

DOI:10.16356/j.1005-2615.2024.05.003

大模型技术的军事应用综述

高志强, 沈佳楠, 姬纬通, 袁梦, 崔隩龙

(武警工程大学反恐指挥信息工程教育部重点实验室(立项), 西安 710086)

摘要: 大模型以强大的自然语言理解、逻辑推理和内容生成等能力,为军事智能发展带来了前所未有的机遇和挑战。本文梳理了大语言模型的架构建模、适配微调、提示学习等技术,以及典型多模态大模型技术的发展现状,深度分析了国内外主流大模型军事应用平台特点,设计了大模型军事应用的典型任务场景和能力需求,探讨了大模型技术军事应用面临的挑战,展望了军事领域大模型应用的共性关键技术发展,以期提升基于大模型技术的重大军事应用设计与实战研究提供参考。

关键词: 大模型; 军事应用; 指挥控制; 军事智能; 人在回路

中图分类号: TP18 **文献标志码:** A **文章编号:** 1005-2615(2024)05-0801-14

Review of Military Applications of Foundation Model Technology

GAO Zhiqiang, SHEN Jianan, JI Weitong, YUAN Meng, CUI Xiaolong

(Key Laboratory of Counter-Terrorism Command & Information Engineering (Approved), Engineering University of PAP, Xi'an 710086, China)

Abstract: Foundation model has brought unprecedented opportunities and challenges to the military field due to its powerful capabilities in natural language understanding, logical reasoning, and text generation. This paper reviews related technologies of large language models, including architectural and modeling, adaption and fine-tuning, and prompting and learning, as well as the current status of the development of typical multimodal foundation models. It conducts a deep analysis of the characteristics of mainstream domestic and international military application platforms that utilize foundation models, and designs typical task scenarios and capability requirements for military applications of large models. Then, challenges of military foundation models are discussed. Finally, the key common technologies in military applications are summarized. Hopefully, our work can be beneficial for the design of significant military applications based on foundation model technologies and researches of actual combat.

Key words: foundation model; military application; command and control; military intelligence; human in the loop

2017年,Google推出的Transformer架构^[1]极大地推动了自然语言处理(Natural language processing, NLP)领域的发展。此后,国外的Speech Transformer^[2]、BERT^[3]、GPT^[4]、DALL-E^[5]、BART^[6]和LLaMA^[7]等模型陆续发布,国内也相继发布了ERNIE^[8]、GLM^[9]和百川大模型^[10]等大

模型产品。2022年底,OpenAI公司发布的ChatGPT,其强大的生成能力与广泛适用性将大模型的热度推上高潮。

除了民用领域,大模型技术的军事领域应用已初见端倪^[11],美国科技公司Palantir于2023年推出军事人工智能平台,利用大模型融合战场实时信息

基金项目: 国家社会科学基金(2022-SKJJ-C-093);武警部队科技创新团队创新研究项目(ZZKY20222103)。

收稿日期: 2024-04-04; **修订日期:** 2024-06-25

通信作者: 高志强,男,副教授, E-mail: 1090398464@qq.com。

引用格式: 高志强,沈佳楠,姬纬通,等. 大模型技术的军事应用综述[J]. 南京航空航天大学学报, 2024, 56(5): 801-814.
GAO Zhiqiang, SHEN Jianan, JI Weitong, et al. Review of military applications of foundation model technology[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2024, 56(5): 801-814.

和基础军事数据,为指挥官提供军事决策、命令下达和作战监控能力;美国 Scale AI 公司推出“多诺万(Donovan)”军事决策辅助系统,利用基于大语言模型的生成式 AI 程序提出行动方案建议,增强作战人员对作战环境的理解。国内渊亭科技发布了面向国防领域的天机大模型,利用军事术语常识、数字战场图谱、战略条令法规和战例演训记录等海量数据,完成了基座大模型的监督指令微调 and 强化学习反馈训练。

目前,大模型技术的军事应用已成为应对常态化外部军事风险和强敌侵权挑衅,塑造全局信息优势、决策优势、体系优势以及军事智能优势的重要方向。本文旨在梳理大模型的关键技术、典型的大模型军事应用平台特点及其发展方向,为大模型技术在军事领域的深度应用和军事智能化研究提供参考。

1 大模型技术

适用军事领域的大模型主要包括大语言模型(Large language model, LLM)和多模态大模型(Multimodal foundation model, MFM)。前者是基于大量军事文本数据训练而得到的深度神经网络模型,可理解军事术语,生成具有自然语言特征的

军事文书,满足智能问答、文本摘要和军语翻译等多种自然语言任务;后者融合了文本、语音、图像和视频等多模态军事数据进行学习和训练,可以同时处理自然语言、视觉、语音等多种复杂多模态任务。

1.1 大语言模型技术

大语言模型技术主要涉及架构建模、预训练、适配微调和提示学习等方面。目前,基于 Transformer 架构的预训练语言模型在自然语言处理任务中取得了巨大的成功,已成为 LLM 的主流技术范式。

1.1.1 架构建模

目前,LLM 的常见架构以 Transformer 为基础,如图 1 所示。Transformer 可根据上下文动态分配不同位置的注意力权重,并在处理长序列的同时关注序列的所有位置,其变体策略主要涉及取代后置层标准化(Layer norm, LN)^[12-14]、采用 GELU^[15] 及 SwiGLU^[16] 激活函数、采用 ALiBi^[17] 和 RoPE^[18] 编码等。LLM 的建模策略大致分为掩码语言建模、自回归语言建模和序列到序列建模 3 种^[19-20]。其中,掩码语言建模基于编码器的双向模型,从输入序列中随机遮盖文本片段,然后基于上下文预测被遮盖的文本片段,例如, $P(x_k|x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_n)$ 表示给定上下文 $\{x_1, x_2, \dots, x_n\}$ 中单词 x_k 的概率^[21]。

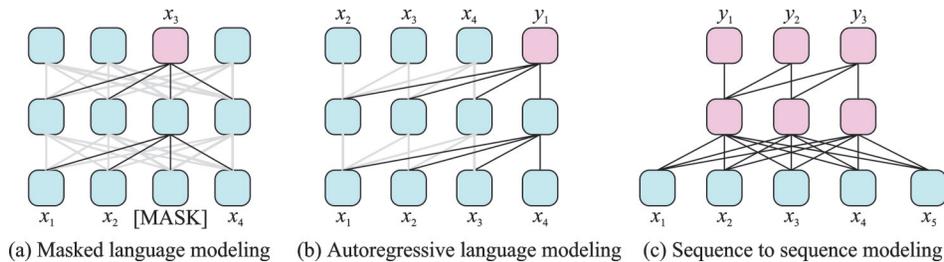


图1 大语言模型架构的建模策略

Fig.1 Modeling strategy for LLM architecture

自回归语言建模基于 Transformer 的解码器,可以捕捉序列的依赖关系和上下文信息,具有顺序生成、一步预测和任务依赖的特点,旨在预测下一词(例如,图 1(b)中的 y)的条件概率或将概率 $P(x)$ 分配给词序列 $x=(x_1, x_2, \dots, x_n)$ ^[22]。序列到序列建模则同时使用 Transformer 的编码器和解码器,输入序列通过编码器转化为固定大小的中间向量,编码器的输出作为解码器的初始隐藏状态,解码器根据之前生成的元素和编码器输出生成下一个元素,直到生成整个序列^[23]。

此外,LLM 的预训练利用大量无标签数据来学习通用的语言表示和知识。预训练数据的来源分为通用文本数据和专用文本数据。前者包括公开书籍和网页等,具有类型多样、规模庞大、常见且较易获取等特点;后者包括行业专业术语和代码

等,具有较强的专业性,可以增强 LLM 解决特定领域任务的能力。预训练数据的数量、质量和分布会影响 LLM 的训练效果。其中,数据量的影响与模型参数的规模法则类似,较大的预训练数据量一般会带来更好的效果,而且缺乏充足的预训练数据,LLM 训练无法达到最优效果^[24-26]。

1.1.2 适配微调

LLM 的适配微调主要分为指令微调、参数高效微调和对齐微调。前者在特定任务上利用少量指令数据训练预训练模型以适应新任务需求,常用指令构建方式包括公开数据集格式化生成的指令^[17,27]、LLM 自动生成的指令^[28]和基于人类需求标注的指令^[29-31]。参数高效微调遵循 Delta Tuning 框架^[32],旨在减少微调参数的同时,保持 LLM 的下游任务表现,包括 3 种类型:(1)在 LLM 上添加

额外的模块或参数,但仅微调新增部分^[33-35]; (2) 基于规范的方法,不引入新参数,仅指定部分参数为可训练参数^[36-37]; (3) 将部分模型参数重参数化到低维度参数空间中,只优化低维空间中的近似参数。如图 2 所示,LoRA^[38]通过添加低秩约束来近

似每层的更新矩阵 W ,其中 $A \in \mathbb{R}^{d \times r}$ 和 $B \in \mathbb{R}^{r \times d}$ 是 LoRA 低秩适应的权重矩阵,秩 r 远小于 d 。DoRA^[39]将预训练权重分解为幅度分量 $m \in \mathbb{R}^{1 \times d}$ 和方向分量 $V/\|W\| \in \mathbb{R}^{d \times d}$,并利用 LoRA 方法对方向分量进行微调,进一步提升微调效率的 $1/\|V+\Delta V\|$ 。

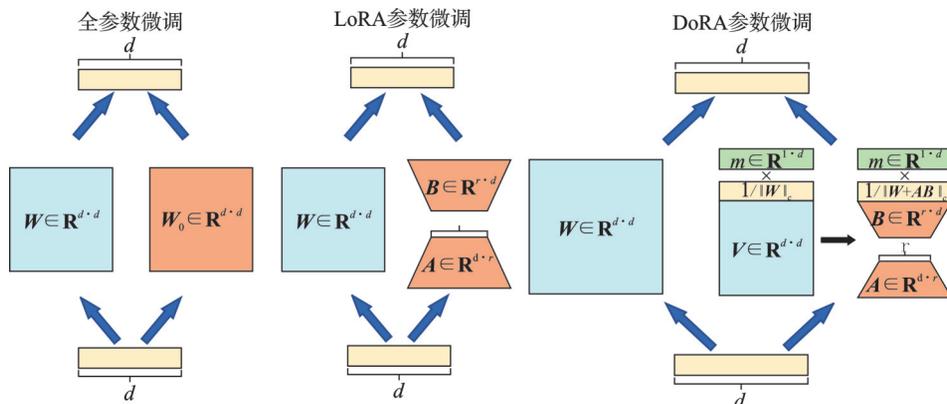


图 2 全参数微调、LoRA 参数微调、DoRA 参数微调对比
Fig.2 Fine-tuning comparison of full, LoRA and DoRA parameters

对齐微调主要考虑有用性、诚实性和无害性等标准,旨在使 LLM 的行为符合人类期望的价值观或偏好,本质是基于人类反馈的强化学习 (Reinforcement learning from human feedback, RLHF)。通常,RLHF 分为收集人类反馈^[39]、训练奖励模型^[40]和强化学习微调等步骤。

1.1.3 提示学习

LLM 的提示学习通过设计合适的提示引导模型的自监督学习任务,主要分为上下文学习 (In-context learning, ICL)^[41]和思维链 (Chain-of-thought, COT) 提示^[42]。影响模型 ICL 能力的重要因素主要有模型规模和任务设计^[43-44]。如图 3 所示,COT 通常被认为是 LLM 规模达到一定程度后的能力涌现,常用的设置方法是对问题使用多个推理路径^[45-47]。

1.2 多模态大模型技术

多模态大模型是军事大模型的另一项支撑技术。2020 年,Google 发布的 Vision Transformer (ViT)^[48]加速了多模态大模型发展进程。MFm 融合了文本、图像、视频等多模态数据,在军事智能交互、智能标图绘图等场景具有重要应用价值。

1.2.1 多模态大模型分类

适用于军事场景的 MFm 主要分为 3 种:面向理解任务的 MFm、面向生成任务的 MFm 和兼顾理解和生成能力的 MFm。前者通常基于 Transformer 的编码器实现,如图 4 所示,根据不同模态特征是否由同一 Transformer 处理,又分为单流和多流结构。其中,单流结构中不同模态特征拼接后由共享的 Transformer 处理,典型代表有

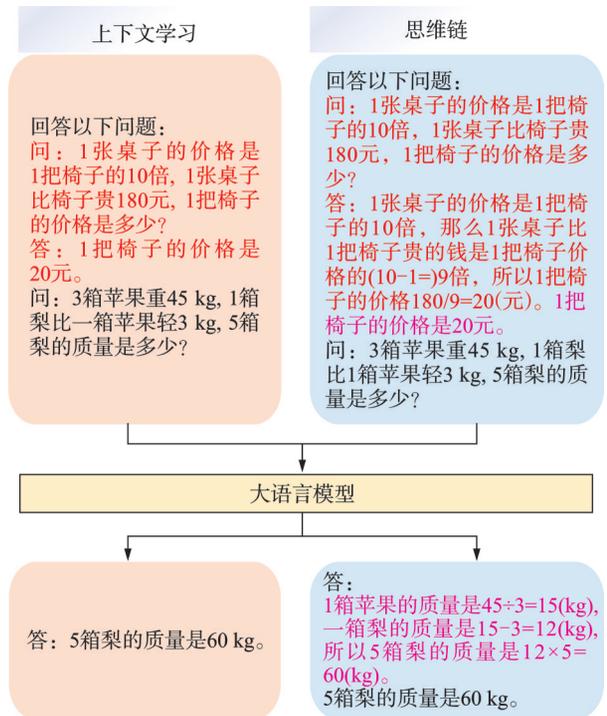


图 3 ICL 和 COT 对比
Fig.3 Comparison of ICL and COT

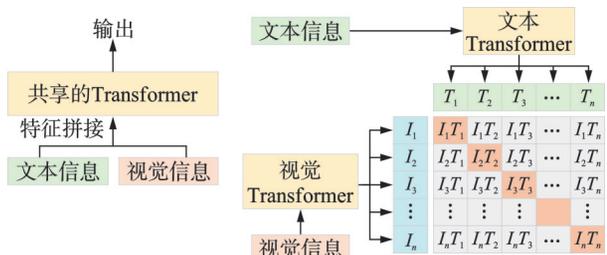


图 4 单流和多流结构对比
Fig.4 Comparison of single-stream and multi-stream architectures

VL-BERT^[49]、UNITER^[50]等。多流结构中不同模态由独立的Transformer编码处理,存在特征上的交互融合机制。典型代表有ViLBERT^[51]和CLIP^[52]等。

面向生成任务的MFM通常基于Transformer解码器,分为序列生成模型和扩散模型。前者的典型代表有DALL-E和CogVLM^[53]等,扩散模型则通过连续添加高斯噪声破坏训练数据,再学习反转噪声以恢复数据^[54]。兼顾理解任务和生成能力的MFM联合了Transformer架构的编码器和解码器,编码器通过双向注意力机制来学习数据的理解

能力,解码器通过单向注意力机制学习生成能力^[55-57]。

1.2.2 多模态大模型的关键技术

如图5所示,MFM的关键技术主要分为4种^[58]:多模态指令微调(Multimodal instruction tuning, M-IT)、多模态上下文学习(Multimodal in-context learning, M-ICL)、多模态思维链(Multimodal chain of thought, M-COT)、LLM辅助视觉推理(Large language model-aided visual reasoning, L-AVR)。其中,M-IT主要包括多模态指令收集和模态桥接^[59]。

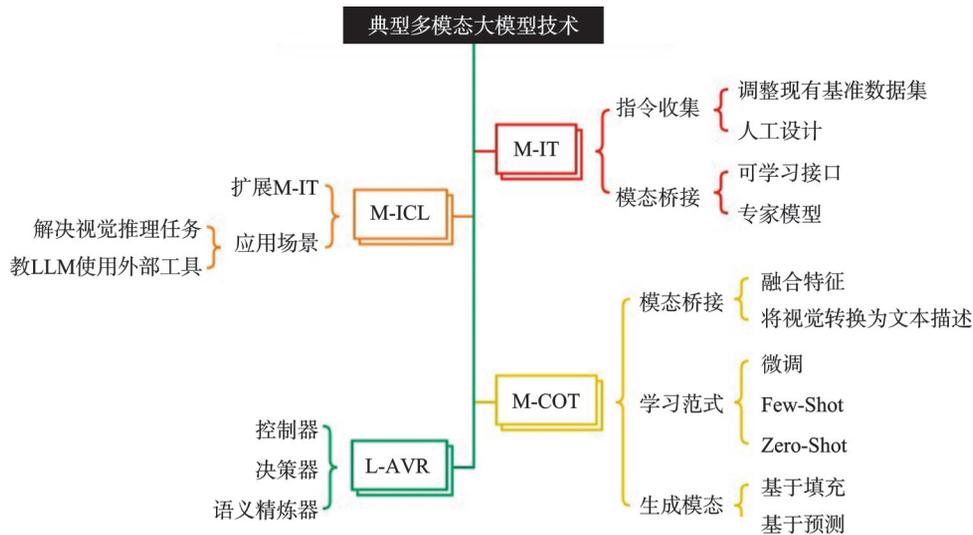


Fig.5 Typical MFM technologies

M-ICL主要解决各种视觉推理任务和指导LLM使用外部工具,可通过扩展多模态指令调优模板实现,在推理时添加1组上下文中的样本作为演示集,但模型对添加样本的方式很敏感^[60]。M-COT在学习范式上和LLM类似,有微调、Few-Shot和Zero-Shot三种方式,通过基于填充或基于预测的模态来构建思维链。在多模态领域,CoT面临的重大问题仍是填补模态差距,常用的模态桥接方式包括融合特征或将视觉转换为文本描述。前者将视觉嵌入映射到单词嵌入空间,然后将映射的嵌入作为提示,发送给其他语言模型,引发多模态思维链推理^[61-63]。

L-AVR受到LLM工具增强的启发,主要有3类:(1)模型充当控制器,将复杂任务分解为简单的子任务,并分配给适当的工具和模块^[64];(2)模型充当决策者,复杂任务以多轮迭代的方式解决,模型需要总结当前上下文和历史信息,并决定当前步骤可用的信息是否足以回答问题或完成任务,然后对答案进行组织和总结,以方便用户的方式呈现;(3)模型作为语义精炼器,通常将信息整合到一致

性好并且流畅的自然语言句子中,或根据不同的特定需求生成文本。

结合军事大模型构建、训练数据集、军事文书理解与生成等能力需求,以及大模型在架构建模、预训练、适配微调、提示学习和多模态等方面的技术特点,汇总比对结果如表1所示。

2 大模型军事应用平台

尽管国外已开展军事大模型系统的相关研究,但总体来看该方向尚处于探索试用阶段,公开的军事理论成果和实证应用研究较为匮乏。本小节对典型的大模型军事应用平台的技术特点、体系架构和发展成熟度进行对比研究。

2.1 Palantir AIP平台

2023年4月Palantir公司发布了面向国防应用的人工智能平台(Artificial intelligence platform for defense, AIP)。AIP平台核心理念是通过信息系统集成方式构建AI操作系统,并将GPT-4和BERT等多种军事人工智能模型或专用大模型系统集成到受控的网络环境中,进而利用大模型来整

表 1 大模型技术汇总比对

Table 1 Summary and comparison of foundation model technology

技术分类	主要方向与技术特点	典型代表与潜在军事应用能力
大语言模型技术	架构改进与建模策略:标准化层、激活函数、位置编码和注意力机制等改进;掩码语言建模、自回归语言建模和序列到序列建模等;实现训练稳定性、外推能力、泛化能力、长序列建模和低计算复杂度等	DeepNorm、GELU、SwiGLU、ALiBi、RoPE、PaLM、BERT、RoBERTa、LLaMA 和 T5 等;军事大模型建模
	预训练数据与策略:高质量数据集、训练策略;实现泛化能力、高效计算	Common Crawl、WikiQA 和 OpenWebText 等;军事大模型训练及训练数据集构建
	适配微调:指令微调、参数高效微调和对齐微调;实现下游任务适配	InstructGPT、BLOOMZ-P3、LoRA、DoRA 和 RLHF;军事大模型适配微调
	提示学习:上下文学习、思维链;实现自监督学习、涌现	GPT-3、零样本 COT;军事大模型提示学习
多模态大模型技术	单流、多流架构:实现不同模态特征拼接与交互融合	VL-BERT、UNITER、ViLBERT 和 CLIP;军事多模态大模型建模
	面向任务和生成能力的 MFM:基于 Transformer 编码器、解码器、序列生成模型和扩散模型	DALL-E、CogVLM、DALL-E2、LDM、VL-T5、Unified VLP 和 Sora;基于军事多模态大模型的理解与生成
	多模态指令微调、上下文学习、思维链和 LLM 辅助视觉推理	Flamingo、BLIP-2、ScienceQA 和 VISPROG 等;军事多模态大模型技术适配

合战场实时信息和军事基础数据,为指挥官提供军事决策、命令下达和作战监控功能,实现优化决策流程、缩短决策时间和获取最优作战方案,保障作战质量的目标,并提供本地化部署、AI 工具驱动和相关决策调用能力,以期在高度受控和敏感的环境下确保合法地将大模型用于军事实践。

如图 6 所示,AIP 平台包括数据层、控制层和交互终端层 3 层架构。其中,数据层包含单兵、军事单元和装备等数据信息,并可控制各单元间的数据流和数据访问权限等;控制层的用户可对模型、认证、资源和第三方应用等基础配置和控制流进行管理;交互终端层的用户可以使用 AIP 平台的左侧对话窗口进行交互,右侧的态势窗口查看战场情况,以及将生成的行动方案等内容发送给指挥链上的其他军事人员。



图 6 AIP 平台架构

Fig.6 Architecture of AIP

下面以应对敌方坦克集结的行动方案生成为例,介绍 AIP 平台的实际应用。当军事人员收到 AIP 平台的告警提示“敌方正在附近集结”后,军事人员立刻向 AIP 平台请求更多详细情报,以及敌方部队的可能编队预测。AIP 平台可以辅助军事人员获得更多精准情报,分析该地区的敌方部队信息,并生成可能的敌方编队。在得知目前态势后,军事人员第一时间做出决策,要求无人机进行现场侦察。侦察中,MQ-9 无人机捕捉到 1 辆 T-80 坦克的存在。接着,军事人员向 AIP 平台询问如何进行应对。如图 7 所示,利用大模型的生成能力,AIP 平台生成了 3 种可能的行动方案,具体包括使用 F-16、远程火炮或标枪导弹攻击 T-80 坦克,这些行动方案随即被自动发送到指挥链上,以支持更高层次的决策方案制定。

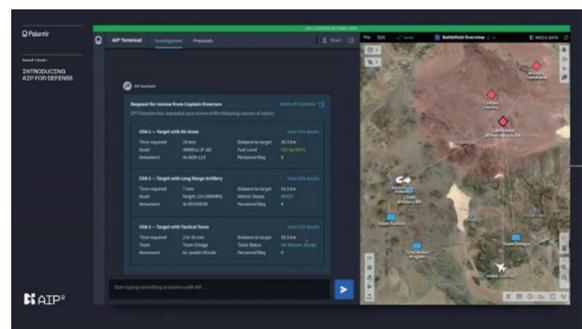


图 7 AIP 平台的行动方案生成

Fig.7 Generation of course of action for AIP

除了大模型本身的军事语义理解、行动方案生成能力,AIP 平台在军事应用中涌现的大模型能力还包括极大跨度的部署能力,例如机密网络、战术边缘设备等,以及大模型自定义的多权限、多角色

防护控制机制,最终形成安全的AIP平台数字操作记录。通过AIP平台的自动分析能力,甚至可以在任务执行前确保附近部队拥有足够的弹药,并自动干扰敌方系统,这为未来战场决策和作战能力生成提供了更为可靠和灵活的基础,但该能力尚未成熟,仍处于探索阶段。

2.2 Scale AI Dnovan 平台

Donovan平台是美国Scale AI公司于2023年5月发布的大模型军事决策辅助系统,部署于美国陆军第18空降师的加密网络,是旨在增强分析人员和作战人员对作战环境的理解,并提出行动方案建议的生成式AI程序,可以调用GPT-3.5和LLaMA 2等大模型完成决策分析任务。

如图8所示,Donovan平台包括数据引接层、系统管理层和行动层。其中,数据引接层可以连接所有用户授权的网络数据或数据源,并能分析非结构化文本数据、结构化数据和地理信息情报等多模态数据,以及PPT、电子邮件和PDF等文件数据;系统管理层可以调整审核模型参数,进行用户行为监视和访问控制检查,以确保大语言模型输出易于理解且受控的结果;行动层提供与大语言模型的交互界面,可以生成各种指令、可视化地图和任务报告,并为外部系统分配任务。



图8 Donovan平台架构

Fig.8 Architecture of Donovan

作战人员和情报分析人员通过与Donovan平台的大语言模型进行交互,可快速生成各种指令、可视化协同作战图和任务报告,并自动为外部作战系统生成空中任务分派命令建议。Donovan系统的工作流程包括引接数据、理解数据、信息检索、生成行动方案、持续学习和改进模型。在引接电子邮件、情报报告、命令和卫星图像等数据后,Donovan平台可定制和微调大语言模型,根据用户需求理

解、组织数据,交互式检索数据和回答作战问题,并生成行动方案(Course of action, COA)、总结报告等,以及根据部队需求不断调整模型和持续学习。目前,Donovan平台的主要应用情况如表2所示。

表2 Donovan平台的主要应用情况

Table 2 Main application of Donovan

军种	使用单位/活动	主要功能	时间
陆军	第18空降师	<ul style="list-style-type: none"> •加速战场态势理解 •支持决策制定,整合实战数据 •支持联合全域指挥控制,缩短计划周期 	2023年5月
空军	第6次全球信息优势演习	<ul style="list-style-type: none"> •辅助决策 •获取目标信息 •支持火力打击任务 	2023年7月
⋮			

在受控情况下,Donovan平台可以与其他大语言模型进行数据交换,除访问本地数据外,还可通过云端或API访问系统,完成情报查询、起草行动方案和规划任务等军事行动,并在事后通过用户反馈和Scale数据引擎增强模型性能。特别地,如图9所示,Donovan平台在交互回答中可以使用引文,并实况追溯数据来源。

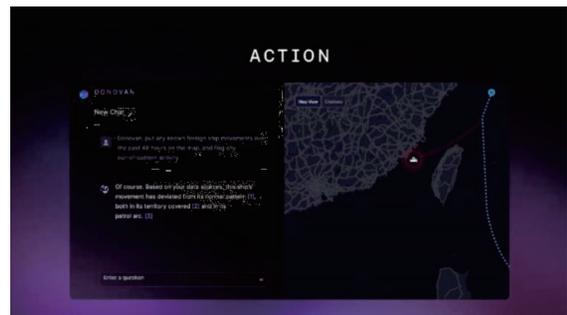


图9 Donovan使用引文回答界面

Fig.9 Interface of Donovan using of citations in its answers

2.3 “艾武大模型+”系统

2023年8月,武警工程大学反恐指挥信息工程教育部重点实验室(立项)提出并组织研发的“艾武大模型+”系统获中国国防科学技术信息学会“智研”杯大模型国防应用邀请赛“应用创意陈述”“应用潜能展示”双赛道第一名,其核心架构是“虚拟人+大模型+大数据+大系统”,旨在实现指挥信息系统的智能交互、语义理解和任务规划等能力。如图10所示,大数据是领域专用、公开和外部等数据,以及涵盖我情、敌情和战场环境等精确领域语料的集合;大模型通过预训练、指令微调、外挂知识库等,形成大语言模型的军语理解、任务规划和决策建议等能力;大系统为多任务大数据应用系统及服务,可以对齐调用系列决策分析(任务)小模型(矩阵);虚拟人的具身为“艾武”虚拟参谋,是系统

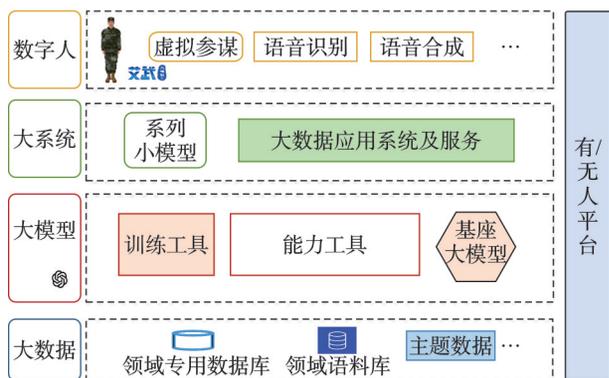


图 10 “艾武大模型+”系统架构
Fig.10 “Aiwu LLM+” system architecture

的人机交互接口,具备语音和文本等智能交互能力,可拓展支持“云-边-端”的有/无人平台。

为达到模型、数据和系统的“1+1+1>3”的智能涌现效果,“艾武大模型+”系统围绕大模型的智能交互、语义理解、任务规划、系统执行、结果数据反向输入、精准内容生成以及反馈评估等功能,从以下方面对模型进行了调优:(1)基于虚拟参谋“语音识别与合成”的综合集成,使大模型具备多模态

的人机交互能力;(2)基于“指挥专用知识库”的提示工程训练,使大模型具备指挥领域的语义理解能力;(3)基于大系统的“大模型+系列小模型+可视化”模式,使大模型具备任务规划和系统执行能力;(4)基于大数据平台的“精准结果数据”反向输入,使大模型具备外挂知识库和精准内容生成能力;(5)基于指挥员评价的“循环反馈提示工程”,使大模型具备自学习的自我提升能力。

此外,渊亭科技发布了面向国防领域的天机军事大模型,旨在打造具备复合能力的作战规划“智能参谋”,但其本质仍是内置了适应军事场景的 ChatGLM、LLaMA 和 BLOOM 等多个基座模型,采用大模型带小模型(1+N)方式,集成实时防务情报智库、作战仿真推演、智能虚拟参谋、情报信息中心以及数字战场图谱等多功能模块的应用平台。天机军事大模型的主要特色为防务行业实践中积累的海量军事数据和情报分析能力。综上所述,基于公开信息的典型大模型军事应用平台架构特点、技术优势、应用场景以及发展阶段情况汇总结果如表 3 所示。

表 3 典型大模型军事应用平台分析

Table 3 Analysis of typical platform using foundation models for military applications

平台/系统	架构特点	技术优势	应用场景	发展阶段
Palantir AIP 平台	数据层、控制层和交互终端层	支持大跨度部署;多权限合法防护;支持受控网络环境集成	行动方案生成	尚未成熟
Scale AI Donovan 平台	数据引接层、系统管理层和行动层	支持 LLaMA 2、GPT-3.5 调用和数据交换;支持云端或 API 访问	态势理解、引文问答	部分应用
“艾武大模型+”系统	虚拟人+大模型+大数据+大系统	架构可扩展性强;具身智能;大模型+系列小模型+可视化	智能化指挥	概念开发、试点应用
天机·军事大模型	未发布	内置 ChatGLM、LLaMA 和 BLOOM 等基座模型;大模型带小模型(1+N)方式	情报分析、策略仿真	仅发布

3 大模型军事应用场景和能力需求

结合公开资料和相关保密规定,下面以大模型适用的国防安全突发事件、非战争军事行动为任务场景,侧重任务筹划和任务理解能力,具体以反介入/区域拒止力量联动筹划任务和地震灾害抢险救援行动为例,梳理基于大模型的军事任务关键处置流程和阶段划分,总结军事应用平台的系统能力需求、大模型技术的基础能力需求以及智能应用能力需求。

3.1 典型任务场景

(1)反介入/区域拒止力量联动筹划任务

针对强敌的反介入/区域拒止战略,综合运用军事防御系统和相关能力,将强敌优势干预力量限制在预定区域外,或者显著降低强敌在特定区域的作战能力,形成能战、胜战的强大战略制衡态势。通常,区域反介入/区域拒止力量联动筹划任务包

括情报搜集整编、确定决心意图、处置与威慑和复盘总结等阶段,结合大模型的联动处置任务流程如图 11 所示。

在情报搜集整编阶段,深入搜集挖掘敌方情报和潜在威胁目标,利用大模型的军事文本、图像等多模态情报的理解能力,抽取军事动向、意图和实力水平等关键信息;在确定决心意图阶段,基于情报分析结果,利用大模型的要害理解能力研判拒止区域范围,生成联动处置决心方案;在处置与威慑阶段,根据决心和意图,利用大模型制定处理力量配系建议,辅助调动部署联合指挥相关力量,形成强大有效的区域拒止威慑,并持续向大模型动态反馈更新情报;在复盘总结阶段,汇总沉淀情报和任务数据,形成大模型所需的联动处置任务知识库,优化拒止策略和兵力编组,提高未来拒止行动的有效性和灵活性。

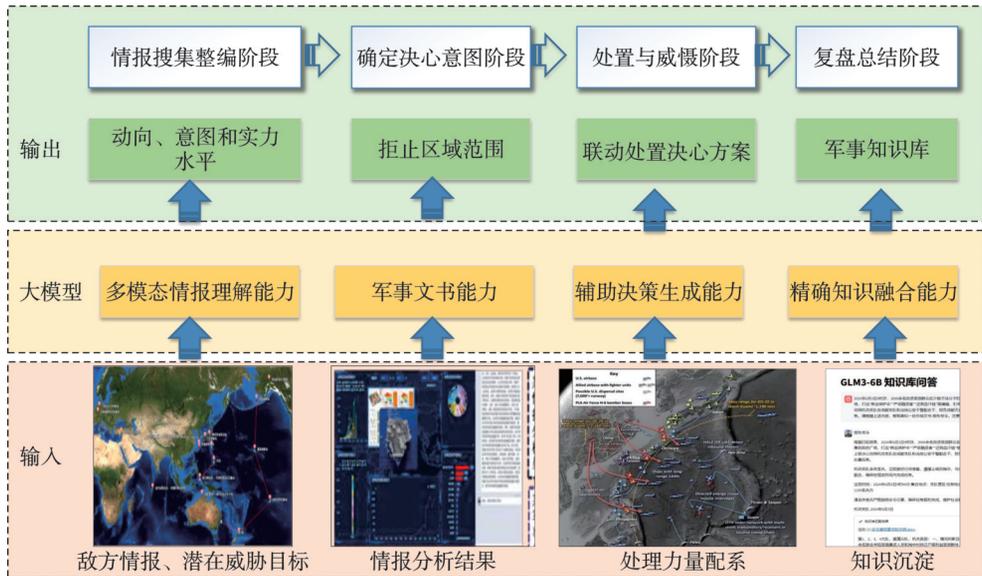


图11 结合大模型的联动处置任务流程

Fig.11 Process of disposal of missions integrated with LLMs

(2)地震灾害抢险救援行动

中国位于环太平洋和欧亚地震带的交汇部位,地震频度高、强度大、震源浅且分布广。作为典型的非战争军事行动样式,地震灾害抢险救援行动与

联动筹划任务不同,面临突发灾情未知、次生灾害救援处置难度大等难题,一般包括先期救援、紧急驰援、现场救援和助民归建等阶段,结合大模型的地震灾害抢险救援行动流程如图12所示。

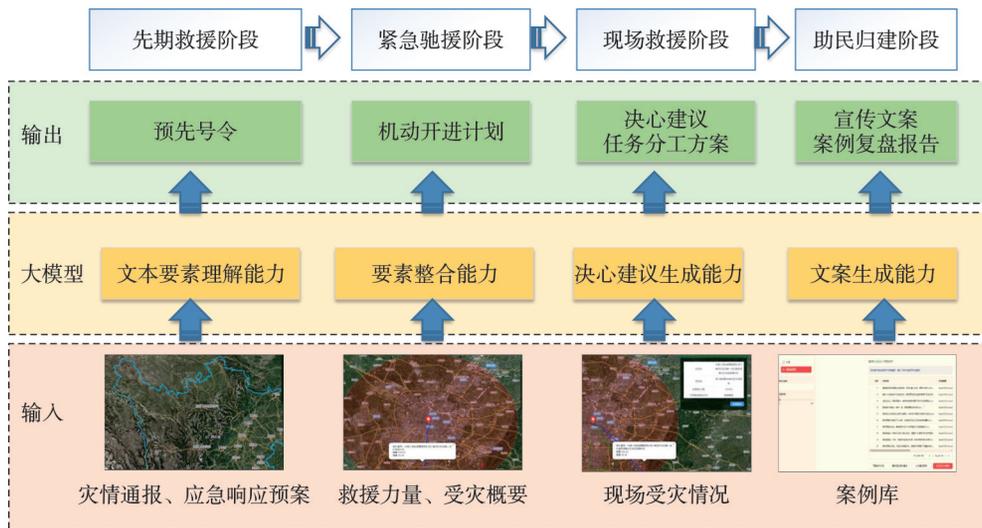


图12 结合大模型的地震灾害抢险救援行动流程

Fig.12 Process of emergency rescue operations for earthquake disasters integrated with LLMs

在先期救援阶段,利用大模型迅速分析灾情通报、应急响应预案,形成预先号令,支持震中附近救援力量第一时间展开救援;在紧急驰援阶段,利用大模型生成机动开进计划,协助专业救援力量、增援力量和大型专业工程装备分批向震中驰援;在现场救援阶段,利用大模型分析震中“72 h黄金期”灾区救援形势,形成决心建议和任务分工分组方案;在助民维稳阶段,汇集民情、社情和舆情,利用大模型生成进村入户宣传文案“军民鱼水情”,支持形成救援行动案例复盘报告。

3.2 能力需求

针对典型的国防安全突发事件,以及非战争军事行动中地震灾害抢险救援等行动等军事任务场景需求,自底向上,可以归纳出大模型军事应用平台应具备症候早期预警、情报快速整编与要图标绘、任务交互理解规划以及要报自动生成等系统能力,其核心能力支撑为大语言模型和多模态大模型技术,以及系列任务小模型,进而兼顾技术和军事需求,形成军事应用所需的语音交互、军语理解、态势理解、任务规划、情报分析、文书编辑、决策生成、

作战标图以及评估成长等基础能力,对标一线指挥员、参谋人员、战略决策人员的3类军事用户需求,形成面向军事任务场景的军事智能应用能力,最终形成的大模型军事应用能力需求矩阵如图 13 所示。

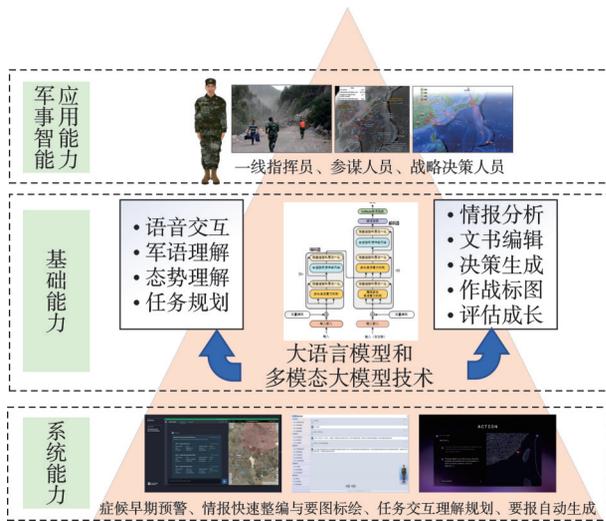


图 13 大模型军事应用能力需求矩阵

Fig.13 Capability requirement matrix for military applications of FMs

下面对应一线指挥员、参谋人员和战略决策人员等角色,从智能化指挥控制、智能参谋业务和认知域作战角度对大模型军事应用能力需求进行概述。(1)智能化指挥控制方面,美国成立了“利马”工作组,负责在国防部范围内评估、协调和使用生成式人工智能技术。美国空军在第 6 次全球信息优势 (Global information dominance experiments, GIDE) 演习中已经将大模型投入初步应用^[65];美国海军陆战队的 Hermes 大模型被用于理解其他国家治国方略等内容,帮助一线指挥员从多个维度观察战场^[66];阳东升等^[67]将大模型融入到指挥与控制 (Command and control, C2) 领域并对应用场景和筹划-准备-执行-评估 (Planning, readiness, execution, assessment, PREA) 环,观察-判断-决策-行动 (Observe, orient, decide, act, OODA) 环做出了分析。(2)在智能参谋业务方面,随着大模型技术在文本生成、推理和视觉多模态领域能力的提升,参谋人员利用多模态大模型学习设置好的模板,可以在极短时间内生成军用文书^[68];大模型具备的推理和运算能力能够支持参谋人员在输入战斗计算问题之后立刻得到答案,不再需要进行传统的纸质运算,其强大的视觉能力能够自动生成图像,可用于作战标图中;此外,大语言模型可辅助参谋人员分析作战需求,生成行动预案,大大缩短作战准备和实施周期。(3)近年的军事案例中,认知域作战中信息传播和操纵在战略决策中发挥了重要

作用,将大模型生成的内容大量投放于互联网可以塑造“人造”信息环境优势^[69],形成隐晦的政治倾向、可塑的价值取向、伪装性和麻痹性强的舆论诱导^[70],精确攻击公众获取信息的新型民用设施等^[71],甚至可以达到战略性的军事收益。

4 面临挑战与发展方向

随着大模型技术的持续发展,军事大模型应用研究既具有民用领域的开源广、迭代快等特点,又面临高质量军事语料不足、军事基座大模型评价体系未统一、安全问题多元等瓶颈与挑战。

4.1 高质量军事语料不足制约大模型军事理解能力

任何大模型的构建都依赖于庞大的语料库进行深度训练,优质语料数据不仅能够提升大模型处理复杂任务的能力,更是模型语义理解和泛化能力的基石,也是克服大模型“幻觉”问题的重要基础。如表 4 所示,LLAMA-3 更是使用了 15 TB 的文本数据进行训练。当前各种通用领域的语料库和数据集建设较为完善,甚至高质量新数据即将用完,亟需拓展“合成数据”。然而,相较于民用领域,军事领域数据密级高、获取难度大和格式不统一问题突出,非专业人士对语料数据标注和分类难度极大,尤其军事语料相关基础性工作人工投入大,“人在回路”相关技术发展较为缓慢。军事领域无论是在语料库还是军事相关数据集的建设方面,都明显滞后,这为大模型技术在军事领域应用带来了严峻的数据挑战,极大地限制了大模型军事理解能力的准确性和可靠性。

表 4 部分大模型预训练数据体量

Table 4 Pre-training data volume of some LLMs			
模型	发布年份/年	参数量/B	预训练数据量
PanGu- α ^[72]	2021	200	1.1 TB
Gopher ^[26]	2022	280	300 B tokens
GPT-3 ^[73]	2020	176	300 B tokens
Baichuan ^[74-75]	2023	7/13	1.4 TB tokens
QWEN ^[76]	2023	1.8/7/14/72	3 TB tokens
LLaMA-3 ^[77]	2024	8/70	15 TB tokens

由于通用大模型存在军事知识盲区,因此高质量军事语料是大模型预训练和微调,以及信息准确性的重要支撑。尤其是军事领域数据跨网隔离,难以统筹实现一体化的标注、管理、共享和运用,数据质量不高、分布不均衡以及数据孤岛林立,加之低质训练语料中潜在偏见和公平性问题,严重制约了军事大模型的构建和性能。值得关注的是,合成数据、规模定律和检索增强生成等可作为实现安全可控的高质量专用语料构建、避免军事理解能力中知识遗忘和“幻觉”问题的重要方向。

4.2 基座大模型评价体系未统一 影响军事大模型部署推广

大模型技术百花齐放,榜单层出不穷,怎样选取基座模型,选取什么基座模型成为军事大模型应用的一大难题。军事基座大模型的选择涉及众多因素,包括模型性能、参数规模、部署成本、是否开源、安全性、可靠性以及支持的应用场景等,不仅是技术问题,也是工程和制度规范问题,传统模型评价主要依据特定任务测试集上的模型性能表现考查准确率、召回率等指标,但军事大模型评价仅靠通用分类或生成数据集无法完整体现军语理解与运用能力。尤其,零/少样本情况下,单一类型或少数任务驱动的基准评测,难以度量军事大模型的真实问答、推理、生成和运用等能力,需要考虑挑战性测试、人工参与程度、贴近实战场景以及多模态信息综合运用等需求。

因此,必须紧密结合军事数据的独特时空特性和军事业务需求,解决军事领域知识高度封闭条件下大模型军事语义理解和精准生成能力的评价问题,现有民用大模型“不懂军语,军味不足”,亟须将军事领域知识融入大模型的评价中,并在实验验证中,对满足基本条件的大模型进行多轮能力评估,全面考察不同模型在军事应用中的性能优劣和成本效益。具体实施中,可以结合大量相关军事专家评价、增加安全风险评估,分阶段形成适用于军事领域的大模型应用评价标准,为后续的应用实践提供指导。

4.3 安全问题多元限制军事大模型的“托底”应用

随着大模型技术的广泛应用,其数据、模型、版权、伦理等安全问题愈发突出,甚至从经典安全领域拓展到数据偏见与公平性等方面,尽管军事大模型应用并未显性涉及数据隐私的泄漏风险,但基座大模型往往基于公开民用数据训练,数据信任、价值观表达仍存在风险,尤其“幻觉”问题易于诱发大模型在军事应用方面的安全风险。因此,如何在关键军事数据流转受限、军语内容规范严格条件下,保证军事大模型的“托底”安全和可靠应用亟待解决。

由于军事领域的特殊安全保密要求,军事大模型交付后必须具有完善的高安全性保障,轻易集成无法保证安全可控的大模型会带来涉密信息泄露、决策误导等一系列重大军事危害,开放、共享、训练、使用和内容生成都需要以可控大模型为前提,价值观对齐、安全评估基准和可解释性安全等都需要满足物理隔离环境下的私有化部署,进而细化军事任务分类,以平衡强对抗干扰等极端条件下网络隔离对大模型性能的影响。

军用语料安全问题涉及来源安全、内容安全和标注安全,需要在收集军事语料数据时保证来源的可追溯性,对语料内容的审核过滤时,需要建立科学且具有实用性的标注规则,突破“人在回路”的高质量数据集标注;模型安全方面涉及基座大模型安全和生成内容安全,需要选取能够离线部署的安全可控模型,确保大模型运行在受控的网络环境下,并定期对模型输出的内容安全性、军事语义倾向性进行评价分析;安全措施和评估方面涉及模型使用军事人员、军事场景等,以及大模型军事需求反馈更新、性能升级和安全性评估等,需要形成核心监管人员、健全组织模式和安全性评估机制。此外,如何将大模型技术融入到既有的军事研究方法体系,探索大智能化背景下大模型技术在多领域交叉场景中的作用,也是值得思考的问题。

4.4 军事大模型技术的发展方向

大模型技术在军事领域应用具有典型的系统集成特征,结合智能化指挥的发展趋势,军事大模型技术的发展方向主要涉及面向军事业务流程的大模型带小模型架构、基于军用语料的军事大模型训练、面向军语理解的人在回路提示工程优化等。

4.4.1 面向军事业务流程的大模型带小模型架构技术

目前采用深度神经网络架构的大模型可解释性差,军事应用所需的可靠性、可信性和安全性难以保证。尽管大模型具有强大的理解、推理、生成和学习等能力,但并不具备复杂军事任务规划、军事资源调度以及军事任务处置执行等能力。针对特定军事业务精细化训练的高性能小模型,可以具备专用业务流程的军事场景定制化能力。因此,针对军事业务流程,利用大模型的语义理解能力,抽取分解任务要素,对齐匹配业务小模型,形成大模型带小模型的协同联动架构,实现军事任务全局规划与局部过程的统一,提升大模型的可解释性和军事领域适配性。如图14所示,大模型的多模态能力可以整合各类军事情报数据,通过军语理解、要素抽取和任务分解,对齐子任务需求与PREA和OODA等业务小模型的语义描述,形成精准匹配调用、大模型+小模型的智能协同,驱动决策建议的精准内容生成,覆盖PREA过程和OODA阶段的智能化指挥链路。

4.4.2 基于军用语料的军事大模型训练技术

军事应用场景是典型的、复杂的庞大封闭系统体系,尤其需要定制化改造开源民用大模型以适应军事智能需求。然而,通用大模型的训练效率依赖于强大硬件基础设施、高质量标注语料,以及人类价值观对齐反馈,对于军事大模型训练场景难以直

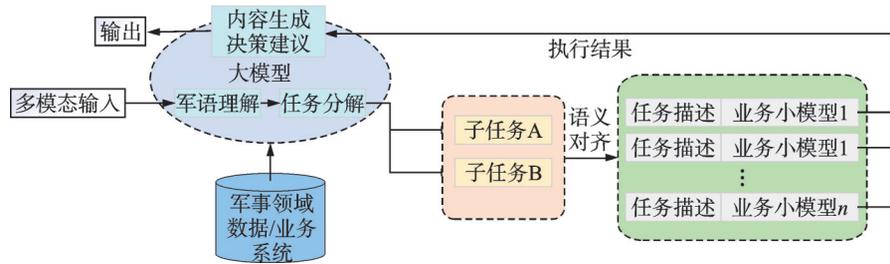


图14 大模型带小模型架构

Fig.14 Integrated architecture of foundation models with business models

接迁移。因此,以约束条件下的高质量专用语料为基础,以大模型军语理解和精准生成能力为目标,亟需形成支撑大模型军事化训练的高质量专用问答语料和方法体系,并探索基于军用语料的“二次预训练-监督微调-偏好优化反馈”的大模型军事化训练方案,需考虑如下问题:(1)如何利用语言建模整合高质量专用军事语料和部分公开通用领域问答语料资源,对大模型进行军事化二次预训练,使大模型能够预先融合更多的通用知识和军事专用知识,以便提升监督微调效果;(2)如何解决传统“预训练+参数微调”方式的灾难性遗忘、通用性弱问题,以及“预训练+提示学习”在具体军事任务上应用效果偏弱问题;(3)如何界定军事场景的偏好评价,定义待训练的生成模型和参考模型,完成偏好优化训练,实现大模型军语理解和精准建议生成能力的涌现迁移。

4.4.3 面向军语理解的人在回路提示工程优化技术

军语是通用人工智能从自然语言向军事应用拓展的关键载体,由于军语来源、类型、密级以及任务差异较大,通用大模型在军语理解上存在困难,外挂知识、人类反馈等工具可以引导大模型生成正确内容,在一定程度上缓解大模型的“幻觉”。提示工程通过提示为数据嵌入额外的上下文,可以有效地适应特定领域任务。但在优化大模型的军语理解能力时,思维链、递归提示等仅以前缀形式附加在军事任务数据上,对特定军事场景的学习能力有限。在大模型的训练和优化过程中,军事人员的参与和引导具有重要作用。人在回路的提示工程有助于大模型对军事属性的认知,对齐军事领域价值,尤其基于指挥员评价的“循环反馈提示”,可以使大模型具备自学习自优化的自我提升能力。指挥员的指令内容包含评价机制,并记录完整的语义理解、任务规划、系统执行过程以及内容生成质量,双向优化对齐情况,通过循环反馈和评价奖励,不断提高大模型的自学习自优化能力和指挥决策建议的水平,为后续能交互、能决策、能学习以及自成

长的演化奠定基础。

5 结 论

大模型技术是军事智能发展的变革性力量。本文从大模型技术出发,梳理了适用于军事智能应用场景的大语言模型技术和多模态大模型技术的脉络。一方面,总结了Transformer架构的改进方向和LLM的建模策略,分析了预训练的数据影响和策略,介绍了指令微调、参数高效微调和对齐微调等3类适配微调技术,以及上下文学习和思维链提示两类提示学习方法;另一方面,从多模态大模型分类、关键技术角度探究了典型多模态大模型技术。基于上述技术成果,对比分析了国内外典型的大模型军事应用平台的架构特点、发展成熟度,从国防安全突发事件、非战争军事行动等军事任务场景和应用需求角度,构建了涵盖系统支撑能力、基础能力和军事智能应用能力的大模型军事应用能力需求矩阵,从高质量军事语料、军事基座大模型评价和安全等方面讨论了大模型军事应用面临的挑战,以及军事大模型技术的发展方向。

总之,目前民用大模型技术发展较为成熟,以军事智能为导向、军事应用为驱动的军事大模型技术发展前景广阔,迎接军事大模型技术面临的挑战,把握大模型军事应用的发展方向,势必会助力掌握信息化智能化战争特点规律,增加新域新质作战力量比重,提升重大军事任务的智能判断决策效能。

参考文献:

- [1] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5998-6008.
- [2] DONG L, XU S, XU B. Speech-Transformer: A no-recurrence sequence-to-sequence model for speech recognition[C]//*Proceedings of 2018 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. Calgary, AB, Canada: IEEE, 2018: 5884-5888.
- [3] DEVLIN J, CHANG M W, LEE K, et al. BERT:

- Pre-training of deep bidirectional Transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis, Minnesota: Association for Computational Linguistics, 2019: 4171-4186.
- [4] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training[EB/OL].(2018-06-11)[2023-12-10]. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.
- [5] RAMESH A, PAVLOV M, GOH G, et al. DALL-E: Creating images from text[EB/OL].(2021-01-05)[2023-12-10]. <https://openai.com/research/dall-e>.
- [6] LEWIS M, LIU Y, GOYAL N, et al. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension[EB/OL].(2019-10-29)[2023-12-11]. <https://arxiv.org/abs/1910.13461>.
- [7] TOUVRON H, LAVRIL T, IZACARD G, et al. LLaMA: Open and efficient foundation language models[EB/OL].(2023-04-02)[2023-12-11]. <https://arxiv.org/abs/2302.13971>.
- [8] SUN Y, WANG S, FENG S, et al. ERNIE 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation[EB/OL].(2021-07-05)[2023-12-12]. <https://arxiv.org/abs/2107.02137>.
- [9] DU Z, QIAN Y, LIU X, et al. GLM: General language model pretraining with autoregressive blank infilling[EB/OL].(2022-03-17)[2023-12-12]. <https://arxiv.org/abs/2103.10360>.
- [10] YANG A, XIAO B, WANG B, et al. Baichuan 2: Open large-scale language models[EB/OL].(2023-09-20)[2023-12-12]. <https://arxiv.org/abs/2309.10305>.
- [11] 沈弼龙. 大模型技术的军事应用[N]. 解放军报, 2023-04-11.
- [12] DING M, YANG Z, HONG W, et al. Cogview: Mastering text-to-image generation via transformers[J]. Advances in Neural Information Processing Systems, 2021, 34: 19822-19835.
- [13] ZHANG B, SENNRICH R. Root mean square layer normalization[J]. Advances in Neural Information Processing Systems, 2019, 32: 12360-12371.
- [14] WANG H, MA S, DONG L, et al. DeepNet: Scaling Transformers to 1 000 layers[EB/OL].(2022-03-01)[2023-12-12]. <https://arxiv.org/abs/2203.00555>.
- [15] HENDRYCKS D, GIMPEL K. Gaussian error linear units (GELUs)[EB/OL].(2023-06-06)[2023-12-12]. <https://arxiv.org/abs/1606.08415>.
- [16] RAMACHANDRAN P, ZOPH B, LE Q V, et al. Searching for activation functions[EB/OL].(2017-10-27)[2023-12-12]. <https://arxiv.org/abs/1710.05941>.
- [17] WANG Y, MISHRA S, ALIPOORMOLABASHI P, et al. Super-natural instructions: Generalization via declarative instructions on 1600+ NLP tasks[C]//Proceedings of the Empirical Methods in Natural Language Processing 2022. Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022: 5085-5109.
- [18] SU J, LU Y, PAN S, et al. RoFormer: Enhanced Transformer with rotary position embedding[J]. Neurocomputing, 2024, 568: 127063.
- [19] CHOWDHERY A, NARANG S, DEVLIN J, et al. PaLM: Scaling language modeling with pathways[J]. Journal of Machine Learning Research, 2023, 24: 1-87.
- [20] LIU P, YUAN W, FU J, et al. Pre-train prompt and predict a systematic survey of prompting methods in natural language processing[J]. ACM Computing Surveys, 2022, 55: 1-35.
- [21] LIU Y, OTT M, GOYAL N, et al. RoBERTa: A robustly optimized bert pretraining approach[EB/OL].(2019-07-26)[2023-12-12]. <https://arxiv.org/abs/1907.11692>.
- [22] JURAFSKY D, MARTIN J H. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition[EB/OL].(2024-02-03)[2024-03-30]. <https://web.stanford.edu/~jurafsky/slp3/>.
- [23] RAFFEL C, SHAZEER N, ROBERTS A, et al. Exploring the limits of transfer learning with a unified text-to-text Transformer[EB/OL].(2019-10-23)[2024-03-30]. <https://arxiv.org/pdf/1910.10683.pdf>.
- [24] HOFFMANN J, BORGEAUD S, MENSCH A, et al. Training compute-optimal large language models[EB/OL].(2022-03-29)[2024-03-30]. <https://arxiv.org/abs/2203.15556>.
- [25] DU N, HUANG Y, DAI A M, et al. GLaM: Efficient scaling of language models with mixture-of-experts[C]//Proceedings of the 39th International Conference on Machine Learning. Baltimore, Maryland, USA: PMLR, 2022: 5547-5569.
- [26] RAE W, BORGEAUD S, CAI T, et al. Scaling language models: Methods, analysis & insights from training Gopher[EB/OL].(2022-03-29)[2024-03-30]. <https://arxiv.org/abs/2112.11446>.
- [27] IYER S, LIN X V, PASUNURU R, et al. Optimal: Scaling language model instruction meta learning through the lens of generalization[EB/OL].(2023-01-30)[2024-03-30]. <https://arxiv.org/abs/2212.12017>.
- [28] HONOVICH O, SCIALOM T, LEVY O, et al. Unnatural instructions: Tuning language models with (almost) no human labor[EB/OL].(2022-12-19)[2024-03-30]. <https://arxiv.org/abs/2212.09689>.
- [29] CHUNG W H, HOU L, LONGPRE S, et al. Scaling instruction-finetuned language models[EB/OL].(2022-12-06)[2024-03-30]. <https://arxiv.org/abs/>

- 2210.11416.
- [30] MUENNIGHOFF N, WANG T, SUTAWIKA L, et al. Crosslingual generalization through multitask fine-tuning[EB/OL].(2023-03-29)[2024-03-30]. <https://arxiv.org/abs/2211.01786>.
- [31] SCAO T L, FAN A, AKIKI C, et al. BLOOM: A 176b-parameter open-access multilingual language model[EB/OL].(2023-06-27)[2024-03-30]. <https://arxiv.org/abs/2211.05100>.
- [32] DING N, QIN Y, YANG G, et al. Delta tuning: A comprehensive study of parameter efficient methods for pre-trained language models[EB/OL].(2022-03-15)[2024-03-30]. <https://arxiv.org/abs/2203.06904>.
- [33] HOULSBY N, GIURGIU A, JASTRZEBSKI S, et al. Parameter-efficient transfer learning for NLP[C]//Proceedings of the 36th International Conference on Machine Learning. Long Beach, California, USA: PMLR,2019: 2790-2799.
- [34] LI X L, LIANG P. Prefix-tuning: Optimizing continuous prompts for generation[C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. [S.l.]: Association for Computational Linguistics ,2021: 4582-4597.
- [35] LESTER B, AL-RFOU R, CONSTANT N. The power of scale for parameter-efficient prompt tuning[EB/OL].(2021-09-02)[2024-03-30]. <https://arxiv.org/abs/2104.08691>.
- [36] LEE J, TANG R, LIN J. What would elsa do? freezing layers during transformer fine-tuning[EB/OL].(2019-11-08)[2024-03-30]. <https://arxiv.org/abs/1911.03090>.
- [37] ZAKEN E B, RAVFOGEL S, GOLDBERG Y. Bit-fit: Simple parameter-efficient fine-tuning for Transformer-based masked language-models[EB/OL].(2022-09-05)[2024-03-30]. <https://arxiv.org/abs/2106.10199>.
- [38] HU E J, SHEN Y, WALLIS P, et al. LoRA: Low-Rank adaptation of large language models[EB/OL].(2021-10-16)[2024-03-30]. <https://arxiv.org/abs/2106.09685>.
- [39] LIU S Y, WANG C Y, YIN H, et al. DoRA: Weight-decomposed low-rank adaptation[EB/OL].(2024-02-14)[2024-03-30]. <https://arxiv.org/abs/2402.09353>.
- [40] 夏润泽, 李丕绩. ChatGPT 大模型技术发展与应用[J]. 数据采集与处理, 2023, 38(5): 1017-1034.
- XIA Runze, LI Piji. Large language model ChatGPT: Evolution and application[J]. Journal of Data Acquisition and Processing, 2023,38(5): 1017-1034.
- [41] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901.
- [42] WEI J, WANG X, SCHUURMANS D, et al. Chain-of-thought prompting elicits reasoning in large language models[J]. Advances in Neural Information Processing Systems, 2022, 35: 24824-24837.
- [43] MIN S, LEWIS M, ZETTLEMOYER L, et al. Metaicl: Learning to learn in context[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle, United States: Association for Computational Linguistics, 2022: 2791-2809.
- [44] SHIN S, LEE S, AHN H, et al. On the effect of pre-training corpora on in-context learning by a large-scale language model[C]//Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Seattle, United States: Association for Computational Linguistics,2022: 5168-5186.
- [45] LI Y, LIN Z, ZHANG S, et al. Making large language models better reasoners with step-aware verifier[EB/OL].(2023-03-24)[2024-03-30]. <https://arxiv.org/abs/2206.02336>.
- [46] FU Y, PENG H, SABHARWAL A, et al. Complexity-based prompting for multi-step reasoning[EB/OL].(2023-01-30)[2024-03-30]. <https://arxiv.org/abs/2210.00720>.
- [47] ZHANG Z, ZHANG Z, LI M, et al. Automatic chain of thought prompting in large language models[EB/OL].(2022-10-07)[2024-03-30]. <https://arxiv.org/abs/2210.03493>.
- [48] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16*16 words: Transformers for image recognition at scale[C]//Proceedings of the 9th International Conference on Learning Representations. Austria: OpenReview.net,2021: 1-22.
- [49] SU W, ZHU X, CAO Y, et al. VL-BERT: Pre-training of generic visual-linguistic representations[EB/OL].(2020-02-18)[2024-03-30]. <https://arxiv.org/abs/1908.08530v4>.
- [50] CHEN Y C, LI L, YU L, et al. Uniter: Universal image-text representation learning[C]//Proceedings of Computer Vision—ECCV 2020: 16th European Conference. Glasgow, UK: Springer, 2020: 104-120.
- [51] LU J, BATRA D, PARIKH D, et al. Vilbert: Pre-training task-agnostic visiolinguistic representations for vision-and-language tasks[J]. Advances in Neural Information Processing Systems, 2019, 32: 1-11.
- [52] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//Proceedings of the 38th International Conference on Machine Learning. Baltimore, Maryland, USA:PMLR,2021: 8748-8763.
- [53] WANG W, LV Q, YU W, et al. CogVLM: Visual

- expert for pretrained language models[EB/OL]. (2024-02-04) [2024-03-30]. <https://arxiv.org/abs/2311.03079>.
- [54] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2022: 10684-10695.
- [55] CHO J, LEI J, TAN H, et al. Unifying vision-and-language tasks via text generation[C]//Proceedings of the 38th International Conference on Machine Learning. Baltimore, Maryland, USA: PMLR, 2021: 1931-1942.
- [56] ZHOU L, PALANGI H, ZHANG L, et al. Unified vision-language pre-training for image captioning and VQA[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 13041-13049.
- [57] BROOKS T, PEEBLES B, HOLMES C, et al. Video generation models as world simulators[EB/OL]. (2024-02-15) [2024-03-30]. <https://openai.com/research/video-generation-models-as-world-simulators>.
- [58] YIN S, FU C, ZHAO S, et al. A survey on multimodal large language models[EB/OL]. (2023-06-23) [2024-03-30]. <https://arxiv.org/abs/2306.13549>.
- [59] LIU H, LI C, WU Q, et al. Visual instruction tuning[EB/OL]. (2023-12-11) [2024-03-30]. <https://arxiv.org/abs/2304.08485>.
- [60] DONG Q, LI L, DAI D, et al. A survey on in-context learning[EB/OL]. (2023-06-01) [2024-03-30]. <https://arxiv.org/abs/2301.00234>.
- [61] ALAYRAC JB, DONAHUE J, LUC P, et al. Flamingo: A visual language model for few-shot learning[J]. Advances in Neural Information Processing Systems, 2022, 35: 23716-23736.
- [62] LI J, LI D, SAVARESE S, et al. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models[EB/OL]. (2023-06-15) [2024-03-30]. <https://arxiv.org/abs/2301.12597>.
- [63] LU P, MISHRA S, XIA T, et al. Learn to explain: Multimodal reasoning via thought chains for science question answering[J]. Advances in Neural Information Processing Systems, 2022, 35: 2507-2521.
- [64] GUPTA T, KEMBHAVI A. Visual programming: Compositional visual reasoning without training[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2023: 14953-14962.
- [65] HARPER J. Pentagon testing generative AI in "global information dominance" experiments[EB/OL]. (2023-07-14) [2024-03-30]. <https://defensescoop.com/2023/07/14/pentagon-testing-generative-ai-in-global-information-dominance-experiments/>.
- [66] JENSEN B, TADROSS D. How large-language models can revolutionize military planning[EB/OL]. (2023-04-14) [2024-03-30]. <https://warontherocks.com/2023/04/how-large-language-models-can-revolutionize-military-planning/>.
- [67] 阳东升, 卢经纬, 李强, 等. 超大预训练模型在指挥控制领域的应用与挑战[J]. 指挥与控制学报, 2023, 9(2): 146-155.
- YANG Dongsheng, LU Jingwei, LI Qiang, et al. Issues and challenges of ChatGPT-like large scale pre-trained model for command and control[J]. Journal of Command and Control, 2023, 9(2): 146-155.
- [68] 郭忠伟. 作战文书自动生成理论及方法研究[D]. 南京: 南京理工大学, 2003.
- GUO Zhongwei. Research on the theory and method of automatic generation of combat documents[D]. Nanjing: Nanjing University of Science and Technology, 2003.
- [69] SAFETY T. Disclosing state-linked information operations we've removed[EB/OL]. (2023-12-02) [2024-03-30]. https://blog.twitter.com/en_us/topics/company/2021/disclosing-state-linked-information-operations-we-ve-removed.
- [70] 陈昌孝, 李浩, 王梓晗, 等. 认知域作战新利器——ChatGPT 认知剖析及对策[J]. 军事文摘, 2023, 11: 28-32.
- CHEN Changxiao, LI Hao, WANG Zhihan, et al. A new weapon for cognitive domain warfare: ChatGPT cognitive analysis and countermeasures[J]. Military Literature, 2023, 11: 28-32.
- [71] NEMKOVA P, UBANI S, OLCAY P, et al. Detecting human rights violations on social media during Russia-Ukraine War[EB/OL]. (2023-06-06) [2024-03-30]. <https://arxiv.org/abs/2306.05370>.
- [72] ZENG Wei, REN Xiaozhe, SU Teng, et al. PanGu- α : Large-scale autoregressive pretrained chinese language models with auto-parallel computation[EB/OL]. (2021-04-26) [2024-03-30]. <https://arxiv.org/abs/2104.12369>.
- [73] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901.
- [74] BaichuanAI. Baichuan-7B[EB/OL]. (2023-09-06) [2024-03-30]. <https://github.com/baichuan-inc/Baichuan-7B>.
- [75] BaichuanAI. Baichuan-13B[EB/OL]. (2023-09-06) [2024-03-30]. <https://github.com/baichuan-inc/Baichuan-13B>.
- [76] AlibabaCloud. QWEN[EB/OL]. (2023-11-30) [2024-03-30]. <https://github.com/QwenLM/Qwen>.
- [77] Meta-llama. LLaMA-3[EB/OL]. (2024-04-18) [2024-03-30]. <https://github.com/meta-llama/llama3>.