

人工智能写作中小说标题生成的算法研究

刘 赛

摘 要:算法研究是人工智能写作领域研究的基础,目前小说创作算法研究还处于肇始阶段,非常不成熟。但是基于已有成果进行大小标题的人工智能创作,则是非常可能的,且可为未来小说创作算法研究提供经验。在小说标题生成的算法研究中,语义空间算法、修辞算法和遗传算法具有内容生产的可推广性。标题生成的三种算法研究,可以打通人工智能写作中叙事语法层和语义算法层的深层逻辑关系,具有较大研究价值。

关键词:人工智能;算法研究;标题生成

中图分类号:H315 **文献标识码:**A **文章编号:**1002-7343(2019)04-0038-06

如今中国数字社会的日均流量产值可达19万亿G,随着人工智能技术和大数据可用性的不断改进,投放市场的文学流量也日益庞大。在这种发展背景下,如何收集和优化小说出版的数据资源,如何为自己的小说创作一组大小标题,让它们在上述流量大战中胜出?人工智能可能会发挥越来越多的作用,它使小说与读者的交互得到大数据的支撑,使内容生产获得更深层次的数据加工和挖掘,尤其是可以应用于小说人物设计、情节策划。作为小说内容创作和策划的最精要部分,标题与小说内容紧密联系,以其简约而凝练的特点,对小说主题进行适应性概括,从而提升小说吸引力。本文通过搜索、输出和筛选环节,展示了小说标题生成的算法运用,并讨论了人工智能写作算法中的生成问题。

一、一种自动创建小说标题的算法模型

何谓算法?“算法,是指通过数学的思想和方法,提供解决问题的思路,运用编程的手段来完成问题的解决或处理。”^①对于本文来说,就是借用数字符号的算法模型来解决标题生成问题。人工智能的标题创作需要以下步骤:首先,建立单词的语义搜索空间;其次,以各种输出函数表达式建立神经元,本文中主要构建比喻句式的输出函数表达式;再次,每两个神经元节点为一个加权值,赋予输出函数不同加权值,进而依此值导出单词排列信息;最后,导入遗传算法进行评估,经过筛选符合目标的目标词和修饰词,最终形成完整标题。人工智能标题创作借鉴了神经网络算法,其算法模型主

作者简介:刘赛,1990年生,山东枣庄人,上海大学文学院博士研究生。电子邮箱:327594479@qq.com。

①郑舸:《浅议计算机算法认识》,《通讯世界》2018年第12期。

要包含语义空间算法(搜索)、比喻修辞算法(输出)和遗传算法(筛选)。这种神经网络式的算法模型具有深度学习特征,它使新一代 AlphaGo 迅速击败人类围棋高手。这一代的 AlphaGo 采用神经网络算法,以策略网络和价值网络两个模块组成的蒙特卡洛树搜索,其优势在于可以针对筛选后的落棋点进行深度计算,降低了搜索的维度。简而言之,就是通过优化数据库中的搜索方式,来达到一局面一搜索,一直到最终的胜负^①。这和从前根据围棋规约编写围棋程序不同,不再是仅仅模拟旗手们的下棋策略,而是不断增加对语义空间算法的改进。类似例子还有 Libratus 在为期 20 天、12 万张牌的扑克比赛里击败了四位来自世界各地的著名扑克选手。

二、基础算法与数据库:语义空间的算法构建

正如人脑具有记忆功能,语义空间算法作为基础算法,承担提供数据库支持的功能。在数据为王的信息化社会里,国家与市场可以使用语义空间算法为公文、新闻、小说快速提供语料库词汇和语段支撑,例如在 2017 年四川九寨沟地震中,中国地震台网官方微信将地震灾情快速向公众传播,该稿件生产仅用时 25 秒^②。而且,语义空间算法不仅为目标词提供可替换数据库,后续环节的数据搜索、语段输出、筛选和评估都要在其中展开。语义空间算法是大数据和人工智能的完美结合,它一方面不断补充与完善,另一方面具有不同类型与使用功能。

在人文科学领域,语义空间算法为其研究带来了极大便利。人文学者可以通过语义空间算法搭建人文数字研究平台,并利用平台查找人文资料,分析人文数据,甚至做个性化研究。国外在这方面的数据库搭建较为完善,例如基于网络的包含 20 亿组单词语义空间数据库 ukWaC²;中文圈语义空间数据库做的时间较长、完善度较高的,有中国历代人物传记资料库(CBDB),这是由哈佛大学费正清研究中心、北京大学中国古代史研究中心等联合搭建的语义空间资料库。此外,还有数位人文学术研究平台(DocuSky),它由台湾大学数位人文研究中心、资讯工程学系数位典藏与自动推论实

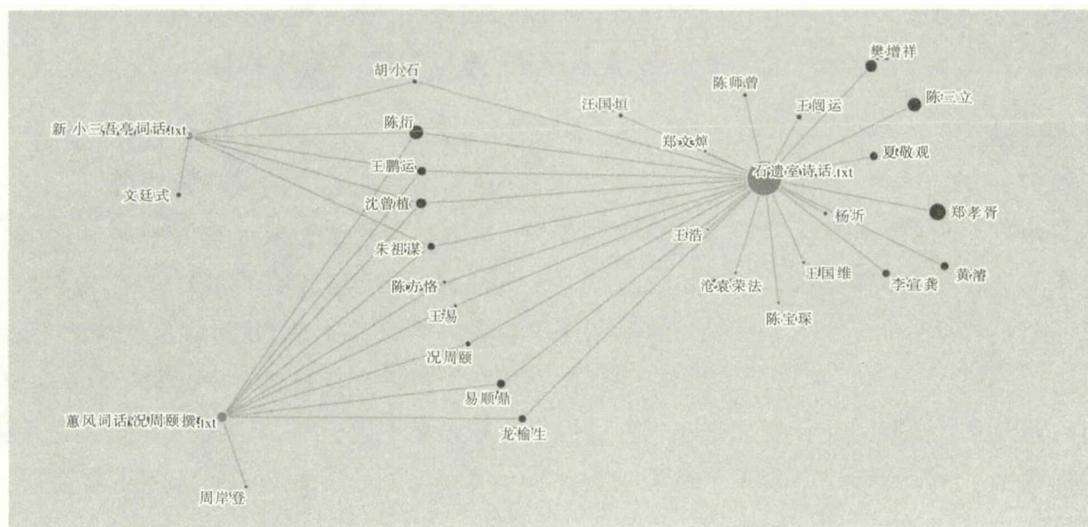


图 1 《石遗室诗话》《蕙风词话》与《小三吾亭词话》书中涉及人物关系图

^①程思雨、林锋:《计算机围棋 AlphaGo 算法对人类围棋算法的影响》,《中国科技信息》2019 年第 2 期。

^②白龙、林楹:《人工智能背景下“机器写作”在灾难报道中的应用和发展趋势》,《吉林广播电视大学学报》2018 年第 12 期。

实验室规划,具有“堪輿与分析”“时空资讯”“视觉化”等功能,旨在建设一个开源链接与友善互动的数字人文研究环境,自主且自由地融合数字人文科技进行研究的数据平台。在图1中,笔者利用 Docusky 数位人文平台绘制了《石遗室诗话》《蕙风词话》《小三吾亭词话》三本书涉及人物的关系图,为各书对不同人物的关注度做数据展示。

小说标题的语义空间算法,首先选定目标单词(例如“英雄”)和形容词(例如“悲剧的”),构建语义空间(两种语义空间),其次生成标题骨架,因为标题骨架是短语的解析树,因此要构建标题骨架语义空间(便于标题骨架的生成),这样做的目的是重用有效标题的句法结构。在小说标题中,可以观察到重用有效标题结构的例子如《为了忘却的纪念》和《为了一夜的爱》。除此之外,修辞算法和遗传算法也要用到语义空间算法,修辞算法通过搜寻语义空间,产生多种表达的结果,之后遗传算法通过语义空间,筛选和评估留下有趣表达效果的标题。

要说明的是,为填充标题骨架,需构建两种语义空间:“亚文化语义空间”和“大众语义空间”。“亚文化语义空间”包含网民偏爱的词汇符号,“大众语义空间”包含所有可填充拟标题骨架的词汇符号。目标词在“亚文化语义空间”搜索相关词(相关词和目标词之间需有比喻关系),搜索中要确保相关词的抽象性。这个抽象系数算法已由 Glucksberg^①在2001年提出,经过确定抽象系数,在大众语义空间中筛选出相关词,并最终填充标题骨架。因此,对于输出标题骨架来说,其最终形态既要有“亚文化语义空间”的趣味性,也要符合主流文化价值观。

无论以小说诗歌、媒体新闻为代表的內容创作,以资料整理、数据分析为代表的数字人文研究,还是指向文化建设和政治建设的国家管理,社会对于人工智能创作的需求已经越来越多。尽管语义空间算法在不断更新完善,数据库类型和内容也在不断多样化,但是对于人工智能创作而言,提出并实践语义空间算法只是第一步,人工智能创作的核心并不在语义空间算法,制约人工智能创作的要害是对于修辞算法的理解与开发。对于社会整体需求来说,目前修辞算法的发展还是杯水车薪,随着人们对于内容创作的要求提高,修辞算法的研究与开发已经迫在眉睫。

三、深层算法与叙事语法:修辞句式的算法构建

小说标题,是表现小说内涵的令人难忘的语段,通常用于小说的指称和媒介传播活动。在小说标题中,比喻、比拟、夸张等修辞手法经常出现,而人工智能创作的修辞算法研究,不仅包含句法的具体表达形式,还要深入语段的深层意义中,如何利用算法完成优秀的修辞表达,是人工智能创作的重要研究范畴。2014年,Miller.D.W 和 Toman.M 就已经开始了相关研究,他们对2390个小说标题做数据分析,发现其中92%的标题至少包含一个修辞手法,并针对这些修辞手法进行排列对比,发现其中最常见修辞手法是比喻,最常出现的修辞组合是比喻和押韵^②。Reece 是标题领域的研究专家,他提出,当读者回忆小说的时候主要依赖于小说标题,而不是小说的主题内涵^③。而且在标题中,不同的修辞手段对于读者会有不同的影响,含有修辞手法的标题更有说服,更加优

①Glucksberg.S. *Understanding Figurative Language: From Metaphor to Idioms* Oxford: Oxford University Press, 2001.

②Miller D.W and Toman.M, "An Analysis of Rhetorical Figures and other Linguistic Devices in Corporation Brand Slogans." *Journal of Marketing Communications*, 22.5(2014): 474-493.

③Reece, B.B, "What Makes a Slogan Memorable and Who Remembers It." *Journal of Current Issues & Research in Advertising*, 16.2(1994): 41-57.

秀^①。因此,构建新颖有趣的小说标题表达,对于小说来说是一种数据资源优化的方式,谁能够最早在修辞算法的研究中拔得头筹,谁就能占据未来人工智能小说市场的先机。

权重逻辑是修辞算法中的核心,也是修辞表达效果的重要参数。拿比喻修辞算法来看,使用比喻句式算法,首先,构建与目标单词(本体单词)相关的语义空间。其次,通过筛选本体和喻体的形容词^②,搜寻喻体单词,形成比喻修辞的表达式。具体说来,即通过权重逻辑,提取出本体相关的修饰词,再通过修饰词的权重逻辑,提取喻体单词,最终将喻体单词纳入表达式中。Granroth 等人曾使用 Meta4meaning 语义空间^③来构建语义模型,并以此找到本体喻体在语义上的相关单词,例如,在比喻句式“时间就是金钱”中,“时间”(本体单词)的相关词是“珍贵”,而“珍贵”又与“金钱”(喻体单词)关联度很高,于是“金钱”就被筛选出来放置到表达式中。

对于人文学者特别是创意写作的研究者来说,人工智能创作中的修辞算法是一项富于挑战性的任务。创意写作中的叙事语法是旨在建立一套共同叙事模式、规则和符号系统^④,所以叙事语法研究是内容生产层面的研究。从内容生产层面来看,语义空间算法是一种句法层的算法,它做的是人工智能创作算法的形式开发,而修辞算法通过提高话语表现力,以此增加对受众吸引力,进一步深入人工智能创作算法的语义层。语义层研究是为了构筑话语所指的意义集域,这个意义集域正是创意写作叙事语法研究的归结点和落脚点。人工智能创作里面的算法,和一切别的人工智能算法一样,仍继续保持在它的生产里,但却在一种特定的叙事语法中产生它的内容。所以,人工智能创作算法研究和创意写作叙事话语研究,二者之间具有共性,极有研究价值。

人工智能创作的理论研究会有很多路径,要想把握其算法层面的创作意义,我们可以在人工智能创作与创意写作的联系里去寻求。笔者尝试提出一种身心关系,即以创作算法为实践形式,以叙事语法为理念内容,反映创意写作其内在学科逻辑,即文学创作与技术实践的密切关系,这种关系是写作理念社会化的自为发展,这在人工智能创作算法研究中,可以视为核心部分。

四、评估机制与筛选机制:遗传算法构建

遗传算法是 1969 年由美国霍兰德教授根据生物进化论提出,经后人总结整理而成的一种模拟进化过程并求解极值的自适应人工智能算法^⑤,它能够运用于一切具有自然选择特性的人工智能实践问题。由前文可知,小说标题创作在经过语义空间算法和修辞算法后,需要在语义空间中继续进行结果筛选和效果表达评估,这是一个多结果选择过程,那些语义准确、结构完整、修辞表达效果良好的生成标题将被保留,不符合标准的、表达效果不佳的或者因算法问题而产生突变的生成标题将被淘汰,直到最终产生最优的结果,因此遗传算法也适用于人工智能的小说标题创作。

在进行遗传算法前,小说标题经历了以下算法步骤:第一,要导入目标词,构建词汇语义空间;

①Burgers,C,Konijn,E.A,Steen,G.J and Iepma,M.A.R."Making Ads Less Complex,Yet More Creative and Persuasive: The Effects of Conventional Metaphors and Irony in Print Advertising." *International Journal of Advertising*,2015,16.2 (1994):.41-57。

②Richards,I. A, *The Philosophy of Rhetoric*, Oxford University Press, 1936.

③Xiao,P,Alnajjar,K,Granroth-Wilding,M,Agres,K and Toivonen, *Meta4meaning:Automatic metaphor in-terpretation using corpus-derived word associations*, ICC, 2016.

④葛红兵:《小说类型学的基本理论问题》,上海:上海大学出版社 2012 年版。

⑤葛继科、邱玉辉、吴春明、蒲国林:《遗传算法研究综述》,《计算机应用研究》2008 年第 10 期。

第二,构建标题骨架语义空间,筛选输出标题骨架;第三,通过权重逻辑和语义空间搜索,导出本体词汇的语义词(形容词和喻体词汇);最后,通过遗传算法进行筛选和评估,将最佳词汇填充标题骨架,并最终选出最佳表达效果的小说标题。要注意的是,在遗传算法中,以面为单位的搜索,比以点为单位的搜索,更能发现全局最优解^①。因为在算法模型的运行中,整个群体执行了多个方向的筛选,它们要在代数、权值系数和抽象系数等参数上做出不同评估,以此支持这些方向上的信息构成和交换。

遗传算法的筛选要符合国家的内容审核机制。小说标题作为高效传达信息的语段,在各类数字媒介平台上被传播,其规则 and 标准既要合乎读者的接受心理,也要符合写作的规则,因此标题创作的算法规约要在筛选和评估环节重点考虑市场与人文性。一方面筛选环节要删除任何不可被市场接受或无效的表达式,其标准是:生成标题中的目标词与修饰词要和积极情绪之间呈现正相关性,并删除含有暴力、宗教、色情、政治等敏感词汇。另一方面评估环节则被用于最大限度优化标题的叙事深度,其标准是:与小说主题呈现正相关性,语言规范要正确,关键词要呈现,修辞方法要正确使用。

包括遗传算法在内的算法开发和研究要遵循这样的前提:算法原理要公开透明,且筛选和评估机制要重点考察,尤其是具有多代选择的遗传算法或者具有自动化决定能力的相关算法。尤瓦尔·赫拉利在《未来简史》中把算法上升到具有扩张性和普遍性的哲学地位,甚至预言人工智能的发展会将地球带回到少数精英创造历史的时代。2017年5月起,欧盟要求所有人工智能算法解释其输出原理,这意味着在欧盟,闭环的深度学习算法成为非法的算法。这不是危言耸听,算法之所以被推崇到这种地位,是因其具有不可忽视的地方。康德在《纯粹理性批判》中认为,人类能认识到现象,而永远认识不到本体。如果说通过人工智能算法使得人类进入了数据海洋时代,那么未来人工智能算法的深度学习以及机器的自我学习会使得人类进入数据宇宙时代,那这样的世界人类是否还可以认识具有普遍性的现象?随着人工智能算法深度学习,如人工智能的遗传算法,模拟遗传代数不是几十代,而是几万代甚至是几十万代来筛选结果时,我们人类还如何确保最终得到的结果不带有“突变基因”结果?当有了人类意识不到的改变后,人工智能自我默认并规定新的存在物,这种存在也许将会超越人类的认知。这种认知的不足是很现实的,假如未来的人工智能制造问题或者产生新知识的速度加快,人类修补问题的速度和学习新变化的速度能跟上吗?这对于人工智能来说,它意味着人工智能捕捉到并使用了这些人类所认识不到的新东西后,那还能说人工智能没有意识吗?人工智能对于世界的理解与人类会越来越不同,人工智能最终也会产生哲学意义上的意识,相较于弱人工智能的自动化决定,这才是真正意义上的黑箱。假如这些情况出现了,那将是人类未曾踏足之地,也可能是人类也无力追寻之地。

五、结语

在本文中,笔者描述了一个具有修辞表达效果的小说标题的自动生成算法模型。该模型首先采用语义空间算法来构建标题的语料库和语义搜索模型,它在内容层面上属于句法层的基础算法研究。其次,通过修辞算法来使得标题的表达式更有趣味性,修辞算法以提高话语表现力增加对受众

^①秦可凡:《人工智能中遗传算法的研究与应用》,《中国战略新兴产业》2017年第8期。

吸引力,它在内容层上属于语义层的深层算法研究。通过句法层和语义层的意义集域指向,笔者进一步分析了人工智能算法和创意写作叙事语法之间的重要关系。最后,使用多目标选择的遗传算法,使标题在语义空间中继续进行结果筛选和效果表达评估,以此得到最优的结果标题。因此,在逻辑层次上语义空间算法是内容构建的基础算法,修辞语法是内容生成深层算法,而遗传算法是内容修正的筛选和评估算法。另外,在对多目标多代选择处理的遗传算法描述中,我们认识到不能忽视人工智能算法深度学习可能带来的危机。也许,随着算法的不断研究和发展,算法会从错误中自我学习,形成“黑箱”。黑格尔说:“意识本质上乃是客观事物的内容的产生者,于是主观的思维本质上是主动的。”在模拟人类思维和决策,不断深度研发算法的人工智能时代,同样可以做到客观事物内容生产的机器算不算意识呢?算是主观思维吗?和人类意识有区别吗?至少在人类中心主义的社会中,当算法产生了无意识的行为,则算法的理性概念会消解,由此带来的不确定性我们并不一定能够承受。因此,我们既要关注人工智能为自动创作带来的便利,也不能忽视对人工智能算法的筛选和评估。

Research on Algorithm of Fiction Title Generation in Artificial Intelligence Writing

Liu Sai(Shanghai University)

Abstract: The algorithm is the basis of the research in the field of artificial intelligence writing. At present, the research on fiction creation algorithm is still very immature. However, artificial intelligence creation based on existing results for large and small titles is very possible and can be used for future fiction creation. Algorithm research provides experience. In the algorithm research of fiction title generation, semantic space algorithm, rhetorical algorithm and evaluation algorithm have the generalization of content production. The research of three algorithms for title generation can open up the deep logical relationship between the narrative grammar layer and the semantic algorithm layer in artificial intelligence writing, which has great research value.

Key words: Artificial Intelligence; Algorithm Research; Title Generation.