

# 大模型与知识图谱

陈玉博<sup>1,2</sup>, 郭少茹<sup>1</sup>, 刘康<sup>1,2,3</sup>, 赵军<sup>1,2</sup>

<sup>1</sup>中国科学院自动化研究所复杂系统认知与决策实验室

<sup>2</sup>中国科学院大学人工智能学院

<sup>3</sup>北京智源人工智能研究院

{yubo.chen, shaoru.guo, kliu, jzhao}@nlpr.ia.ac.cn

## 摘要

知识图谱作为一种重要的知识组织形式，常被视为下一代人工智能技术的基础设施之一，引起了工业界和学术界的广泛关注。传统知识图谱表示方法主要使用符号显式地描述概念及其之间的结构关系，具有语义清晰和可解释性好等特点，但其知识类型有限，难以应对开放域应用场景。随着大规模预训练语言模型（大模型）的发展，将参数化的大模型视为知识图谱成为研究热点。在这一背景下，本文聚焦于大模型在知识图谱生命周期中的研究，总结分析了大模型在知识建模、知识获取、知识融合、知识管理、知识推理和知识应用等环节中的研究进展。最后，对大模型与知识图谱未来发展趋势予以展望。

**关键词：** 大模型；知识图谱；神经符号学习

## Large Language Models and Knowledge Graphs

Yubo Chen<sup>1,2</sup>, Shaoru Guo<sup>1</sup>, Kang Liu<sup>1,2,3</sup>, Jun Zhao<sup>1,2</sup>

<sup>1</sup>The Laboratory of Cognition and Decision Intelligence for Complex Systems

<sup>2</sup>School of Artificial Intelligence, University of Chinese Academy of Sciences

<sup>3</sup>Beijing Academy of Artificial Intelligence

## Abstract

As an important form of knowledge organization, knowledge graphs are widely recognized as one of the foundational infrastructures for the next generation of artificial intelligence technologies, receiving considerable interest from both industry and academia. Traditional methods for representing knowledge graphs mainly employ symbolic representations to explicitly describe concepts and their relationships, with clear semantics and good interpretability. However, these methods have limited coverage of knowledge types, making it challenging to apply them in open-domain scenarios. With the development of large pre-trained language models (large language models), most researchers have considered parameterized large language models as knowledge graphs. Thus, this paper focuses on the research of the life cycle of knowledge graphs in large language models. Specifically, we summarize the related work on knowledge modeling, knowledge acquisition, knowledge fusion, knowledge management, knowledge reasoning, and knowledge application. Finally, we anticipate the future development trends of large language models and knowledge graphs.

**Keywords:** Large Language Models, knowledge graphs, Neural Symbolic Learning

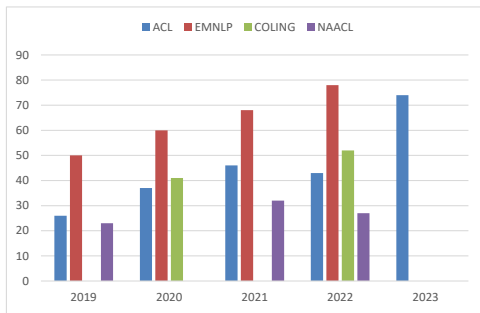


Figure 1: Knowledge相关论文数量

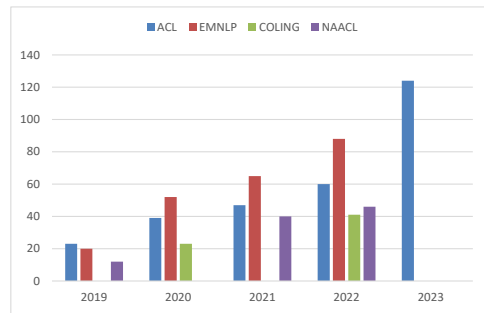


Figure 2: Language Model相关的论文数量

## 1 引言

1977年，在第五届国际人工智能会议上，图灵奖获得者爱德华·费根鲍姆提出了“知识工程”的概念，确立了知识工程在人工智能中的重要地位(Feigenbaum, 1977)。新时代，随着网络技术的发展，积累了规模巨大的互联网大数据和行业大数据。为了有效利用这些信息，将大数据转化为大知识成为一项迫切需求，这也给知识工程研究提出了新的挑战，即大数据知识工程。知识图谱(Knowledge Graph)正是一种应对大数据知识工程挑战的范式。

知识图谱由Google于2012年提出，是用来支持从语义角度组织网络数据，从而提供智能搜索服务的知识库。具体来说，知识图谱是一种比较通用的语义知识形式化描述框架，它是以三元组为基本语义单元，以有向标签图为数据结构，从知识本体和知识实例两个层次，对世界万物进行体系化、规范化描述，并支持高效知识推理和语义计算的大规模知识系统(赵军et al., 2018)。知识图谱极大地推动了语义网、自然语言处理、数据库等相关技术的发展，它被认为是下一代人工智能技术的基础设施之一，受到工业界和学术界的广泛关注。

传统知识图谱使用符号化的方法来进行知识表示，旨在将各种知识对象(如实体、事件、属性和关系等)记作具体符号，并以某种结构组织起来，通过符号匹配(如索引和检索)和数理演算(如谓词推理)等形式化方法完成各种语义计算任务。符号化表示方法能够显式地描述知识，具有语义清晰和可解释性好等特点，但其知识类型有限，且大多数逻辑推理规则需要通过人工编写和校验的方式获取，难以应对开放域应用场景。

近年来，大规模预训练语言模型(大模型)的发展引起了知识图谱领域的广泛关注。大模型可以从海量无标注数据中自动挖掘知识，并将知识以隐式的方式存储在参数化的模型中。相关研究工作表明，大模型涵盖了丰富的知识类型，包括语言学知识(Liu et al., 2019)、世界知识(Petroni et al., 2019)和常识知识(Li et al., 2022)等，使得模型具备了强大的知识表达能力。与此同时，大模型无需事先定义类型，能够灵活应用于开放域场景。在知识图谱的构建和应用中，越来越多的任务开始采用大模型进行建模，并取得了一定的效果。因此，本文聚焦于参数化大模型在知识图谱生命周期中的研究。

本文对近五年(2019年至2023年)自然语言处理领域国际会议(ACL、EMNLP、COLING和NAACL)的研究趋势进行了分析。图1和图2分别展示了与知识图谱和语言模型相关的论文数量在自然语言处理国际会议中的变化趋势<sup>0</sup>。如图所示，与知识图谱和语言模型相关的研究逐年增加，这表明知识图谱和语言模型在当前的自然语言处理研究领域中既是热点问题，也是核心问题，同时反映了研究者们对将知识图谱和语言模型应用于自然语言处理任务的关注和投入。

同时，本文围绕知识图谱生命周期各个环节对ACL2019至ACL2023的论文进行了统计，包括知识建模、知识获取、知识融合、知识管理、知识推理和知识应用，如图3所示。通过图中的统计结果可以发现，对知识图谱生命周期各个环节的研究呈现出明显的增长趋势，重点关注知识获取与知识应用。此外，图4展示了ACL2023论文的词云图，从中可以观察

©2023 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

<sup>0</sup>统计信息来源于ACL Anthology (<https://www.aclweb.org/anthology/>)。值得注意的是，2020年NAACL会议未召开，COLING会议每两年举办一次，且截至目前仅公布了ACL2023的论文录用列表。

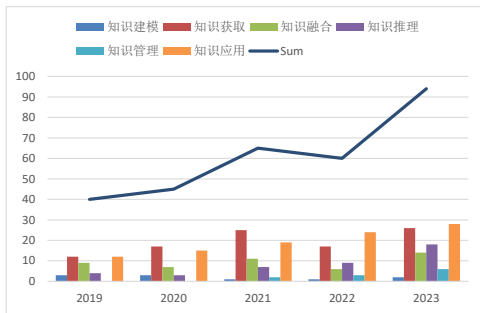


Figure 3: 知识图谱生命周期统计

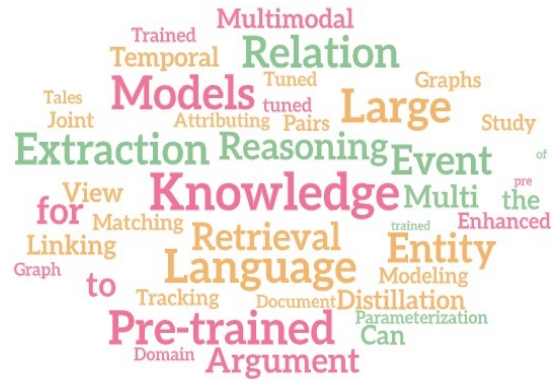


Figure 4: ACL2023词云

到“Knowledge”、“Language Models”等相关研究引起了极大关注，表明知识图谱和语言模型在当前研究中的重要地位。

大模型逐渐成为人类知识的全新载体，将参数化的大模型视为知识图谱成为当前的研究趋势。在此背景下，本文聚焦于大模型在知识图谱生命周期各个环节中的研究，总结分析了大模型在知识建模、知识获取、知识融合、知识管理、知识推理和知识应用等环节的研究进展。

## 2 知识建模

知识建模，也称知识体系构建或本体建模，旨在构建一个本体对目标知识进行描述。该本体定义了知识的类别体系、类别下所属的概念、概念所具有的属性以及概念之间的语义关系。目前，实体和事件是知识图谱中典型的知识类型。因此，本节将分别介绍基于大模型的实体本体建模和事件本体建模的相关研究。

实体本体是指将实体作为知识单元，通过描述实体的概念、属性以及它们之间的分类关系来刻画客观世界中事物的静态规律。Jullien et al. (2022)采用提示学习和微调方法探测大模型的概念知识。类似地，Peng et al. (2022)也使用了提示学习和微调方法，不仅探测大模型中的概念知识，还包括属性知识和概念之间的关系。此外，Wu et al. (2023)同样采用提示学习方法，在探测概念、属性及概念关系同时，进一步探测了属性之间的关系。这些研究工作表明，大模型在一定程度上能够建模实体本体知识，但在关系建模方面存在一定困难，并且词语共现可能会引发概念幻觉等问题。

事件本体是指将事件作为认知单元，通过描述事件类型、事件论元以及它们之间的语义关系来刻画现实世界中事物的动态变化规律。针对事件本体建模，目前大多数工作聚焦于大模型在事件类型建模（概念）、原子事件模式建模（概念及属性）和事件图模式建模（概念之间语义关系）三个方面的研究。事件类型建模是指利用大模型自动地从原始文本中发现新的事件类别。其中，Edwards and Ji (2023)使用大模型建模文本表示，并通过注意力机制聚合类别特征，从而归纳出新的事件类别。原子事件模式建模旨在从多个相似事件实例中归纳出一个模板。为了实现这一目标，Tang et al. (2023)采用情景学习方法，引导大模型构建面向不同领域的事件模式。事件图模式建模是指从一组相关事件实例中归纳出事件之间的模式和关系，Li et al. (2023)利用提示学习引导大模型生成特定场景下的关键事件及事件之间的关系。此外，还可以通过人与大模型协作的方式，由人工干预大模型事件本体建模过程，生成符合人类认知的事件模式(Zhang et al., 2023b)。目前基于大模型的事件本体建模主要关注时序关系，无法对事件本体丰富关系进行完整建模。同时，由于事件本体的复杂性，基于大模型自动构建的事件本体质量难以保证。

## 3 知识获取

知识获取旨在从非结构化数据中抽取结构化知识。根据人们的认知过程，可将获得的知识分为语言学知识、世界知识和常识知识。因此，本节将分别介绍基于大模型语言学知识获取、世界知识获取和常识知识获取的相关研究。



语言学知识是指词性、句法结构以及词语之间的关系等方面的知识。由于语言学知识通常需要采用特定的标注方式，因此通常使用有训练的探测方法来获取这些知识<sup>1</sup>。例如，Liu et al. (2019)通过整合多种语言结构预测任务（如句法分块、命名实体识别等），在预训练的隐层表示基础上重新训练分类器，探测大模型中与语言结构相关的知识。而Jain and Anke (2022)则通过将词汇之间的上下位关系转化为自然语言提示模板，来探测大模型中的词汇关系知识。相关研究表明，大模型的隐层表示中包含了一定程度的词性和句法知识，并且能够捕捉一些词汇关系。然而，大模型对于自然语言提示模板比较敏感，并且其语言学知识并不完备(Rogers et al., 2020)。

世界知识是指与特定实体和事件相关的事实性知识。由于世界知识通常可以用自然语言句子表达，并且在训练语料中广泛存在，因此对于世界知识的探测通常采用无训练的方法。其中，最具代表性的探测方式是LAMA (Language Model Analysis) (Petroni et al., 2019)，它将三元组或问答对形式的世界知识转化为自然语言填空的形式，通过预测正确答案在词表中的排位来评估大模型对世界知识的掌握程度。在LAMA探针实验的基础上，还衍生出了一系列相关研究，如自动生成提示语(Jiang et al., 2020)和使用连续向量作为提示语(Qin and Eisner, 2021)等，通过解决提示语选择等问题来进一步探测大模型中的世界知识。相关研究表明，大模型掌握了一定量的世界知识，但模型结果受到多种因素的干扰，对其处理世界知识的机制并不清晰，因此对于世界知识的探测实验需要更加严谨的设置和评估方法。

常识知识是指人们默认掌握的关于物理世界和人类社会的概括性知识。在常识知识探测研究中，通常采用无训练的方法。然而，与世界知识的设定不同，由于常识知识的多样性以及难以用单个标记进行填空的形式进行考察，常识知识的探测往往需要采用打分对比判断的形式(Zhou et al., 2020; Li et al., 2022)或句子排序等方式(Lin et al., 2021)，通过分析大模型在判断句子是否符合常识方面的能力，来探测大模型所具备的常识知识。此外，Bosselut et al. (2019)通过微调大模型获取常识知识，West et al. (2022)采用情景学习策略提示大模型生成常识知识，Wang et al. (2022)通过优化生成和答案过滤方法，辅助大模型在生成质量较差的条件下获取大规模高质量常识知识。相关研究表明，大模型掌握了一定程度的常识知识，但其对答案分布的拟合可能导致探测偏差。因此，如何客观地评估大模型的常识知识获取能力是一个难点问题。

## 4 知识融合

知识融合旨在对不同来源、不同语言和不同结构的知识进行融合，进而对已有知识图谱进行补充、更新和去重。由于知识图谱往往是由不同机构或个人构建的，其设计和构建并不统一，从而导致了异构性和冗余性的问题。因此，如何发现和建立不同知识图谱之间的关联成为各个领域亟需解决的重要问题。从融合的对象看，知识融合包括本体融合和实例融合。因此，本节将分别介绍本体融合和实例融合的相关研究。

本体融合是指将两个或多个异构知识本体进行融合，将相同的概念、属性和关系进行连接。基于大模型的本体融合主要关注于利用大模型获取本体的向量表示，并使用两个本体向量之间的相似度作为本体融合的依据。目前，主要采用对大模型进行微调的方法来获取本体的向量表示，从而完成实体本体融合(He et al., 2022a; He et al., 2022b)和事件本体融合(Guo et al., 2023)。相关研究表明，在本体融合任务中，大模型主要用于完成基本的向量表示，而更多的应用潜能尚未得到充分研究。因此，探索基于大模型的本体融合方法对于进一步提升本体融合效果具有重要价值。

实例融合是指对两个不同知识图谱中的知识实例（实体实例、关系实例）进行融合的过程。类似于本体融合，实例融合通常采用基于语义匹配的方法。这种方法利用大模型将知识图谱中的实例表示为低维向量，并通过计算实例向量之间的相似度来判断实例之间的语义关联关系。为了获得更好的实例融合性能，大模型在对实例建模过程中整合了多维度的实例信息，例如名称、描述、属性和结构信息，以获取语义丰富的低维向量(Tang et al., 2020; Yang et al., 2019)。此外，Zhao et al. (2023)提出了一种将实例对齐任务转化为文本蕴含任务的方法，将基于提示构建的实例对序列输入到大模型中，从而充分捕捉实例之间的语义关联。目前的实例对

<sup>1</sup>知识获取方法可以分为有训练方法和无训练方法。有训练方法指冻结模型参数，提取隐藏层表示，并添加额外的分类模块，在知识相关任务上进行训练，通过模型表现来评估其对知识的掌握程度；无训练方法则常常利用自然语言提示构建填充问题，使模型预测正确答案的似然分数，并与其他答案进行比较。

齐方法主要依赖于知识图谱的结构信息。然而，在现有的知识图谱中存在大量的长尾实例，这些实例的邻接实体通常只有一个或两个，缺乏丰富的结构信息。因此，如何有效地利用大模型来融合长尾实例，是一项具有挑战性的任务。

## 5 知识管理

知识管理旨在实现对知识图谱的持久化存储，以及对目标知识的高效检索。大模型通过对海量文本数据的训练来学习知识，并以隐式的方式存在。在大模型中，知识是如何存储的？又是如何进行更新的呢？围绕这两个问题，本节将分别介绍基于大模型的知识定位和知识编辑相关研究。

知识定位指的是探索大模型中“知识”的存储位置和访问机制。目前的主流观点认为，大语言模型中的前馈网络模块（Feedforward Neural Network, FFN）起到了知识存储的作用(Geva et al., 2021)。具体而言，可以将Transformer的前馈网络模块视为一个键值存储器（Key-Value Memory）。在这个存储器中，每个神经元的键向量用于识别输入中的语言或知识模式，可以看作是模式探测器，而对应的值向量则扮演着知识生成器的角色，代表着该模式的知识向量。相关的研究通过实验分析对这一结论提供了支持，例如Dai et al. (2022)使用归因方法验证了前馈网络模块隐藏层中存在与事实性知识相关的特定神经元，而Meng et al. (2023)则从因果关联的角度论证了前馈网络模块与事实性知识的关联。相关研究为理解大模型中的知识定位提供了一种新的视角，揭示了知识是通过参数化的方式存储在模型中的。然而，目前的研究主要集中在处理结构化知识，对于其他形式的知识，目前的方法还存在一定的局限性。

知识编辑指的是对隐含在语言模型参数中的知识进行有针对性的更新。该任务的目标是在更新特定知识的同时，尽量减少对其他知识的破坏。主要的方法可以分为超网络方法和定向知识编辑方法。超网络方法依赖于数据驱动的训练，通过训练一个超网络，有针对性地修改模型中的事实知识。具体来说，超网络替代优化器生成模型参数的更新量，并使用约束项来确保在超网络训练过程中，特定知识更新成功的同时减少对其他知识的破坏(Cao et al., 2021)。由于大模型的参数规模巨大，设计高效的超网络方法仍然是一个挑战。定向知识编辑方法是一种基于“键值存储”假设，用于直接定位和修改模型参数的方法。在该方法中，假设Transformer的前馈网络模块负责存储知识，并且每个键都有对应的值。通过使用表达相同键含义的不同提示，可以定位到存储了相关知识的神经元，并对其对应的值进行修改(Meng et al., 2022; Meng et al., 2023)。目前定向知识编辑方法与三元组形式耦合较强，在处理如过程性知识、数学知识、语言知识等某情况，存在一些限制。

## 6 知识推理

知识推理旨在采用推理手段发现已有知识中隐含的知识。通过知识建模、知识获取和知识融合，可以构建一个可用的知识图谱。然而，由于数据的不完备性和稀疏性，很难通过抽取或融合方法来填补缺失的知识。因此，需要采用推理的手段来发现已有知识中隐含的知识。大模型从海量无标注数据中学习丰富的知识，如果能从大模型已有知识中推导出新的知识，将有助于知识图谱的构建。为此，研究人员致力于探索大模型推理能力，而推理能力的关键在于思维链（Chain of Thought, CoT）技术。因此，本节围绕思维链介绍大模型在知识推理方面的相关研究。

思维链是参考人类解决复杂问题的一种推理方式，将问题分解为一系列中间问题，并逐步解决这些问题以获得最终结果。思维链推理技术通过向大模型展示少量样例并解释推理过程，在回答提示时也显式展示推理过程，从而引导模型输出更准确的结果(Wei et al., 2022)。零样本思维链（Zero-shot-CoT）是思维链的一种衍生形式，通过在问题结尾附加提示语句，如“Let’s think step by step”，激励大模型生成一个回答问题的思维链，从而产生答案(Kojima et al., 2022)。在这基础上，Wang et al. (2023b)通过自洽性（Self-consistency）方式改进了思维链，该方法利用多数投票的思想生成多个思维链，然后选择多数答案作为最终答案，以提高思维链的性能。上述方法通过简单的提示或精心设计的样例来激发大模型生成中间推理步骤，但生成的证据常常会出现错误，导致不准确和不可靠的推理链。因此，Wang et al. (2023a)提出了知识链（Chain-of-Knowledge）提示方法，通过生成显式的三元组结构知识证据来引导大模型进行推理，从而有效提升文本推理任务的性能。为了更好地适应需要预测的复杂推理任务，Yao et al. (2023)提出了思维树（Tree of Thoughts, ToT）方法，大模型通过对不同的推理路径进行



评估来决定下一步的行动方案，并且在必要时可以向前或向后追溯，以实现全局决策。尽管常规文本提示在一定程度上可以促进模型的推理能力，但存在逻辑上的歧义，可能导致错误的答案。而三元组结构知识提示可以进一步提升推理性能。因此，需要将文本提示与结构特征相融合，以使大模型能够生成可靠且具体的推理过程。

## 7 知识应用

知识应用是指将知识应用于具体任务和应用场景的过程。大模型在解决各种自然语言处理任务方面展现了巨大的潜力，并在某种程度上为通用人工智能铺平了道路。本节将介绍大模型在自然语言理解和自然语言生成任务方面的应用。

自然语言理解的目标是让计算机能够像人类一样理解自然语言。自然语言理解任务包括文本分类、自动问答等。文本分类是指将文本映射到预先给定的某一类别或某几类别主题的过程。针对该任务，Sun et al. (2023)提出了一种基于大模型的渐进推理方法来进行文本分类。该方法利用提示语引导大模型捕捉关键词、语气等表层线索，并进一步基于这些线索通过推理对文本进行分类。Zhang et al. (2023c)通过设计清晰而直接的提示，引导大模型完成传统的情感分类任务、基于方面的情感分析和多维主观文本分析。自动问答要求系统根据对文本的理解给出问题的答案，是衡量机器自然语言理解程度的重要指标。Tan et al. (2023)将大模型自身的知识作为知识库，衡量其在传统基于知识问答任务上的性能。Huang et al. (2023)提出了一种基于答案反馈的情景学习方法，通过将更正的答案记录作为反馈信息，构建大模型的增强提示，以提高其在问答任务中的性能。由于大模型强大的泛化能力，在面对分布外的数据或非常少的训练数据时可以提供帮助(Yang et al., 2023)，但在复杂的自然语言理解任务，如基于方面的情感分析、多跳问答等任务上，仍然存在一定挑战。

自然语言生成的目标是生成连贯、有意义且与上下文相符的高质量文本。它包括两类主要任务：序列转化任务和开放式生成任务。序列转化任务旨在将输入文本转换为新的符号序列。例如，Liu et al. (2023)使用大模型生成训练数据，以指导摘要模型的学习。而Zhang et al. (2023a)则通过使用提示词模板和情景学习方法来引导大模型完成机器翻译任务。开放式生成任务是从头开始生成文本或符号，以准确匹配输入的描述。例如，Yang et al. (2022)利用递归提示和调整策略来引导大模型自动生成成长篇故事。而Yang et al. (2022)则利用大模型从海量数据中学习到的丰富知识，有效生成个性化新闻和新闻摘要等内容。由于大模型具备强大的生成能力和创造力，因此在许多生成任务中展现出更高的优越性(Yang et al., 2023)。然而，当大模型生成文本时，它无法自行判断生成结果的准确性，这导致了所谓的幻觉性问题。因此，解决大模型中的幻觉性问题成为一个重要的挑战。

## 8 总结与展望

本文聚焦于大模型在知识图谱生命周期的研究，并总结分析了大模型在知识建模、知识获取、知识融合、知识管理、知识推理和知识应用等环节中的研究进展。

由于大模型能够自动从无标注数据中学习知识并以参数形式存储，因此能灵活地应用于开放域场景。然而，大模型也存在一些根本性问题和限制。首先，大模型缺乏显式的知识存储结构，这使得知识的组织和检索变得困难。其次，大模型有时会受到所谓的“幻觉”现象的影响，导致生成不准确的输出。最后，大模型纯粹的数据驱动学习方式使得其缺乏可解释性。相比之下，知识图谱以显式的方式描述知识，并且知识的表达基于清晰的结构和严谨的描述逻辑，因此具有确定性。知识图谱中的节点和边都有具体的含义，使得知识的解释和理解更加容易。然而，知识图谱的构建困难，其泛化性能不如大模型，也难以生成新的事实或表示未知的知识(Pan et al., 2023; Cao et al., 2023)。因此，将知识图谱与大模型相结合可以相辅相成，相互促进，从而取得更好的效果。具体地：

大模型助力知识图谱构建。在构建知识图谱的过程中，利用大模型可以提高知识的抽取和注入效率，增强知识图谱的覆盖范围和泛化性能。

知识图谱辅助大模型进行知识校准和提升可解释性。大模型可以借助知识图谱提高知识的准确性，同时利用知识图谱来解释大模型的知识推理过程，增强模型的可解释性。

大模型和知识图谱协同工作。充分利用数据驱动和知识驱动的优势，提升自然语言处理各项任务的能力，促进人工智能技术的发展。

大模型的快速发展为知识图谱的研究注入了新的活力，两者相结合研究的前景非常值得期待。

## 参考文献

赵军, 刘康, 何世柱, and 陈玉博. 2018. 知识图谱. 高等教育出版社.

- Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: commonsense transformers for automatic knowledge graph construction. In Anna Korhonen, David R. Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 4762–4779. Association for Computational Linguistics.
- Nicola De Cao, Wilker Aziz, and Ivan Titov. 2021. Editing factual knowledge in language models. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 6491–6506. Association for Computational Linguistics.
- Boxi Cao, Hongyu Lin, Xianpei Han, and Le Sun. 2023. The life cycle of knowledge in big language models: A survey. *CoRR*, abs/2303.07616.
- Damai Dai, Li Dong, Yaru Hao, Zhifang Sui, Baobao Chang, and Furu Wei. 2022. Knowledge neurons in pretrained transformers. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 8493–8502. Association for Computational Linguistics.
- Carl Edwards and Heng Ji. 2023. Semi-supervised new event type induction and description via contrastive loss-enforced batch attention. In Andreas Vlachos and Isabelle Augenstein, editors, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2023, Dubrovnik, Croatia, May 2-6, 2023*, pages 3787–3809. Association for Computational Linguistics.
- Edward A. Feigenbaum. 1977. The art of artificial intelligence: Themes and case studies of knowledge engineering. In Raj Reddy, editor, *Proceedings of the 5th International Joint Conference on Artificial Intelligence. Cambridge, MA, USA, August 22-25, 1977*, pages 1014–1029. William Kaufmann.
- Mor Geva, Roei Schuster, Jonathan Berant, and Omer Levy. 2021. Transformer feed-forward layers are key-value memories. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 5484–5495. Association for Computational Linguistics.
- Shaoru Guo, Chenhao Wang, Yubo Chen, Kang Liu, Ru Li, and Jun Zhao. 2023. Eventoa: An event ontology alignment benchmark based on framenet and wikidata. In *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics*.
- Yuan He, Jiaoyan Chen, Denvar Antonyrajah, and Ian Horrocks. 2022a. Bertmap: A bert-based ontology alignment system. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelfth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pages 5684–5691. AAAI Press.
- Yuan He, Jiaoyan Chen, Hang Dong, Ernesto Jiménez-Ruiz, Ali Hadian, and Ian Horrocks. 2022b. Machine learning-friendly biomedical datasets for equivalence and subsumption ontology matching. In Ulrike Sattler, Aidan Hogan, C. Maria Keet, Valentina Presutti, João Paulo A. Almeida, Hideaki Takeda, Pierre Monnin, Giuseppe Pirrò, and Claudia d’Amato, editors, *The Semantic Web - ISWC 2022 - 21st International Semantic Web Conference, Virtual Event, October 23-27, 2022, Proceedings*, volume 13489 of *Lecture Notes in Computer Science*, pages 575–591. Springer.
- Zixian Huang, Jiaying Zhou, Gengyang Xiao, and Gong Cheng. 2023. Enhancing in-context learning with answer feedback for multi-span question answering. *CoRR*, abs/2306.04508.

- Devansh Jain and Luis Espinosa Anke. 2022. Distilling hypernymy relations from language models: On the effectiveness of zero-shot taxonomy induction. In Vivi Nastase, Ellie Pavlick, Mohammad Taher Pilehvar, José Camacho-Collados, and Alessandro Raganato, editors, *Proceedings of the 11th Joint Conference on Lexical and Computational Semantics, \*SEM@NAACL-HLT 2022, Seattle, WA, USA, July 14-15, 2022*, pages 151–156. Association for Computational Linguistics.
- Zhengbao Jiang, Frank F. Xu, Jun Araki, and Graham Neubig. 2020. How can we know what language models know. *Trans. Assoc. Comput. Linguistics*, 8:423–438.
- Maël Jullien, Marco Valentino, and André Freitas. 2022. Do transformers encode a foundational ontology? probing abstract classes in natural language. *CoRR*, abs/2201.10262.
- Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. 2022. Large language models are zero-shot reasoners. In *NeurIPS*.
- Xiang Lorraine Li, Adhiguna Kuncoro, Jordan Hoffmann, Cyprien de Masson d’Autume, Phil Blunsom, and Aida Nematzadeh. 2022. A systematic investigation of commonsense knowledge in large language models. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 11838–11855. Association for Computational Linguistics.
- Sha Li, Ruining Zhao, Manling Li, Heng Ji, Chris Callison-Burch, and Jiawei Han. 2023. Open-domain hierarchical event schema induction by incremental prompting and verification. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*.
- Bill Yuchen Lin, Seyeon Lee, Xiaoyang Qiao, and Xiang Ren. 2021. Common sense beyond english: Evaluating and improving multilingual language models for commonsense reasoning. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 1274–1287. Association for Computational Linguistics.
- Nelson F. Liu, Matt Gardner, Yonatan Belinkov, Matthew E. Peters, and Noah A. Smith. 2019. Linguistic knowledge and transferability of contextual representations. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 1073–1094. Association for Computational Linguistics.
- Yixin Liu, Alexander R. Fabbri, Pengfei Liu, Dragomir Radev, and Arman Cohan. 2023. On learning to summarize with large language models as references. *CoRR*, abs/2305.14239.
- Kevin Meng, David Bau, Alex Andonian, and Yonatan Belinkov. 2022. Locating and editing factual associations in GPT. In *NeurIPS*.
- Kevin Meng, Arnab Sen Sharma, Alex J. Andonian, Yonatan Belinkov, and David Bau. 2023. Mass-editing memory in a transformer. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.
- Shirui Pan, Linhao Luo, Yufei Wang, Chen Chen, Jiapu Wang, and Xindong Wu. 2023. Unifying large language models and knowledge graphs: A roadmap. *CoRR*, abs/2306.08302.
- Hao Peng, Xiaozhi Wang, Shengding Hu, Hailong Jin, Lei Hou, Juanzi Li, Zhiyuan Liu, and Qun Liu. 2022. COPEN: probing conceptual knowledge in pre-trained language models. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 5015–5035. Association for Computational Linguistics.
- Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick S. H. Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander H. Miller. 2019. Language models as knowledge bases? In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 2463–2473. Association for Computational Linguistics.



- Guanghui Qin and Jason Eisner. 2021. Learning how to ask: Querying lms with mixtures of soft prompts. In Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tür, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou, editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, pages 5203–5212. Association for Computational Linguistics.
- Anna Rogers, Olga Kovaleva, and Anna Rumshisky. 2020. A primer in bertology: What we know about how BERT works. *Trans. Assoc. Comput. Linguistics*, 8:842–866.
- Xiaofei Sun, Xiaoya Li, Jiwei Li, Fei Wu, Shangwei Guo, Tianwei Zhang, and Guoyin Wang. 2023. Text classification via large language models. *CoRR*, abs/2305.08377.
- Yiming Tan, Dehai Min, Yu Li, Wenbo Li, Nan Hu, Yongrui Chen, and Guilin Qi. 2023. Evaluation of chatgpt as a question answering system for answering complex questions. *CoRR*, abs/2303.07992.
- Xiaobin Tang, Jing Zhang, Bo Chen, Yang Yang, Hong Chen, and Cuiping Li. 2020. BERT-INT: A bert-based interaction model for knowledge graph alignment. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 3174–3180. ijcai.org.
- Jialong Tang, Hongyu Lin, Zhuoqun Li, Yaojie Lu, Xianpei Han, and Le Sun. 2023. Harvesting event schemas from large language models. *CoRR*, abs/2305.07280.
- Chenhao Wang, Jiachun Li, Yubo Chen, Kang Liu, and Jun Zhao. 2022. Cn-automic: Distilling chinese commonsense knowledge from pretrained language models. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 9253–9265. Association for Computational Linguistics.
- Jianing Wang, Qiushi Sun, Nuo Chen, Xiang Li, and Ming Gao. 2023a. Boosting language models reasoning with chain-of-knowledge prompting. *CoRR*, abs/2306.06427.
- Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023b. Self-consistency improves chain of thought reasoning in language models. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-thought prompting elicits reasoning in large language models. In *NeurIPS*.
- Peter West, Chandra Bhagavatula, Jack Hessel, Jena D. Hwang, Liwei Jiang, Ronan Le Bras, Ximing Lu, Sean Welleck, and Yejin Choi. 2022. Symbolic knowledge distillation: from general language models to commonsense models. In Marine Carpuat, Marie-Catherine de Marneffe, and Iván Vladimir Meza Ruíz, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022*, pages 4602–4625. Association for Computational Linguistics.
- Weiqi Wu, Chengyue Jiang, Yong Jiang, Pengjun Xie, Ru Li Kewei Tu, and Jun Zhao. 2023. Do plms know and understand ontological knowledge? In *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics*.
- Hsiu-Wei Yang, Yanyan Zou, Peng Shi, Wei Lu, Jimmy Lin, and Xu Sun. 2019. Aligning cross-lingual entities with multi-aspect information. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 4430–4440. Association for Computational Linguistics.
- Kevin Yang, Yuandong Tian, Nanyun Peng, and Dan Klein. 2022. Re3: Generating longer stories with recursive reprompting and revision. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022*, pages 4393–4479. Association for Computational Linguistics.

- Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Bing Yin, and Xia Hu. 2023. Harnessing the power of llms in practice: A survey on chatgpt and beyond. *CoRR*, abs/2304.13712.
- Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. 2023. Tree of thoughts: Deliberate problem solving with large language models. *CoRR*, abs/2305.10601.
- Biao Zhang, Barry Haddow, and Alexandra Birch. 2023a. Prompting large language model for machine translation: A case study. *CoRR*, abs/2301.07069.
- Tianyi Zhang, Isaac Tham, Zhaoyi Hou, Jiaxuan Ren, Liyang Zhou, Hainiu Xu, Li Zhang, Lara J. Martin, Rotem Dror, Sha Li, Heng Ji, Martha Palmer, Susan Windisch Brown, Reece Suchocki, and Chris Callison-Burch. 2023b. Human-in-the-loop schema induction. *CoRR*, abs/2302.13048.
- Wenxuan Zhang, Yue Deng, Bing Liu, Sinno Jialin Pan, and Lidong Bing. 2023c. Sentiment analysis in the era of large language models: A reality check. *CoRR*, abs/2305.15005.
- Yu Zhao, Yike Wu, Xiangrui Cai, Ying Zhang, Haiwei Zhang, and Xiaojie Yuan. 2023. From alignment to entailment: A unified textual entailment framework for entity alignment. *CoRR*, abs/2305.11501.
- Xuhui Zhou, Yue Zhang, Leyang Cui, and Dandan Huang. 2020. Evaluating commonsense in pre-trained language models. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, pages 9733–9740. AAAI Press.