符号主义与联结主义：从常识推理的视角看

Symbolism and Connectionism: A Perspective from Commonsense Reasoning

徐超 /XU Chao　周北海 /ZHOU Beihai

（山西大学哲学学院，山西太原，030006）
(School of Philosophy, Shanxi University, Taiyuan, Shanxi, 030006)

摘　要：长久以来，符号主义方法一直占据着常识推理领域的主流地位，而联结主义方法一直被批评不具备复杂推理能力。然而，随着以ChatGPT为代表的大语言模型的兴起，联结主义方法在常识推理方面的表现超越了符号主义方法，展现出了复杂推理的能力。从常识推理的视角对符号主义和联结主义方法进行了对比分析，首先介绍了两种方法解决常识推理问题的基本思路，并分析二者所面临的问题；然后通过一个实例来解释符号主义方法所面临的困难，并与联结主义方法进行对比；最后从常识推理表现、知识表示方法、知识获取方式、推理方式、可解释性和创造性方面对两种方法进行了综合对比，并探讨了两种方法融合的可能性。

关键词：符号主义　联结主义　常识推理　语言模型　ChatGPT

Abstract: The symbolic approach has dominated the field of commonsense reasoning for a long time, while the connectionist approach has been criticized for lacking complex reasoning capabilities. However, with the rise of large language models represented by ChatGPT, the performance of the connectionist approach in common sense reasoning has surpassed that of the symbolic approach, demonstrating complex reasoning abilities. This paper compares and analyzes the symbolic and connectionist approaches from the perspective of commonsense reasoning. We first introduce the basic ideas of the two approaches in addressing commonsense reasoning problems and analyze the challenges they face. Then, we explain the difficulties encountered by the symbolic approach through a case study and compares it with the connectionist approach. Finally, it comprehensively compares the two approaches in terms of commonsense reasoning performance, knowledge representation, knowledge acquisition methods, reasoning methods, interpretability, and creativity, and explores the possibility of integrating the two approaches.

Key Words: Symbolism; Connectionism; Commonsense reasoning; Language models; ChatGPT

中图分类号：N031; G201　DOI: 10.15994/j.1000–0763.2025.11.005　CSTR: 32281.14.jdn.2025.11.005

引　言

常识推理是人类智能的核心特征之一。常识推理是指人类获取世界场景中某些方面的信息，并依据常识知识或对世界运作方式的理解对该场景的其他方面进行推理的过程。[1] 自人工智能诞生以来，研究者们致力于采用不同的

基金项目：国家社会科学基金一般项目“基于意象图式理论的人类推理逻辑研究”（项目编号：23BZX126）。
收稿日期：2024年8月21日
作者简介：徐　超（1987–）男，山东新泰人，山西大学哲学学院副教授，研究方向为逻辑与人工智能、逻辑与认知、人工智能哲学。Email: chao.xu@sxu.edu.cn
周北海（1955–）男，浙江乐清人，山西大学哲学学院教授，研究方向为符号逻辑、逻辑与人工智能。Email: zhoubh@sxu.edu.cn

方法来模拟人类常识推理过程。其中，符号主义和联结主义作为两个主要的认知建模框架，引起了广泛的关注和讨论。[2]-[7]

符号主义和联结主义源自不同的哲学传统。符号主义源自理性主义传统，强调符号和规则在思维和认知过程中的重要性，认为符号是智能的本质，主张使用符号来表示知识，并将推理和问题解决作为智能的核心特征；而联结主义源自经验主义传统，强调从数据中学习，认为智能可以通过神经网络的连接来实现，并将学习作为智能的核心特征。前者可以视为对人类心灵的模拟；而后者可以视为对人类大脑的模拟。[2] 这两种不同的方法在常识推理领域展现出了各自的优势和局限性。

符号主义方法主张使用符号操作来模拟人类的常识推理过程，大致可以分为逻辑方法和非逻辑方法。逻辑方法作为符号主义方法的主流方法，其背后的基本假设是人类的大部分推理都可以分析为演绎或者近似演绎的推理过程；[8] 非逻辑方法是指那些不严格依赖逻辑来表示和处理知识的方法。符号主义方法在一些领域取得了显著的成功，例如，分类推理、时态推理、关于动作和变化的推理以及定性推理。[9] 然而，在面对复杂的常识推理任务时，其表现不尽如人意。

联结主义方法主张通过模拟人类大脑的神经网络连接来存储和处理知识，从而让机器实现常识推理。尽管联结主义在模拟一些复杂的感知任务上表现出色，如图像识别、自动驾驶等，但在常识推理领域一直备受质疑，被认为缺乏复杂推理能力。[10]-[12] 然而，随着以ChatGPT为代表的大语言模型的兴起，联结主义方法在常识推理方面的表现超越了符号主义方法，展现出了复杂推理的能力。

本文聚焦于符号主义和联结主义方法在常识推理领域的对比研究。第二节和第三节将分别介绍基于符号主义和基于联结主义的常识推理解决方案并分析各自所面临的问题；第四节通过实例来解释和分析两种方法的表现；第五节对这两种方法从整体上进行对比，并探讨两种方法融合的可能性；第六节总结全文。

一、符号主义方法

直观上，人类之所以能够进行常识推理，是因为人类具有常识知识，而且能够利用常识知识进行推理。因此，要让机器模拟人类的常识推理能力，需要为机器构造常识知识库，并设计相应的推理机制。依据知识表示和推理方法的不同，符号主义方法可以分为逻辑方法和非逻辑方法，前者使用逻辑语言来表示常识，并基于逻辑方法来构建推理机制；后者使用结构化的知识表示方法，并基于自然语言理解来构建推理机制。

1. 基于逻辑的常识推理方法

要基于逻辑方法构造常识推理系统，需要构建基于逻辑表示的常识知识库和逻辑推理机制。与此同时，人类所面临的常识推理问题是使用自然语言描述的，而机器使用逻辑语言来进行推理，因此还需要让机器具备自动形式化的能力，即将自然语言自动地翻译为逻辑语言。综上，基于逻辑方法的常识推理系统包括三个部分：自动形式化系统、常识知识库以及逻辑推理机制，如图1所示。

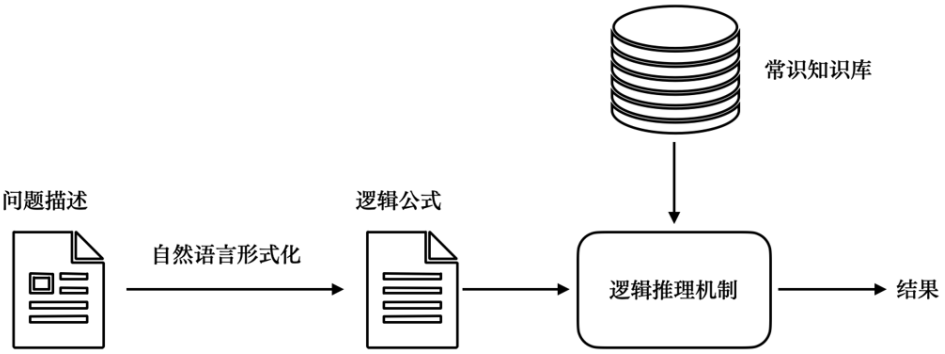

图1 基于逻辑方法的常识推理系统结构图

Cyc项目是逻辑方法在人工智能领域应用的一个典范。Cyc项目创建了一种人工语言CycL来描述知识。CycL是一阶逻辑语言的扩张，量词允许加在谓词、函数和句子前面，因此，本质上一种高阶语言。[13]CycL包括逻辑符号和非逻辑符号，其中逻辑符号包括变项、逻辑联结词和量词；非逻辑符号包括常项、函数和谓词。例如，“所有人都会死”的一阶表达式为∀x(Person(x)→Mortal(x))，CycL语言表达式如下所示：

```
(#$implies
    (#$isa ?person #$Person)
    (#$isa ?person #$Mortal))
```

其中，#$implies表示逻辑联结词蕴涵，(#$isa X Y)表示X是集合Y的实例，?person表示变项；#$Person和#$Mortal分别表示谓词“人”和“会死的”。到2017年，Cyc项目共包括约150万个项和2450万条知识。这些知识涉及因果、时间、空间、物质、意图、矛盾、不确定性、信念、情感和计划等内容。[14]

在解决具体的常识推理问题时，需要从海量的常识知识中挑选出与当前问题相关的常识知识。否则，需要遍历所有的常识知识，这会导致推理效率非常低，无法在合理时间内给出回答。关于世界的知识是基于人们对于世界的认识，然而这种认识并非严格确定的，时常存在着不一致的情况。按照经典逻辑的推理规则，矛盾可以推出一切，由此导致无法得出正确的结论。为解决这两个问题，Cyc引入了“微理论”（microtheory）概念，将基于相同假设、相同主题和相同来源的知识划分为一个微理论，每一个断言至少包含在一个微理论中。[15]通过对断言进行分类，一方面可以解决知识整体不一致的问题；另一方面只在相关微理论中进行推理可以提高推理效率。

Cyc中的推理是都是演绎推理，其推理的每一步都有严格的逻辑证明。Cyc采用归结原则（resolution principle）实现自动推理。在确定了推理范围后，需要遍历所有的断言以得到正确答案。Cyc提供了深度优先和广度优先两种答案搜索算法。深度优先算法更为有效且占用内存少，但是容易进入一个无穷支，然后无限循环下去；广度优先优点是可以得到这个答案最简单的证明，但是如果目标答案的层数较深，则需要存储大量的信息。利用归结原则理论上可以得到正确的答案，但是在实际应用中，为提升推理的效率和语言的表达力，Cyc最终放弃了推理的完备性。[14]

2. 非逻辑的常识推理方法

类似于逻辑方法，非逻辑的常识推理方法同样需要构建常识知识库，并构建相应的推理机制。在知识表示方面，不像逻辑方法受限于严格的逻辑语法，非逻辑的常识知识表示方法更加灵活，通过使用语义网络或者知识图谱来展示实体间的语义关系，使其更容易理解。除此之外，由于使用自然语言来表示知识，其在知识获取方面也更具优势，目前基于机器学习的本体学习技术，可以从文本数据或非结构化数据中自动学习常识知识来构建知识库。

常识知识库又被称为本体知识库、语义知识库、语义网络或知识图谱。目前学界已经构建了大量的常识知识库，例如，WordNet、FrameNet、VerbNet、ConceptNet、YAGO、DBpedia[16]–[21]和谷歌知识图谱等。这些知识库从理论基础到实际的构造方式都有很大的不同。WordNet是基于认知科学和语言学的研究成果构建的；VerbNet和FrameNet是基于语言学的研究成果构建的；ConceptNet、DBpedia、YAGO和谷歌知识图谱则是从具体的工程应用出发来构建的。WordNet、VerbNet和FrameNet最初完全人工构建的，后利用一些信息技术来辅助人工构建；DBpedia、YAGO和谷歌知识图谱是利用语义抽取技术自动构建的；ConceptNet最初是采用众包的形式收集数据，后又采用信息处理技术进行结构化。

在推理机制方面，区别于逻辑方法的演绎推理方式，非逻辑方法主要采用归纳的推理方式，包括基于图结构的推理、基于规则学习的推理以及基于表示学习的推理。[22]前两种推理方法显式地定义了推理学习的特征，属于显式的推理方式；而基于表示学习的推理算法在学习向量表示的过程中，自动学习推理所需的特

征，是一种隐式的推理方式。值得说明的是，基于表示学习的推理方法属于联结主义的推理方法。

逻辑方法与非逻辑方法并非严格区分的。由于非逻辑方法构建的常识知识库是一种结构化的表示，可以相对容易地翻译为逻辑公式，因此，也可以采用逻辑方法对其进行推理。由此，也将面临将常识推理问题和常识知识自动形式化的问题。

3. 符号主义方法所面临的问题

符号主义方法面临的第一个难题是知识表示不足的问题。人类在与世界的互动过程中不断获得和更新关于世界的知识，使其大致符合世界的真实情况。然而，这些知识并非都可以采用符号来表示，特别是对于一些模糊和不确定性的知识。虽然目前构建了诸多的常识知识库，但这些知识库并未能构建关于世界的合适描述，来满足常识推理任务的需要。

符号主义方法面临的第二个难题是如何选择相关的前提参与推理。一个相对完备的常识知识库至少要具备上千万条甚至上亿条常识知识，面对一个具体常识推理问题，若将所有的常识知识参与推理，则会导致搜索空间过大，从而使得推理过程缓慢，甚至无法在合理时间内完成。目前符号主义方法尚未对这一问题提供良好的解决方案。Cyc项目通过构建微理论来缩减推理的范围，这种方案又引发了如下两个问题：（1）Cyc项目中的有2万多个微理论，这些微理论多达50层，而且存在诸多冗余的子类型关系，如何将常识知识正确地划分到不同的微理论是一个难题；（2）一个常识推理问题可能同时涉及多个微理论，如何选择合适的微理论参与推理是其所面临的第二个难题。在自动定理证明领域，目前已经开发了诸多工具来辅助前提的选择，例如，布兰切特（J. C. Blanchette）等提出的“锤子”方法利用启发式搜索来选择与当前推理相关的前提。[23] 然而，常识推理的前提选择比自动定理证明更为复杂，一方面，常识知识库的规模要远大于自动定理证明的知识库；另一方面，相较于定理知识，常识知识具有不确定性、模糊性和上下文依赖性，因而在前提选择方面面临更大的挑战。

逻辑方法面临的第三个难题是自然语言自动形式化的问题。当前自然语言形式化的方法大致划分为以下八种：基于规则的方法、基于句法分析的方法、基于实例的方法、基于归纳逻辑编程的方法、基于统计机器学习的方法、基于深度学习的方法、基于语言模型的方法以及基于语义分析的方法。[24] 目前，这些方法能够将具体领域的自然语言正确地翻译为形式语言，但是常识推理所涉及的是通用领域，目前尚未有统一的解决方案能够将任意的自然语言翻译为形式语言。

尽管非逻辑方法并不面临自然语言形式化的问题，但由于采用自然语言表示，因此面临着一词多义和表达多样性的问题。例如，ConceptNet中的一个项可能具有多个含义，比如，“apple”既指代一种水果，也指代电脑品牌。要使用ConceptNet中的常识进行推理，首先需要确定语境中的单词具体是哪种含义。与此同时，同一个意义可以采用不同的表达方式，这会造成匹配困难，即问题中出现的单词无法准确地定位到知识库中的概念。尽管当前的常识知识库中包含了单词间以及部分词组间的同义关系，但是由于词组的组合具有复杂性和多样性，因而很难构建一个包含所有同义表达的知识库。由此，也就造成了在推理时概念匹配的困难。

二、联结主义方法

人工智能界构建了诸多的常识推理任务来评测系统的常识推理能力。尽管联结主义方法在这些任务上已经取得非常优异的表现，[25]-[27] 但一直被认为不具有常识推理能力。然而，随着ChatGPT等大语言模型的兴起，这种认知正逐渐被改变。本节以GPT系列模型为例，来分析联结主义方法解决常识推理问题的基本思路以及所面临的问题。

1. 基于语言模型的常识推理方法

语言模型是语言中标记序列概率分布的模

型，对于给定的标记序列，可以预测下一个出现的（或缺失的）标记。标记可以是单词、字符甚至是字节。语言模型在大规模文本数据上进行预训练，使模型学习到语言的语法和语义特征。学习到的知识存储在模型的参数中，因而是一种隐式的知识表示。目前主流的语言模型都是基于Transformer模型的自监督学习来训练，并且通过自注意力机制有效地捕捉句子成分间的依赖关系。GPT系列模型作为一种自回归（auto-regressive）语言模型，其工作机制是依据输入的序列逐个生成输出的序列。具体来说，将标记序列$w_1 \cdots w_n$输入语言模型，预测下一个标记为w_{n+1}；然后将$w_1 \cdots w_n\ w_{n+1}$作为输入预测下一个标记为w_{n+2}；重复上述过程，直到输出完整的标记序列。

语言模型应用于常识推理问题经历了“预训练 + 微调”和“预训练＋提示”两种不同的范式。GPT-1模型采用“预训练＋微调”的模式，预训练结束后，使用常识推理的数据集对其进行微调。一个常识推理问题一般包括一个上下文z，一个问题q和一个可能的答案集$\{a_k\}$。将上下文、问题和每个可能的答案连在一起便得到多个文本序列$[z; q; \$; a_k]$，由此对每个这样的序列进行单独处理，通过一个Softmax层得到每个答案的概率分布。[28]一般概率高的即为正确答案。很显然这样的微调方式仅适合处理有备选项的常识推理问题。

当模型规模和训练数据增大到一定程度后，语言模型具备了一定的上下文学习能力。GPT-1模型有大约1.17亿参数，训练数据有约5GB；GPT-3模型具有1750亿参数，训练数据达到570GB。前者并不具有上下文学习能力，而后者具有。当语言模型具备上下文学习能力后，不再需要对语言模型进行微调，仅通过提示便可让语言模型完成下游任务。在常识推理任务上，GPT-3未经微调便达到了非常优异的表现。然而，在测试GPT-3在这些任务上的表现时，需要对测试任务进行改造，将常识推理任务改造为概率计算问题。以Winograd模式挑战任务[29]为例，此任务的目标是确定句子中代词的指称。例如，在句子“The city councilmen refused the demonstrators a permit because they feared violence”（市政府拒绝给示威者颁发游行许可证，因为他们担心暴力事件）中，依据常识我们可以确定“they”指代的是the city councilmen。在测试GPT-3时，先将两个备选项“the city councilmen”和“the demonstrators”代入到“they”的位置，得到两个句子，然后计算这两个句子的概率，概率高的即为正确答案。这种解决方案被认为并没有在推理，塞巴（W. S. Saba）批评这种方法不仅具有误导性，而且会有损那些需要真正基于理解的研究领域。[12]尽管GPT-3在诸多的常识推理任务上取得了优异表现，这并不意味着其已经具备了常识推理能力。例如，虽然GPT-3模型已经在物理常识推理任务上取得了优异表现，但无法处理像“我把奶酪放进冰箱，它会融化吗？”这样的常识推理问题。[27]然而，随着语言模型的不断进化，目前像ChatGPT这样的大语言模型已经展现出了非常强的常识推理能力。

从GPT-3到ChatGPT的进化过程，有两个中间版本的语言模型，Codex[30]和InstructGPT，[31]如图2所示。Codex是在GPT-3的基础上采用了Github上的代码进行训练而得到的；InstructGPT是在Codex的基础上，通过应用向人类意图对齐技术进一步优化而来的；ChatGPT和InstructGPT仅在训练数据方面存在微小差异。ChatGPT与GPT-3相比，主要的改变是使用了代码训练和向人类意图对齐技术。

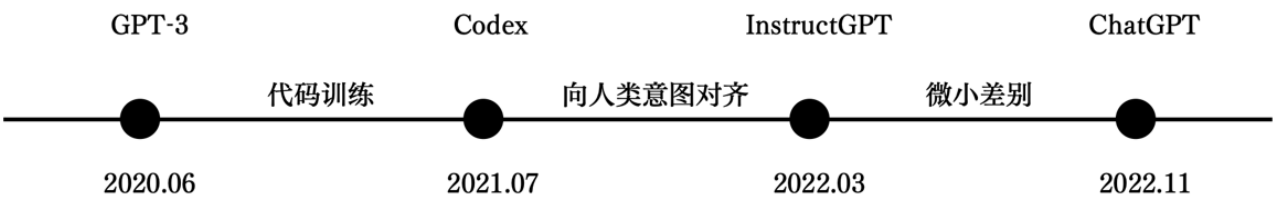

图2 从GPT-3到ChatGPT的进化

向人类意图对齐使得语言模型能够更好地理解人类的意图和指令，产生人类期望的输出。以Winograd模式挑战为例，ChatGPT可以通过对话式的方式直接回答问题，并且能对答案给出合理的解释，如图3所示。相较于GPT-3模型将Winograd问题转换为概率计算问题的“推理”方式，ChatGPT更有可能被认为具有常识推理能力。

目前并未有证据表明向人类意图对齐会带来常识推理能力的提升，相反甚至可能会影响语言模型在常识推理任务上的表现，这种情况也被称为“对齐税”（alignment tax）。例如，InstructGPT模型在经过向人类意图对齐后，在SQuAD、DROP和HellaSwag常识推理任务上的表现反而变差了。[31]-[34]GPT-4技术报告也表明，GPT-4在各项考试中所展现的能力可能主要来自于预训练阶段，人类反馈的强化学习过程对于其表现影响不大。[35]

语言模型在常识推理任务上的表现与人类的提问方式相关。魏杰森（J. Wei）等提出了思维链（chain of thought）的概念，并通过实验发现当在提示中加入推理的中间步骤示例时，语言模型在常识推理任务上的表现会得到大幅提升。[36]同时，发现了思维链推理能力是模型规模增大后所涌现出的一种能力，涌现能力是指小模型中未出现而在大模型中出现的能力。因此，常识推理能力的提升与模型规模呈现正相关的关系。小岛（T. Kojima）等进一步发现无需在提示中加入包含中间推理步骤的示例，仅加入“Let’s think step-by-step”这句咒语便可以提升语言模型的常识推理表现。[37]

代码训练可以提升语言模型的常识推理能力。小岛等发现在未经代码训练的GPT-3模型引入思维链推理后，在多步算术推理测试集上的表现提升了16.3%；而在经过代码训练的InstructGPT模型在此任务上的表现提升了61%。[37]马达恩（A. Madaan）等发现经过代码训练的语言模型在结构化常识推理任务上的表现要优于自然语言训练的模型。[38]关于为何代码微调会提升语言模型的推理能力，尚未见有明确的解释，一种可能的直观解释是程序代码中包含了推理的模式和思想，例如，面向过程编程类似人类解决问题的过程；面向对象编程中的类、继承、多态的思想也与人类的推理过程相似。[39]语言模型可以从代码中学到相关的推理模式。

2. 联结主义方法所面临的问题

联结主义方法具有不可解释性，体现在知识表示的不可解释和推理过程的不可解释。联结主义方法采用隐式的知识表示，知识存储在模型的参数中，并不是直接可见的，因此很难从模型中直接获取和解释知识。OpenAI采用GPT-4对GPT-2模型中的神经元进行解释，但仅有非常少量的神经元可以得到可靠的解释，而且这一研究还存在诸多理论和实践上的限制。[40]由于模型的参数是固定的，一旦预训练完成，知识的更新和修改需要重新进行预训练和微调，而无法像符号主义方法一样灵活地更新知识。联结主义方法采用下一个词预测的方式来完成推理的过程，这种简单的机制为何能

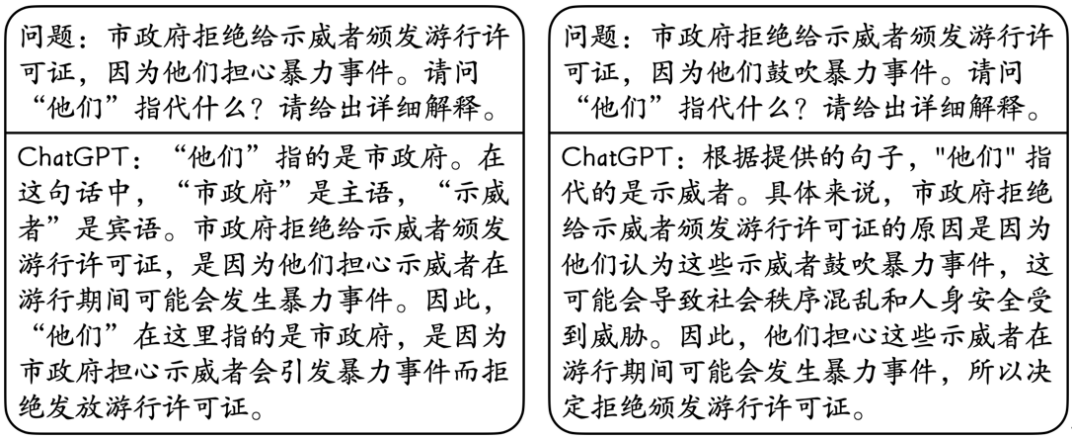

图3 ChatGPT回答Winograd模式问题实例

产生常识推理能力，目前尚未见有明确的解释。

目前语言模型在常识推理问题上的表现并不稳定。使用同一问题在对ChatGPT进行多次提问时，会得到不同，甚至截然相反的答案。这是由于语言模型采用采样（sample）的方式来生成答案，每次采样的不同会生成不同的答案。为了解决这一问题，利用“自一致性”（self-consistency）方法，通过生成多个答案，从中选择最一致的回答，可以有效缓解回答的随机性问题。[41]语言模型对提问方式是敏感的，采用不同的提问方式可能得出完全不同的答案，使用思维链的提问方式可以提升语言模型在常识推理任务上的表现。除此之外，大语言模型作为通用模型，其模型的更新需要平衡其各种能力，在增强其他能力的同时可能会降低其常识推理的表现。通过对ChatGPT的演化情况进行研究，发现ChatGPT随着版本的更新在自然语言推理任务上的表现变差，例如，Winograd推理任务上的表现从81.69%（2023年1月版本ChatGPT）降到了71.83%（2023年3月版本ChatGPT）。[42]GPT-4在识别质数测试中也出现了准确率骤降的情况，从97.6%（2023年3月版本GPT-4）降到2.4%（2023年6月版本GPT-4），而ChatGPT从2023年3月到2023年6月的更新却带来了准确率的大幅提升。[43]

大语言模型存在幻觉（hallucination）问题，在回答问题时可能生成虚构的信息，而不是基于已有的知识和数据，这可能会对其常识推理表现造成影响。从OpenAI发布的技术报告看，GPT-4模型在减少幻觉方面相较于ChatGPT有较大改进。[35]但是，目前来看语言模型的幻觉问题无法完全消除。目前大语言模型在构建前提、中间结论和最终结论之间的因果关联方面，仍存在问题。由此，忠实推理（faithful reasoning）逐渐成为大语言模型研究的重要方向之一。

联结主义方法在某些常识推理领域仍面临一些挑战。博尔吉（A. Borji）收集了ChatGPT一些推理失败的实例，涉及空间推理、时间推理、物理推理、心理推理、常识推理等。[44]方艺珍（Y. Bang）等对ChatGPT的推理能力进行评测，发现其在10个推理领域的准确率仅有64.33%，因此认为其是一个不可靠的推理器。[45]由于这些测试是基于ChatGPT的早期版本，因此，有些不能解决的问题会由于ChatGPT的升级变得可以解决。特别地，GPT-4相较于ChatGPT在常识推理方面有非常大的提升，因此，这些评测结果可能并不适用于GPT-4。一些研究结果表明尽管GPT-4的表现要明显优于ChatGPT，并且二者在分布外数据集以及需要逻辑推理的自然语言推理任务上表现不佳。[46]当前大语言模型在数学推理、逻辑推理、因果推理、抽象推理及心灵推理方面仍存在诸多不足。[47]值得说明的是，由于ChatGPT和GPT-4的表现受到提问方式的影响，因此，当在上下文中给出更多的实例或者引入思维链推理会进一步提升其评测的表现。

三、实例分析

莱韦斯克（H. J. Levesque）提出了“鳄鱼能参加长跑障碍赛吗？”（could a crocodile run a steeplechase?）的问题。[48]普通人可能从未设想过这样的问题，但是当人们听到这一

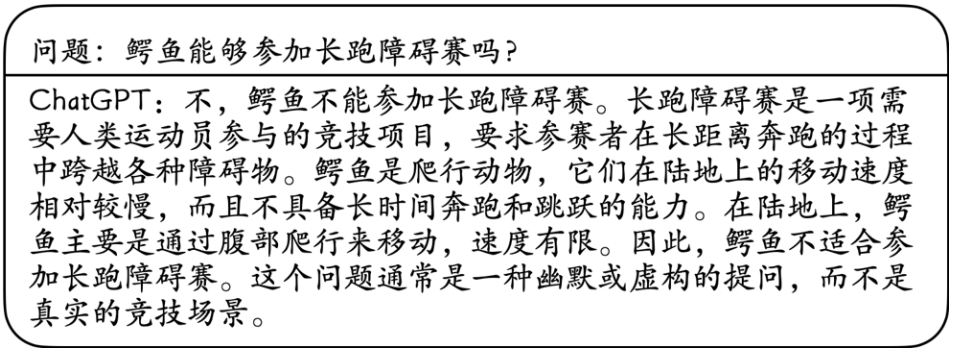

图4 ChatGPT回答问题实例

问题时，可以依据常识对其作出回答。目前ChatGPT已经能够就这一问题给出合理的回答，如图4所示。本节借用此实例来分析符号主义方法在解决常识推理问题时所面临的困难，并与联结主义方法进行对比。

要采用符号主义方法解决“鳄鱼能够参加长跑障碍赛”，首先要具有关于“鳄鱼”和“长跑障碍赛”的常识，然而，“鳄鱼”的百科知识中并不涉及“长跑障碍赛”；同时“长跑障碍赛”的百科知识中也不会提及“鳄鱼”。人类在回答这一问题时，会设想参加长跑障碍赛的基本条件，并分析鳄鱼是否具有这样的条件，由此得出结论鳄鱼不能参加长跑障碍赛。然而，在目前已有的常识知识库并不能涵盖上述推理过程所需的常识，因此，存在知识表示不足的问题。相较之下，联结主义方法由于采用海量的数据进行训练，在大部分的常识推理中并不面临知识表示不足的问题。

ChatGPT的回答中涉及到诸多的概念和常识，如“参加”“长距离奔跑”“障碍物”“移动速度”“爬行”“跳跃”“鳄鱼不具备长时间奔跑能力”“鳄鱼不具备跳跃能力”“鳄鱼通过腹部爬行来移动”等。现在假设常识知识库中包含充足的知识，如何依据问题描述（“鳄鱼能参加长跑障碍赛吗”）从海量的常识知识库中将上述知识选择出来，并组合成合理的论证是符号主义方法面临的另一个问题。相较之下，联结主义方法在回答常识推理问题时，大部分情况下可以选择相关的知识来构成合理的论证。大语言模型的这种能力目前仅仅已知与其注意力机制相关，其具体机制尚不清楚。

若采用逻辑方法解决此问题，还面临着自然语言自动形式化的问题，需要将问题描述和常识知识翻译为逻辑表达式，然而目前尚未有统一化的方法实现这种翻译。除此之外，逻辑方法的推理机制主要是演绎的，然而如果我们分析ChatGPT的回答会发现，这一回答并非完全演绎的，即如果我们将这一回答形式化为一阶公式，并不能必然推出我们期望的结论。

还有一种可能的设想是将这种类型的知识直接加入常识知识库。假设我们允许将“鳄鱼不能参加长跑障碍赛”这样的知识加入常识知识库。当面临其他相似的问题时，我们还需要加入与此相似的知识。例如，鳄鱼能打乒乓球吗？鳄鱼能打篮球吗？鳄鱼能做研究吗？鳄鱼能把水从瓶子里倒到杯子里吗？鸡能参加长跑障碍赛吗？苍蝇能参加长跑障碍赛吗？马能参加长跑障碍赛吗？假设有n种动物，有m种人类能够从事的活动，那这种组合就有n*m种。由此，会造成知识爆炸的问题。如果采用封闭世界语义，即所有未出现在数据库中的事实都是假的，虽然可以简化知识表示和推理过程，但是可能会导致错误的结论。例如，对于“鸵鸟能跨栏吗？”这样的问题，依据鸵鸟和跨栏的常识，我们可以推理出鸵鸟可以跨栏。然而，依据封闭世界假设，若这样的事实不包含在数据库中，则会推理出鸵鸟不能跨栏。

四、两种方法的对比

本节主要对上述两种方法从常识推理表现、知识表示方式、知识获取方式、推理方式、可解释性以及创造性方面进行对比，并探讨两种方法融合的可能性。

1. 常识推理表现

联结主义方法在所有的常识推理任务上都取得了最优表现。对于一些大规模的常识推理测试集，目前仅有基于联结主义方法构造的系统；对于相对小规模的常识推理任务，如Winograd模式挑战，尽管有符号主义方法的解决方案，但相较于联结主义方法表现较差。由于符号主义方法存在知识表示不足、前提选择以及自动形式化问题，目前仅在面向特定领域的常识推理任务中取得了良好表现。

2. 知识表示方式

符号主义方法通过符号和规则来表示知识，知识被显式地表示为符号和符号之间的关系，这种知识表示形式更为清晰和易于解释，但在表示模糊和不确定性知识时存在困难。联结主义则是通过神经网络的连接权重来表示知识，是一种隐式的表示方式。同时，知识以分散的形式编码在模型的不同部分，能够有效地

捕捉知识的模糊性和不确定性。

3. 知识获取方式

符号主义方法的知识库构建方式主要分为人工构建、众包构建和基于数据挖掘或本体学习技术的自动化构建方式。虽然目前可以在一定程度上自动构建知识库，但在实际应用中，通常需要进行人工审核和干预，以确保知识库的质量和准确性。联结主义方法主要采用自监督学习的方式来进行训练，使用大规模的未标记数据进行学习，无需人工标注。

4. 推理方式

符号主义方法利用推理规则来对符号进行操作从而实现推理。逻辑方法主要基于逻辑推理机来实现推理；非逻辑方法主要通过图推理和规则学习的方式进行推理。联结主义方法将输入数据传递到神经网络中，然后神经网络中的每个神经元会根据输入数据和连接权重计算其激活状态，并将结果传递给下一层神经元，经过多个层次的信息处理和传递，从而得出最终的推理结果。

5. 可解释性

由于符号主义的知识表示是显式的，并且推理过程是基于符号和规则进行的，因此可以很容易地解释模型的推理过程。由于联结主义的知识表示是分布式和隐式的，并且推理过程是通过数据在神经网络中的传递和处理进行的，因此很难对模型推理过程进行精确解释。

6. 创造性

符号主义方法的推理过程依赖规则和逻辑，无法生成超出规则和知识库内容之外的回答，因而缺乏创造性和灵活性；联结主义方法通过预测下一个词的方式来输出答案，能够依据不同的场景和数据，给出具有创造力的答案。

综上所述，两种方法在处理常识推理问题方面各有优劣，但目前联结主义方法在表现上相较于符号主义方法具有绝对优势。一个自然的想法将两种方法综合起来，构造兼具两者优点的统一框架。这一想法可以追溯到20世纪90年代，这种方法也被称为神经符号（neural-symbolic）方法。神经符号方法的目标是探索两种方法的优势，通过结合神经网络的学习能力和符号方法的严格、可解释的推理过程，来解决人工智能领域的问题。

在大语言模型出现之前，联结主义在感知相关的任务上取得了显著的成就，如图像识别、自动驾驶、语音识别等，但被认为并不具备复杂推理能力；而符号主义方法被认为在可解释性和推理方面更具有优势。因此，符号主义方法在解决常识推理问题方面被寄予厚望。符号主义的常识推理方法涉及自然语言形式化、常识知识库构建和推理机制三个部分，利用深度学习技术可以辅助这些过程的实现，从而提升了其在常识推理任务上的表现。与此同时，构建的常识知识库提供了丰富且精确的外部信息，在探索改进深度学习的鲁棒性、可解释性、弱标注学习等方面起到了关键作用。神经符号系统在推理效率、泛化能力和可解释性方面相较于单纯的符号系统或深度学习模型都有较大优势。[49]

大语言模型出现之后，传统认知正在被改变，深度学习方法被认为具有复杂推理能力。[50]依靠海量数据的学习，大语言模型在常识推理问题上展现出了非常强的鲁棒性，不仅能够对现实世界的场景进行推理，还能够对一些设想的荒谬场景进行推理。大语言模型虽然展现出了强大的常识推理能力，但其存在幻觉和推理能力欠缺等问题。目前，通过将传统的知识图谱与大语言模型集成，在减轻大语言模型的幻觉，增强其推理能力方面展现出了明显的效果。虽然思维链提示方法可以在很大程度上提升大语言模型的推理能力，但在处理依赖符号表达和严格推理规则的逻辑推理方面仍然可能遇到困难。通过将符号表达式和逻辑规则融入到大语言模型的思维链中，不仅可以提升其推理表现，而且推理更加忠实和可解释。[51]

结　语

本文介绍了符号主义和联结主义在解决常识推理问题上的基本方法以及所面临的问题，并对二者在常识推理表现、知识表示方式、知识获取方式、推理方式、可解释性和创造性方

面进行了对比。从目前的分析来看，符号主义方法在现有的理论框架下面临着知识表示不足、前提选择及自动形式化等问题，这也使其目前无法为常识推理提供一种可行的解决方案。联结主义方法在诸多方面都相较于符号主义方法具有较大的优势，而且展现出了巨大的发展潜力。因此，本文认为在未来一段时间内联结主义方法将成为常识推理研究的主流解决方案。联结主义方法还需要就知识更新、增强推理稳定性、减少幻觉以及提升可解释性等方面进一步展开研究。与符号主义方法的结合可以在解决这些问题方面起到促进作用。

[参 考 文 献]

[1]Mueller, E. T. *Commonsense Reasoning: An Event Calculus Based Approach*[M]. Waltham: Morgan Kaufmann, 2015.

[2]Dreyfus, H. L., Dreyfus, S. E. 'Making a Mind Versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint'[J]. *Daedalus*, 1988, 117(1): 15–43.

[3]Pinker, S., Mehler, J. *Connections and Symbols*[M]. Amsterdam: MIT Press, 1988.

[4]Smolensky, P. 'Connectionist AI, Symbolic AI, and the Brain'[J]. *Artificial Intelligence Review*, 1987, 1(2): 95–109.

[5]DyerM, G. 'Connectionism Versus Symbolism in High-Level Cognition'[A], Horgan,T., Tienson, J. (Eds.) *Connectionism and the Philosophy of Mind*[C], Dordrecht: Kluwer Academic Publishers, 1991, 382–416.

[6]Dinsmore, J. *The Symbolic and Connectionist Paradigms: Closing the Gap*[M]. New York; Hove: Psychology Press, 2014.

[7]Goel, A. 'Looking Back, Looking Ahead: Symbolic Versus Connectionist AI'[J]. *AI Magazine*, 2022, 42(4): 83–85.

[8]McDermott, D. 'A Critique of Pure Reason'[J]. *Computational Intelligence*, 1987, 3(1): 151–160.

[9]Davis, E., Marcus, G. 'Commonsense Reasoning and VCommonsense Knowledge in Artificial Intelligence'[J]. *Communications of the ACM*, 2015, 58(9): 92–103.

[10]LeCun, Y., Bengio, Y., Hinton, G. 'Deep Learning'[J]. *Nature*, 2015, 521(7553): 436–444.

[11]Li, H. 'Deep Learning for Natural Language Processing: Advantages and Challenges'[J]. *National Science Review*, 2017, 5(1): 24–26.

[12]Saba, W. S. 'A Simple Machine Learning Method for Commonsense Reasoning? A Short Commentary on Trinh & Le'[J]. ArXiv Preprint, ArXiv: 1810.00521, 2018.

[13]Matuszek, C., WitbrockM., Cabral, J., et al. 'An Introduction to the Syntax and Content of Cyc'[A], *Papers from the 2006 AAAI Spring Symposium*[C], Technical Report SS-06-05, 2006, 44–49.

[14]Lenat, D. B. 'CYC: A large-scale Investment in Knowledge Infrastructure'[J]. *Communications of the ACM*, 1995, 38(11): 33–38.

[15]Lenat, D. B., Guha, R. V. 'The Evolution of CycL, the Cyc Representation Language'[J]. *ACM SIGART Bulletin*, 1991, 2(3): 84–87.

[16]Fellbaum, C. 'Word Net'[A], Poli, R., Healy, M., Kameas, A. (Eds.) *Theory and Applications of Ontology: Computer Applications*[C], London; New York: Springer, 2010, 231–243.

[17]Fillmore, C. J., Lee-Goldman, R., Rhodes, R. 'The FrameNet Constructicon'[A], Boas, H. C., SagI, A. (Eds.) *Sign-based Construction Grammar*[C], CSLI Publications, 2012, 309–372.

[18]Schuler, K. K. 'VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon'[D]. Philadelphia: University of Pennsylvania, 2005.

[19]Speer, R., Chin, J., Havasi, C. 'ConceptNet 5.5: An Open Multilingual Graph of General Knowledge'[A], Proceedings of the 31st AAAI Conference on Artificial Intelligence[C], 2017, 4444–4451.

[20]Suchanek, F. M., Kasneci, G., Weikum, G. 'YAGO: Acore of Semantic Knowledge'[A], Williamson, C., Zurko, M. E. (Eds.) *Proceedings of the 16th International Conference on World Wide Web*[C], New York: Association for Computing Machinery, 2007, 697–706.

[21]Lehmann, J., Isele, R., Jakob, M., et al. 'DBpedia-A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia'[J]. *Semantic Web*, 2015, 6(2): 167–195.

[22]王昊奋、漆桂林、陈华钧. 知识图谱：方法，实践与应用[M]. 北京：电子工业出版社，2019.

[23]Blanchette, J. C., Kaliszyk, C., PaulsonL, C., et al. 'Hammering Towards QED'[J]. *Journal of Formalized Reasoning*, 2016, 9(1): 101–148.

[24]徐超. 基于认知的自然语言自动形式化研究[M]. 北京：中国社会科学出版社，2023.

[25]Yang, Z., Dai, Z., Yang, Y., et al. 'XLNet: Generalized

Autoregressive Pretraining for Language Understanding'[J]. *Advances in Neural Information Processing Systems*, 2019, 32: 5753–5763.

[26] Liu, Y., Ott, M., Goyal, N. 'RoBERTa: A Robustly Optimized BERT Pretraining Approach'[J]. ArXiv Preprint, ArXiv:1907.11692, 2019.

[27] 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.

[28] Radford, A., Narasimhan, K., Salimans, T., et al. 'Improving Language Understanding by Generative Pretraining'[EB/OL]. OpenAI, https://openai.com/index/language-unsupervised. 2018–06–11.

[29] Levesque, H. J., Davis, E., Morgenstern, L. 'The Winograd Schema Challenge'[A], *Proceedings of Thirteenth International Conference on the Principles of Knowledge Representation and Reasoning*[C], Rome, 2012, 552–561.

[30] Chen, M., Tworek, J., Jun, H. et al. 'Evaluating Large Language Models Trained on Code'[J]. ArXiv Preprint, ArXiv: abs/2107.03374, 2021.

[31] Ouyang, L., Wu, J., Jiang, X., et al. 'Training Language Models to Follow Instructions with Human Feedback'[J]. *Advances in Neural Information Processing Systems*, 2022, 35: 27730–27744.

[32] Rajpurkar, P., Jia, R., Liang, P. 'Know What You Don't Know: Unanswerable Questions for SQuAD'[A], Gurevych, I., Miyao, Y. (Eds.) *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*[C], Melbourne: Association for Computational Linguistics, 2018, 784–789.

[33] Dua, D., Wang, Y., Dasigi, P., et al. 'DROP: A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs'[A], Burstein, J., Doran, C., Solorio T. (Eds.) *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*[C], Minneapolis: Association for Computational Linguistics, 2019, 2368–2378.

[34] Zellers, R., Holtzman, A., Bisk, Y., et al. 'HellaSwag: Can A Machine Really Finish Your Sentence?'[A], Korhonen, A., Traum, D., Màrquez, L. (Eds.) *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*[C], Florence: Association for Computational Linguistics, 2019, 4791–4800.

[35] OpenAI. 'GPT-4 Technical Report'[EB/OL]. https://cdn.openai.com/papers/gpt-4.pdf, 2023–03–27.

[36] Wei, J., Wang, X., Schuurmans, D., et al. 'Chain-of-thought Prompting Elicits Reasoning in Large Language Models'[A], Koyejo, S., Mohamed, S., Agarwal, A., et al. (Eds.) *Proceedings of the 36th International Conference on Neural Information Processing Systems*[C], New York: Curran Associates Inc., 2022, 24824–24837.

[37] Kojima, T., Gu, S. S., Reid, M., et al. 'Large Language Models Are Zero-shot Reasoners'[A], Koyejo, S., Mohamed, S., Agarwal, A. et al. (Eds.) *Proceedings of the 36th International Conference on Neural Information Processing Systems*[C], New York: Curran Associates Inc., 2022, 22199–22213.

[38] Madaan, A., Zhou, S. Alon, U., et al. 'Language Models of Code Are Few-shot Commonsense Learners'[A], Goldberg, Y., Kozareva, Z., Zhang, Y. (Eds.) *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*[C], Abu Dhabi: Association for Computational Linguistics, 2022, 1384–1403.

[39] Fu, Y., Peng, H., Khot, T. 'How Does GPT Obtain Its Ability? Tracing Emergent Abilities of Language Models to Their Sources'[EB/OL]. https://yaofu.notion.site/How-does-GPT-Obtain-its-Ability-Tracing-Emergent-Abilities-of-Language-Models-to-their-Sources-b9a57ac0fcf74f30a1ab9e3e36fa1dc1, 2022–12–11.

[40] Bills, S., Cammarata, N., Mossing, D., et al. 'Language Models Can Explain Neurons in Language Models'[EB/OL]. https://openai.com/index/language-models-can-explain-neurons-in-language-models/. 2023–05–09.

[41] Wang, X., Wei, J., Schuurmans, D., et al. 'Self-consistency Improves Chain of Thought Reasoning in Language Models'[A], *Proceedings of the Eleventh International Conference on Learning Representations*[C], New York: Curran Associates Inc, 2023.

[42] Tu, S., Li, C., Yu, J., et al. ChatLog: Recording and Analyzing ChatGPT Across Time'[J]. ArXiv Preprint, ArXiv:2304.14106, 2023.

[43] Chen, L., Zaharia, M., Zou, J. 'How is ChatGPT's Behavior Changing Over Time?'[J]. *Harvard Data Science Review*, 2024, 6(2): 1–47.

[44] Borji, A. 'A Categorical Archive of ChatGPT Failures'[J]. ArXiv Preprint, ArXiv:2302.03494, 2023.

[45] Bang, Y. 'A Multitask, Multilingual, Multimodal

Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity'[A], Park, J., Arase, Y., Hu, B., et al. (Eds.) *Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics*[C], Nusa Dua: Association for Computational Linguistics, 2023, 675–718.

[46]Liu, H., Ning, R., Teng, Z., et al. 'Evaluating the Logical Reasoning Ability of ChatGPT and GPT-4'[J]. ArXiv Preprint, ArXiv: 2304.03439, 2023.

[47]Ma, Y., Zhang, C., Zhu, S. C. 'Brain in a Vat: On Missing Pieces Towards Artificial General Intelligence in Large Language Models'[J]. ArXiv Preprint, ArXiv: 2307.03762, 2023.

[48]Levesque, H. J. 'Logic and the Complexity of Reasoning'[J]. *Journal of Philosophical Logic*, 1988, 17: 355–389.

[49]Yu, D., Yang, B., Liu, D., et al. 'A Survey on Neural-Symbolic Learning Systems'[J]. *Neural Networks*, 2023, 166: 105–126.

[50]徐超. 创造力、涌现能力与经验主义的胜利[J]. 认知科学, 2023, 7（1）: 29–43.

[51]Xu, J., Fei, H., Pan, L., et al. 'Faithful Logical Reasoning Via Symbolic Chain-of-thought'[A], Ku, L., Martins, A., Srikumar, V. (Eds.) *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*[C], Bangkok: Association for Computational Linguistics, 2024, 13326–13365.

［责任编辑　王巍 谭笑］