

DOI: 10.3969/j.issn.1671-0673. 2026. 04. 011

融合多智能体与描述逻辑的军事本体构建方法

吕恒钊, 车 森

(信息工程大学, 河南 郑州 450001)

摘要: 针对大语言模型 (LLMs) 在构建军事本体时易产生知识幻觉和逻辑冲突的问题, 提出一种基于多智能体的协作本体构建方法 (MACOD)。该方法通过角色解耦, 将复杂的知识抽取任务分配至建模师、逻辑专家与测试员 3 个专业智能体, 并引入符号推理机作为严格的后置约束机制, 对生成的本体图谱进行全局一致性审查。依托军事条令数据开展的对照与消融实验显示, 相较于单一模型模式, 多智能体协作架构显著提升了本体概念与关系抽取的覆盖率及准确度; 同时, 符号推理机制有效识别并拦截了逻辑矛盾, 降低了系统的幻觉率。

关键词: 大语言模型; 多智能体协作; 本体构建; 军事条令; 知识抽取

中图分类号: TP182

文献标识码: A

文章编号: 1671-0673(2026)04-0457-09

Military Ontology Development Method Fusing Multi-Agent and Description Logic

LYU Hengzhao, CHE Sen

(Information Engineering University, Zhengzhou 450001, China)

Abstract: To address the issues of knowledge hallucination and logical conflicts that large language models (LLMs) tend to produce when constructing military ontologies, a multi-agent collaborative ontology development (MACOD) method is proposed. Through role decoupling, the complex knowledge extraction tasks are assigned to 3 professional agents: modeler, logic expert, and tester. The symbolic reasoner is introduced as a strict post-constraint mechanism to check the global consistency of the generated ontology graph. Experiments based on military doctrine data, including comparative and ablation studies, show that compared with single-model approaches, the multi-agent collaborative architecture significantly improves the coverage and accuracy of ontology concept and relationship extraction. At the same time, the symbolic reasoning mechanism effectively identifies and blocks logical contradictions, reducing the system's hallucination rate.

Key words: LLMs; multi-agent collaboration; ontology development; military doctrine; knowledge extraction

随着现代战争向信息化、智能化加速演进, 战场数据激增引发的信息过载已成为制约指挥决策效能的核心瓶颈。海量非结构化军事条令文本 (如美军 JP 3-0《联合战役与作战》^[1]、JP 5-0《联合计划》等) 蕴含联合作战关键逻辑与规则, 其向机器可理解、可推理的结构化知识转化, 是实现从信息

优势向认知与决策优势跨越的关键。本体 (ontology) 作为领域知识的形式化规范, 能精准表达概念、属性和公理约束, 是构建指挥信息系统语义互操作与智能推理能力的基础。但军事领域知识具有逻辑刚性强、约束复杂和容错率低的特点, 传统构建方法与新兴智能化手段均面临着严峻挑战。

收稿日期: 2026-04-09; 修回日期: 2026-06-02

基金项目: 国家自然科学基金 (42371438)

作者简介: 吕恒钊 (2002—), 男, 硕士生, 主要研究方向为软件工程、人工智能与大数据。E-mail: 1594705009@qq.com

早期军事本体构建多依赖人工专家或半自动化规则模板。例如,基于斯坦福七步法的手工构建虽能保障逻辑严谨性,但高度依赖专家隐性知识,构建周期长、成本高,难以适配军事知识体系的快速迭代^[2]。为提升效率,研究者们提出多种自动化技术。文献[3]设计融合本体特征的BiLSTM-CRF军事事实识别模型;文献[4]构建美军《任务清单》本体;文献[5]和文献[6]分别研究军事信息服务与作战行动序列核心本体的建模和推理方法;基于需求的螺旋反馈法^[7]、面向专家示例的子本体抽取算法^[8]和规则与预训练模型结合的军事术语知识图谱构建方法^[9]也相继涌现。这些研究推动了军事知识结构化进程,但在处理深层语义逻辑与跨文本长距离依赖关系时仍存在局限。

近年来,以大语言模型(Large Language Models, LLMs)为代表的生成式AI技术,为知识工程带来新范式。LLMs具备强大的语义理解与泛化能力,在少样本甚至零样本场景下展现出优异的信息抽取潜力。文献[10]提出的上下文学习与语义检索增强方法,显著提升大模型零样本推理能力;InstructionNER方法^[11]通过指令微调实现高效实体识别;OntoKG和AI Ontology^[12]等框架利用LLMs辅助构建概念层次结构;文献[13]探索能力问题(Competency Question, CQ)驱动本体生成的可行性。在文本生成与处理领域,基于提示工程的实体关系抽取^[14]和特定格式化文档微调生成研究^[15],印证了LLMs在规范化文本输出上的优势;学术文献挖掘领域中,SciBERT、Sentence-BERT^[16]和关系图卷积网络(Relational Graph Convolutional Network, R-GCN)^[17]用于优化引文推荐^[18],这些关于深层语义表征与全局信息聚合的研究思路^[19],也为处理军事条令复杂语义提供了有益参考。

然而,单一LLM直接应用于军事本体构建存在不可忽视的幻觉风险,易生成违背条令原意的虚假知识,如虚构跨军种指挥关系、混淆作战指挥权(Combatant Command Authority, COCOM)和行动控制权(Operational Control, OPCON)的互斥约束,且现有模型缺乏领域刚性逻辑的内嵌式校验机制,难以满足军事应用的严谨性要求。

与生物医学、电子商务等通用领域本体构建相比,上述问题在军事领域尤为突出。军事本体具有4个方面的显著特殊性:1)术语定义的条令刚性,军事概念由条令严格界定,不允许模型基于语义相似性自由泛化,任何偏离条令原意的生成均

构成幻觉;2)逻辑约束的绝对互斥性,如COCOM与OPCON的互斥、指挥链的严格传递性等刚性约束一旦违反,将直接导致决策逻辑崩溃,容错空间极小;3)关系类型的领域特异性,“指挥—隶属”“权限委托—权限行使”等关系附带方向性、传递性等复合约束,超出通用关系抽取模型的建模能力;4)语料的封闭性与稀缺性并存,条令文本高度凝练但总量有限,通用领域依靠海量语料驱动的范式难以直接迁移。这些特殊性决定了军事本体构建需要一种能同时满足条令忠实性、逻辑严谨性与关系完备性的构建方法。

多智能体系统(Multi-Agent System, MAS)为突破单一模型认知局限、强化逻辑约束提供了思路。其通过模拟人类团队协作模式,拆分任务并交互协作,可显著提升复杂任务解决能力。文献[20]构建的“眼—脑—手”三维能力框架,文献[21]基于观察—评估—决策—行动循环(Observe-Orient-Decide-Act Loop, OODA)构建的层级化多智能体指挥控制(Command and Control, C2)模型,均验证了该架构在动态博弈与任务规划中的有效性。但在知识工程领域,尤其如何将描述逻辑(description logic)推理机深度嵌入多智能体协作流程以实现实时逻辑纠错,相关研究仍较欠缺。

针对上述挑战,提出一种多智能体协作本体构建方法(Multi-Agent Collaborative Ontology Development, MACOD)。该方法的核心创新点如下:1)提出“角色解耦—闭环博弈”的多智能体协作机制。区别于现有的单一LLM端到端生成范式,MACOD将本体构建任务解耦为建模、逻辑校验和测试3个认知独立的子任务,通过“抽取—校验—反思”闭环博弈实现智能体间的对抗性协作,从架构层面抑制单模型注意力分散导致的幻觉生成。2)设计符号推理机深度嵌入的双重逻辑校验机制。不同于将推理机仅作为后置评估工具的常规做法,MACOD将HermiT推理机^[22]嵌入生成链路内部,在逻辑专家输出后与本体形式化后分别执行校验,形成“生成中拦截+生成后验证”的双重防线,将不可控的神经网络幻觉转化为边界清晰的确定性修正项。3)构建面向军事条令的全流程本体构建与评估框架。以美军JP 3-0中的C2条令为对象,建立从预处理、自动构建、形式化到多维度量化评估的完整技术链路,并构建包含概念、关系、公理和实例的军事C2领域本体。

MACOD的设计针对军事本体的4个特殊挑战

均做出了针对性的回应。1)针对术语定义的条令刚性,建模师智能体的提示词中嵌入了“仅抽取条令原文明确提及的概念,严禁生成无条令依据的内容”这一硬约束,并要求每个抽取元素标注条令出处,从源头将生成行为锚定于条令文本,抑制自由泛化导致的幻觉;2)针对逻辑约束的绝对互斥性,HermiT推理机双重校验机制专门面向军事公理中的互斥声明、传递性约束等进行形式化验证,任何违反刚性约束的输出均被自动拦截并触发逻辑专家的反思修正;3)针对关系类型的领域特异性,在CQ引导表与建模师提示词中预定义了军事领域特有的关系模板,引导模型按领域关系模式

进行定向抽取,而非依赖通用关系分类器;4)针对语料的封闭性与稀缺性,多智能体协作架构通过角色分工与闭环博弈,以少量高质量条令文本为输入,即可完成高覆盖率的本体构建,降低了对大规模标注数据的依赖。

1 所提方法

针对军事条令本体构建中效率提升与逻辑严谨性难以兼顾的核心痛点,以全流程闭环管控、分阶段逻辑校验和角色化分工协同为设计原则,构建数据预处理、多智能体协作、本体形式化、实例验证和多维度评估的MACOD框架,如图1所示。

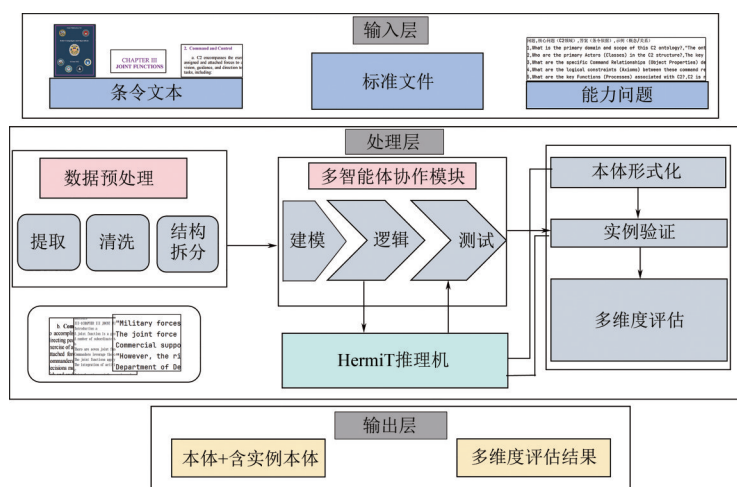


图1 MACOD框架

该框架自上而下划分为输入层、处理层和输出层,形成标准化输入、自动化生成与校验、规范化输出和全维度评估的技术链路。1)输入层以美军JP 3-0《联合战役与作战》(2022版)第3章联合职能第2节C2文本为核心数据源,同步结合领域专家构建的基准本体和CQ列表,为本体构建提供标准化输入与权威校验基准。2)处理层为框架核心,通过角色解耦的多智能体协作模块完成本体核心结构的抽取、优化和场景适配,同步深度集成HermiT推理机双重校验机制,在生成全流程拦截逻辑冲突、保障本体逻辑严谨性。3)输出层则负责输出符合OWL 2 DL规范的标准化Web本体语言(Web Ontology Language, OWL)格式本体、含实例的实战化验证本体,以及覆盖基础质量、逻辑效能、应用价值的多维度量化评估结果。

1.1 数据集预处理与基准本体构建

实验以美军JP 3-0《联合战役与作战》为核心数据源,精准提取第3章联合职能第2节C2相关内容,原始文本总量约2.3万字。采用正则表达式与

人工校验结合的方式完成预处理,去除文本中的页眉页脚、格式符和冗余内容,按句子边界拆分为独立语义单元,最终得到1 286条有效句子,作为实验标准化输入。依据斯坦福七步法,由领域专家构建C2领域基准本体,作为实验评估的金标准。该本体严格锚定JP 3-0条令原文,包含10个核心概念、10个关键关系与26条军事专属公理。由于军事条令的高度凝练性,该基准本体侧重于构建C2的深层逻辑骨架,并向下辐射支撑25个具体作战实例的实例化断言,形成了小而精、高约束的军事强逻辑本体验证基准。同时,设计5类核心CQ,以SPARQL查询执行结果为验证标准,确保本体的实战推理能力,如表1所示。

1.2 角色解耦的多智能体协作框架设计

底层驱动引擎均采用通义千问Qwen-Max模型构建,通过系统提示词完成建模师、逻辑专家和测试员的角色固化、职责界定、输出规范和强制约束,确保每个智能体仅聚焦单一核心任务,避免多任务并行导致逻辑混乱与幻觉生成,如表2所示。

表1 CQ引导表

序号	CQ描述	预期本体支撑要求
CQ1	查询所有拥有 COCOM 权限的指挥官及其控制的作战区域	需包含 Commander“COCOM”“OperationalArea”类,exercises-Authority“controlsArea”对象属性
CQ2	验证 COCOM 与 OPCON 是否满足互斥约束	需包含“COCOM disjointWith OPCON”公理,支持 HermiT 推理机冲突检测
CQ3	查询某 C2 系统(如网络 C2 系统)支持的指挥权限类型	需包含 C2System 子类 CyberC2System,supportsAuthority 对象属性
CQ4	查找某作战区域(如太平洋责任区)对应的所有下属部队	需包含 OperationalArea 子类 AOR_Pacific,hasSubordinateForce 对象属性
CQ5	验证指挥链传递性(如作战司令→联合部队司令→下属部队)	需包含“hasCOCOM rdfs:subPropertyOf hasOPCON”传递公理,支持实例关系推导

表2 多智能体提示词示例

智能体角色	智能体提示词约束
建模师	角色设定:你是军事条令本体构建领域的专业建模师,仅负责在给定的美军 JP 3-0 条令文本中抽取本体结构。 抽取规则: 1)仅抽取条令原文中明确提及的概念类、对象属性、数据属性,严禁生成无条令依据的内容; 2)每个抽取的元素必须精准标注条令来源(如“JP 3-0 Chapter 3 Section 2 Line X”); 3)必须优先覆盖给定的 10 个核心概念与 10 个核心对象属性。 输出规范:严格输出 JSON 格式(仅含 classes, object_properties, data_properties 字段),每个元素必须包含名称、父类/定义域值域、条令来源、描述。不得添加任何解释性文本。
逻辑专家	角色设定:你是军事本体领域的逻辑专家,负责对本体草稿进行逻辑优化,并处理 HermiT 推理机的报错反馈。 校验规则: 1)依据 JP 3-0 条令原文,补充军事专属公理(重点保障 COCOM 与 OPCON 互斥、指挥链传递等刚性约束); 2)反思与纠错机制。当你接收到外部 HermiT 推理机返回的冲突日志(如类层级冲突、公理违背)时,必须优先定位导致冲突的源属性或公理,并依据条令逻辑进行修正或删减,直至打破逻辑僵局。 输出规范:输出优化后的完整 JSON 本体框架,并明确标注出所有新增/修改公理的条令依据。禁止生成缺乏条令证据的虚构约束。
测试员	角色设定:你是军事本体领域的测试工程师,负责依据 CQ 列表与生成的本体框架,完成查询闭环验证。 测试规则: 1)为给定的每一条 CQ 生成符合标准规范的 SPARQL 查询语句(命名空间统一为 http://jp30-c2-ontology.org/#); 2)对照生成的 JSON 本体框架执行静态依赖检查,判定本体中是否真实存在支撑该 SPARQL 执行的类、属性与公理。 输出规范:输出固定格式的测试报告,每条包含:CQ_ID、SPARQL 语句、覆盖状态(覆盖/未覆盖)。若状态为“未覆盖”,必须指出缺失的具体元素名称。不得私自修改本体框架。

为克服传统大语言模型在长文本处理中的幻觉,本框架摒弃了缺乏交互的线性流水线设计,构建基于“抽取—校验—反思”闭环博弈(closed-loop game of extraction-validation-reflection)的多智能体协作机制。通过 Qwen-Max 底座,配合结构化提示词(structured prompt),实现角色能力的逻辑隔离与认知协同。具体运行机制如下。

1)基于认知隔离的本体草图生成。建模师智能体被严格限制在封闭域抽取的认知边界内,仅负责将非结构化条令解析为 JSON 格式的概念与关系网络,不进行主观逻辑推断,从而从源头压制模型幻觉。

2)基于环境反馈的自主纠错博弈。这是多智能体协作的核心环节。逻辑专家智能体不再是被动接收数据的节点,而是与外部环境(HermiT 推理

机)形成交互闭环。当 HermiT 检测到类层级冲突或公理违背时,不再是单纯的日志报错,而是将冲突类型与原始公理转化为反思提示词(reflection prompt),反向注入给逻辑专家智能体。逻辑专家需依据反馈进行自我纠正(self-correction),直至打破逻辑僵局,实现从神经网络生成到符号逻辑验算的跨模态交互。为避免多智能体陷入无限纠错死循环,系统设定最大反思重试次数为 3。该阈值的设定基于以下经验依据:实验调试阶段的观测表明,绝大多数可由 LLMs 自主修正的逻辑冲突(如简单的子类声明错误、值域类型不匹配)在前两次重试内即可消解,第 3 次重试主要用于处理涉及多条公理联动的复合冲突;超过 3 次后,剩余未消解冲突通常源于跨章节条令间的深层隐含矛盾,已超出模型当前上下文窗口的认知能力,继续

重试仅增加时间开销而无质量增益。该设定与多智能体自我反思领域的主流实践一致,例如文献[23]提出的Reflexion框架将反思迭代上限设为3,文献[24]提出的Self-Refine方法同样采用3~4次迭代上限,这表明该量级是LLMs自我纠错的效能饱和和区间。若超出阈值仍未消解冲突,则保留冲突现场并标记为待人工介入,交由测试阶段输出。

3) 实战化场景覆盖测试。测试员智能体作为最终质量把控者,接收优化后的本体与CQ列表,通过生成SPARQL查询进行对抗性测试。若查询失

败,则记录断点,为人工提供精准的故障诊断坐标。

1.3 HermiT推理机双重校验机制

HermiT推理机是符合OWL 2 DL标准的主流推理工具,相较Pellet等同类工具,在军事本体互斥关系、传递性约束等复杂公理校验场景中,具备运算速度快、冲突信息定位清晰和Python集成便捷的优势。本文将其深度集成至多智能体协作流程,设计逻辑专家输出后、本体形式化后的双重校验机制,在控制流程复杂度的同时,最大化实现逻辑冲突的早期检测。图2为处理层交互样式。

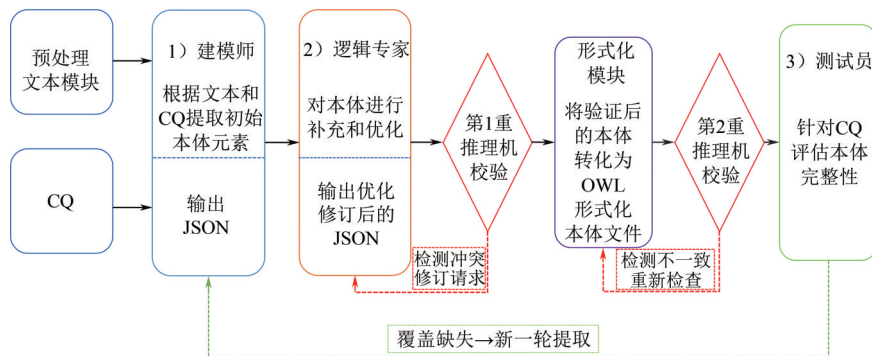


图2 处理层交互

第1重校验设置于逻辑专家智能体输出优化本体框架之后、测试员智能体执行之前,核心目标为排查类层级与公理冲突,避免测试员智能体基于错误本体生成无效SPARQL查询。实施流程为通过owlready2库将JSON本体框架转化为符合OWL 2 DL规范的临时OWL文件,开发基于Python的结构化解析脚本,进行字段映射。调用HermiT推理机执行一致性校验,重点检测类层级冲突、公理冲突和属性约束错误3类问题,最终生成含冲突

类型、位置和条令参考依据的校验日志。HermiT推理机检验多智能体系统生成的冲突,其生成的冲突日志作为高密度反馈信号,直接触发逻辑专家智能体的反思重构机制。这种“生成—检验—纠错”的自动化微循环,极大过滤显性逻辑错误,将不可控的模型幻觉转化为数量极少的确定性修正项,从而为后续环节扫清逻辑障碍。该校验以机器自动推理替代人工逻辑检查,可提前拦截前序环节的显性逻辑缺陷。图3为冲突转化示例。

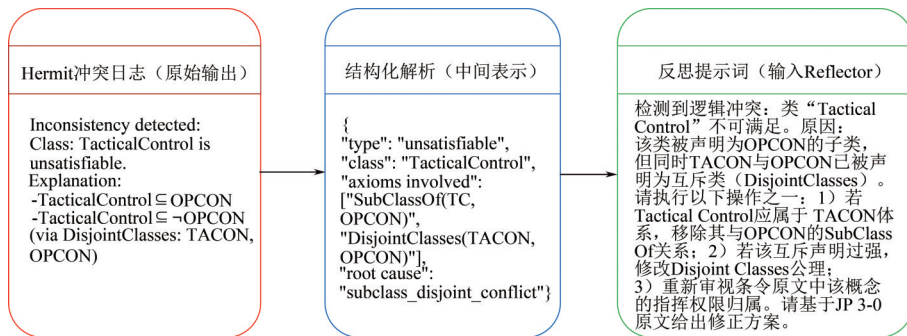


图3 冲突转化示例

本文设计一套基于规则的冲突日志解析与反思提示词生成机制。当HermiT推理机检测到本体不一致时,owlready2库会抛出结构化异常信息,系统通过Python脚本对异常信息进行正则解析,提取冲突类型(conflict_type)、涉及实体(involved_entities)和触发公理(triggering_axiom)3个关键字段。

随后,基于预定义的冲突类型—提示词映射模板库,将符号化的冲突信息转化为逻辑专家智能体可理解的自然语言反思提示词。该模板库覆盖类层级冲突、公理矛盾和属性约束错误3类核心冲突场景,其具体的OWL异常特征和反思提示词模板如表3所示。

表3 HerMiT推理机冲突示例

冲突类型	OWL异常特征	反思提示词模板
类层级冲突 (DisjointViolation)	InconsistentOntologyException 中包含 DisjointClasses 关键字	“HerMiT 检测到逻辑冲突:类‘{EntityA}’与类‘{EntityB}’已被声明为互斥 (DisjointClasses),但当前本体中存在使二者同时为真的路径。请依据 JP 3-0 条令原文,判断应删除哪条子类声明或修改哪条公理以消解冲突。”
公理矛盾 (AxiomContradiction)	异常信息中包含 SubClassOf 与 DisjointWith 的循环引用	“HerMiT 检测到公理矛盾:公理‘{Axiom1}’与公理‘{Axiom2}’产生逻辑环路。请定位条令中对‘{Property}’的原始定义,保留有条令依据的公理,删除或修正缺乏依据的约束。”
属性约束错误 (RangeViolation)	异常信息中包含 ObjectPropertyRange 不匹配	“HerMiT 检测到属性值域违规:属性‘{Property}’的声明值域为‘{DeclaredRange}’,但实际关联了类型为‘{ActualType}’的实体。请核查条令中该属性的适用对象,修正值域声明或删除错误关联。”

第2重校验设置于本体形式化完成之后、实例构建之前,核心目标为验证形式化本体的 OWL 语法合规性与逻辑闭环完整性。实施流程为:加载生成的 OWL-Turtle 文件,先校验命名空间、类/属性/公理声明的 OWL 2 DL 语法规范性,再执行全量逻辑校验,重点验证核心基准公理的生效情况;若存在语法错误或公理缺失,则返回形式化环节修正,校验通过则进入实例构建环节。该校验为本体标准化后的最终逻辑把关,保障本体的机器可读性和逻辑严谨性。两重校验仅覆盖语法规范与公理声明的有效性,因知识抽取错误导致的语义幻觉,由人工终审环节识别并计入幻觉率指标。

1.4 本体形式化与实例构建

利用 rdflib 库将经过校验的中间态知识转化为符合 OWL 2 DL 标准的 Turtle 格式。严格定义命名空间,并将自然语言描述转化为 rdfs:comment 注释,确保本体的语义透明度。为验证本体的实战推理能力,构建包含 25 个核心实例的虚拟断言盒 (Assertion Box, ABox)。构建原则遵循虚拟性、条令一致性和全链路覆盖。生成包括 CombatantCommander、JointTaskForce、AOR_Pacific 在内的实体,并建立符合术语盒 (Terminology Box, TBox) 约束的关系三元组,如 (Commander_01, exercisesAuthority, COCOM_01),形成闭环的战术逻辑验证环境。

2 实验设计与结果分析

2.1 实验设计

实验选用美军 JP 3-0《联合战役与作战》作为核心数据源,选取其中第3章第2节 C2 文本。作为美军联合作战条令体系的顶层文件,JP 3-0 系统定义了联合作战的核心概念框架,涵盖指挥权限 (COCOM、OPCON、TACON)、作战职能 (指挥控制、火力、情报、机动、防护、保障)、组织结构 (联合部队司令部、组成部队、职能部队) 及其相互关系,是

军事 C2 领域概念密度最高、逻辑约束最复杂的单一文献之一。

为确保实验数据的标准化,本文执行严格的预处理流程。首先,精准提取 C2 相关章节,涵盖指挥权限、C2 流程和作战区域等核心子域;其次,通过正则表达式清洗非语义字符,并按语义完整性将其拆分为 1 286 条有效句子;最后,依据斯坦福七步法构建“基准本体”作为评估基准 (ground truth),包含 10 个核心概念、10 个关键关系和 26 条军事公理。为验证 MACOD 的优越性,实验设置 3 组对比方案,严格遵循单一变量原则,仅改变构建方法,保持数据集与评估标准一致。表 4 为实验分组设计方案。

表4 实验分组设计方案

组别	方法名称	核心特征	验证目标
实验组	MACOD	多个智能体+ HerMiT 双重校验	验证 MACOD 的有效性 与鲁棒性
对照组 1	单一智能体 抽取 (参考 Ontogenia)	Qwen-Max 单智能体+ 无过程校验	评估基线性能, 验证无协作模式 的局限性
对照组 2	基准本体	Protégé 手工编辑	确立本体质量的 理论上限
对照组 3	零样本	大模型直接生成+ 不引入 CQ 引导	展示基础 构建方案

对照组 1 采用与文献[13]中 Ontogenia 类似的 CQ 引导式单模型抽取策略,使用 Qwen-Max 作为底层 LLM。具体而言,借鉴 Ontogenia 中以 CQ 驱动本体生成的核心思路,即向单一 LLM 提供完整的 CQ 列表与领域文本,由模型一次性生成本体结构。与原始 Ontogenia 的主要差异在于:1) 底层模型由 GPT-4 替换为 Qwen-Max,以确保与实验组的模型变量一致性;2) 输入文本由通用领域替换为军事条令文本;3) 未复现 Ontogenia 中的迭代优化模块,仅保留其单轮生成的基础范式,以严格对照多智能体协作带来的增量贡献。使用与实验组一致的全量 CQ 引导和条令文本输入,提示词仅包含本体

构建基础要求(无多智能体分工约束和逻辑校验引导),模型参数设置:温度系数(temperature)为 0.3,最大输出长度为 2 048,无过程校验环节。

对照组 2 由本文依据斯坦福七步法独立构建。构建过程严格锚定 JP 3-0《联合战役与作战》原文,每个概念与关系均标注条令出处,并严格对照 CQ 进行逐条验证。使用 Protégé 5.5.0 工具完成构建,历经多轮迭代修订。该本体作为严格条令锚定、过程可溯的人工构建参考基准,其质量通过条令符合度与 CQ 覆盖度双重保障,用于客观评估自动化方法与权威条令要求的契合程度。

对照组 3 采用零样本直接生成策略,使用与实验组相同的 Qwen-Max 模型,仅通过通用本体构建提示词引导模型从条令文本中一次性抽取概念、关系与公理,不引入 CQ 引导、多智能体协作或推理机校验,该策略代表了当前基于 LLMs 进行本体构建的传统基础范式。

基础质量包括召回率 R 、精确率 P 、 $F1$ 值,衡量本体构建的基础性能,计算公式为:

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100\%$$

(1)

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \times 100\%$$

(2)

$$F1 = 2 \times \frac{P \times R}{P + R} \times 100\%$$

(3)

式中: N_{TP} 为真正例数,即系统正确抽取且与基准本体匹配的概念或关系数量; N_{FN} 为假负例数,即基准本体中存在但系统未能抽取的概念或关系数量; N_{FP} 为假正例数,即系统生成但在基准本体中不存在的概念或关系数量。

逻辑效能包括推理无冲突率和幻觉率,衡量本体的逻辑严谨性与抗幻觉能力,其中:推理无冲突率 $=1-(\text{推理机检测出的矛盾公理数})/(\text{总公理数})\times 100\%$;幻觉率 $=(\text{虚构概念数}+\text{虚构关系数})/(\text{总构建元素数})\times 100\%$ 。虚构元素的判定标准为:该概念或关系在源条令文本中完全无对应表述,亦无法通过条令已有内容经合理推理得出。若某概念在条令中存在但其属性约束或关系方向被错误构建,则归类为构建错误而非幻觉。

应用验证包括 CQ 覆盖率,即成功响应 CQ 的比例,衡量本体的实战应用价值。

效率指标包括人工校准耗时,衡量方法的构建效率。

2.2 量化结果对比分析

表 5 展示了 4 组实验在 7 项核心指标上的统计结果,其中空白表示该指标项不适用于基准组。实验数据表明,实验组在各项关键性能上均显著优于对照组 1,且在逻辑一致性等关键指标上逼近对照组 2。

表 5 实验核心指标对比结果

组别	基础质量			逻辑效能		应用验证	效率指标
	召回率 R /%	精确率 P /%	$F1$ 值/%	推理 无冲突率/%	幻觉率/%	CQ 覆盖率/%	人工校准 耗时/h
实验组	90.00	63.64	74.56	90.91	2.40	80.00	4.50
对照组 1	60.00	25.00	35.29	75.00	15.80	55.00	6.50
对照组 2				100.00	0	100.00	
对照组 3	50.00	15.20	23.31	60.00	60.00	20.00	9.00
相比对照组 1 提升幅度	+30.00 百分点	+38.64 百分点	+39.27 百分点	+15.91 百分点	-13.40 百分点	+25.00 百分点	-30.77%

实验结果显示,实验组的 $F1$ 值达到 74.56%,相较于对照组 1 的 35.29% 有较大幅度提升,提升了 39.27 个百分点。尽管引入多智能体框架在处理速度上相比单一模型略有损耗,但其精确率 P 提升了 38.64 个百分点。这反映出逻辑专家智能体的二次校验机制能够有效过滤单一智能体产生的易混淆错误。

在应用验证层面,对照组 1 在面对复杂推理场景时表现出一定局限性,CQ 覆盖率为 55.00%。相比之下,实验组通过多智能体的分步推理,实现了 80.00% 的 CQ 覆盖率,表明 MACOD 生成的本体能

更好地支撑领域复杂查询。CQ3 的覆盖案例可直观体现 MACOD 的严谨性:实验组能严格依据条令原文,准确构建 CyberC2System_USCYBERCOM 与 supportsAuthority Operations 的关联;而对照组 1 在处理复杂关联时,易生成类似 supportsAuthority COCOM 这种缺乏事实依据的幻觉断言,违背了条令逻辑。

此外,针对实验组 90.91% 的推理无冲突率,进一步分析发现,未能自动消解的极少数冲突,多源于跨章节条令间深层隐含逻辑的冲突(如非显式的指挥权转移),而绝大多数常规逻辑冲突已被

HermiT 双重校验机制有效拦截。这一结果表明,本框架在一定程度上将大模型不可控的内生幻觉,转化为了边界清晰、易于定位的可控偏差。配合后期人工校准耗时降低 30.77% 的数据,整体反映了 MACOD 在兼顾本体构建质量与人工协同效率方面的可行性。

2.3 消融实验分析

为进一步量化多智能体协作与 Hermit 校验的独立贡献,设置两组消融实验,温度参数(temperature)为 0.3,以保障生成的稳定性。实验组采用完整的神经符号双向驱动架构,多角色协同并结合 Hermit 进行闭环纠错。消融组 1 保留建模师、逻辑专家和测试员的协同架构,但切断逻辑专家与 Hermit 推理机的交互闭环,完全依赖 LLMs 自身的内在逻辑进行构建。消融组 2 移除多智能体架构,由单一智能体独立完成初始抽取后,直接接收 Hermit 的冲突日志进行一次性的后处理修正。实验结果如表 6 所示。

表 6 消融实验结果对比 单位:%

组别	推理无冲突率	CQ 覆盖率
实验组	90.91	80.00
消融组 1	87.20	76.30
消融组 2	85.00	60.00

实验显示,消融组 2 的推理无冲突率停滞在 85.00% 且伴随 CQ 覆盖率下滑,而本文实验组则提升至 90.91% 并维持了较高的 CQ 覆盖率。这一现象与大模型的认知超载机制相关:当单一智能体在同一窗口内处理条令抽取与消解报错双重任务时,易因注意力分散而采取妥协性的破坏性删减,导致知识流失。与之相对,本文架构通过任务解耦促进了认知卸载,建模师智能体侧重抽取,逻辑专家智能体侧重修正。这表明,在神经符号协同框架下,适度的角色分工能在一定程度上缓解单模型的注意力瓶颈,从而在处理逻辑冲突时较好地兼顾本体知识的完整性。

3 结束语

针对 C2 领域本体构建中人工成本高、单一智能体易产生幻觉和逻辑冲突的问题,提出一种基于多智能体认知解耦与符号逻辑客观锚定的神经符号协同构建方法 MACOD。研究表明,单纯依靠大模型难以内化军事条令的绝对互斥约束,而单纯依靠符号推理机又缺乏对非结构化文本的理解力。尽管本研究取得了预期成果,但仍存在多源

条令融合能力不足、智能体动态协作机制有待优化、少量幻觉需人工终审等局限性。未来将重点拓展多源联合作战条令的一体化融合构建,引入协调者智能体与反思调试模块优化动态协作架构,结合检索增强生成(Retrieval-Augmented Generation, RAG)技术进一步抑制幻觉。

关于方法的可扩展性,MACOD 的模块化架构为处理更大规模数据集提供了天然的扩展路径。在横向扩展层面,输入层的文本分块与 CQ 引导机制支持将大规模条令语料按章节或语义单元拆分后并行处理,各分块独立经过多智能体协作流程后,在输出层通过本体对齐与合并操作整合为统一本体,智能体间的交互开销不随语料总量线性增长。在纵向扩展层面,当领域概念规模增大时,Hermit 推理机的一致性检测复杂度将随公理数量增加而上升,这是描述逻辑推理的固有计算瓶颈。对此,可采取分模块增量校验策略,即按子域分别构建并校验局部本体,最终合并时仅对跨域公理执行全局一致性检测,以控制单次推理规模。后续研究将在多篇条令联合构建的场景下验证上述扩展策略的实际效果。

参考文献:

[1] Joint Chiefs of Staff. Joint campaigns and operations: JP 3-0[R]. Washington, USA: Joint Chiefs of Staff, 2022: 71-85.

[2] 邹焯翰,冯畅赫,程光权,等.面向军事条令条例的本体构建技术[J].指挥与控制学报,2019,5(1):47-54.

[3] 齐玉东,丁海强,吴晋豫,等.融合本体特征的 BiLSTM-CRF 军事实体识别模型[J].兵器装备工程学报,2020,41(5):118-123.

[4] 易文安,王亮,曾皓.美军联合作战理论的逻辑体系构建[J].国防科技,2022,43(5):49-56.

[5] 张清辉,杨楠,梁政.任务驱动的军事信息服务知识推理研究[J].火力与指挥控制,2021,46(5):64-70.

[6] 程恺,陈刚,尹成祥,等.作战行动序列核心本体建模及其推理方法[J].系统工程与电子技术,2018,40(4):805-814.

[7] 刘梦超,王玉玫,吴亚非,等.基于本体的军事装备知识建模及分析[J].计算机与现代化,2021(1):76-80.

[8] 阮书鹤,钟林辉,高荣锦,等.面向专家示例的 Stack Overflow 本体构造和推理研究[J].计算机应用研究,2023,40(12):3736-3741.

[9] 黄伟春,肖刚,杨健,等.基于本体的军事术语知识图谱构建方法[J].指挥控制与仿真,2023,45(5):10-17.

[10] 卢银鹏,郭凯威,卢记仓,等.基于上下文学习和语义检索增强的零样本立场检测方法[J].信息工程大学学报,2025,26(2):196-202.

[11] Wang L W, Li R M, Yan Y, et al. InstructionNER: a

- multi-task instruction-based generative framework for few-shot NER [PP/OL]. arXiv (2022-03-08) [2026-02-09]. <https://doi.org/10.48550/arXiv.2203.03903>.
- [12] Joachimiak M P, Miller M A, Caufield J H, et al. The artificial intelligence ontology: LLM-assisted construction of AI concept hierarchies[J]. *Applied Ontology*, 2024, 19(4):408-418.
- [13] Lippolis A S, Saeedizade M J, Keskiärrkkä R, et al. Ontology generation using large language models[M]//Curry E, Acosta M, Poveda-Villalón M, et al. *The semantic web (ESWC 2025)*. Cham, Switzerland: Springer, 2025: 321-341.
- [14] 李文,李秀霞,尹晓甜.基于提示工程的大语言模型实体关系抽取效果对比研究[J/OL]. *图书馆杂志*, 2025-09-12. <https://link.cnki.net/urlid/31.1108.G2.20250912.1600.002>.
- [15] 兰晨曦,沈宗亮,冯建周,等.融合智能体与大语言模型微调的格式化文档生成研究[J/OL]. *计算机工程*, 2026-01-04. <https://link.cnki.net/doi/10.19678/j.issn.1000-3428.0252901>.
- [16] Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-networks [C]//Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP). Stroudsburg, USA: ACL, 2019:3982-3992.
- [17] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks [M]//
- Gangemi A, Navigli R, Vidal M-E, et al. *The semantic web (ESWC 2018)*. Cham, Switzerland: Springer, 2018: 593-607.
- [18] 张晓娟,马乐.基于上下文语义与全局信息的局部引文推荐研究[J]. *情报学报*, 2025, 44(10):1272-1286.
- [19] Jebari C, Herrera-Viedma E, Cobo M J. Context-aware citation recommendation of scientific papers: comparative study, gaps and trends[J]. *Scientometrics*, 2023, 128(8): 4243-4268.
- [20] 吴永和,姜元昊,陈圆圆,等.大语言模型支持的多智能体:技术路径、教育应用与未来展望[J]. *开放教育研究*, 2024, 30(5):63-75.
- [21] 孙毅,郑雨,黄海燕,等.多循环嵌套的大语言模型多智能体指挥控制过程[J]. *指挥与控制学报*, 2024, 10(6): 732-739.
- [22] Glimm B, Horrocks I, Motik B, et al. HermiT: an OWL 2 reasoner[J]. *Journal of Automated Reasoning*, 2014, 53(3):245-269.
- [23] Shinn N, Cassano F, Berman E, et al. Reflexion: language agents with verbal reinforcement learning [PP/OL]. V4. arXiv (2023-10-10) [2026-05-29]. <https://doi.org/10.48550/arXiv.2303.11366>.
- [24] Madaan A, Tandon N, Gupta P, et al. Self-Refine: iterative refinement with self-feedback [C]//Proceedings of the 37th conference on neural information processing systems. Red Hook, USA: Curran Associates Inc., 2023: 46534-46594.
- (编辑:李志豪)
- ~~~~~
- (上接第456页)
- [9] Guo Z R, Xia L H, Yu Y H, et al. LightRAG: simple and fast retrieval-augmented generation [C]//Findings of the association for computational linguistics: EMNLP 2025. Stroudsburg, USA: ACL, 2025: 10746-10761.
- [10] Yu H Q, McQuade F. RAG-KG-IL: a multi-agent hybrid framework for reducing hallucinations and enhancing LLM reasoning through RAG and incremental knowledge graph learning integration. [PP/OL] arXiv (2025-03-14) [2026-01-28]. <https://doi.org/10.48550/arXiv.2503.13514>.
- [11] Zhang Y T, Chen L X, Li S H, et al. Way to specialist: closing loop between specialized LLM and evolving domain knowledge graph [C]//Proceedings of the 31st ACM SIGKDD conference on knowledge discovery and data mining V.1. New York, USA: ACM, 2025:1996-2007.
- [12] Yu D H, Zhang S, Ng P, et al. DecAF: joint decoding of answers and logical forms for question answering over knowledge bases [C]//Proceedings of the 11th international conference on learning representations. Appleton, USA: ICLR, 2023. DOI: 10.48550/arXiv.2210.00063.
- [13] Jiang J H, Zhou K, Zhao W X, et al. UniKGQA: unified retrieval and reasoning for solving multi-hop question answering over knowledge graph [C]//Proceedings of the 11th international conference on learning representations. Appleton, USA: ICLR, 2023. DOI: 10.48550/arXiv.2212.00959.
- [14] Liu Y, Yavuz S, Meng R, et al. Uni-parser: unified semantic parser for question answering on knowledge base and database [C]//Proceedings of the 2022 conference on empirical methods in natural language processing. Stroudsburg, USA: ACL, 2022:8858-8869.
- [15] Lan Y S, Wang S H, Jiang J. Knowledge base question answering with topic units [C]//Proceedings of the twenty-eighth international joint conference on artificial intelligence. California, USA: IJCAI, 2019: 5046-5052.
- [16] Cao S L, Shi J X, Yao Z J, et al. Program transfer for answering complex questions over knowledge bases [C]//Proceedings of the 60th annual meeting of the association for computational linguistics. Stroudsburg, USA: ACL, 2022:8128-8140.
- [17] Chen Y R, Li H Y, Qi G L, et al. Outlining and filling: hierarchical query graph generation for answering complex questions over knowledge graphs [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(8): 8343-8357.
- [18] 杨大伟,周刚,卢记仓,等.基于知识表示学习的知识图谱补全研究综述[J]. *信息工程大学学报*, 2021, 22(5): 558-565.
- (编辑:李志豪)