

面向大模型智能体的动态本体技术：概念、实现框架与展望

赵翔，曾维新，廖增华，刘丽华，黄宏斌

国防科技大学大数据与决策国家级重点实验室，湖南 长沙 410073

摘要

随着大模型应用由知识问答向智能体任务执行拓展，复杂业务环境下任务执行的可靠性与可控性成为亟待解决的问题。动态本体通过统一组织对象语义与执行语义，为贯通业务数据、任务理解与任务执行提供了新的技术路径。围绕动态本体技术，系统分析了其从静态概念建模、本体演化更新、过程语义建模向动态本体协同的发展脉络，概述了由数据层、动态本体层和工作流层构成的技术实现框架，并结合执行示例说明了动态本体的作用机制。在此基础上，从动态本体构建、异构数据语义映射以及动态本体与工作流协同演化等三个方面探讨了关键研究问题，并对未来发展方向进行展望。

关键词

动态本体；大模型智能体；工作流；任务执行

中图分类号：TP391

文献标志码：A

Dynamic ontology technology for large language model agents: concepts, implementation framework, and prospects

ZHAO Xiang, ZENG Weixin, LIAO Zenghua, LIU Lihua, HUANG Hongbin

National Key Laboratory of Big Data and Decision, National University of Defense Technology, Changsha 410073, China

Abstract

As large language model applications expand from knowledge question answering to agent-based task execution, the reliability and controllability of task execution in complex business environments have become urgent issues. By uniformly organizing object semantics and execution semantics, dynamic ontology provides a new technical path for connecting business data, task understanding, and task execution. Focusing on dynamic ontology technology, its development trajectory from static conceptual modeling, ontology evolution and updating, and process semantic modeling to dynamic

ontology collaboration was systematically analyzed. A technical framework consisting of the data layer, dynamic ontology layer, and workflow layer was outlined, and the mechanism of dynamic ontology was illustrated through an execution case. On this basis, key research issues were discussed from three perspectives: dynamic ontology construction, semantic mapping of heterogeneous data, and the collaborative evolution of dynamic ontology and workflows, followed by an outlook on future research directions.

Key words

dynamic ontology, large language model agent, workflow, task execution

1 引言

近年来,大语言模型(large language model, LLM, 以下简称大模型)的能力边界持续拓展。早期大模型主要应用于自然语言理解与文本生成^[1-2]。随着模型规模扩大、训练数据增长和学习方法演进,其能力逐步扩展至知识检索^[3]、代码生成^[4]、多模态处理^[5]、复杂推理^[6]与任务规划^[7-10]等方面。上述能力的发展使大模型具备了执行真实任务的潜力,并逐渐成为支撑智能体感知、理解、决策与行动的通用能力底座。

大模型能力的增强推动了其应用形态由知识问答向任务执行转变。知识问答场景主要围绕用户问题完成信息检索与答案生成,典型流程为“查询—检索—回答”,重点关注事实准确性、知识可追溯性和语义一致性^[3]。随着大模型逐步成为智能体的核心决策模块,其应用目标由回答问题逐渐转向完成特定任务^[11-12]。在这一阶段,系统不仅需要理解用户目标,还需要进行任务规划、调用外部工具^[13],并根据执行结果持续调整行动。因此,大模型应用的典型流程逐步转变为“目标—规划—行动—反馈”,系统关注重点也相应扩展至对象识别、执行约束、流程组织和反馈

闭环。

然而,大模型智能体从能够执行任务到稳定完成任务仍存在明显差距,突出表现为复杂、动态业务环境下任务执行可靠性不足,难以持续生成符合语义约束和业务规则的行动^[14-15]。首先,对象语义与任务执行难以稳定绑定,智能体可能无法准确识别任务涉及的业务对象、角色及其关联关系。其次,操作执行条件缺乏显式约束,工具调用的作用对象、输入输出、前置条件、业务规则和结果校验尚未得到统一描述,容易引发误调用或越界执行。最后,执行过程难以形成有效闭环,状态跟踪、异常纠偏与经验沉淀缺乏衔接,制约了任务再规划和执行经验复用。因此,提升大模型智能体任务执行的可靠性,仍需重点解决对象绑定、执行约束与流程闭环等关键问题。

动态本体为解决上述问题提供了新的语义支撑机制。动态本体是指在传统本体所表达的对象语义基础上,进一步引入函数、操作等执行语义,并通过工作流组织任务执行过程的语义运行中枢。其“动态性”主要体现在两个层面:一是对象层面的状态转移,即定义并支撑对象从一种状态向另一种状态变化;二是任务执行过程的动态反馈调整,即依据执行反馈持续调整后续工作流。具体而言,在对象层面,动态本体将多源异构数据组织为语义明确的业务对象,实现任务与对象的稳定绑定。

在任务执行层面，动态本体通过定义函数和操作的作用对象、输入输出、前置条件及业务规则，为工具调用提供语义约束；同时，根据业务状态和执行结果持续跟踪任务进展并调整后续 workflows，支撑异常纠偏和经验复用。由此，动态本体贯通数据、任务与执行，成为支撑大模型智能体可靠任务执行与闭环演化的语义运行中枢。

围绕大模型智能体时代的动态本体技术，本文分析了动态本体的概念与发展脉络，概述了由数据层、动态本体层和工作流层构成的技术实现框架，并结合执行示例说明其在任务执行中的作用机制；最后，探讨动态本体面临的关键研究问题及未来发展方向。

2 动态本体技术演进

本体(ontology)的概念源于哲学中的本体论，旨在研究事物的存在本质及其分类与组织方式。随着本体被引入计算机与人工智能领域，其内涵逐步转化为对领域概念的系统化表达^[16-17]。如图 1 所示，从相关研究与实践的发展脉络看，本体建模范式可概括为静态概念建模、本体演化更新、过程语义建模和动态本体协同四个阶段。其建模范围由对象语义逐步扩展至过程语义和执行语义，功能定位也由知识描述逐步转向支撑系统运行与任务协同。

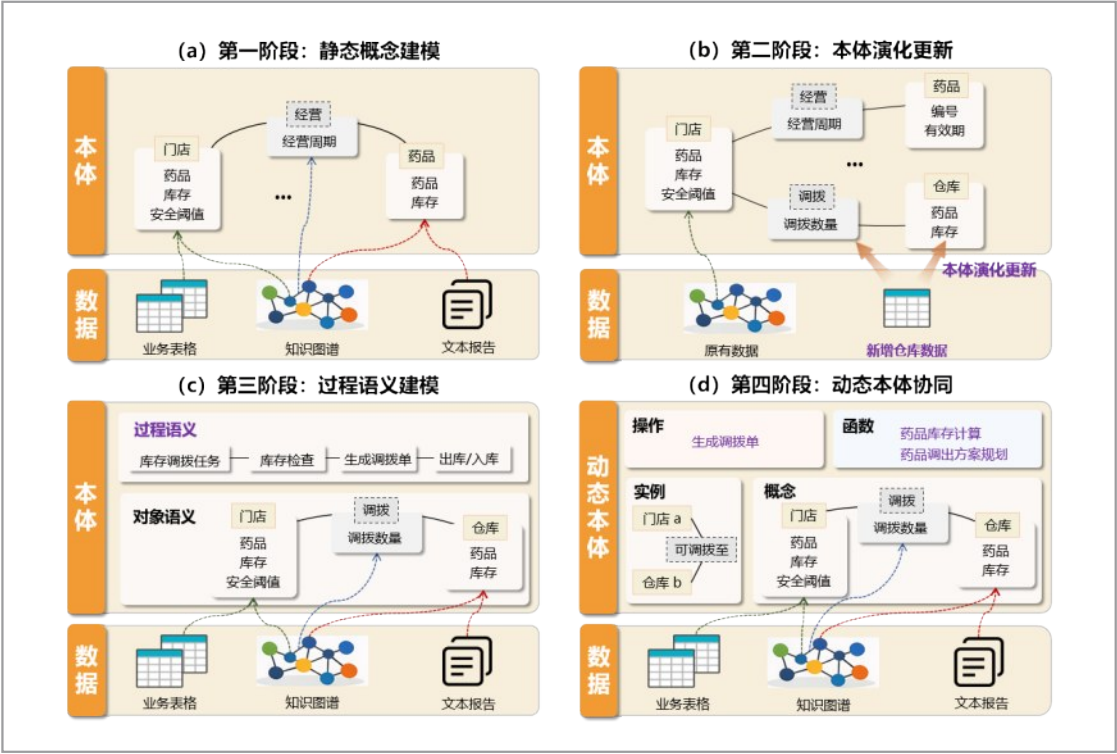


图 1 本体建模范式的四阶段演进

2.1 静态概念建模

静态概念建模旨在对领域知识进行统一、规范的语义描述，是本体建模的基础形态。本体最初被定义为“概念化的显式规范说明”^[16]，后续研究进一步强调其形式化与共享性^[17]。该阶段主要围绕对象、属性和关系等语义要素展开：对象用于抽象具有共同特征的领域实体，属性用于描述对象特征，关系用于刻画对象之间的语义关联。如图 1(a)所示，在药品零售领域，可以定义门店、药品等对象类型，使用药品库存和经营周期等属性描述其状态，并通过经营等关系表达对象之间的关联。在此基础上，业务表格、知识图谱和文本报告等多源异构数据可映射至统一的领域语义结构，为知识组织、语义查询和基础推理提供支撑^[18]。

静态概念建模通常以相对稳定的领域知识为前提，主要回答领域中“包含哪些对象”以及“对象之间具有何种关系”。当新的对象、属性或关系出现时，既有本体难以及时适应领域数据的变化。同时，该阶段尚未将任务过程和执行语义纳入建模范围，因而难以描述任务如何组织和执行。总体而言，静态概念建模为系统建立了基本的语义骨架，但其对领域变化和动态任务执行的支撑能力仍然有限。

2.2 本体演化更新

随着领域知识、业务需求和数据结构持续变化，一次性构建的静态本体难以长期保持完整性与适用性。本体演化更新逐渐成为本体研究的重要方向，其核心是在保持既有语义连续性的基础上，根据领域知识、业务需求和数据的变化对本体进行持续调整^[19-20]。本体演化更新通常包括变更识别、语义要素更新、一致性校验和版

本管理等环节，具体涉及对象、属性和关系的增加、修改或删除^[21-22]。如图 1(b)所示，当药品零售场景接入新的仓库数据时，可以在门店、药品等既有对象的基础上，新增仓库对象和调拨关系，并建立门店、药品与仓库之间的语义关联，从而使本体能够适应领域知识与数据结构的变化。该阶段推动本体由一次性建模产物转变为可持续维护和更新的语义模型。

该阶段的动态性主要体现为对象、属性和关系等语义结构随领域知识变化而持续调整，尚未将任务过程和操作执行逻辑纳入建模范围。因此，本体演化更新虽然增强了本体对领域变化的适应能力，但仍难以描述业务任务的组织与执行。

2.3 过程语义建模

随着业务任务复杂性不断提高，仅对领域对象及其结构变化进行建模，已难以满足任务组织与执行逻辑的表达需求。过程语义建模由此产生，其核心是将任务、操作、执行条件及其依赖关系纳入本体表示，推动本体由描述领域对象及其关系拓展至描述任务如何组织与执行。OWL-S^[23]和 WSMO^[24]等研究将服务功能、输入输出、前置条件、执行效果及组合过程引入本体表示，为网络服务的发现、调用与组合提供了语义基础。过程语义建模主要围绕任务分解、操作描述和流程组织展开，通过顺序、分支、循环和依赖等关系，将多个操作组织为完整的任务过程。如图 1(c)所示，库存调拨任务可以分解为库存检查、生成调拨单和出库/入库等步骤，并描述各步骤之间的执行顺序和依赖关系。由此，本体的建模范围由对象及其结构变化扩展至任务过程，具备了描述流程结构与执行逻辑的能力。

该阶段的动态性主要体现为将任务步骤、执行顺序和依赖关系纳入本体表示，使本体能够刻画任务过程及其状态变化。然而，过程语义与对象语义仍相对分离，任务步骤尚未与具体业务对象、底层数据和实时运行状态建立统一关联，相关流程通常也需要预先定义。因此，其动态性仍以任务过程语义描述为主，尚未实现数据、对象、任务与执行的贯通。

2.4 动态本体协同

过程语义建模虽然将任务过程纳入本体表示，但对象语义、过程语义与任务执行之间仍未实现有效贯通。为协同支撑业务理解与任务执行，本体建模视角开始由专家驱动的静态知识描述转向面向领域业务的语义组织与运行支撑。动态本体协同由此产生，其核心并非简单扩展本体的表达范围，而是围绕业务运行所需的概念、实例、函数、操作和工作流构建统一的语义组织框架，推动本体由知识与过程的描述工具演进为连接业务数据与任务执行的运行中枢。

以 Palantir Foundry Ontology 为代表的工业实践将对象、属性和关系等对象语义与函数、操作等执行语义进行统一组织。如图 1(d)所示，在药品零售领域，除定义门店、药品和仓库等对象，以及门店 a、仓库 b 等实例外，还可以封装药品库存计算、药品调出方案规划等函数，并定义生成调拨单等操作，从而为任务解析、能力匹配与执行编排提供统一的语义依据。

相较于过程语义建模，动态本体协同的动态性进一步体现为对象语义与执行语义的协同联动：函数和操作的调用以及工作流的运行支撑对象由一种状态向另一种状态转移；任务执行结果则作为反馈，为

后续工作流的动态调整提供依据。围绕具体业务需求，动态本体统一组织业务对象、运行状态、执行能力及其约束关系，使任务与对象、操作规则和执行过程建立关联，从而形成“任务执行—结果反馈—流程调整”的运行闭环。由此，动态本体实现了由结构演化和过程描述向运行协同的拓展，贯通数据、对象、任务与执行，成为支撑业务运行和大模型智能体任务协同的语义运行中枢。

3 动态本体技术的实现框架

动态本体位于领域数据与业务之间，是连接多源业务数据与业务工作流的语义中间层。如图 2 所示，其技术实现框架由数据层、动态本体层和工作流层构成。数据层承载多源异构数据；动态本体层将底层数据组织为统一的业务对象，并通过概念、实例、函数和操作表达对象语义与执行语义；工作流层则基于动态本体提供的语义要素与执行约束，完成任务解析、执行调度和反馈演化。通过三层协同，该框架实现了领域数据、业务语义与任务执行的贯通，为大模型智能体提供语义明确、规则可控、过程闭环的任务执行支撑。

3.1 数据层：多源异构数据

数据层是动态本体技术实现框架的基础，主要承载与业务运行相关的原始数据和知识资源。如图 2 底部所示，数据层包括业务表格、数据库等结构化数据，知识图谱等图结构数据，以及文档、日志等半结构化或非结构化数据。不同类型的数据从业务对象、关联关系和状态变化等维度记录业务运行情况，为动态本体构建提供

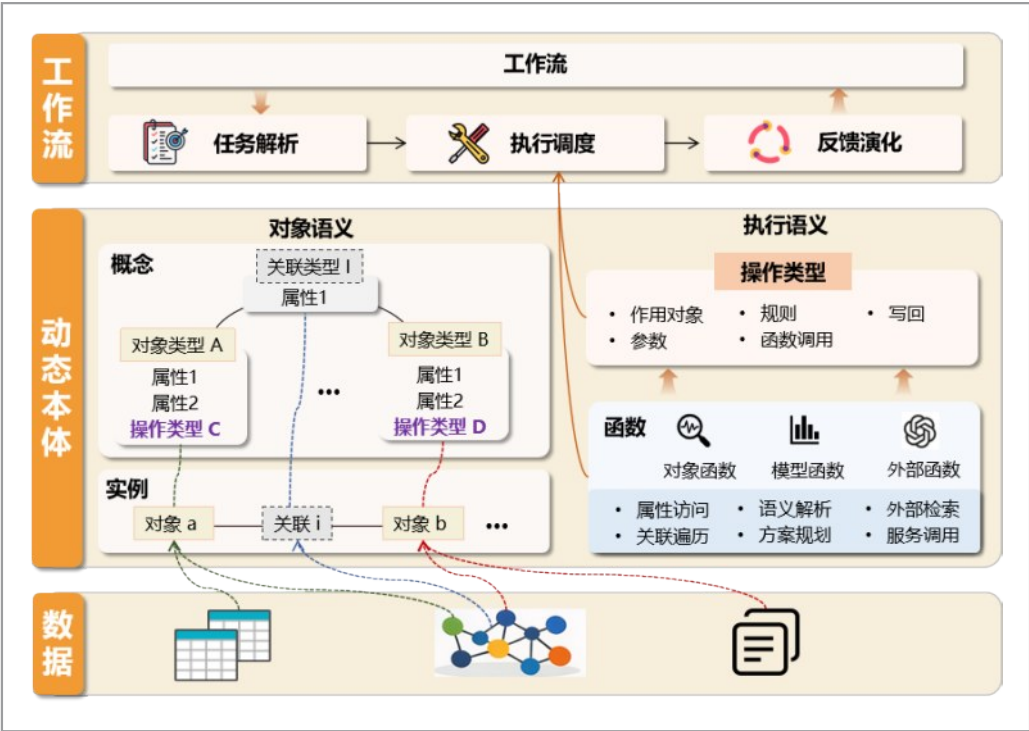


图 2 动态本体技术实现框架

数据来源。

然而，多源业务数据通常分散于不同系统，数据模式、表示形式和语义粒度存在差异。同一业务对象的信息可能分布在多个数据源中，不同数据源也可能使用不同名称或结构表示相同概念。因此，原始数据难以直接形成统一、完整的业务对象表示，需要通过动态本体层的语义映射与实体对齐，将分散、异构的数据组织为具有明确对象、属性、关系和状态的业务实例^[25]，为上层任务解析与执行编排提供数据基础。

3.2 动态本体层:语义运行中枢

动态本体层是整个技术实现框架的核心，承担底层数据语义化和执行能力组织两项功能。如图 2 中部所示，该层由对象语义和执行语义构成。对象语义包括概念

和实例，其中，概念用于定义对象、属性和语义关联，实例由底层数据映射形成，用于表示具体业务对象及其运行状态。由此，分散的数据记录被组织为具有明确业务含义的对象语义。

执行语义主要由函数和操作构成。函数用于封装属性访问、关系遍历、模型调用和外部服务等基础能力；操作则面向具体业务行为，可定义作用对象、输入参数、执行规则、函数调用和状态写回等内容。通过将函数、操作与相应对象关联，动态本体能够明确“对什么对象、在何种条件下、调用何种能力、产生何种结果”，为任务执行提供统一的语义依据和规则约束。由此，动态本体层实现了对象语义与执行语义的协同组织，成为连接领域数据与任务工作流的语义运行中枢。

3.3 工作流层:任务编排与执行

工作流层主要负责将用户需求和业务目标转化为可执行的任务过程。如图 2 顶部所示，该层包括任务解析、执行调度和反馈演化三个环节。工作流层不直接处理原始数据，而是基于动态本体层提供的概念、实例、函数、操作，对任务步骤及其依赖关系进行组织与编排^[26]。

任务解析负责识别任务目标、业务对象、具体实例、输入参数和执行约束，并将其映射至相应函数、操作或已有工作流；执行调度依据任务步骤及其依赖关系调用相关函数和操作，同时读取业务实例状态并写回执行结果；反馈演化则跟踪任务执行状态，根据执行结果识别异常、调整工作流，并将验证有效的操作和工作流沉淀为可复用的执行能力。三者相互衔接，形成从任务理解、执行控制到反馈演化的闭环。

3.4 动态本体任务执行示例

为说明动态本体在任务执行中的作用机制，以药店毛利波动分析为例，分析其在新任务场景下对语义组织、能力匹配与流程生成的支撑作用。如图 3 所示^①，动态本体中已定义门店、药品和经营等概念，形成门店 a、药品 d 和经营周期等业务实例；在执行语义方面，系统已封装计算销售收入、查询药品成本和计算药品毛利等函数，但尚未形成面向毛利波动归因的专用操作和工作流。

当用户提出“分析本周药店综合毛利率下降的原因”时，系统首先识别门店、药品、时间范围和毛利下降归因等任务要素，并将其映射至动态本体中的相关概念、实例和函数。由于现有工作流中不存在匹配的执行流程，工作流层基于对象语义与

执行语义，组合已有函数，构建药品毛利计算和定位异常药品临时操作，并组织为“药店毛利波动分析”临时工作流。在执行完成后，系统对相关操作和工作流进行验证，并将有效的操作与工作流沉淀为可复用的执行能力。该执行示例表明，动态本体能够通过任务解析、操作构建、执行验证、经验沉淀等步骤，支撑大模型智能体在新任务场景下的动态任务执行。

4 动态本体技术的关键问题

尽管动态本体为大模型智能体的任务执行提供了新的语义支撑机制，但其构建与应用仍面临三个关键问题：一是动态本体包含多类语义要素，人工构建成本较高，难以适应业务与任务的持续扩展；二是业务数据来源多样、结构异构，难以准确映射至统一的概念与实例；三是动态本体与工作流均会随新任务和执行反馈不断变化，二者之间尚缺乏有效的协同演化机制。基于此，本章从动态本体构建、异构数据语义映射以及动态本体与工作流协同演化三个方面探讨关键研究问题及未来研究方向。

4.1 动态本体自动构建

动态本体构建的核心任务是形成兼具稳定语义体系与动态执行能力的本体模型。现有自动化本体学习研究主要关注术语发现、概念识别、分类体系构建和关系抽取等对象语义要素^[27-28]，对函数、操作、执行条件和状态变化等执行语义的建模仍不充分。同时，动态本体涉及概念、实例、函数、操作和约束等多类要素，完全依赖

①为保证执行示例易于聚焦和理解,技术实现框架的数据层被进行了半透明化处理。

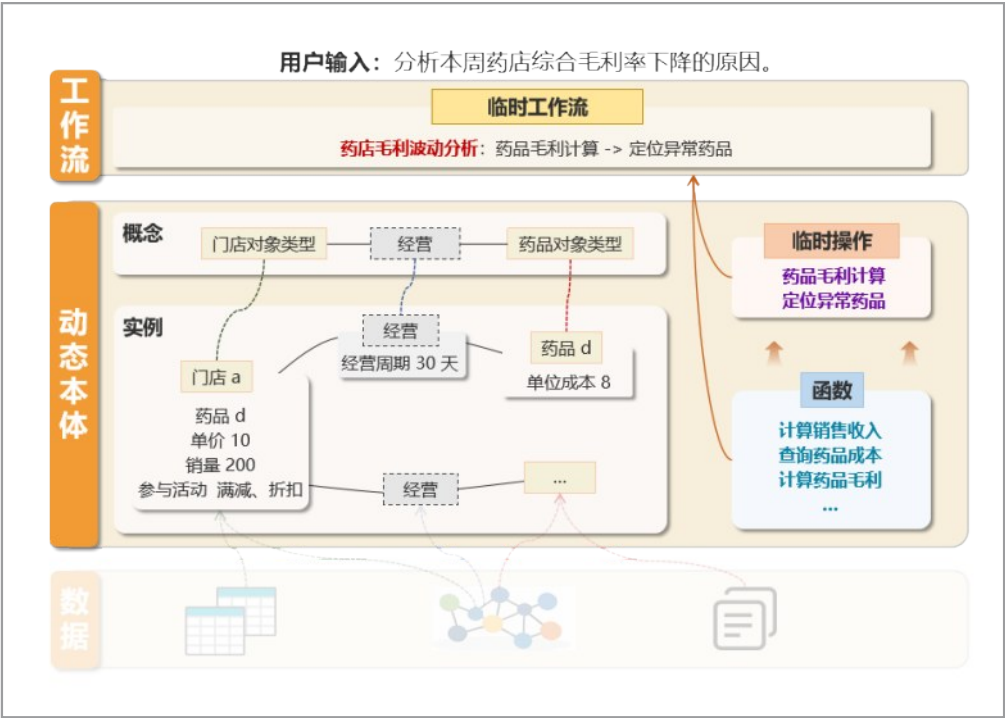


图 3 面向药店毛利波动分析新任务的执行过程

领域专家不仅建模成本较高，也难以适应数据、业务和任务的持续变化。因此，未来可重点关注以下三个研究方向。

(1) 迭代式动态本体构建

动态本体包含概念、实例、函数、操作等多类语义要素，难以在初始阶段实现全面、准确的建模，需要在多轮迭代中不断补充、修正和扩展。未来应构建生成、检查、修正的迭代构建机制，根据新增数据、业务需求和实际应用反馈，识别缺失或冲突的语义要素，持续优化本体结构，从而逐步提升动态本体的构建质量及其对业务任务的覆盖能力。

(2) 神经—符号协同构建

利用大模型开展术语发现、关系归纳、结构补全以及函数和操作生成，能够提高

动态本体的构建效率，但受概率生成机制影响，容易产生概念漂移、关系不稳定和约束不严谨等问题^[29-30]。依赖领域专家和符号规则虽然能够提供明确的语义与逻辑约束，但面临人工成本较高、构建周期较长和扩展效率有限等不足^[31]。因此，单独采用神经方法或符号方法均难以兼顾构建效率、语义准确性与逻辑一致性。

未来可探索神经方法与符号方法协同的构建机制，将用户输入、业务数据及已有本体共同输入大模型，通过意图识别和要素抽取，将模糊意图转化为任务目标、业务对象、属性、关系和约束等结构化信息；对于无法与已有本体匹配但具有上下文或数据证据支持的候选实体，进一步推断其对象类型、属性、关系及相关函数和操作，形成新业务对象候选。该过程可能因意图理解偏差或模型幻觉产生虚构对象、类型误判和关系冲突，并将错误传递至后

续 workflow。因此，新业务对象候选不应直接写入本体，而应通过符号规则进行一致性校验、逻辑约束和结果修正，并对存在歧义或置信度较低的结果进行专家复核。由此形成“大模型生成—符号校验—本体更新”的协同构建闭环，从而兼顾构建效率、语义准确性与逻辑一致性。

(3) 任务驱动的本体结构生成

传统本体构建主要面向领域知识的系统化表达，侧重组织对象、属性和关系等语义要素；而动态本体还需支撑具体任务的组织与执行，其结构应围绕具体任务需求进一步组织。未来可根据任务目标、执行步骤和能力需求，反向识别所需的对象、属性、关系、函数、操作，并建立任务过程与相关语义要素之间的关联，使本体结构与实际业务过程紧密结合，为工具调用、工作流生成和任务执行提供直接支撑。

4.2 异构数据语义映射

异构数据语义映射的核心是将来源不同、结构各异且语义粒度不一致的业务数据映射至统一的本体语义空间。现实业务数据通常分散于业务表格、知识图谱、文档日志等多种载体，同一现实实体的信息也可能分布在不同数据源中。若无法建立稳定、准确的语义映射，动态本体将难以形成完整、一致的业务实例，进而影响对象识别、语义关联和任务调用。因此，未来可重点关注以下三个研究方向。

(1) 以实体为中心的语义映射机制

传统数据映射主要关注字段名称、数据类型和模式结构之间的对齐，而以实体为中心的数据管理^[32]强调围绕现实实体组

织和关联多源数据，有助于将分散于不同数据源的信息聚合为完整的业务实例。在真实业务环境中，同一现实实体的信息可能同时存在于表格、文档、日志和知识图谱中，仅依靠字段匹配难以形成统一、稳定的实体语义。未来应以诸如门店、药品和经营活动等现实实体为中心，对多源信息进行识别、关联与融合，并将其映射至统一的对象语义空间，从而形成完整、一致的业务实例，为对象识别、关系分析和任务调用提供支撑。

(2) 面向连续场景的实体语义集成

真实业务数据并非一次性静态导入，而是随着系统运行持续接入和变化。新增记录、活动和运行日志可能不断补充或修改既有实体的属性、关系和状态。若每次数据变化均重新进行全量映射，将产生较高的计算和维护成本。未来应研究面向连续数据环境的增量式实体语义集成机制，围绕已有业务实例持续开展新数据识别、属性补充、关系更新和上下文整合，使动态本体能够及时、准确地反映业务状态变化。

(3) 跨模态联合语义对齐

现实业务数据除结构化表格外，还包括文本、日志、知识图谱和图像等多种模态，不同模态的数据在表示形式和语义粒度方面存在显著差异。未来应探索跨模态共享语义表示与联合对齐方法^[33]，使同一现实实体在不同数据模态下能够被一致识别和关联。通过跨模态联合语义对齐，动态本体可以接入更加丰富的数据来源，形成更为完整的业务实例，为复杂场景下的任务理解与执行提供语义基础。

4.3 动态本体与工作流的协同演化

动态本体与工作流分别承担语义组织和任务编排功能。动态本体通过概念、实例、函数、操作及约束描述业务对象与执行能力，工作流则基于这些语义要素组织任务步骤和执行逻辑。二者并非相互独立：动态本体的变化会影响工作流的组织与执行，新任务及其执行反馈也会推动动态本体更新。因此，需进一步推动动态本体与工作流相互驱动、协同演化。未来可重点关注以下三个研究方向。

(1) 动态本体—工作流双向反馈机制

现有系统主要由动态本体单向提供概念、实例、函数、操作，工作流据此完成任务编排，二者之间尚缺乏有效的双向反馈。在实际执行过程中，新业务实例、任务需求和执行约束的出现，既可能要求调整工作流结构，也可能推动动态本体更新。未来应建立本体更新、工作流重构与执行验证之间的联动机制：一方面，根据动态本体的变化识别受影响的工作流，并调整相应的任务步骤和执行条件；另一方面，从工作流执行过程及结果中提取新的语义要素和约束，反向补充或修正动态本体，从而保持语义结构与任务流程的一致性。

(2) 面向新任务的函数与操作构建

面对现有动态本体和工作流尚未覆盖的新任务，仅依赖预定义函数和操作难以持续满足执行需求。未来应根据任务目标和执行要求识别缺失能力，优先组合已有函数形成新的操作，并在必要时生成或接入新的函数与操作^[34]。新构建的执行能力经有效性与安全性验证后，可沉淀至动态本体，相关工作流则可固化为可复用工作

流，从而推动系统由依赖固定执行能力向持续扩展执行能力演进。

(3) 面向实时高并发任务的分层增量式协同调度

在实时高并发业务环境中，动态本体的语义检索与状态更新、工作流的生成与重构以及大模型推理等环节可能被频繁触发，由此带来额外的计算开销并增加任务调度延迟。未来可构建分层增量式协同调度机制：一是缓存高频语义查询结果并预编排可复用工作流，仅对受状态变化影响的局部本体要素和流程节点进行增量更新，以减少全局推理与流程重构；二是依据任务复杂度和实时性要求，对符号规则、轻量模型与大模型进行分级调度，并结合异步更新、并行执行和弹性资源分配提升系统吞吐量。此外，应综合评估响应延迟、吞吐量、资源消耗、语义一致性和任务成功率，在实时响应与可靠执行之间取得平衡。

5 总结与展望

随着大模型应用由知识问答向智能体任务执行拓展，如何准确识别业务对象、规范组织执行能力并形成任务闭环，已成为智能体落地应用的关键问题。动态本体通过统一组织对象语义与执行语义，并与工作流协同运行，能够贯通多源业务数据、领域语义与任务执行，为任务解析、执行调度和反馈演化提供语义支撑，已成为保障大模型智能体任务执行可靠性、可控性与可复用性的重要基础设施。

尽管动态本体已展现出良好的应用前景，但研发仍处于初期探索阶段，许多潜力尚未得到充分挖掘和释放。下一步，随

着大模型能力、语义建模方法和智能体执行机制的持续发展，动态本体技术有望进一步提升复杂任务的理解与可靠执行能力，

为大模型智能体在真实业务环境中的规模化应用提供更好更可靠的支撑。



赵翔(1986-),男,博士,国防科技大学大数据与决策国家级重点实验室研究部主任,教授、博导。青年长江学者,湖南省科技创新领军人才,军队青年科技英才。主要研究方向为知识图谱、高级数据管理、世界模型。



曾维新(1995-),男,博士,国防科技大学大数据与决策国家级重点实验室副教授,湖南省芙蓉青年人才,主要研究方向为知识图谱、数据管理。



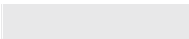
廖增华(2000-),男,2022年获得国防科技大学学士学位,现为国防科技大学博士研究生,研究方向为知识图谱、大模型智能体、智能交互。



刘丽华(1981-),女,博士,国防科技大学大数据与决策国家级重点实验室主任,教授、博导,军队学科拔尖人才,主要研究方向为数据工程、人工智能。



黄宏斌(1975-),男,博士,国防科技大学大数据与决策国家级重点实验室研究员、博导,享受国务院特殊津贴,主要研究方向为数据工程与知识工程、数据挖掘与智能决策。



参考文献：

- [1] Brown T B, Mann B, Ryder N, et al. Language models are few-shot learners [C]//34th Annual Conference on Neural Information Processing Systems, Red Hook: Curran Associates, 2020: 1877–1901.
- [2] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer [J]. Journal of Machine Learning Research, 2020, 21(140): 1–67.
- [3] Lewis P, Perez E, Piktus A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[C]//34th Annual Conference on Neural Information Processing Systems, Red Hook: Curran Associates, 2020: 9459–9474.
- [4] Li Y, Choi D, Chung J, et al. Competition-level code generation with AlphaCode[J]. Science, 2022, 378(6624): 1092–1097.
- [5] Alayrac J B, Donahue J, Luc P, et al. Flamingo: a visual language model for few-shot learning[C]//36th Annual Conference on Neural Information Processing Systems, Red Hook: Curran Associates, 2022: 23716–23736.
- [6] Wei J, Wang X, Schuurmans D, et al. Chain-of-thought prompting elicits reasoning in large language models[C]//36th Annual Conference on Neural Information Processing Systems, Red Hook: Curran Associates, 2022: 24824–24837.
- [7] Huang W, Abbeel P, Pathak D, et al. Language models as zero-shot planners: extracting actionable knowledge for embodied agents[C]//39th International Conference on Machine Learning, Cambridge: PMLR, 2022: 9118–9147.
- [8] Li J W, Gao Y, Yang Y Z, et al. Fundamental capabilities and applications of large language models: a survey[J]. ACM Computing Surveys, 2026, 58(2): 1–42.
- [9] 陶伟, 王健宗, 张旭龙, 等. 大语言模型长文本推断优化技术综述[J]. 大数据, 2025, 11(6): 72–94.
- Tao W, Wang J Z, Zhang X L, et al. Long-context inference optimization for large language models: a survey[J]. Big Data Research, 2025, 11(6): 72–94.
- [10] 许小颖, 廖文杰, 王瀚林. 大语言模型在推荐系统中的应用研究综述[J]. 大数据, 2026, 12(3): 136–149.
- Xu X Y, Liao W J, Wang H L. Survey on the applications of large language models in recommender systems[J]. Big Data Research, 2026, 12(3): 136–149.
- [11] Xi Z H, Chen W X, Guo X, et al. The rise and potential of large language model based agents: a survey[J]. Science China Information Sciences, 2025, 68(2): 121101.
- [12] 姜春宇, 韩晓璐, 王超伦. 智能体技术演进: 场景分析、关键技术与发展趋势[J]. 大数据, 2026, 12(2): 129–138.
- Jiang C Y, Han X L, Wang C L. Evolution of agent technology: scenario analysis, key technologies and development trends [J]. Big Data Research, 2026, 12(2): 129–138.
- [13] Yao S Y, Zhao J, Yu D, et al. ReAct: synergizing reasoning and acting in language models[C]//11th International Conference on Learning Representations, Amherst: OpenReview, 2023: 1–33.
- [14] Liu X, Yu H, Zhang H, et al. Agent-Bench: evaluating LLMs as agents[C]//12th International Conference on Learning Representations, Amherst: OpenReview, 2024: 52989–53046.
- [15] Yao S Y, Shinn N, Razavi P, et al. τ -bench: a benchmark for tool-agent-

- user interaction in real-world domains [C]//13th International Conference on Learning Representations, Amherst: OpenReview, 2025: 9965–10017.
- [16] Gruber T R. A translation approach to portable ontology specifications[J]. Knowledge Acquisition, 1993, 5(2): 199–220.
- [17] Studer R, Benjamins V R, Fensel D. Knowledge engineering: principles and methods[J]. Data & Knowledge Engineering, 1998, 25(1–2): 161–197.
- [18] 徐杨, 王晓峰, 何清漪. 物联网环境下多智能体决策信息支持技术[J]. 软件学报, 2014, 25(10): 2325–2345.
- Xu Y, Wang X F, He Q Y. Internet of Things based information support system for multi-agent decision[J]. Journal of Software, 2014, 25(10): 2325–2345.
- [19] Noy N F, Klein M. Ontology evolution: not the same as schema evolution[J]. Knowledge and Information Systems, 2004, 6(4): 428–440.
- [20] Flouris G, Manakanatas D, Kondylakis H, et al. Ontology change: classification and survey[J]. The Knowledge Engineering Review, 2008, 23(2): 117–152.
- [21] Plessers P, De Troyer O. Ontology change detection using a version log[C]//4th International Semantic Web Conference, Berlin: Springer, 2005: 578–592.
- [22] Gro A, Pruski C, Rahm E. Evolution of biomedical ontologies and mappings: overview of recent approaches[J]. Computational and Structural Biotechnology Journal, 2016, 14: 333–340.
- [23] Martin D, Burstein M, McDermott D, et al. Bringing semantics to Web services with OWL-S[J]. World Wide Web, 2007, 10(3): 243–277..
- [24] Roman D, Keller U, Lausen H, et al. Web service modeling ontology[J]. Applied Ontology, 2005, 1(1): 77–106.
- [25] Sun Z Q, Zhang Q H, Hu W, et al. A benchmarking study of embedding-based entity alignment for knowledge graphs[J]. Proceedings of the VLDB Endowment, 2020, 13(11): 2326–2340.
- [26] Van der Aalst W M P, Ter Hofstede A H M, Kiepuszewski B, et al. Workflow patterns[J]. Distributed and Parallel Databases, 2003, 14(1): 5–51.
- [27] Lo A, Jiang A Q, Li W, et al. End-to-end ontology learning with large language models[C]//38th Annual Conference on Neural Information Processing Systems, Red Hook: Curran Associates, 2024: 87184–87225.
- [28] Babaei Giglou H, D' Souza J, Auer S. LLMs4OL: large language models for ontology learning[C]//22nd International Semantic Web Conference, Cham: Springer, 2023: 408–427.
- [29] Ji Z W, Lee N, Frieske R, et al. Survey of hallucination in natural language generation[J]. ACM Computing Surveys, 2023, 55(12): 1–38.
- [30] 刘泽垣, 王鹏江, 宋晓斌, 等. 大语言模型的幻觉问题研究综述[J]. 软件学报, 2025, 36(3): 1152–1185.
- Liu Z Y, Wang P J, Song X B, et al. Survey on hallucinations in large language models[J]. Journal of Software, 2025, 36(3): 1152–1185.
- [31] Maedche A, Staab S. Ontology learning for the Semantic Web[J]. IEEE Intelligent Systems, 2001, 16(2): 72–79.
- [32] Shen Y Y, Wang X S, Du X Y, et al. Entity-centric data management for the ubiquitous computing era[J]. Science China Information Sciences, 2026, 69(5): 156101.
- [33] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision[C]//38th International Conference on Ma-

chine Learning, Cambridge: PMLR, 2021: 8748–8763.	International Conference on Learning Representations, Amherst: OpenReview, 2024: 54067–54089.
[34] Cai T, Wang X, Ma T, et al. Large lan - guage models as tool makers[C]//12th	

录用日期: XXXX-XX-XX
通信作者: 曾维新(zengweixin13@nudt.edu.cn)
基金项目: 国家自然科学基金资助项目 (Nos. U25B2047, 62272469, 62302513); 湖南省科技创新计划 (No. 2026RC3144)
Foundation Items: The National Natural Science Foundation of China (Nos. U25B2047, 62272469, 62302513), The Science and Technology Innovation Program of Hunan Province (No. 2026RC3144)

