人工智能要考虑到溯因推理※

〔美〕罗仁地[1] 著　姚洲[2] 译

(1.北京师范大学珠海校区 人文和社会科学高等研究院语言科学研究中心,广东 珠海 519087;
2.浙大城市学院 人文学院,浙江 杭州 310015)

摘要: ChatGPT可以搜索大量数据,并根据可能组合搭配的统计概率,拼凑出看似人类生成的文本。这并不令人惊讶,因为所使用的数据本身就是人类创造的文本。但是这些算法并不理解它们正在写什么,也没有超越狭义人工智能的范围。要实现通用人工智能,需要这些算法模仿人类的意义创造能力,这取决于对方对交际者交际意图的溯因推理。溯因推理本质上是对交际者为何做了他们所做之事(可能是语言行为或仅仅只是手势或面部表情,或三者兼具)的猜测。它是一种因果推理,提出某现象的所以然,这在创造一般性的知识以及推断他人执行某种行为时的动机时都有应用,包括在交际中的使用。这种因果推理是我们意识的核心。

关键词:人工智能　ChatGPT　溯因推理　归纳推理　交际行为

DOI:10.19866/j.cnki.cjxs.2023.04.010

最近,关于ChatGPT、GPT-4以及类似的人工智能产品引发了很大的讨论。在ChatGPT出现的头两个月里,超过1亿人使用了该产品。这些算法能够搜索大量的数据(也就是所谓的大型语言模型),并根据可能组合搭配的统计概率(也就是归纳法),生成看似由人类产生的文本。这并不奇怪,因为这些算法所使用的数据本身就是人类创造的文本。但这种结果引发了一些人对这种算法可能带来危险的担忧,好像这个算法正在自我思考一样。然而,这些算法并不真正理解它们所生成的内容。它们只是像镜子一样反映了它们所收集的数据以及与其互动的人所提供的数据,从而反映出人们的个性和欲望。它们的工作基于归纳法,通过访问大量数据来总结模式。然而,它们无法超越这个范围,只能局限于执行特定的任务,比如文本翻译或预测某人的购买偏好。

一、狭义人工智能vs通用人工智能

在人工智能研究的历史中,20世纪50年代到90年代的主导模型是符号人工智能,一种基于语法规则和演绎推理方法的模型。然而,该模型并没有取得很好的成果。进入世纪之交,该领域的学者转向了归纳推理,也称为联结主义方法或机器学习方法,以杰弗里·辛顿(Geoffrey Hinton)为代表(辛顿已经在70年代提出联结主义方法,但当时没有多少人相信行得通)。这种方法产生的结果要好得多,因此归纳推理现在成为主导方法。①谷歌大脑实验室的负责人杰夫·迪恩(Jeff Dean)曾说:"我们不需要语法。"

这种人工智能是一种弱人工智能,也就是指那些只能执行特定任务的非人类系统,无法超越这个范围。而人工智能的下一步发展是发展出强人工

作者简介:罗仁地(Randy J. LaPolla, 1955—),男,美国人,北京师范大学珠海校区人文和社会科学高等研究院语言科学研究中心特聘教授,主要从事意义创造论、汉藏语研究。

姚洲(1991—),男,湖北宜昌人,浙大城市学院人文学院讲师,主要从事知识表示、认知语言学、意义创造论研究。

※本文曾在香港城市大学2023年6月6—8日举办的"大数据时代的语言学、语言应用和翻译研究国际学术研讨会"做过汇报,由原作者罗仁地授权译者翻译成中文在《长江学术》发表,特此致谢!——译者注

①Lewis-Kraus, Gideon. 2016. The great AI awakening. *The New York Times Magazine*, published online 14 December, 2016.

智能，即能够学习、解决问题、适应并改善自身系统的人工智能。这样的系统甚至可以执行超出其设计任务的工作，但前提是机器能够模仿人类的推理能力，并真正创造出有意义的内容。这也是许多科学家对未来的担忧，因为我们将无法控制这样的系统。

超过 1000 名科学家签署了一封信，请求科技公司延迟通用人工智能的开发。杰弗里 • 辛顿最近辞去了谷歌的工作，原因是他反对该公司计划开发通用人工智能，他认为自己在那里工作时无法批评公司（《纽约时报》2023 年 5 月 1 日）。他对自己一生的工作感到后悔。目前，人们已经意识到弱人工智能系统中存在一些固有的偏见问题，但在通用人工智能系统中，这些问题将更加严重，因为它们能够自主决策并超越编程范围进行操作。

如果没有任何伦理或道德监督，后果将是灾难性的。即使是坚定支持科技的伊隆 • 马斯克（Elon Musk）也警告说，高级人工智能可能对人类构成“生存危机”。目前的控制方法是不够的。

然而，我们还没有达到那个阶段。要让机器像人类一样真正创造有意义的内容，它需要模拟人的溯因推理能力，这是人类用来创造各种意义的方法（包括交流）。但是目前机器无法模仿这种能力。溯因推理的关键部分不仅是获取信息，还包括判断信息与所讨论问题的相关性。迄今为止，溯因推理在下一阶段的重要性甚至在专注于这项技术的学者中也没有得到广泛认识，但有一些人正在努力解决这个问题。

二、什么是溯因推理？

推理有两种主要类型：证明性推理（也被称为演绎推理）和非证明性推理；非证明性推理又有两个子类型：归纳推理和溯因推理。归纳推理是对一组数据（或现象）的概括，溯因推理则是提出某现象的所以然（即因果推理）。在证明性（分析性）推理中，前提的真实性保证了结论的真实性，所以它是一个恒真式，比如“所有的人都会死”（前提）和“苏格拉底是人”（前提）会导出“苏格拉底会死”（结论）。但在非证明（综合性）推理中，前提的真实性仅仅使结论的真实性成为可能。正如查尔斯 • 皮尔斯（Charles Pierce）所言：“归纳的本质是从一组事实推断出另一组相似的事实，而假设（即溯因推理——作者注）则是从一类事实推断出另一类事实。”①

溯因推理本质上是对为什么某种现象会以其特定的方式存在的假设（猜测），无论那个现象是自然物体或事件，还是人类行为。这是一种反向的因果推理，即从结果推导出原因。这在创造一般性的知识以及推断他人执行某种行为时的动机时都有应用，包括在交际中的使用。根据皮尔斯的观点，溯因推理是为了解释事实，即创造假设：“只有这三种推理方式，演绎或归纳都不能给我提供任何新想法。除非我能通过假设追根究底，否则我还不如放弃尝试去理解它们。”②保罗 • 格莱斯（H. Paul Grice）更是说溯因推理是我们意识的核心：“意识本质上是从结果到原因的推理。”③溯因推理实际上是一种生存本能，为这个世界“创造意义”，即让我们对所住的环境里的现象有一种主观的了解，比如了解这些现象对我们有益还是有害。

三、溯因推理与交际

当我们进行溯因推理时，我们并不是利用整个环境，而是选择我们认为与理解我们试图解释的现

① Peirce, Charles S. 1878[1992]. Deduction, induction, and hypothesis. *Popular Science Monthly* 13（August 1878）: 470—82. Reprinted in *The Essential Peirce: Selected philosophical writings* Vol. I（1867—1893）, edited by Nathan Houser and Christian Kloesel, 186—199. Bloomington: Indiana University Press, 198.

② Peirce, Charles S. 1900[1985]. *Historical Perspectives in Peirce's Logic of Science*, 2 volumes, C. Eisele（ed.）. Amsterdam: Mouton Publishers, 878—879.

③ Grice, H. Paul and Alan R. White（1989[1961]）. The causal theory of perception. *Proceedings of the Aristotelian Society, Supplementary Volumes*, Vol. 35: 121—168. Oxford University Press, 122. Also published in *Studies in the Way of Words*. Cambridge, MA: Harvard University Press, 1989.

象相关的某些事实和假设。这一环境被称为“理解的环境”（context of interpretation）。正是该环境使得现象对人来说“有意义”。该环境的创建当然完全是主观的，所以结果也是主观的。

溯因推理的一种用途是通过推断某些可观察现象的原因来预测未来的情况。心理学文献中有大量关于预期/预测的心理学证据，实际上，安迪·克拉克（Andy Clark）认为大脑“本质上是预测机器”[①]。当我们试图理解其他人正在做什么和可能做什么时，就可以用这种方法来推断他们在做他们所做的事情时的意图。[②]这也是生存所必需的，也是人跟人互动的基础。

尝试了解其他人在做什么的一种应用是，在他们有目的地试图让你推断他们的意图时推断他们的意图。沟通就是：一个交际者进行一个交际行为（可能是语言行为，也可能只是手势或面部表情，或者三者都有），而对方推断出交际者的交际意图，即交际者为什么做他们所做的交际行为。对方被迫推断特定理解的程度，取决于交际行为在多大程度上制约对方选择获得在该语境中有意义的理解所必需的语境预设。

语言学一直有一种错误的观念，认为意义全在于形式，但事实并非如此。由于交际基于非证明性推理，因此交际本质上是不确定的；语言只是提供限制推理的线索。这不是一个编码解码的过程；它是推断交际者在进行交际行为时的意图。交际不是一件容易发生的事情，正如卡尔·波普尔（Karl Popper）所言：“永远记住，不可能以不被误解的方式说话：总会有人误解你。”[③]

即使是交际行为的识别也需要推理，因此不需要对方熟悉形式，只要对方能够推断交际者的意图即可。例如，图 1 的形式不是人们熟悉的英文字母形式，而是由不同爱尔兰图案组成的。虽然形式陌生，但不妨碍人们认出谷歌公司的目的是让人推理出 Google 这个名字。

此外，我们对意义的创造（与所有事物有关，而不仅仅是语言）与我们所知道的、对我们来说最重要的或我们自己的观察角度有关，比如图 2 左边的

图 1

① Clark, Andy. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36(3), 181—253.

②基于转换器的生成式预训练模型可以预测某些事情，但只是基于概率统计，而不是我们使用溯因推理的方式。

③ Popper, Karl. 1976. *Unended Quest: An Intellectual Autobiography*. London & NY: Routledge, 29.

图 2

图大部分的人会推测出 525，但换一个角度图 2 就会推测出 252。这完全是由角度决定的。[①]

一般来说，交际者还对对方能够理解什么进行推理（猜测），然后使用最有可能促进对方推理过程的交际行为。但是文学、艺术和幽默是例外，其目标是让对方做更多的推理工作，这是因为我们从推理中获得了愉悦，就像我们从满足其他生存本能中获得愉悦一样。

交际可以使用或不使用语言进行。功能性磁共振成像研究表明，非语言和语言交际在大脑的相同区域进行处理，包括那些被称为“布罗卡区”和“韦尼克区”的区域。[②]非语言交际和语言交际之间的区别只是工具或方式的不同，从而导致精确度的差异，就像用手把面包撕成两片和用刀小心地切面包之间的区别一样。语言有助于制约推理过程，使听话者更容易推断（猜测）交际者的意图。我们对语言的认识只是知道词语和结构在过去如何被用来实现某种目的。我们将这种知识作为理解环境的一部分，用于推断说话者使用某些词语的意图，但我们创造的意义将独属于那个特定的环境，并因此扩展了词语和结构的使用。这就是为什么语言一直在变。推理过程可以或多或少地受到制约，但绝不会完全受到制约（即不可能完全确定其意）。许多心理语言学文献的一个问题是假设涉及语言的意义创造是特殊的，但他们很少将其与不涉及语言的意义创造进行比较。我的假设是，如果我们进行这样的研究，我们将不会发现任何差异。

由于交际中的意义创造取决于推断交际者执行某些交际行为的意图，因此在单词或符号的意义上不存在与某些交际者的实际使用脱节的语义。也就是说，一切都是语用。当我们试图了解脱离具体环境的某词、句或事时，我们会创造一个理解环境，在该理解环境中，该词、句或事才可能“有意义”，也就是说，我们通过选择一个赋予它意义的特定框架（理解环境）来创造它的意义。

四、警惕人工智能可能存在的危险

我们需要关注当前这一代的弱人工智能，因为它可能被用来传播虚假信息，而且提供答案的可靠性并不高（例如当它为一名律师“虚构”案例时）。此外，它还反映了其所获取的数据中存在的偏见和态度。

要实现通用人工智能，需要能够模仿人类创造意义的方法，这取决于溯因推理的能力。这有可能

①其实这里的图根本不是数字，但如果我们能创造意义就会创造意义。这跟溯因推理法本能与习惯的性质有关。

② Xu, Jiang, Gannon, Patric J., Emmorey, Karen, Smith, Jason F., & Braun, Allen R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences*, 106(49), 20664—20669.

在几年内实现，特别是因为谷歌、微软和伊隆·马斯克等正在竞相创造通用人工智能。[①]这将是非常危险的，因为我们目前还不知道如何控制它。做这样的研究一定要以人为中心，不能像以前那样为了科技本身而发展科技，那样的研究导致了现在的很多大问题。[②]

正因如此，目前已有超过27000名技术专家签署了前文提到的信件，要求这些公司暂时停止开发这些算法，直到找到相应的控制方法为止。最近，OpenAI、Google、DeepMind、Anthropic和其他人工智能实验室的领导人又签署了第二封信，警告未来的系统可能像大流行病和核武器一样具有致命性，因此“减轻人工智能灭绝人类的风险应成为全球的首要任务”。

此外，澳大利亚首席科学家发布了一份关于生成式人工智能的《快速响应信息报告》（*Rapid Response Information Report*），对这些危险提出了警告，并讨论了应对方法。

与杰弗里·辛顿一样，我也处于纠结之中：一方面我想让人们了解创造意义的方式（因为我在这里提出的观点与许多领域都有关，不仅仅是计算机科学和语言学），另一方面我又担心一旦他们获得成功的秘诀（即溯因推理法）将带来危险和失控。

Considering Abduction in Artificial Intelligence

Randy LaPolla[1]　Trans. Yao Zhou[2]

（1. Institute for Advanced Studies in Humanities and Social Sciences, Beijing Normal University at Zhuhai, Zhuhai 519087, Guangdong, China; 2. School of Humanities, Hangzhou City University, Hangzhou 310015, Zhejiang, China）

Abstract: ChatGPT can search a large amount of data and, on the basis of the statistical probability of possible collocations, can put together something that seems like human-produced text, which should not be surprising, as the data used is human text. But these algorithms do not understand what they are writing, and cannot go beyond Artificial Narrow Intelligence. To achieve Artificial General Intelligence will require algorithms to imitate human meaning creation, which depends on abductive inference of the communicative intention of the communicator by the addressee. Abductive inference is essentially guesses as to why the communicator did what they did(which may be languaging or just gesture or facial expression, or all three). It is causal reasoning, asking why some observed phenomenon is the way it is, and this is applied in creating general knowledge, and in inferring the motivations of others when they perform some action, including its use in communication. This causal reasoning is the core of our consciousness.

Keywords: Artificial Intelligence; ChatGPT; Abduction; Inductions; Communication

责任编辑：杨旭

①杨立昆（Yann LeCun）认为大型语言模型的进一步发展并不会实现通用人工智能。

②〔美〕罗仁地：《以人为中心：交叉研究的必然走向》，《语言战略研究》2023年第3期。