

·专辑:人工智能与图书馆·

编者按:在资本注入、数据集聚、行业布局、技术突破、应用普惠、国家战略……的多重力量驱动下,人工智能(AI)迎来了一个爆发发展期。

在被称为“AI应用元年”的2017年,从AlphaGo惊人的学习能力到“女性公民”Sophia,从阿里“鲁班”设计师到“说人话”的未来智能助理,不可否认,人工智能给予了我们许多好奇、兴奋、惊喜与期望;不可否认,AI同我们每一个人的关系日趋亲密,并不断挑战着我们的感知与认知极限;不可否认,在未来社会,AI不仅将是我们生活中的基础资源,还将成为社会发展的驱动力。

在图情领域,从百度和武大图书馆联合重新定义“借书”,到机器人“图小灵”蹒跚于上图一楼大厅,从“浙思享”在杭城到“思维导图”展现于禅城,怀抱AI的步伐亦坚实又矫健!站在AI时代来临之际,我们认为,AI带给图情界的既是挑战,如图书情报的业务重构与事业转型等,也是机遇,如基于AI的多场景应用、提升的用户个性化深度服务体验等;我们认为,AI时代的到来,也将对图情事业、教育与研究均产生影响,智能情报、智能检索等将真正实现,图情学术研究的方向、工具、范式等也将随着AI时代的到来和发展而发生改变。

为此,我们组织了本期“人工智能与图书情报”专辑,邀请到了吴建中、王世伟、黄晓斌、叶鹰、李广建、黄水清等业内专家学者,分别围绕人工智能时代的图书馆发展对策、图书馆服务重塑、图书馆发展变革、智能信息处理和智能信息分析、智能情报分析系统设计、机器学习图情应用等主题展开学术探讨,研究内容有宏观框架也有微观思辨、有实践应用也有理论分析,是我国目前唯一的一组系统结合了AI与图情研究的文章。

我们相信,本专辑文章必将对AI时代的图情事业发展、业务重构、用户服务、AI技术应用产生积极的学术支撑作用!

人工智能与图书馆

吴建中

(1.上海图书馆 上海 200031)

摘要:文章从人工智能发展的技术环境、人工智能与信息组织以及人工智能在图书馆的应用三个方面,探讨了人工智能对我国图书馆事业发展的影响。文章认为:图书馆在人工智能的应用方面要转变观念,人工智能既是一门工具,也是人类的朋友;图书馆要以更积极的姿态拥抱人工智能,并善于与之一起,共同为激活图书馆的交流功能和服务效能做出贡献。

关键词:人工智能;图书馆;物联网;智慧图书馆;智能分析

中图分类号:TP18;G250.7 文献标识码:A DOI:10.11968/tsyqb.1003-6938.2017107

Artificial Intelligence and Library

Abstract The author discussed the influence of AI on library from the technological environment of AI development, AI and information organization, and the application of AI in library. The author believes that the library should change idea about the application of AI. AI is a tool as well as a friend of human beings. The library should embrace AI with more positive attitude and use AI to activate the communicating function and service function of the library.

Key words artificial intelligence; library; Internet of Things; smart library; intelligent analysis

半个世纪前人工智能似乎还只是一种奢望,这两年已成为街谈巷议的热门话题。中央高度关注人工智能的发展。2015年7月发布的《国务院关于积极推进互联网+行动的指导意见》和2016年3月发布的《国家十三五规划纲要》都提出要大力发展人工智能,2017年更是将其写入《政府工作报告》,并于当年7月颁布《国务院新一代人工智能发展规划》,由此,人工智能在我国进入实质性发展阶段。

人工智能不仅已经走来,而且将深刻地影响着经济、文化以及社会生活等各个领域。据有关预测,2045年约有一半的劳动岗位将由人工智能代替,而在中国这一数字将达到77%^[1]。与此同时,人工智能也将促进新业态、新模式的涌现,并催生新的工作机会。本文将从人工智能发展的技术环境、人工智能与信息组织以及人工智能在图书馆的应用三个方面,探讨人工智能对我国图书馆事业发展的影响,并提出发展和应对建议。

1 人工智能发展的技术环境

人工智能之所以成为近期发展热点,主要源于两方面因素:一是大数据及其处理速度的提升和信息流技术的发展。数十年前人类已进入信息爆炸时代,信息大量涌现且无处不在,但“信息孤岛”现象相当严重,互联网及智能技术的深入发展为信息根据人的需要有序流动和智能利用成为可能;二是脑科学和深度学习的发展。计算机通过大数据进行深度学习和机器学习,逐渐形成数据驱动的人工智能。如今人工智能正在从传统的“用计算机模拟人的智能”,如人机博弈、机器识别、自然语言处理等,向“机器与人融合智能”,如自主无人系统、机器自主创作、智慧城市、智慧医疗、智慧金融、智能制造等的方向迈进^[2]。

人工智能的实现必须有三个前提。首先是互联。从传统互联网到移动互联网,人类走进了一个万物互联的时代。不仅信息被有效地激活起来,而且通过与其他信息的连接和组合,不断产生出新的信息。更重要的是,互联网通过神奇的连接方式将分布于不同空间的信息在同一时间里聚集起来。简单地区分有用信息和无用信息的时代已经过去了,任何包括

结构化的或非结构化的信息都可能具备某种价值,有些信息现在有用,有些信息将来有用,有些曾经被认为无用的信息与其他信息连接起来会变成有用甚至关键的信息。而这一切的关键在于互联;第二是规模。互联是条件,数据是基础,人工智能的实现有赖于庞大的数据量。如对于机器学习而言,数据量越大越能提升模型的精确性;第三是速度,数据传输速度与数据规模是成正比的,机器要模拟人脑,除了提升算法能力以外,还必须要有庞大的数据规模和快速的传输速度支撑。也就是说人工智能要达到更高目标,有赖于更快的传输通道和速度,而5G时代的到来对于人工智能来说真可谓如虎添翼,有人甚至把5G看做是人工智能的钥匙^[3]。

人工智能之所以沉寂了近半个世纪才迅猛发展起来,是因为数据规模有限、数据流通不畅以及信息孤岛等问题严重地制约了它的发展。随着现代科学技术的飞速发展,尤其是大数据、云计算、物联网以及移动互联网的广泛应用,人工智能正步入发展的快车道,其影响必将渗透到社会各个领域、各个层面。总部设于美国的高德纳咨询公司(Gartner Group)每年都要发布年度科技报告,该公司2017年发布的《十大技术趋势》,把人工智能和高级机器学习放在十大技术趋势中的首位^[4]。同样由美国新媒体联盟(New Media Consortium, NMC)发布的《2017年地平线报告:图书馆版》,也将人工智能和物联网看作是未来四至五年内将对图书馆产生深刻影响的技术^[5]。

2 人工智能与信息组织

人工智能离我们并不遥远,这里举两个与人们日常生活有密切关联的例子。如2016年8月,日本广播协会(NHK)发布了一则新闻,披露了日本首次通过人工智能救治一名白血病患者的案例。东京医科研究所用Watson从2000万篇论文和1500万件专利中获得了有关白血病治疗的信息,仅用了十分钟就提出了一个有效治疗的建议^[6];2017年1月,《自然》杂志报道,美国斯坦福大学研究人员利用“卷积神经网络”系统(Convolutional Neural Network)对皮肤癌病例进行诊断,研究人员训练系统分析了近13万张痣、皮疹和其他皮肤病变的图像,然后让系统

与斯坦福大学医学院的21位皮肤科医生对比测试,结果发现系统的精确度与人类医生相当^[7]。这意味着未来我们携带有拍摄功能的手机就可与智能系统互联,达到疾病诊断的目的。这两个事例都表明,人工智能已经悄然来到我们身边。

实际上,具有二十多年历史的搜索引擎系统早就与人工智能结缘,现在一些知名的搜索引擎都在某种程度上应用人工智能以提升和增强自身的信息服务功能,如谷歌于2015年推出的RankBrain现已应用于日常搜索。这些年来,从搜索关键词发展到搜索自然语言,从搜索语言发展到与语音、图像和视频结合,搜索引擎已成为规模最庞大的人工智能实验和应用平台。有学者认为,网络搜索应用人工智能有很好的发展前景,它能借助数据分析获得用户的选择偏好,从而让用户网络搜索的过程更高效,搜索结果更符合用户的真正需求^[8]。

搜索引擎公司之间也在加强合作。2017年6月,美国艾伦人工智能研究所(Allen Institute for Artificial Intelligence)发布了一条新闻,称该公司与微软、谷歌、百度等合作,建立开放科学检索(Open Academic Search, OAS)工作小组,加大智能检索合作的力度。该公司由微软的共同创建者保罗·艾伦建立于2014年,致力于人工智能与计算机科学研究与开发。以前各检索系统都利用各自的人工智能系统进行开发,如今全球最大的几个搜索引擎能携起手来,不仅有利于统一系统和格式,而且能给广大网民带来便利。据该网站信息,我国清华大学也参与了具体开发,其承担的任务是权威档的建设,即对同名同姓的作品等进行识别研究^[9]。

人工智能的发展经历了能存会算(运算智能)、能听会说与能看会认(感知智能)和能理解会思考(认知智能)的三个阶段。机器早已实现了能存会算的运算智能,而且在能听会说与能看会认的感知智能上,机器也已证明在大多数情况下超越人类。有报道称,机器识图的能力已经超越人类,错误率小于5%^[10]。就目前而言,机器在确定性、完整性、封闭性的知识领域,比人类做得更好(如阿尔法狗)。但是,真正意义上的人工智能是能理解会思考,从机器可读机器可懂是基础阶段,而从机器可懂到机器能像人一样

思考,即达到认知智能,机器还有很长的路要走。

这一点从图书馆目录揭示方式演变就可以看得出来,图书馆目录揭示方式经历了三个阶段:书本及卡片目录时代是人可读、机器不可读,机读目录时代是机器可读、一般人不会读,而到了XML目录时代是机器和人都可读,基于XML的语义网技术的发展为迈向机器可懂创造了条件。为了让机器像人一样有智慧,本体语言、知识图谱等应运而生,它们把人类知识表示成计算机能直接使用形式,为机器实现认知智能奠定了重要基础。

如果说20年为一代的话,人类至今已经历了25万代了,从猿到人,进化的过程是相当缓慢、潜移默化的。但人工智能从老鼠的智力,发展到猴子的智力,再达到类似人的智力只不过用了几十年,机器达到强人工智能是近在眼前的事,用不了百年就能够实现超人工智能^[11]。

3 人工智能在图书馆的应用

数据是人工智能与图书馆的基础,也是最大公约数。人工智能一出现就引起了图书馆界高度重视。1990年代,图书馆工作者就梦想借助人工智能提升参考能力,提出参考服务要借助人工智能从书目查询向有针对性解决问题的方向发展^[12]。《国际图联趋势报告》(IFLA Trend Report)从2013年推出以后,每两三年更新一次。在首版报告中,就已经提到人工智能,认为“人工智能的进展使互联网设备能够与语音识别、机器翻译和语音合成相结合,从而支持实时多语言语音翻译”^[13]。2016年报告进一步指出,目前,人工智能既拥有加强已有的图书馆功能的能力,也有代替这些功能的能力。图书馆需要展开创新思维,开发科学动画无国界(Scientific Animations Without Borders)行动等创新型工具^[14]。在《国际图联趋势报告》的文献综述中,人工智能被列为四大技术趋势之一,并提出人工智能对图书馆未来的影响主要有三个方面:(1)下一代超越关键词检索的浏览器和对网页内容的语义分析;(2)综合语音识别、机器翻译、语音合成以支持实时多语言翻译;(3)对多元、复杂网页内容的云服务众包翻译和识别^[15]。此外,美国图书馆协会2016年度十大新闻也将人工智能发展列为

第五位^[16],由此可见人工智能已经进入图书馆发展的议事日程之中。在人工智能研究与应用方面,图书馆是可以大有作为的。

(1)要着力自然语言检索的研究,并探索图情方法在自然语言中的移植和实现途径。兰开斯特(Frederick Wilfrid Lancaster)早就预测过,自然语言将成为情报检索的规范,普通受控叙词表的使用将会衰退,这似乎是肯定无疑的了^[17]。在《21世纪图书馆展望(对话录)》(吴建中著)中,张琪玉教授和侯汉清教授都认为自然语言化是当代检索语言发展最重要的特征和趋势,这一预测在人工智能热的今天得到了证实。20世纪70、80年代情报语言学的重心是规范语言,如今将向自然语言转移,探索情报语言学理论在自然语言处理及其在信息组织和检索中的应用。情报语言学可以为人工智能的深入发展做出应有贡献。

(2)要着力深入内容的智能分析能力。今天,一些人工智能技术和工具如卷积神经网络和协同过滤(Collaborative Filtering)等在智能分析应用上已经比较成熟,这些智能化手段为图书馆情报咨询和参考服务的精准化提供了更多可能性。此外,智能化信息检索工具如语义计量(Semantometrics)和语义学者(Semantic Scholar)等在语义和内容关联方面也做了一些有益尝试。语义计量的特点是通过深入内容的语义连接,将不同文献之间的相似内容进行比较和关联,以此引向新的科学发现^[18],语义学者是在谷歌学术(Google Scholar)之类的学术搜索系统基础上的创新应用,它将深入论文内容,并展示论文中的数据图表及参考文献^[19]。图书馆不仅要积极地将上述已有的成

熟工具运用到图书馆管理与服务之中,还要主动地开发类似系统和工具,以提升自身的服务能力和质量。

(3)要着力探索人工智能在图书馆的应用。人工智能已经广泛应用于智能要求较高的知识活动之中,如日本经济新闻社于2017年1月宣布该社已开始利用人工智能自动发布“财经汇要”(Financial Summaries)^[20]。还有报道人工智能应用于论文评议的案例,如由美国Aries Systems Corporation公司开发的国际上使用最广泛的网络投稿与审稿系统之一的Editorial Manager系统,已运用人工智能工具“元书目计量机器人”(Meta Bibliometric Intelligence)进行论文初审^[21]。目前已经有不少图书馆应用人工智能的成功案例。如英国阿伯里斯特威斯大学的两位学生开发了一款目录查询和位置引导机器人,为读者提供该大学图书馆80万馆藏图书的查询和导引服务^[22];美国康乃狄克州Westport图书馆于2014年9月引进机器人为读者提供信息服务^[23];日本山梨县南都留郡山中湖村创造情报馆也于2015年9月引进机器人作为正式员工^[24]。虽然程度各有不同,但这些应用在提升图书情报业务与服务的效能上走出了可喜的一步,值得推广。

未来图书馆是知识中心、学习中心和交流中心。图书馆引进人工智能技术不是为了替代图书馆员,而是为了丰富和增强知识交流和人际交往。因此,图书馆在人工智能的应用方面也要转变观念,人工智能既是一门工具,也是人类的朋友。图书馆要以更积极的姿态拥抱人工智能,并善于与之一起,共同为激活图书馆的交流功能和服务效能做出贡献。

参考文献:

- [1] 卢泽华.人工智能:全球竞赛中国领跑(网上中国)[N].人民日报(海外版),2017-03-11(10).
- [2] 邓中翰.人工智能已进入2.0时代[EB/OL].[2017-06-06].<http://news.sina.com.cn/c/2017-06-06/doc-ifyfuvpm7569175.shtml>.
- [3] 百家号.未来产业的霸主是人工智能?5G就呵呵了[EB/OL].[2017-03-27].<http://mini.eastday.com/a/170327173440650.html>.
- [4] Gartner.Top 10 Strategic Technology Trends for 2017[EB/OL].[2016-10-14].<https://www.gartner.com/doc/3471559?ref=Site-Search&stkw=top%2010%20trends&fml=search&srcId=1-3478922254>.
- [5] Michele Cummins.Submit your projects for the NMC Horizon Report 2017 Library Edition[EB/OL].[2017-03-23].<http://www.nmc.org/archive/submit-your-projects-for-the-nmc-horizon-report-2017-library-edition/>.
- [6] Munenori Taniguchi.IBMのWatson、わずか10分で難症例患者の正しい病名を見抜く、医師に治療法を指南[EB/OL].

- [2016-08-07].<http://japanese.engadget.com/2016/08/07/ibm-watson-10/>.
- [7] Jessica Hall. Deep learning algorithm diagnoses skin cancer as well as seasoned dermatologists[EB/OL]. [2017-01-25]. <https://www.extremetech.com/extreme/243352-deep-learning-algorithm-diagnoses-skin-cancer-seasoned-dermatologists>.
- [8] 亮晨. 人工智能助力搜索引擎, 让用户搜索过程更精准[EB/OL]. [2017-06-22]. <http://www.citnews.com.cn/news/201706/47416.html>.
- [9] Alan Boyle. Microsoft, Google and Baidu team up with Paul Allen's AI2 on Open Academic Search[EB/OL]. [2017-06-20]. <https://www.geekwire.com/2017/microsoft-google-baidu-team-paul-allens-ai2-open-academic-search/>.
- [10] 沈淑莎. 人工智能何时才能读懂人心[N]. 文汇报, 2017-08-20(3).
- [11] 严炎, 刘星, 编译. 霍金预测: 未来百年内人工智能将接替人类[EB/OL]. [2015-05-15]. http://tech.ifeng.com/a/20150515/41082632_0.shtml.
- [12] Riddick J F. Reference librarians and serial publications in the age of artificial intelligence[J]. The Reference Librarian, 1990, 12(27-28): 281-287.
- [13] 王晶晶, 周伽妹译, 朱强校. 乘风破浪还是随波逐流?——如何在信息狂潮中把握方向(中文版)[EB/OL]. [2013-08-22]. https://trends.ifla.org/files/trends/assets/ifla-trend-report_simplified-chinese.pdf.
- [14] IFLA. IFLA Trend Report 2016 Update[EB/OL]. [2016-08-25]. <https://trends.ifla.org/files/trends/assets/trend-report-2016-update.pdf>.
- [15] IFLA. Technological Trends[EB/OL]. [2017-08-18]. <https://trends.ifla.org/literature-review/technological-trends>.
- [16] ALA. Ten Stories That Shaped 2016[EB/OL]. [2016-12-15]. http://lisnews.org/ten_stories_that_shaped_2016.
- [17] (美) 兰开斯特, 杨劲夫, 等译. 情报检索词汇控制[M]. 北京: 科学技术文献出版社, 1982.
- [18] Petr Knuth and Drahomira Hermannova. Towards Semantometrics: A New Semantic Similarity Based Measure for Assessing a Research Publication's Contribution. D-Lib, [EB/OL]. [2016-11-12]. <http://www.dlib.org/dlib/november14/11contents.html>.
- [19] Nicola Jones. Artificial-intelligence institute launches free science search engine[EB/OL]. [2015-11-02]. <http://www.nature.com/news/artificial-intelligence-institute-launches-free-science-search-engine-1.18703>.
- [20] 日本经济新闻社. AIを活用した「決算サマリー」配信スタート[EB/OL]. [2017-01-25]. <http://www.nikkei.co.jp/nikkei-info/news/press/394.html>.
- [21] Aries Marketing. Artificial Intelligence Integration Allows Publishers a First Look at Meta Bibliometric Intelligence[EB/OL]. [2016-10-17]. <http://www.ariessys.com/views-press/press-releases/artificial-intelligence-integration-allows-publishers-first-look-meta-bibliometric-intelligence/>.
- [22] Pasi William Sachiti and Ariel Ladegaard. Meet Hugh, the robot librarian[EB/OL]. [2016-02-26]. <https://www.aber.ac.uk/en/news/archive/2016/02/title-181095-en.html>.
- [23] WestportLibrary. Robotics at the Library[EB/OL]. [2017-03-22]. <http://westportlibrary.org/about/news/robotics-library>.
- [24] ロボット図書館職員 Pepperから見てくる未来の図書館[EB/OL]. [2017-02-02]. <http://current.ndl.go.jp/e1856>.

作者简介: 吴建中, 男, 上海图书馆研究馆员。