技术的快速突破。AIGC 多模态人工智能代表了人工智能发展的一个重要方向, 拥有跨领域的理解能力, 能够基于自然对话方式理解复杂系统, 通过整合和处理多种数据类型, 不仅推动了 AI 在各个领域的应用, 还提高了智能系统对复杂环境的响应能力, 能将不同领域不同类型的复杂数据和信息进行融合, 快速、高效地生成多领域、高质量的内容, 以实现更加准确、高效的人工智能应用, 以满足经济、社会、生态、生活的多样化需求。AIGC 的应用场景主要集中在内容创作和生产领域。

2.2 AIGC 的技术基础及关键技术

AIGC 的核心在于其强大的理解和创作能力。它通过分析大量的数据, 学习人类的创作风格和表达方式, 从而能够根据用户的需求生成相应内容。

这种技术依赖于深度学习、神经网络和生成式对抗网络 (GAN) 等先进技术。

深度学习: 通过多层神经网络的学习, 提高 AI 的智能水平。通过大量数据的学习, 能够处理复杂的任务, 能够在特定任务上表现出色, 是 AIGC 技术的关键组成部分。

神经网络: 模拟人脑神经元连接, 用于处理和生成数据。

信息整合: 通过深度思考, 突破常规思维边界, 激发思维新潜能, 跨领域地将多种数据源整合, 使系统能够更准确地理解复杂场景和问题, 在处理多元复杂信息时具有准确性和全面性, 为解决复杂问题提供了强有力的了解和认识。

大规模跨境追踪技术: 该技术让利用传感器数据、视频流和语音指令和多个摄像头能同步追踪特定目标, 为管理和决策提供有效支撑。

多模态大模型技术: 能够处理包含文本、图像、音频等多种信息的复杂数据, 能够同时处理和分析多种类型的数据。

丰富的信息表达: 多模态技术可以将不同模态的数据进行联合分析, 从而获得更加丰富、全面的信息表达。

数据互补和一致性: 不同模态的数据往往具有互补性, 通过将它们融合在一起, 可以弥补单一模态数据的不足。同时, 多模态技术还可以保持不同模态数据之间的一致性, 使得分析结果更加准确和可靠。

更智能的决策和交互: 多模态技术可以帮助计算机系统更好地理解人类的语言、图像和声音等信息, 从而实现更智能的决策和交互。

2.3 AIGC 处理多模态问题的基本架构与工作原理

多模态表示学习, 旨在通过定义函数学习不同模态信

息的语义相关性, 并将其映射到同一空间。

多模态对齐, 是实现不同模态之间的有效传递和融合, 包括时间维度对齐和空间维度对齐。

多模态融合, 多模态融合是将多个模态的信息进行整合, 以生成更具意义和语义丰富的输出。

2.4 AIGC 在多模态应用方面的进展

图像生成, 能够生成具有丰富细节和高分辨率的图像, 为图像生成领域带来了新的突破。

视频生成, 该技术通过对视频帧之间的时间序列关系进行建模, 结合多模态信息, 利用循环神经网络 (RNN) 或长短期记忆网络 (LSTM), 对视频帧的时间序列进行建模, 捕捉帧与帧之间的动态变化, 将音频信息与视频帧进行同步对齐, 生成连贯、自然的内容, 生成的视频在视觉和听觉上更加协调。

三维形状生成, 在三维形状生成方面, 扩散模型和神经网络的结合为生成精确、复杂的三维模型提供了新的途径。

2.5 应用价值

AIGC 不仅能够提高内容生产的效率和质量, 还能推动内容创新和新业态的形成。它有助于降低生产成本, 满足个性化需求, 并为数字经济的持续发展注入新的活力, 为各行各业带来革命性的变化。例如, AIGC 可以用于自动生成新闻、文章、音乐、图像、视频、音频等。AIGC 的创建方式多种多样, 例如深度学习、自然语言处理 (NLP) 或生成对抗网络 (GANs)。

2.6 市场前景

多模态 AIGC 正处于快速发展的阶段, 有广泛的应用价值和潜力, AIGC 市场将迎来快速增长。据有关报告预计, 到 2030 年, AIGC 市场规模将超过万亿人民币。

3 AIGC 技术已介入以下领域

AIGC 的应用领域有很多, AIGC 的应用范围极为广泛, 特别是以下行业可能会迅速从 AIGC 中受益:

金融服务、城市管理、智慧商业、轨交运维、交通运输、智慧园区、物流行业、电商平台、医疗行业、法律服务、语音交互、软件开发、电子商务、营销和广告、新闻产业、艺术和设计、教育行业、体育健康、文化娱乐、游戏行业等。例如:

在金融服务领域, 使用 AIGC 自动生成关于市场趋势的分析报告, 准确预测潜在风险, 提出投资建议。

在工业制造领域, 用于产品设计、生产仿真、故障预测、预维护等, 将车间、流水线、设备等映射在虚拟工厂上, 实时监控生产状态, 可以及时发现问题, 提高生产效率和管控水平, 引领制造业转型升级, 提升运营管理效率, 获得独特的生产优势和投资回报率 (ROI)。

在城市管理领域, 用于城市规划、交通管理、应急管理

理,预警城市可能产生的不利影响,并提供合理有效的解决措施。

在建筑工程领域,用于建筑设计,创造出创新的设计方案,以及施工模拟、运维管理等。

在物流行业,通过对真实世界的物流设备进行虚拟,能够依赖大数据和机器学习,对海量物流数据进行分析、优化仓库的存储结构、优化设备的运行效率,并协助决策与维护。

在营销和广告方面,AIGC能够根据消费者的行为和偏好定制广告文案和视觉设计,提高用户的参与和互动和个性化营销的效果。

在教育行业:教师利用AIGC准备课程资料,使教学内容更加丰富和定制化,生成符合学生学习进度和风格的个性化教材和练习题。

在音乐制作方面,能协助音乐家创作旋律或和声,甚至完成整首歌曲的制作。

随着新一轮科技革命和产业变革的高速发展,多模态理论正在向广阔的各个领域发展,展现出巨大的潜力和创造力。

4 多模态 AIGC 的未来发展

4.1 多模态 AIGC 技术研究的未来方向

4.1.1 数据层面的研究方向

多模态数据的获取、标注和管理是AIGC工作难度大,成本高,严重制约了AIGC的发展。不同模态数据的格式、特点和采集方式各异,如何将这些数据进行有效的整合和预处理,是保证模型性能的关键。

4.1.2 模型层面的研究方向

设计能够有效处理多模态信息的模型架构是一个复杂的问题。如何提高模型的泛化能力和可解释性,是亟待解决的重要问题。在泛化能力方面,由于多模态数据的多样性和复杂性,模型在不同数据集和任务上的表现差异较大。可解释性问题则使得用户难以理解模型的决策过程和结果,限制了AIGC在一些对安全性和可靠性要求较高的领域的应用。

4.1.3 多模态数据的高效处理与融合

利用强化学习等技术,研究更加高效的数据处理和融合策略,提高模型对不同模态数据的适应性。实现不同模态信息的深度融合;提高多模态数据的利用效率。例如将先进的传感测量、信息通信、自动控制、人工智能、云计算、大数据、三维可视化等信息技术和生产过程、管理技术相结合,构建数字化、信息化、智能化的管理平台,全面提升生产、管理、运营的信息化、数字化、智能化水平。

4.1.4 可解释性和鲁棒性的模型设计

开发具有可解释性和鲁棒性的AIGC模型,提高模型的可信度和可靠性。通过引入可视化技术和因果推理方法,解释模型的决策过程和结果;研究对抗训练和强化学习等技

术,提高模型的鲁棒性和抗干扰能力。

4.1.5 AIGC 与数字孪生技术的交叉融合

在当前的技术发展背景下,AIGC和孪生技术都展现出了各自独特的优势。AIGC在内容生成、降低成本、创造性内容生成等方面具有显著优势,已衍生出丰富的能力矩阵;而孪生技术在实时数据同步、预测性维护、优化设计与制造等方面表现出色,可以模拟真实场景、可视化操作、实时监测和优化、节省成本和时间、提高精确性和可靠性,将复杂物理系统的结构、状态、行为、功能和性能映射到数字化的虚拟世界。

4.2 多模态 AIGC 应用研究的新领域

发达国家把多模态AIGC技术和数字孪生技术定义为未来发展的关键战略。关键战略技术必须为实现国家安全的重大战略决策服务。就当前而言,以AIGC与孪生技术的融合,开创智能水利、智能农业、智慧医疗的研究刻不容缓。

4.2.1 智能水利

水是生态之基,是人类生存和经济社会可持续发展的重大命脉,是影响国家安全的战略性资源。我国是水资源严重短缺的国家,又是水旱灾害频繁的国家。水资源时空分布不均衡,开发利用效率偏低,特别是大规模开发地下水以及江河湖泊水生态环境的退化,已影响国家的安全和生产生活的民生保障。强化水资源收支评估能力、水资源优化配置和集约利用能力、江河湖泊生态保护治理能力、水旱灾害防御能力等,是多模态人工智能研究应用的广阔领域。

水是以流域为生态位形成的资源。智能水利的建设,关键是流域水资源治理。以AIGC的跨领域交叉整合能力,综合运用遥感RS、地理信息系统GIS、全球定位系统GPS、网络技术、多媒体及虚拟现实等现代高新技术,对全流域的地理环境、自然资源、生态环境、人文景观、社会和经济状态等各种信息进行采集和数字化管理,构建全流域综合信息平台 and 三维影像模型,使各级部门能够有效管理整个流域的经济、社会、生态建设,做出宏观的资源利用与开发规划。

在水资源的精准评价方面,多模态AIGC可以整合空气质量监测设备、水质监测传感器等多种设备采集的数据,以及卫星图像、无人机影像等多模态信息,实现全面动态监测和智能管理,从而攻克水循环时空分布的变化规律,精确获取大气降水、土壤水、地表水、地下水资源现状,分析变化趋势,获取地表水储量、土壤水分、地表温度、植被覆盖等信息,进而反演得到蒸散发、径流等模拟数据,更加精准地评价水量、水质和收支核算,优化配置全流域水资源。

在防汛方面,通过对水文监测数据和卫星图像的融合分析,可以提供洪水淹没范围、水位等洪灾应急监测信息。

在水环境监测方面,利用高光谱遥感技术监测的水质状况,结合传感器数据和无人机拍摄的河流、湖泊影像,多模态AIGC可以准确判断水体的生态变化,水质类型的变化,

以及污染源的位置和扩散范围，及时发布预警信息。

4.2.2 智能农业

农业是“保障粮食安全，端牢中国饭碗”的重大战略性新兴产业，研究农作物的生长规律和过程，以有效的水、肥、土、种、农药和人工管理，大幅度提高作物产量，减少无效的资源浪费和农产品污染，是 AIGC 研究应用的重点领域。可以借鉴生物神经系统的信息处理机制，通过模拟大脑对不同感官信息的整合方式，开发更加智能的多模态数据融合，实现多模态信息的高效处理和融合。推动 AIGC 农业领域的应用

一是对土壤营养要素的检测和优化组合，哪类土壤种植什么作物，施入什么肥料和用量，能达到优质高产。农作物的优质高产，至少需要十多种营养元素。通常使用的肥料，重视氮磷钾，而忽视有机质和矿物质微量元素。通过农作物提供的信息，针对性地研究肥料的科学配置，是 AIGC 的主攻目标之一。

二是对农作物需用水的监测，整合气象降水、地表径流、地下水、土壤含水率、地表温度、地面蒸发、作物类型、植被覆盖度、生长状况、田间蒸散发等数据信息，确定农作物精准灌溉制度（用水量、时间和频次），实现作物高产。

三是实时记录作物信息，从种植、施肥（化肥用量）、灌溉、管理、除草、灭虫（农药名称，用量）采摘、收割、储存、加工、包装、运输、销售、消费等全流程的产品视屏、数据等及时溯源信息。

4.2.3 智慧医院

党的十九大提出实施“健康中国”的重大决策，将维护人民健康提升到了国家战略的高度。在智能医疗领域，首当其冲是建设智慧医院运营管理系统，推动医疗向精准化和

个性化方向发展。以 AIGC 为依托，助力医院实现信息聚合、数字建模，三维映射，搭建一个智能化数字空间，进而通过对多模态医疗数据的综合分析和生成，为疾病诊断、治疗方案制定提供更准确的支持，提升医院运营管理效率，提升医护效率，促进医患沟通，发挥精准医疗优势。在医疗服务方面，根据市民的健康档案、就医记录等数据，为市民提供个性化的健康管理建议和医疗预约。

5 结语

多模态 AIGC 是人工智能前沿技术，有跨领域理解与复杂环境响应能力，为智能产业注入动力，影响众多领域。随着科技发展，其潜力与创造力受重视，正广泛拓展。未来需强化跨领域多学科交叉研究，推动其纵深发展，实现更智能高效的内容生成与应用，引领自然生态与经济社会和谐发展。

水利、农业是超级系统工程，多模态 AIGC 在其领域的应用研究需多领域专家、生产经营者合作，产学研匹配。国家应出台政策规划，组织强大团队，保障衔接，科技攻关。

参考文献

- [1] 岳頔, 张晨康. 多模态场景下AIGC的应用综述, 计算机科学与技术, 2025,19(1)。
- [2] 多模态引领五大领域应用新趋势 搜狐 2024-12-11
- [3] AI多模态模型发展情况解析_财富号 caifuhao.eastmoney.com 2023-12-12
- [4] 人工智能领域多模态的概念和应用场景 www.elecfans.com, 2023-12-15
- [5] 多模态学习研究进展综述 CSDN博客频道,2024-12-11
- [6] 多模态AI技术详解:跨越数据边界的智能未来 CSDN博客频道, 2025-01-10