计算机科学
COMPUTER SCIENCE

基于预训练大模型的行动方案生成方法

颜玉松, 周圆, 王琮, 孔圣麒, 王权, 黎敏讷, 王之元

引用本文

颜玉松, 周圆, 王琮, 孔圣麒, 王权, 黎敏讷, 王之元. 基于预训练大模型的行动方案生成方法[J]. 计算机科学, 2025, 52(1): 80-86.

YAN Yusong, ZHOU Yuan, WANG Cong, KONG Shengqi, WANG Quan, LI Minne, WANG Zhiyuan. COA Generation Based on Pre-trained Large Language Models [J]. Computer Science, 2025, 52(1): 80-86.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240900075

基于预训练大模型的行动方案生成方法

颜玉松[1] 周 圆[2] 王 琮[2] 孔圣麒[1] 王 权[2] 黎敏讷[2] 王之元[2]
1 国防科技大学计算机学院 长沙 410005
2 智能博弈与决策实验室 北京 100000
(yanyusong23@nudt.edu.cn)

摘 要 围绕生成式人工智能赋能指挥决策需求,分析了指挥决策中方案生成问题的难点挑战和新兴预训练大语言模型技术的应用前景,提出了一种基于预训练大模型的作战行动方案生成方法——COA-Gen。首先,为了使生成的行动方案符合目标,设计了多轮方案生成框架;其次,构建了多要素中文提示词模板用于整合海量多源信息;最后,针对特定小领域的数据缺乏问题,引入知识增强技术以提升大模型规划效能。为了验证所提行动方案的效果,制定了基于《星际争霸II》游戏引擎和"虎爪"想定的方案验证环境。实验结果表明,该方法具有较好的鲁棒性,可以较好地依从指挥员意图,验证了大模型用于作战行动方案生成的可行性。此外,不同预训练大模型在相同任务中展现出不同的效果,表明在实际应用中选择不同的预训练大模型可能会生成具有不同风格的行动方案,从而影响最终的行动结果。

关键词:大模型;生成式人工智能;智能决策;指挥与控制;作战行动方案
中图分类号 TP399

COA Generation Based on Pre-trained Large Language Models

YAN Yusong[1],ZHOU Yuan[2],WANG Cong[2],KONG Shengqi[1],WANG Quan[2],LI Minne[2] and WANG Zhiyuan[2]
1 College of Computer,National University of Defense Technology,Changsha 410005,China
2 Intelligent Game and Decision Lab,Beijing 100000,China

Abstract Focusing on empowering the command and control(C2) procedure of generative AI,we analyze the challenges of course of action(COA) generation in C2 and the prospects of pre-trained large language models(LLMs). Then,a COA generation method based on pre-trained LLMs,COA-Gen,is proposed. Firstly,a multi-round generation framework is designed to align the generated plans with objectives. Secondly,a multi-factor prompt templates is constructed to integrate vast amounts of multi-source information. Lastly,knowledge-augmented generation technology is introduced to improve the generation quality of the few-shot military domain. To validate the effectiveness of the generated plans,an emulation environment based on the StarCraft II engine and the "Tiger Claw" scenario is established. The results show the robustness of the method and its alignment with the commander's intention. The feasibility of using LLMs for COA generation has been verified. Additionally,different pre-trained models exhibit varying performances in the same task,indicating that the choice of model in real-world applications can lead to action plans with different styles,thereby affect the ultimate outcomes.

Keywords Large language model,Generative AI,Intelligent decision-making,Command and control,Course of action

1 引言

军事行动方案(Course of Action,COA)生成与规划,主要研究如何构建有效的行动方案,通过有序、高效地执行一系列任务达到某一既定目标[1]。现代战争是包含陆、海、空、天、电磁等在内的多域一体化作战,战场要素日趋复杂,战场态势实时变化。如何从不确定环境下的海量数据中高效筛选出有用信息、敏捷制定行动方案是指挥员面临的一大挑战。

随着计算机技术和人工智能的发展,涌现出了多种方法来辅助指挥员解决这一难题。传统方法基于规则生成行动方案,以专家编写的规则集为依据指导系统生成行动方案,虽然在特定任务上表现良好,但在面对复杂、不确定性较高的环境时,规则的编写和维护成本较高且难以覆盖所有情况;基于强化学习(Reinforcement Learning,RL)的方法可以让智能系统通过"试错"来学习最优的行动策略[2],在复杂环境下表现出色,但模型训练收敛难度随问题规模的增加呈指数递增;基于

到稿日期:2024-09-12 返修日期:204-10-14
基金项目:国家自然科学基金青年科学基金(62102442);国家自然科学基金(62402500)
This work was supported by the Young Scientists Fund of the National Natural Science Foundation of China(62102442) and National Natural Science Foundation of China(62402500).
通信作者:周圆(yuaanzhou@outlook.com)

生成对抗网络（Generative Adversarial Network，GAN）和变分自编码器（Variational Autoencoder，VAE）等机器学习模型的生成式方法的优势在于生成的方案具有一定的创造性和灵活性，但是其训练较为复杂，同时需要克服模式坍缩带来的样本多样性等问题，面临落地挑战[3-4]。

预训练大语言模型（Pre-trained Large Language Model，下文简称 LLM）的兴起标志着人工智能的重大进步[5-8]。预训练大模型在各领域的成功应用，为智能决策和规划带来了新的思路和方法。预训练大模型基于超大参数规模的神经网络模型架构学习大规模数据中的语义信息和模式，具备语义理解和生成文本的通用基础能力，并可通过结合微调、人机对齐、思维链、反思机制、检索增强等技术[9-11]进行能力增强，其在行动方案生成领域的潜力逐渐凸显。

本文通过整合预训练大模型自身的语义理解和文本生成能力，构建了一个多轮行动方案生成框架，并设计了一种包含多要素的中文提示词模板。进一步地，结合知识增强技术，提出了 COA-Gen 方法，该方法能够实现作战行动方案的生成与规划。在即时策略游戏《星际争霸Ⅱ》的“虎爪”想定场景中，通过多个实验验证了 COA-Gen 方法的有效性和鲁棒性。

2 相关工作

2.1 军事行动方案生成与规划

STRIPS（Stanford research Institute Problem Solver）[12]是用于军事行动方案生成与规划的经典框架，其通过形式化的逻辑语言来描述环境的初始状态、目标状态以及一组操作，并生成从初始状态到目标状态的动作序列，即行动方案。虽然 STRIPS 为复杂的军事行动方案生成与规划提供了严谨的理论基础，但仍存在复杂度增加、难以描述时间、资源等不足等问题。此外，层次任务网络（Hierarchical Task Network，HTN）[13]通过将复杂的任务分解为更小的、可管理的子任务来生成和规划行动方案。它首先定义军事任务的总体目标，接着将其逐层分解为更具体的任务，这一过程会一直持续直至得到可由基本单位执行的基本行动指令。HTN 可以系统化地处理多层次的复杂任务，增强方案的灵活性和鲁棒性，但其复杂性和计算成本随着任务层次结构的增加而显著增加，难以实现复杂的现代战争背景下的军事行动方案的高效生成与规划。同时，HTN 方法依赖于专家知识来定义和分解任务，这对指挥与控制人员的专业水平提出了更高的要求。

尽管 STRIPS 和 HTN 等经典规划方法在处理结构化问题时表现出色，但难以适应具有高复杂性、不确定性和大规模信息等显著特征的现代军事场景。对此，越来越多的研究者开始探索人工智能方法在方案生成与规划方面的应用。Sarcia[14]指出，机器学习和人工神经网络的方法经过大规模数据训练后可以显著提升军事行动方案的生成与规划效率。Schwartz 等[15]深入研究了人工智能方法协助指挥与控制人员迅速制定和优化多样化的行动方案，以应对现代战场的复杂性。

2.2 预训练大模型

预训练大模型通常指参数量在十亿及以上的大语言模型。依托于大规模文本语料库与自监督预训练技术，LLM 展现出强大的文本语言表达普适性，并在应对复杂任务挑战时展现出卓越的性能[16]。此外，鉴于视觉、音频等多模态信息在理解世界方面同样至关重要，越来越多的多模态预训练大模型[17]受到广泛关注。

根据模型的开放性，LLM 主要分为两类：闭源大模型和开源大模型。其中，闭源大模型指训练数据、模型架构等具体实现细节不公开的大模型。以 OpenAI 推出的 GPT 系列[5]为代表的模型获得了最为广泛的认可，其中 GPT-3.5 和 GPT-4 模型以其强大的性能表现成为当今众多领域研究者的首选 LLM。与闭源大模型相反，开源大模型向公众公开模型架构、训练数据、参数等模型细节，并允许研究人员和开发者自由访问、修改和微调。当前，深受研究者青睐的开源大模型有 Qwen 系列[6]、LLaMA 系列[7]、GLM 系列[8]等。

通常，开源大模型具有更强的可定制性，在数据隐私与保护方面展现出更大的安全优势，闭源大模型则往往表现出更大的性能优势。值得注意的是，由于开源社区的蓬勃发展，两者的性能差距正在不断缩小，开源大模型甚至在某些领域实现了反超。

2.3 基于预训练大模型的决策与规划

基于思维链、思维树、少样本提示、反思机制等技术，预训练大模型已被证明在多种任务场景中能够表现出强大的生成、推理、规划和决策能力。其主要特性如下：

1）海量多源数据处理：随着战争科技飞速进步，未来战场中包含的数据信息会更加多元复杂，其来源将不仅限于传统的空中、陆地和海上，还将扩展至太空、网络空间等新兴领域。预训练大模型能够有效整合和分析来自不同信源、不同类型的文本、图像、视频、音频等大量数据，提供更全面和及时的数据支持。

2）敏捷响应临机更新：行动方案的制定需要依靠具有专业知识和丰富经验的行业专家，这将带来高额的成本。同时，在真实战场中抓住稍纵即逝的机会往往可以给予对方沉重的打击，具有比对方更快作出有效行动策略的决策能力至关重要。大模型本身具有的海量训练知识与高效的推理机制可有效支撑方案精准有效前提下的敏捷响应和临机更新。

3）人机对齐迭代优化：预训练大模型通常具有对话能力，可实现人在回路中的方案迭代优化。在对话交互过程中，指挥与控制人员扮演评论家的角色，不断向预训练大模型提出改进建议和需求。同时，大模型会根据这些建议进一步优化生成的行动方案，最终生成令人类满意的结果。

考虑到上述预训练大模型在方案生成与规划方面的优势，越来越多的研究者开始在多种任务场景中研究预训练大模型的方案生成与规划能力。在《我的世界》游戏中，Wang 等[18]基于 GPT-4 设计了一种名为 Voyager 的具身智能体。Voyager 利用自身的自动课程、技能库和迭代提示机制等能力，连续生成和规划当前状态下最佳的行动方案并执行，成功完成了解锁技能树、自主探索世界等任务。在家庭任务场景中，Ahn 等[19]设计了 SayCan 框架，利用 LLM 和可供性函数共同生成和规划行动方案以完成人类设定的任务目标，同时确保机器人与世界环境对齐以及减轻“幻觉”；在军事方面，Lamparth 等[20]在模拟真实战争场景下，从多个角度分析和

验证预训练大模型等 AI 技术应用于军事战略决策的可行性，表明大模型生成与规划的方案与人类决策具有较高一致性。Goecks 等[21]提出了 COA-GPT 框架，通过分析战场环境、友方和敌方单位部署等信息，结合指挥与控制人员设定的任务目标，高效地生成行动方案。

综上，预训练大模型在方案生成与规划方面展现出了强大的潜力，可为用户快速生成行动方案提供参考。

3 基于大模型的行动方案生成方法

本文提出的 COA-Gen 方法由 4 部分组成：多轮行动方案框架、多要素提示词、领域知识增强以及行动方案验证平台。该方法确保生成的军事行动计划既具备可执行性又遵循用户指令。

3.1 多轮行动方案框架

以 COA-GPT 为例，现有的基于 LLM 的行动方案生成方法以单轮框架为主。如图 1 所示，即对于军事目标只做一次行动方案生成，并基于该方案执行相关动作。无论目标是否达成，大模型都不再制定后续方案。在实际军事行动过程中，执行单轮方案会面临两个问题：1）单轮方案基于初始状态制定军事行动计划，无法确保达成目标；2）战场情况瞬息万变，基于初始状态的单轮方案无法预知这些信息，不能有效捕捉战场的实时概况。因此，本文提出多轮方案框架。

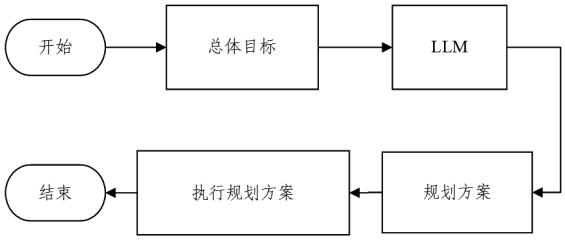


图 1 单轮方案框架

Fig. 1 Framework of single-round scheme generation

整体框架如图 2 所示，即基于目标制定行动方案，基于该方案执行相关动作。动作执行完毕后观察目标是否达成，如果目标已经达成，则无后续动作。如果目标未达成，则继续下一轮方案生成，以此类推，直至达成总体目标。

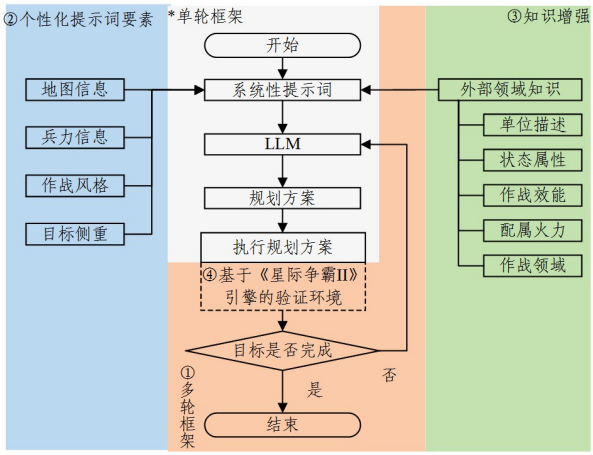


图 2 多轮方案框架

Fig. 2 Framework of multi-round scheme generation

3.2 多要素提示词

大模型通过预训练获得世界知识，其本身是历史世界知识的聚合体，自身具备一定的规划能力。但是，这种能力是隐式的，需要特定的提示词进行激发。大语言模型接收的提示词不同会导致输出结果存在巨大差异。因此，应设计合理有效的提示词，确保大语言模型跟随用户意图，输出符合预期的结果。同时，考虑到用户的语言习惯，本文设计的提示词模板以中文进行交互，极大地便捷了指挥员的使用。

整体上，提示词由两部分组成：系统性提示词和个性化提示词。系统性提示词作为固定的内容在每轮规划开始时拼接到个性化提示词前共同作为 LLM 的输入，其中描述了 LLM 功能、输出命令格式规定等内容，如图 3 所示。个性化提示词则与当前具体方案生成环境相关，描述了当前态势、任务等内容。由于现实中的军事行动方案是基于多源信息制定，主要包括地图信息、兵力信息、作战风格以及目标侧重等，因此，个性化提示词必须囊括以上信息。通过分析基于这些输入信息生成的军事行动计划可以评估大语言模型遵循提示词的指令能力，从而进一步证明大语言模型在生成军事行动计划方面的能力。图 2 给出了使用地图信息、兵力信息、作战风格以及目标侧重信息扩充提示词后的多轮方案框架。

您是一名军事指挥官助手。您的用户是军事指挥官，您的主要职责是帮助他们制定军事行动计划。
军事指挥官会在您开始制定军事行动计划之前，告知您任务目标、地形信息以及可用的友军和敌军资产。
根据这些信息，您将制定若干行动方案（由指挥官指定），以便他们能够与您一起对其进行反复推敲，并选择他们最喜欢的方案。
为了确保行动计划的完整性，需要您对每一个友军单位分配指令，不对敌军单位分配指令，可选择的指令有两种，attack_move_unit和engage_target_unit，指令的描述如下所示：
1) attack_move_unit(unit_id, target_x, target_y):
attack_move_unit表示指挥友军单位unit_id，移动到坐标(target_x, target_y)的位置，如果沿途遇到敌军，需要和敌军交战。
2) engage_target_unit(unit_id, target_id,target_x, target_y):
engage_target_unit表示指挥友军单位unit_id与敌军单位target_id进行交战，如果敌军单位超出友军射程，友军单位将在交战前移动到敌军单位(target_x, target_y)的位置。
请您牢记，一定要为每一个友军单位分配指令(attack_move_unit或engage_target_unit)，不为敌军单位分配指令。生成的计划以JSON格式输出，下面是一个具体的例子：
{**此处填写希望LLM输出的行动方案的格式样例，包含两个示例单位**},
在上述的JSON例子中，只模拟为两个友军单位分配指令，在实际情况中，需要根据友军单位的数量，为每一个友军单位分配指令。

图 3 系统性提示词

Fig. 3 System prompt

3.3 知识增强生成

大模型的规划能力通常“涌现”自大规模通用数据集，对于军事这一特定领域，训练数据的缺乏将导致数据计划中存在“模型幻觉”的风险。因此，研究者们提出使用领域知识增强生成（Knowledge-Augmented Generation，KAG）技术[22]来提升生成方案质量。KAG 技术通过检索外部领域知识，弥补大模型在军事领域的不足，以此减少模型幻觉，生成更合理的军事行动计划。本文中，

将包括单位描述、状态属性、作战效能、配属火力的作战领域外部领域知识作为背景知识注入提示词中，参与到军事行动计划制定的过程中（见图 2），相关提示词模板如图 4 所示。

我需要制定一个单一的军事行动方案，以实现以下任务目标：
{此处叙述①当前需要达成的任务目标；②作战风格；③侧重目标等}，
该任务发生在以下地图/地形中：
{此处表述当前的战场环境，包括①关键地形信息；②关键建筑物信息；③地形与作战方式的制约关系等}
上一轮执行命令的历史信息为：
{此处填写上一轮命令执行前后敌我对象信息列表}，
以下JSON对象中定义了当前可用的友军和敌军单位，包括它们各自的标签(id)、名称(name)和位置(position)以及兵种概括信息。
{此处描述①敌我对象的当前状态；②兵种信息的领域知识的客观表征，包括对象概述、作战效能、配属火力、作战领域等}，
请基于地图信息，实现我们的任务目标，切记，为每一个友军单位分配指令，一定要按照我们规定的格式输出军事行动方案。

图 4 个性化提示词

Fig. 4 Customized prompts

3.4 方案验证平台

生成的行动方案的效果需要通过仿真环境验证。验证环境应与军事场景一致或尽可能反映作战执行逻辑，同时应兼顾获取和实现的便利性。本文利用即时策略游戏《星际争霸 II》的游戏引擎为基础构建方案验证环境。《星际争霸 II》中可设置敌对的双方来模拟敌我对抗环境。其地图编辑器中可设置不同地形、地貌要素，以便模拟实际战场环境。同时其拥有丰富多样的战斗单元、建筑物等，可有效模拟真实作战场景想定中的多域力量编组。《星际争霸 II》引擎在本文方法框架中所处的位置如图 2 所示，具体想定设置将在 4.1 节中详细介绍。

4 实验与结果

4.1 实验设置

本文使用即时策略游戏《星际争霸 II》引擎结合“虎爪”想定[21]作为验证环境，以展示和验证所提方法的结果及有效性。游戏想定场景的设置、状态信息的获取、行动方案的执行通过 burnysc2 库中相应 API 实现。

游戏想定场景俯视图如图 5 所示，其中地表类型包含平原、河流、桥梁。我、敌对象分别表示为俯视图中的绿色、蓝色单位。想定中包含的单位数量、游戏引擎与实际场景的单位类别对应关系如表 1 所列。LLM 接收提示词输入，生成军事行动计划。游戏引擎解析该计划并执行相关动作，基于动作结果计算奖励，评估计划的有效性。

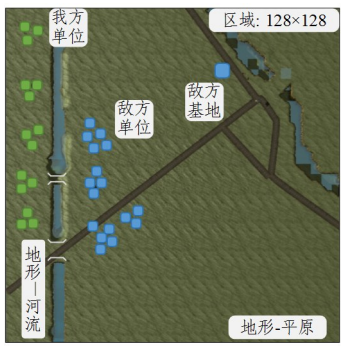


图 5 游戏想定场景俯视图（电子版为彩图）

Fig. 5 Top view of game scenario

提示词包含资源池信息、用户意图信息以及任务目标信息。资源池信息涉及分派兵力，目前兵力的数量可分为 6 兵力、9 兵力、12 兵力以及 15 兵力，分别由“调用 6 个友军兵力”“调用 9 个友军兵力”“调用 12 个友军兵力”以及“调用 15 个友军兵力”描述。用户意图信息主要涉及两个方面：一方面是分散兵力，实现目标，由“分配兵力，多区域战斗”描述；另一方面是集中兵力，实现目标，由“集中兵力”描述。任务目标信息主要也涉及两个方面：一是以击杀敌军作战单位为主，摧毁基地为辅，由“击杀全部敌人，不考虑基地”描述；二是以摧毁基地为主，击杀敌军作战单位为辅，由“击杀全部敌人，摧毁主基地”描述。完整的提示词由资源池信息、用户意图信息以及任务目标信息拼接而成。

表 1 敌我双方兵力配置详细信息

Table 1 Detailed information on military configuration of both sides

阵营	游戏兵种	实际兵种	数量
我方	Ghost	侦察兵	1
	Marauder	炮兵	1
	Hellion	机械化步兵	3
	SiegeTank	坦克	8
	VikingFighter	战斗机	2
敌方	Marine	普通步兵	1
	Reaper	反装甲单位	1
	Hellion	机械化步兵	12
	SiegeTank	坦克	2
	VikingFighter	战斗机	1
	CommandCenter	基地	1

LLM 输出的军事行动计划包含的动作有两种：attack_move_unit(id, target_x, target_y) 和 egage_target_unit(id, target_id, ta-rget_x, target_y)。attack_move_unit 表示指挥友军 id，移动到坐标(target_x, target_y)的位置，如果沿途遇到敌军，需要和敌军交战；engage_target_unit 表示指挥友军 id 与敌军单位 target_id 进行交战，如果敌军单位超出友军射程，友军单位将在交战前移动到敌军单位(target_x, target_y)的位置。

奖励的定义如下：每击杀一个敌军单位，奖励加 10；每牺牲一个友军单位，奖励减 10；一个友军单位从桥的左侧到达右侧，奖励加 10；一个友军单位从桥的右侧到达左侧，奖励减 10。通过计算最终获得奖励评估 LLM 生成的军事行动计划的有效性。

本文设计了 6 种不同的实验来评估 COA-Gen 方法，分别为单轮多轮对比实验、不同大模型对比实验、不同提示词指令遵循评估、KAG 有效性实验、与传统强化学习方法对比，以及单轮与多轮框架的 token 消耗对比。

4.2 单轮 vs 多轮方案生成效果

图 6 展示了以 gpt-3.5-turbo 作为大模型生成单轮和多轮军事行动方案，在不同兵力条件下的敌军与友军伤亡率、奖励值以及基地存活率。

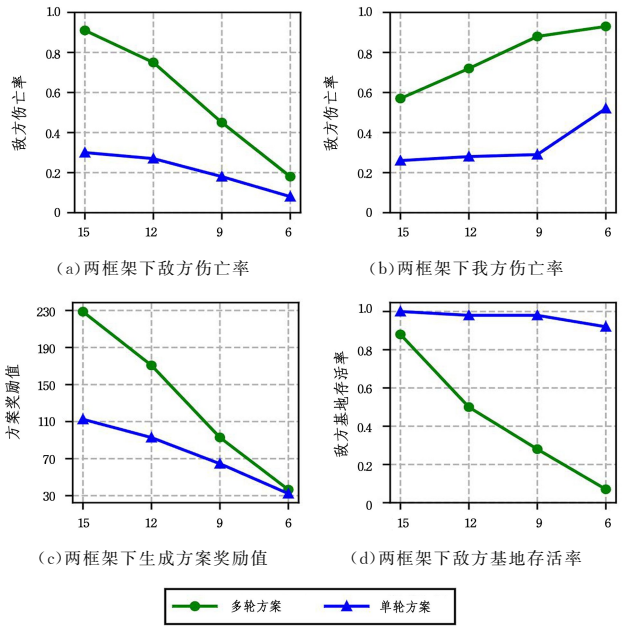


图 6 单轮 vs 多轮框架下的效果对比

Fig. 6 Effect comparison between single-round and multi-round frameworks

多轮方案中，友军利用前轮信息能有效击杀敌军和摧毁基地，从而提高奖励，降低敌军基地存活率，表现均优于单轮方案。同时，为了取得更好的效果，多轮方案也增加了敌我交战次数，导致双方伤亡增加。综合来看，多轮方案能生成更佳的军事行动计划。

图 7 给出了某次单轮、多轮方案的结果。如图所示，友军调用 15 兵力，集中兵力击杀敌军，摧毁基地的作战轨迹。在单轮框架下，兵力仅集中在坐标(19,63)，无后续动作，未能击杀敌军或摧毁基地，效果不佳。多轮方案下，第一轮集中兵力到达桥(19,36)处，第二轮集中兵力攻击基地，路途中遇到敌军并发起战斗，并成功击杀敌军和摧毁基地。行动轨迹证明了多轮方案优于单轮方案。

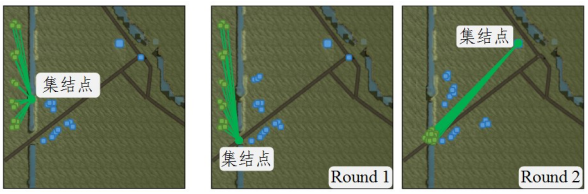


图 7 单轮、多轮框架下方案行动可视化

Fig. 7 Visualization of COA under single-round and multi-round frameworks

4.3 不同大模型的方案生成效果

大模型自身的性能差异对生成的军事行动计划效果至关重要。为了进一步验证在不同模型下，多轮方案相较于单轮方案的鲁棒性，选择 gpt-3.5-turbo，Claude-3-haiku 以及 Deepseek-v2 进行实验。考虑到敌军伤亡增加也包含基地的摧毁，因此多模型对比的评价指标选择敌军伤亡率、友军伤亡率和总奖励。由于不同兵力和不同提示词在实验中表现出一致性，因此不同大模型对比实验只设计不同的兵力。结果图 8 所示，gpt-3.5-turbo，Claude-3-haiku，Deepseek-v2 的多轮方案的敌军伤亡率和奖励值显著高于单轮方案，和 4.2 节的结论一致，说明了本文提出的方法在不同模型下表现出性能优势的鲁棒性。

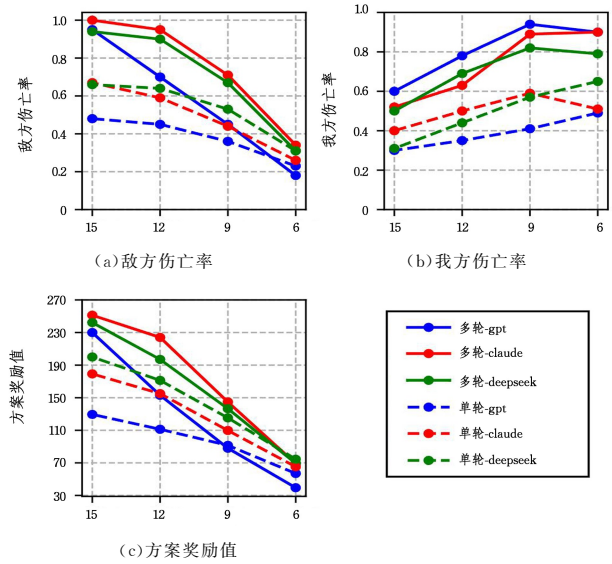


图 8 不同大模型下生成方案评分对比

Fig. 8 Comparison of generative scheme ratings with different large models

4.4 提示词指令跟随效果

不同提示词的指令意图存在较大差异，导致生成的军事行动计划也不相同。本实验设置 4 种提示词：1)击杀全部敌人，不考虑基地；2)击杀全部敌人，摧毁主基地；3)分配兵力，多区域战斗，击杀全部敌人，摧毁主基地；4)集中兵力，击杀全部敌人，摧毁主基地。

不同提示词下指令跟随效果可视化结果如图 9 所示。

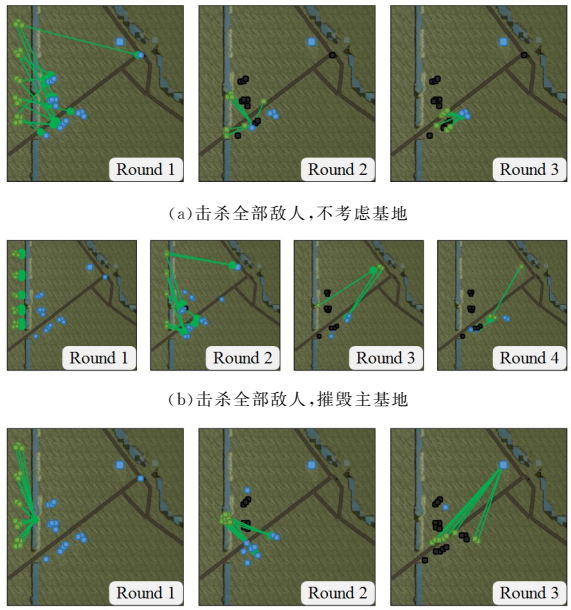


图 9 不同提示词指令跟随效果可视化

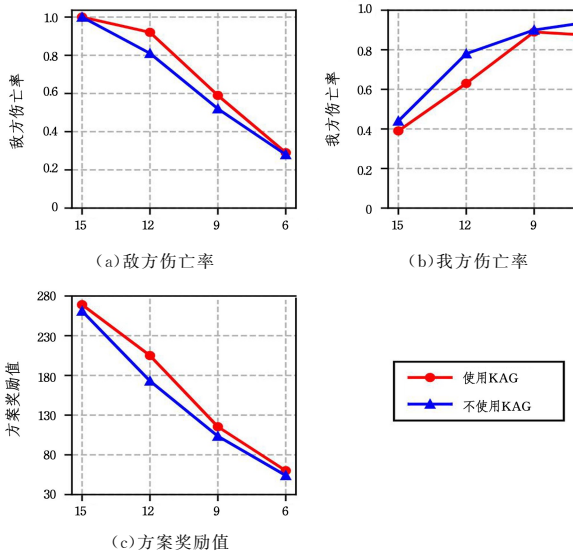
Fig. 9 Visualization of instruction following effect of different prompt words

图 9(a)的提示词是“击杀全部敌人，不考虑基地”，LLM 生成的计划以击杀敌军单位为主要任务；图 9(b)的提示词是“击杀全部敌人，摧毁主基地”，LLM 在第二轮就分配兵力摧毁敌军基地；图 9(c)的提示词是“分配兵力，多区域战斗，击杀全部敌人，摧毁主基地”，LLM 在第二轮对兵力进行分配，和不同区域的敌军进行战斗；图 7(b)的提示词是“集中兵力，击杀全部敌人，摧毁主基地”，LLM 两轮都是集中兵力，摧毁敌军基地，在前进过程中会遭遇敌军发生战斗，击杀敌军。实验结果表明，LLM 严格遵循了提示词指令，生成符合用户预期的行动方案。

4.5 基于 KAG 的方案生成效果

为了缓解模型幻觉，生成更合理的行动方案，我们在现有框架下增加 KAG 技术，除了增加基本军事信息外，还给出了作战单元的详细参数，包括属性、生命值、伤害、防御力、移动速度、武器信息等，期望 LLM 在行动生成中针对不同敌军类型，对友方兵种的选择更合理。为了验证 KAG 技术的有效性，在 gpt-3.5-turbo 模型下，对多轮行动方案生成进行了消融实验。

实验结果如图 10 所示，对于不同兵力数量作战，采用 KAG 技术均可以不同程度地提高敌军伤亡率，同时给友军更好的保护，总奖励也有所增加。实验表明，KAG 技术可以辅助 LLM 生成更合理高效的行动方案。



(a)敌方伤亡率 (b)我方伤亡率

(c)方案奖励值

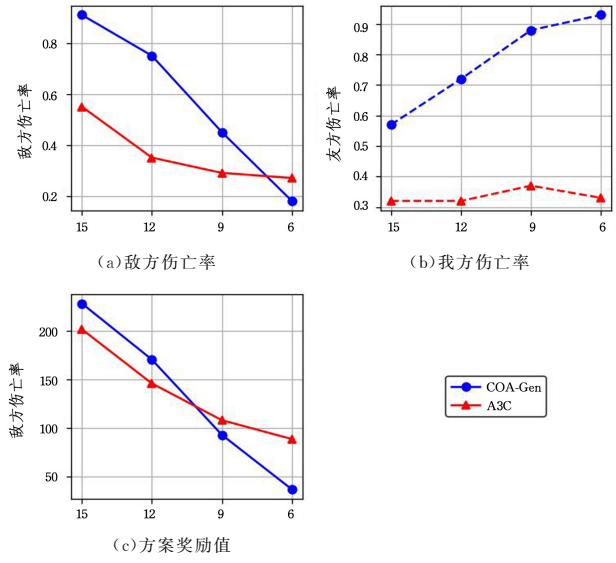
图 10 是否采用 KAG 技术生成方案评分对比

Fig. 10 Comparison of generative approach rating whether adopt KAG technology or not

4.6 与传统强化学习方法对比

A3C(Asynchronous Advantage Actor-Critic)算法[23]是一种强化学习的分布式算法，由 DeepMind 团队于 2016 年提出。A3C 算法的提出突破了过去方法的计算瓶颈，加快了收敛速度，提升了策略表现，在许多游戏任务中表现出色。在本节中，将本文提出的方法 COA-Gen 与 A3C 方法进行实验对比。

图 11 给出了 COA-Gen 与 A3C 方法在 3 种不同评估指标下的对比结果。从敌方伤亡率指标来看，随着友方人数的增加，两种方法都能对敌方造成更大的损失，但整体上，COA-Gen 方法优于 A3C 方法。结合友方伤亡率数据，可以得出 COA-Gen 方法比 A3C 方法更具进攻性的结论，这可能与我们给出的指令内容有关。当友方人数超过一定程度时，COA-Gen 方法的奖励值始终优于 A3C 方法，但两者之间的差距不大。我们推测 A3C 更倾向于采取“过桥”动作来获得奖励，这是一种保守的得分策略，可能不符合指挥人员的指令要求。



(a)敌方伤亡率 (b)我方伤亡率

(c)方案奖励值

图 11 COA-Gen 与 A3C 方法对比结果

Fig. 11 Comparison results of COA-Gen and A3C methods

4.7 单轮与多轮框架的 token 消耗对比

在探讨方案生成方法时，不仅要着眼于方案本身的有效性和可行性，还必须深入分析生成这些方案所需投入的成本。在大模型的应用场景中，通常采用 token 数量作为衡量成本的关键指标。在选择方案生成方法时，必须综合考虑其有效性与成本效益，力求在保证方案质量的同时，实现对成本的有效控制。

表 2 列出了单轮与多轮框架在不同兵力下生成方案所需的 token 数量对比。与经验知识相同，多轮框架需要消耗更多的 token 数量，并且多轮框架消耗的 token 数量平均是单轮框架的 3.8 倍。尽管如此，正如 4.2 节所述，多轮框架在生成方案的质量和效果上整体优于单轮框架生成的方案。因此，我们在选择合适的方案时需要综合考虑多方面的因素，仔细权衡成本与效益之间的关系，以确定最适合特定应用场景的方法。

表 2 单轮与多轮框架的 token 消耗对比

Table 2 Comparison of token consumption generated by single-round and multi-round frameworks

框架	6 兵力	9 兵力	12 兵力	15 兵力
单轮	667	818	1473	2874
多轮	2353	3797	6784	7468

结束语 本文面向现代复杂动态战场下的军事行动方案生成与规划问题，分析了现有方法的优缺点；结合作战能力需求和大模型技术的发展，提出了基于预训练大模型的军事行动方案生成与规划方法。具体包括：针对动态变化的战场环境，提出了 COA-Gen 方法，这是一种多轮方案生成框架；为挖掘预训练大模型自身的规划能力，验证了不同提示词下的

指挥员意图跟随效果；为提升基于通用数据训练的大模型在军事领域的效能，提出了 KAG 知识增强技术。与传统强化学习方法相比，COA-Gen 方法具有明显优势。以上关键技术和方法在基于“虎爪”想定和《星际争霸 II》游戏引擎的环境中进行了有效性和鲁棒性的验证。

未来的工作主要包括以下两点：1）真实作战场景中，态势信息通常以多种数据类型的形式出现。后续考虑在本文表示的基础上，将多模态态势信息作为预训练大模型的输入，以进一步挖掘模型决策潜力、提升行动方案生成的准确性。2）实际作战中行动方案的生成往往由多个要素共同作业形成，后续将考虑引入基于大模型的多智能体架构，使方案生成结果更符合实际指控流程和指挥员价值。

参 考 文 献

[1] ZHANG Y X. Research on Modeling and Optimization Methods for Military Mission Planning under Uncertainty[D]. Changsha: National University of Defense Technology, 2014.

[2] WAYTOWICH N, HARE J, GOECKS V G, et al. Learning to guide multiple heterogeneous actors from a single human demonstration via automatic curriculum learning in StarCraft II [C]// Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications IV. SPIE, 2022, 12113: 283-293.

[3] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[4] KINGMA D P, WELLING M. Auto-encoding variational bayes [J]. arXiv: 1312.6114, 2013.

[5] 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.

[6] BAI J, BAI S, CHU Y, et al. Qwen technical report[J]. arXiv: 2309.16609, 2023.

[7] TOUVRON H, LAVRIL T, IZACARD G, et al. Llama: Open and efficient foundation language models [J]. arXiv: 2302.13971, 2023.

[8] ZENG A, LIU X, DU Z, et al. Glm-130b: An open bilingual pre-trained model[J]. arXiv: 2210.02414, 2022.

[9] 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.

[10] SHINN N, CASSANO F, GOPINATH A, et al. Reflexion: Language agents with verbal reinforcement learning[J]. Advances in Neural Information Processing Systems, 2024, 36: 8634-8652.

[11] HUANG Y, HUANG J. A Survey on Retrieval-Augmented Text Generation for Large Language Models[J]. arXiv: 2404.10981, 2024.

[12] FIKES R E, NILSSON N J. STRIPS: A new approach to the application of theorem proving to problem solving[J]. Artificial Intelligence, 1971, 2: 189-208.

[13] TATE A, DRABBLE B, DALTON J. The use of condition types to restrict search in an AI planner[C]// Proceedings of the AAAI Conference on Artificial Intelligence. AAAI, 1994: 1129-1134.

[14] SARCIA S A. Organizing Structures and Information for Developing AI-enabled Military Decision-Making Systems[C]// 2023 IEEE International Workshop on Technologies for Defense and Security (TechDefense). IEEE, 2023: 455-460.

[15] SCHWARTZ P J, O'NEILL D V, BENTZ M E, et al. AI-enabled wargaming in the military decision making process[C]// Artificial Intelligence And Machine Learning for Multi-Domain Operations Applications II. SPIE, 2020, 11413: 118-134.

[16] LUO J Z, SUN Y L, QIAN Z Z, et al. Overview and Prospect of Artificial Intelligence Large Models [J]. Radio Engineering, 2023, 53(11): 2461-2472.

[17] BAI J, BAI S, YANG S, et al. Qwen-vl: A versatile vision-language model for understanding, localization, text reading, and beyond[J]. arXiv: 2308.12966, 2023.

[18] WANG G, XIE Y, JIANG Y, et al. Voyager: An open-ended embodied agent with large language models [J]. arXiv: 2305.16291, 2023.

[19] AHN M, BROHAN A, BROWN N, et al. Do as I can, not as I say: Grounding language in robotic affordances[J]. arXiv: 2204.01691, 2022.

[20] LAMPARTH M, CORSO A, GANZ J, et al. Human vs. machine: Language models and wargames[J]. arXiv: 2403.03407, 2024.

[21] GOECKS V G, WAYTOWICH N. Coa-gpt: Generative pre-trained transformers for accelerated course of action development in military operations[C]// 2024 International Conference on Military Communication and Information Systems. IEEE, 2024: 1-10.

[22] HU S, HUANG T, LIU L. Pok\'eLLMon: A Human-Parity Agent for Pok\'emon Battles with Large Language Models[J]. arXiv: 2402.01118, 2024.

[23] MNIH V. Asynchronous Methods for Deep Reinforcement Learning[J]. arXiv: 1602.01783, 2016.

YAN Yusong, born in 2001, Ph.D candidate. His main research interests include reinforcement and intelligent decision and so on.

ZHOU Yuan, born in 1993, Ph.D, assistant researcher. Her main research interests include machine learning and intelligent decision.

（责任编辑：杨雪敏）