大语言模型技术和研究热点

杨璐
南京航空航天大学 金城学院，江苏 南京 211156

DOI: 10.61369/TACS.2025090045

摘　　要： OpenAI 推出 ChatGPT 启动了大模型元年，大语言模型的应用和技术发展进入了密集阶段。大语言模型的相关技术包括模型架构、预训练技术、模型微调、提示词、模型压缩、多模态融合。大模型的应用领域也非常宽泛，除了 ChatGPT 这样的聊天机器人，还可以应用于办公助手类产品、代码助手类产品、教育知识类产品、搜索引擎和推荐系统、企业业务定制化等。随着 OpenAI 发布的通用人工智能发展的5个阶段，大模型的发展前景也非常广阔，可以从模型架构、训练方法、模型压缩、应用创新、安全隐私等方向对大语言模型技术进行研究改进。

关 键 词： 大语言模型；Transformer；模型微调；人工智能

Large Language Model Technology and Research Hotspots

Yang Lu
Jincheng College, Nanjing University of Aeronautics and Astronautics Nanjing, Jiangsu 211156

Abstract: The launch of ChatGPT by OpenAI marked the beginning of the era of large models, ushering in an intensive phase of application and technological development for large language models. Relevant technologies for large language models include model architecture, pre-training techniques, model fine-tuning, prompts, model compression, and multimodal fusion. The application areas of large models are also very broad. Besides chatbots like ChatGPT, they can be applied to office assistant products, code assistant products, educational knowledge products, search engines and recommendation systems, and customized enterprise business solutions. With the five stages of general artificial intelligence development released by OpenAI, the prospects for large models are also very promising, and research and improvements in large language model technology can be conducted in areas such as model architecture, training methods, model compression, application innovation, and security and privacy.

Keywords: large language model; Transformer; model fine-tuning; artificial intelligence

引言

大语言模型（大模型）是近几年人工智能领域热点技术之一。所谓大模型是指拥有超大规模参数（通常在十亿个以上）、具有复杂计算结构的机器学习模型。2022年是大模型元年，在这一年，OpenAI 公司推出了对话聊天机器人 ChatGPT，可以让 AI 跟人进行自然的交流。ChatGPT 除了具备出色的自然语言处理技术，还拥有上下文对话能力，支持文章写作、诗词生成、代码生成等能力。对比传统的搜索引擎，ChatGPT 针对用户的问题，跳过了网页浏览和整合这两步，直接提供了精准答案，大大节省了用户的浏览、对比和整合时间。ChatGPT 的优点是高效（短时间生成大量高质量文本）、通用和自适应（可应用于不同领域，如智能客服、教育、医疗等）、人性化（生成的文本自然流畅）。ChatGPT 这些优点来源于其背后的技术 -- 大语言模型。

表1 Foundational GPTs

Model	Architecture	Parameter count	Training data	Release data
Original GPT(GPT-1)	12-level, 12-headed Transformer decoder(no encoder), followed by linear-softmax.	117 million	BookCorpus:4.5GB of text, from 7000 unpublished books of various genres.	Jun.11,2018
GPT-2	GPT-1,but with modified normalization	1.5 billion	WebText:40GB of text,8milllion documents, from 45 million webpages upvoted on Reddit.	Feb. 14,2019
GPT-3	GPT-2，but with modification to allow larger scaling	175 billion	570GB plaintext,0.4 trillion tokers. Mostly CommonCrawl, WebText,English Wikipedia, and two books corpora(Books1 and Books2)	Jun 11,2020

GPT-4	Also trained with both text prediction and RLHF; accepts both text and images as input.Further details are not public.	Undisclosed	Undisclosed	Mar.14,2023

斯坦福大学发布的《2024年人工智能指数报告》（Artificial Intelligence Index Report2024）显示，2023 年产业界、学术界研发的基础模型达 140 多个，在图像分类、视觉推理、英语理解等多项基准测试中超越人类。据称 GPT-4 在律师资格模拟考试中的分数超过 90% 的人类考生 [1]。

一、大模型相关技术

大模型在处理数据时采用的是“预训练 --> 微调 --> 推理”三个处理步骤。首先用大量样本对大模型进行预训练，让模型具备语法和语义知识；然后针对专门的领域，用一些带标签的小样本数据对模型进行微调，让模型在某个专门的领域更专业；最后根据用户的输入推理得到生成相应的输出。在这三步中，预训练涉及到特征提取模型（Transformer 模型）、预训练技术、基于人类反馈的强化学习、模型压缩技术、多模态融合。微调涉及到模型微调、基于人类反馈的强化学习。推理涉及到特征提取模型（Transformer 模型）、Prompt。

1. 预训练。预训练过程包括数据收集与预处理、模型选择、预训练和微调。模型选择指选择特征提取的模型架构，如 Transformer 模型。预训练使用无标签数据学习语言结构和语义。以 ChatGPT 举例，ChatGPT 是基于 Transformer 架构，通过大数据预训练学习通用特征，然后将这些特征应用于计算机视觉，自然语言处理等领域。其原理是通过海量数据提取语言知识和语义信息。

（1）训练数据的准备

对收集的文本、图像、音频等结构化和非结构化的通用数据、专业数据进行质量过滤、去除冗余、消除隐私、统一格式等清洗处理。将训练数据划为训练集、验证集和测试集，分别用于训练模型、验证模型性能和测试模型泛化能力 [2]。

（2）Transformer 模型

在自然语言处理领域，有三种特征提取模型架构：卷积神经网络（Convolutional Neural Networks,，CNN），循环神经网络（Recurrent Neural Network，RNN）和 Transformer。大语言模型一般都采用 Transformer 模型。它由编码器和解码器组成，引入了自注意力机制。自注意力机制是其创新处，使其在处理序列数据时能同时考虑输入序列中的所有位置，允许模型根据输入序列中的不同部分来赋予不同的注意权重，从而更好地得到语义关系，且可以实现快速并行处理 [3]。

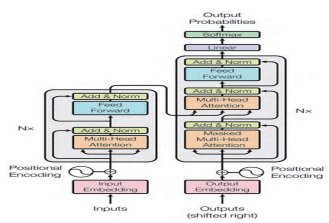

图 1 Transformer 架构

2. 微调是针对具体任务使用有标签的数据调整模型参数。

（1）模型微调

模型微调是指用少量带标签的数据对预训练模型进行再次训练，以适应特定任务。在这个过程中，模型的参数会根据新的数据分布进行调整。即在模型经过大规模数据集的预训练后，再在特定任务的特定数据集上进行精细调整，从而将预训练期间获得的广泛知识导入到具体应用中。微调技术主要包括基于经典参数微调、高效参数微调、提示微调、人类反馈的强化学习（RLHF）等方式 [4]。

（2）基于人类反馈的强化学习（RLHF，Reinforcement Learning fromHuman Feedback）

RLHF 是一种训练和微调大语言模型的方法，使其能够正确遵循人类指令。RLHF 创建了一个奖励系统，将人类的判断作为引导模型行为的奖励信号，让奖励模型评估特征提取模型（如 Transformer）的输出，并返回奖励信号，然后用奖励信号来优化其参数，以指导模型哪种响应更符合人类偏好。

（3）模型压缩技术

模型压缩技术包括权重裁剪、量化和知识蒸馏等，通过这些技术能显著减小模型的大小、优化其性能，降低了存储和计算负担，提高了部署效率和便捷性，同时又能保持模型性能。

（4）多模态融合

多模态融合技术通过整合来自不同模态的数据，如文本、图像、音频等，实现了对信息的全面、准确捕捉，极大地提升了模型的感知和理解能力。多模态融合技术包括数据预处理、特征提取和融合算法等步骤。融合算法是指将提取出的特征进行整合，生成更全面、准确的特征表示。

3. 推理：

（1）Prompt 指提示词，该技术通过给大模型提供一个或多个提示词或短语，指导模型生成符合要求的输出。Prompt 根据使用场景可以分为四种：Zero-Shot Prompt（在零样本场景下使用），Few-Shot Prompt（在少样本场景下使用），Chain-of-thought prompt（用于推理复杂任务），Multimodal prompt（将不同模态的信息融合形成多模态的提示）。设计合适的 prompt 可以提升大模型的准确率和可靠性。

二、大模型应用

目前，大模型已经在很多领域产品化了，除了我们在引言中

提到的 ChatGPT 聊天机器人，还出现了一些其他的应用。

1. 聊天机器人产品：以 OpenAI 公司的 ChatGPT 作为典型代表，这也是大语言模型应用的首款产品。能够基于在预训练阶段所见的模式和统计规律生成回答，还能根据聊天的上下文进行互动，真正像人类一样来聊天交流。它强大的自然语言处理能力和多模态转化能力使之可用于多个场景和领域 [5]。

2. 搜索引擎和推荐系统：通过深度学习算法，对用户的搜索意图进行准确理解，提供更精准的搜索结果和个性化推荐内容。如 Spotify 推出的 AI 播放列表功能，让用户通过书面提示生成个性化的音乐列表。

3. 企业业务定制化大模型：在工业领域、医药领域、管理领域等针对专业问题，提供定制化的服务。如开源免费软件 ChatDoctor 是一款医疗助手软件，该软件用 50 多万条真实医患对话对 LLaMA 模型进行微调 [6]；用户只需描述症状，ChatDoctor 就会像真人医生一样询问其他症状与体征，然后给出初步诊断和治疗建议 [1]。

三、大模型研究热点和发展前景

OpenAI 曾经发表基于大模型技术的通用人工智能的 5 个发展阶段：第一阶段，聊天机器人 Chatbots，具备对话能力的 AI；第二阶段，推理者 Reasoners，具备人类的推理水平能解决很多复杂难题；第三阶段，智能体 Agents，不只是推理，还能执行全自动化业务的智能体；第四阶段，创新者 Innovators，能协助人类完成新发明的 AI；第五阶段，组织者 Organizations，可以自动执行组织全部业务的 AI。目前，OpenAI 公司认为自己的产品处于第一阶段，但即将迈入第二阶段。

根据这 5 个发展阶段，可以看出大模型技术有长远的发展前景，需要解决和研究的问题还是很多的。

1. 模型架构的创新，尤其是与行业应用相结合。Transformer 虽然已取代 CNN，RNN 成为特征工程的主流框架，但还可以进一步改进与优化，如结合 CNN 和 RNN 的优点，构建混合型架构；支持多种输入类型的多模态大模型等。

2. 训练方法的改进。因为大模型训练样本和参数规模都很大，因此提高训练效率也很重要。如可以考虑采用分布式训练或大规模并行计算提高训练效率。

3. 模型压缩与加速。如新的蒸馏技术和高效的推理加速方法。

4. 应用领域的创新：大模型虽然已有一些应用，但都处于起步阶段。针对大模型的特点，除了行业应用创新外，还可以对现有应用进行升级改造，如基于大模型的文本生成、机器翻译、问答系统。

5. 安全性和隐私性。大模型训练需要大量的数据支持，但很多数据涉及到机密以及个人隐私问题，如客户信息、交易数据等。因此需要对大模型和数据进行隐私保护，以防止数据泄露、滥用和模型攻击。

四、总结

大模型作为人工智能的新兴技术和热点，已经受到了广泛的关注，目前国内外基于大模型的创新产品也层出不穷，且短时间内积聚了大量的用户。受益于大模型技术，这些产品给用户带来了极大的便利。纵观大模型技术和应用现状，发展前景广阔。

参考文献

[1] 刘聚海，胡玥，姜喆，等 . AI 大模型综述——兼论 AI 赋能不动产登记的基本思路 [J]. 自然资源信息化，2024,(04):1-18.

[2] 白培发，黄宗浩，王奕 . 大模型在智慧医院的应用研究综述 [J]. 计算机应用与软件，2024,41(07):1-5+19.

[3] 张钦彤，王昱超，王鹤羲，等 . 大语言模型微调技术的研究综述 [J]. 计算机工程与应用，2024,60(17):17-33.

[4] 任福继，张彦如 . 通用大模型演进路线 [J]. 科技导报，2024，42 (12)：44-50.

[5] 孙长秋，杜长斌，李菲，等 . 人工智能及大模型技术研究 [J]. 通信管理与技术，2024,(03):43-47.

[6] 张建云，孙璐萌 . 马克思主义理论视域下 ChatGPT 的功能、本质及意义 [J]. 兰州学刊，2023,(10):5-15.

[7] 骆卫华，刘群，白硕 . 面向大规模语料的语言模型研究新进展 [J]. 计算机研究与发展，2009(10):9.DOI:CNKI:SUN:JFYZ.0.2009-10-016.

[8] 程立海，崔荣国，董瑾，等 . 自然资源和国土空间大数据技术应用框架 [J]. 地球信息科学学报，2024，26(4):881-897.DOI:10.12082/dqxxkx.2024.230637.

[9] 武千千 . 基于大语言模型和知识图谱增强的建筑领域问答技术研究 [D]. 广东工业大学，2025.

[10] 张钦彤，王昱超，王鹤羲，等 . 大语言模型微调技术的研究综述 [J]. 计算机工程与应用，2024,60(17):17-33.

