

Research on the Application of Lightweight LLMS in Real-time Military Decision-Making

Xiaolong Wang Qinfu Li Xiuquan Yu* Jun Cai

Chinese Academy of Electronic Sciences, Beijing, 100041, China

Abstract

In today's era of rapid development of information technology, the military field has higher requirements for the precision and efficiency of real-time decision-making. With its unique advantages, lightweight large models have demonstrated great application potential in real-time military decision-making. This article delves deeply into the core technologies and advantages of lightweight large models, as well as their application scenarios in real-time military decision-making, such as intelligence analysis and situation awareness, generation and evaluation of combat plans, command and control, and coordinated operations. It also analyzes the challenges faced during the application process, including issues related to data quality and security, model performance and reliability, and adaptation to military applications. Corresponding countermeasures are proposed with the aim of providing theoretical support and practical references for promoting the intelligent development of military decision-making.

Keywords

Lightweight Large Model; Real time military decision-making; Model compression; Model optimization; multimodal

轻量化大模型在实时军事决策中的应用研究

王晓龙 李钦富 于修全* 蔡俊

中国电子科学研究院, 中国·北京 100041

摘要

在信息技术迅猛发展的今天, 军事领域对实时决策的精确性和效率性要求更高。凭借独特的优势, 轻量化大模型在实时军事决策中显示出了巨大的应用潜力。本文深入探讨了轻量化大模型的核心技术、优势, 及其在实时军事决策中情报分析与态势感知、作战方案生成与评估、指挥控制与协同作战等应用场景, 分析了应用过程中面临的数据质量与安全问题、模型性能与可靠性问题、军事应用适配问题等挑战, 并提出了相应的应对策略, 旨在为推动军事决策智能化发展提供理论支持和实践参考。

关键词

轻量化大模型; 实时军事决策; 模型压缩; 模型优化; 多模态

1 引言

在现代战争中, 战场信息呈现多源异构、高动态等特点。军事决策者需要在极短的时间内对这些信息进行分析、判断并做出科学合理的决策, 以掌握战争主动权。传统的军事决策方法在面对如此复杂和快速变化的信息时, 往往难以满足实时性和准确性的要求。

大模型技术的出现为军事决策带来了新的思路^{[1][3]}, 但由于军事应用场景的特殊性, 对模型的部署和运行效率有

严格限制^{[2][6]}, 轻量化大模型应运而生。轻量化大模型在保持强大语言理解和生成能力的同时, 通过优化模型结构、参数压缩等技术手段, 降低了模型的计算量和存储需求, 使其能够在资源有限的军事装备上高效运行, 为实时军事决策提供有力支持。

2 轻量化大模型概述

2.1 轻量化大模型核心技术

2.1.1 模型压缩与优化

参数剪枝: 通过移除冗余参数、对性能影响较小的连接或神经元, 减少模型体积(参数数量)、提升推理速度, 适用于战场终端设备的部署^[4]。

模型优化: 采用更简洁、高效的神经网络架构, 如 MobileNet、ShuffleNet 等轻量级卷积神经网络结构, 减少卷积层的参数数量、优化卷积核的大小和连接方式等, 在降低模型复杂度的同时, 保证一定的精度。

【作者简介】 王晓龙(1989-), 男, 中国天津人, 硕士, 助理工程师, 从事智能决策系统设计、分布式计算、大模型技术应用等研究。

【通讯作者】 于修全(1978-), 男, 中国黑龙江青冈人, 本科, 高级工程师, 从事云计算、大数据、人工智能等研究。

量化技术：采用 LLM.int8() 混合精度量化方法，将模型参数从 32 位浮点压缩至 8 位整数，降低内存占用、加速计算；同时，通过 SmoothQuant 方法降低异常值对精度的影响。

知识蒸馏：以教师模型为指导，训练学生模型，将教师模型的知识迁移至学生模型，保留关键推理能力，将多模态大模型 (MLLMs) 压缩为可部署于无人机的轻量版本。^{[3][6]}

2.1.2 多模态信息处理优化

视觉 Token 压缩：通过多尺度信息融合或动态 Token 选择，减少图像与视频数据的计算负载。

高效架构设计：采用 Mamba 结构替代传统 Transformer，利用状态空间模型 (SSM) 高效处理长序列数据。

2.2 轻量化大模型优势

低资源消耗：轻量化大模型所需的计算资源和存储资源大幅减少，能够在便携式军事设备、无人机、车载终端等资源受限的硬件平台上运行，以满足机动、灵活等军事应用的要求。^[4]

快速推理：由于模型复杂度降低，推理过程中的计算量减少，使得模型能够快速对输入数据进行处理和分析，输出决策建议，满足实时军事决策对时间的严格要求。^[5]

适应性强：可以根据不同的军事应用场景和硬件条件，灵活调整模型的轻量化程度，在保证决策准确性的基础上，实现最优的资源利用和性能表现。

3 轻量化大模型在实时军事决策中的应用场景

3.1 情报分析与态势感知

多源情报融合分析：轻量化大模型可以对卫星侦察、无人机侦察、地面传感器、开源情报等多源异构情报进行快速融合和分析，提取关键信息。通过对情报的深度理解和关联分析，为军事决策者提供全面、准确的战场态势图^[6]。

异常行为检测：轻量化大模型可以对海量的战场数据进行实时监测，识别出异常的军事活动模式，支持及时发现异常行为，为军事决策者提前预警、采取相应应对措施提供支持。

3.2 作战方案生成与评估

作战方案快速生成：轻量化大模型通过学习大量的历史战例和军事理论知识，结合当前战场态势和作战目标，运用智能算法生成包含兵力运用、火力分配、基本战法等关键要素的作战方案，为军事决策者提供丰富的决策选项^[6]。

作战方案评估与优化：轻量化大模型从多个维度对生成的作战方案进行作战效能、风险程度、资源消耗等方面评估，支持通过自然语言指令快速构建虚拟战场、模拟战场态势、分析其优劣势，为军事决策者提供动态决策依据。

3.3 指挥控制与协同作战

实时指挥决策辅助：轻量化大模型可以作为作战过程中的实时指挥决策辅助工具^[7]，为军事指挥官提供实时的战场信息分析、决策建议和风险提示。

多军兵种协同作战：轻量化大模型支持构建统一的作战信息交互平台，实现各军兵种之间的信息快速传递和理

解，优化协同作战流程；同时，支持对各军兵种作战任务和行动的智能规划与协调，提高多军兵种协同作战的效率和效果。

4 面临挑战与对策

4.1 数据质量与安全问题

4.1.1 面临挑战

数据质量：战场数据获取受到战场环境、传感器性能等因素影响，数据可能存在噪声、缺失、错误等问题；低质量的数据会影响轻量化大模型训练效果和决策准确性，如何进行有效的预处理和质量控制是一个关键问题。

数据安全：在数据使用全流程中，确保作战计划、军事装备参数等敏感数据的安全性和保密性至关重要。防止数据泄露、篡改和恶意攻击，是轻量化大模型在军事应用中面临的重要安全挑战。^{[2][6]}

4.1.2 应对策略

提升数据质量：围绕数据清洗、去噪、填补缺失值等方面，建立机制手段和预处理流程，提升军事数据质量；利用数据增强技术对图像数据进行旋转、缩放、裁剪等操作，扩充数据量、增强模型泛化能力。

确保数据安全：采用加密技术进行军事数据存储和传输，确保数据保密；建立形成数据访问权限管理机制；部署网络安全防护，部署防火墙、入侵检测系统等安全设备，防止数据遭受外部攻击。

4.2 模型性能与可靠性问题

4.2.1 面临挑战

模型性能：在轻量化过程中，如何在降低模型复杂度的同时，保证模型的准确性和泛化能力，实现模型性能与资源消耗的最佳平衡，是一个技术难题。过度轻量化可能导致模型性能大幅下降，无法满足军事决策的精度要求，复杂度较高又会影响受限资源设备的运行效率。^[6]

可靠性验证：军事决策对模型的可靠性要求极高，任何决策失误都可能导致严重后果。然而，目前对于轻量化大模型的可靠性验证方法还不够完善，如何建立科学有效的模型可靠性评估体系，确保模型在复杂多变的战场环境下能够稳定、可靠地运行，是应用过程中需要解决的重要问题。

4.2.2 应对策略

优化模型性能：结合参数压缩、知识蒸馏等多种轻量化技术，实现模型性能与资源消耗的平衡；采用自适应学习率调整、正则化等训练技巧，提高模型训练效果和泛化能力；优化模型推理算法，提高推理效率。

验证可靠性：建立包括模型准确性、稳定性、鲁棒性等多维度的模型可靠性评估指标体系。采用模拟仿真等方式对模型进行验证和测试，通过大量的实验数据评估模型在不同场景下的可靠性。

4.3 军事应用适配问题

4.3.1 面临挑战

军事场景复杂性：战场环境具有高度的复杂性和不确

定性，与模型训练时所使用的模拟场景存在较大差异。如何使轻量化大模型更好地适应真实的军事场景，准确理解和处理各种复杂的战场信息，是应用过程中面临的一大挑战。^[9]

军事人员接受度：轻量化大模型应用于军事决策，部分军事人员对新技术存在疑虑或不熟悉，如何提高军事人员对轻量化大模型的接受度和使用能力也是一个关键问题。

4.3.2 应对策略

适配军事应用：加载真实战场数据，对轻量化大模型进行针对性训练和优化，使其更好地适应军事场景的复杂性和不确定性。利用强化学习技术让模型不断学习和进化，提高其应对复杂情况的能力。

培训人员：开展军事人员的轻量化大模型技术培训课程，了解基本原理、功能和使用方法。通过案例分析、模拟操作等方式，提高军事人员运用模型进行辅助决策的能力。

5 典型案例^[8]

5.1 边端部署的指挥决策系统

应用场景：在通信受限或网络不稳定的战场环境中，轻量化大模型支持离线部署，帮助指挥官实时分析情报并生成行动方案。

案例：唐山海事局本地化部署 DeepSeek-R1 大模型，在无网络条件下实现海事应急决策支持，军事领域可嵌入前线指挥车或单兵设备中，类比应用生成作战路线规划建议等。

5.2 智能后勤与供应链优化

应用场景：实时优化物资配送路线，预测装备故障，确保前线资源的高效调度。

案例：美国陆军采用 Scale Donovan 大语言模型进行后勤管理，通过轻量化部署在加密网络中，快速分析部队移动数据并生成补给方案。

5.3 战场医疗分诊与健康监测

应用场景：在野战医疗场景中，轻量化模型可辅助快速诊断伤员病情并推荐治疗方案。

案例：美国空军研究实验室主导开发轻量化医疗分诊模型，并在智能眼镜、手表等穿戴设备搭载轻量化模型监测士兵健康状态，通过分析战场实时数据，提供优先级分诊建议，缩短救治时间。

6 未来展望

轻量化大模型在实时军事决策中的应用未来可聚焦以下发展方向。

6.1 技术优化方面^[4]

联合优化技术深入发展：采用联合优化技术，综合考虑模型精度、计算复杂度、内存占用、推理速度等因素，基于加权求和、Pareto 前沿、强化学习等方法，让轻量化大模型在资源受限的军事设备上能更加高效地运行。

与边缘计算等融合深化：与边缘计算、雾计算等技术

进行深度融合，将轻量化大模型部署靠近数据源的边缘节点，减少数据传输延迟、实现快速实时决策。

6.2 应用拓展方面^[4]

多域作战融合：轻量化大模型将更广泛地应用于跨域态势感知、多域任务规划与协同等，实现各作战域之间的高效信息共享和协同决策。

无人作战系统集成化：轻量化大模型将支持大规模无人集群的自主协同作战，实现无人车、机、艇等在复杂战场环境下自主编队与任务分配、避障与目标打击等。

6.3 性能提升方面^[6]

可解释性增强：发展可解释性技术，使轻量化大模型的决策过程和依据更加透明，便于军事人员理解和信任模型的输出结果，将模型决策与人类决策更好地进行结合。

实时自适应能力提升：具备更强的实时自适应能力，根据战场动态变化、任务调整以及敌方策略改变等，支持实时自动调整模型参数和决策策略，确保决策性能最佳。

7 结论

轻量化大模型通过技术创新与场景适配，在实时军事决策中具有广阔的应用前景，能够有效提升军事决策的准确性和时效性，为现代战争的胜利提供有力支持。

尽管在应用过程中面临数据质量与安全、模型性能与可靠性、军事应用适配等诸多挑战，但通过采取相应的解决策略，可以逐步克服这些困难，推动轻量化大模型在军事领域的广泛应用和深入发展。

未来，随着轻量化大模型的技术优化、应用拓展和性能提升，轻量化大模型在军事决策中的应用将更加成熟和完善，也必将促进军事智能化发展。

参考文献

- [1] 康强,罗晓泽,马进兆. 优化建模在军事航海实时模拟训练系统中的应用研究[C]//第四届全国虚拟现实与可视化学术会议论文集.大连:大连海事大学出版社, 2004:34-38.
- [2] 姚雨,宋春林,邵江琦. 无人机航拍军事车辆实时检测及定位算法[J]. 兵工学报,2024,45(z1):354-360.
- [3] 齐大伟,贺筱媛,胡晓峰,等. 军事通信路径规划动态模型研究[J]. 指挥与控制学报,2015,1(2):203-207.
- [4] 蔡磊,孟宪波,韩冬梅,等. 大模型在军事垂直领域的应用[J]. 舰船科学技术,2024,46(5):171-175.
- [5] 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.
- [6] 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.
- [7] 姚奕,陈朝阳,杜晓明,等. 多模态知识图谱构建技术及其在军事领域的应用综述[J]. 计算机工程与应用,2024,60(22):18-37.
- [8] 董华辉,马晓. ChatGPT在军事领域的潜在应用分析[J]. 舰船电子工程,2024,44(5):14-17,54. [9] 廖伟,王翔宇,李宏伟. 人工智能在军事领域的应用综述[J]. 智能安全,2024,3(1):106-112.