

机器学习界所抵制的,是旧的知识分享方式

冷眼观

张盖伦

国际知名期刊《自然》顺应人工智能发展势头,准备发布新子刊《自然·机器学习》,收割一波论文。结果刚刚宣布正式接受投稿不久,就被一纸声明弄得很难堪。机器学习社群表示,这期刊和自己三观不合,他们决定不投稿、不评审、不编辑。总而言之,就是要公开抵制。

俄勒冈州立大学网站上的公开信道出了原因:几乎所有重要的机器学习发表平台都不对发表和获取文章收费,《自然》的新期刊沿袭旧制度,是种倒退。在旧制度里,跨国期刊出版商是这么赚钱的:他们从学者处获得知识产品的版权,再将

其打包高价出售给科学共同体;他们通过销售科学知识产品攫取高额利润,用商业化运作作为学术传播交流筑起高墙。在传统期刊发过论文的学者都经历过这个悖论:既要付费发表,又要付费获取。用句文绉绉的话说,学界苦出版商久矣。

有研究者认为,数字学术资源的市场垄断和产品价格的持续攀高之间形成了恶性循环,甚至导致了学术交流危机。最先被危机笼罩的,自然是那些经费较少的机构。简单粗暴一点来讲,越穷的科研机构,买的数据库越不全;在这类科研机构工作的研究人员,想要获取科技信息就越费劲。那些草根科研人员,就被拦在了学术信息的商业屏障之后,这叫学术交流中的“数字鸿沟”。

但是,资本要最大限度的追逐利润,出版商自然也就将知识视为商品。学术界轰轰烈烈开展的

“开放获取”运动,正是为了挫挫大型出版集团的锐气,也是为了拆掉学术交流的高墙。它倡导知识的公有性理念,认为论文发表可以依托网络媒介,人们可以以分享的方式传播自己的作品,科学家论文的版权不应该无偿或廉价转让给出版商。

当然,还有不少高影响因子的期刊依然走传统路线,毕竟在这些期刊发表论文,于现行的学术评价机制下是一种“刚需”,它们的地位暂时无可撼动。

《自然》的新子刊也准备“走在老路上”,只是机器学习界跳出来,对其当头敲了一棍。因为,机器学习这个领域还比较年轻,本身就拥抱开放共享,受传统期刊“荼毒”甚少,也没什么历史包袱。

要知道,早在2001年,《机器学习期刊》的编委会就全体辞职,转而成立了一份零成本开放获取期刊《机器学习研究》。十七年前他们就认为,

时代不同了,互联网这么便捷,为什么还要人为给知识传播制造障碍?所以,十七年后,机器学习界更没有理由再欢迎一份封闭获取期刊。

而且,机器学习界的选择很多。他们有各类国际会议,也有足够数量的开放获取期刊。少了一个自然子刊,他们仍有平台和空间去拓展自己的学术影响,去进行无障碍的学术交流。要让这群在Github社区上进行开源代码分享的人忍受传统期刊的运作风格,想来也确实困难。

学界和出版界的关系错综复杂,可以说是“相爱相杀”。这次,机器学习界的科学共同体在用实际行动夺回话语权:在我的地盘,你就得听我的。《自然》会妥协吗?“开放获取”运动会再下一城吗?新的知识共享机制应该如何建立呢?想到后续进展,还真有点小期待呢。

好机友

AI当教练 手把手教你科学运动

本报记者 华凌



近日,北京开了一家“人工智能健身房”。有了AI加持的健身房有什么不一样?AI真能取代人类教练,帮助人们健身吗?科技日报记者实地探访了一番。

初来这家健身房的健身者需佩戴一个蓝色智能手环,用于收集个人运动数据,还可让设备随之自动调整状态,再站到一台采集身体数据的电脑屏幕前,无感扫描读取全身指标信息,建立个人信息档案。

“这里的健身设备专门从德国引进,在整个运动过程中,有一套精密的人工智能数据分析系统‘运筹帷幄’,实时判定用户动作速率、重量等数据,有针对性地更新训练方案。训练结束后,系统会将每次运动数据记录并形成分析报告。”智能健身房VENTO合伙人、2017年国际男子健美大赛冠军朱睿说。

人工智能应用于运动训练将如何发挥作用?中关村双创园创新创业企业北京元太体育首席技术官李浩告诉记者:“人工智能技术可从运动前、运动中和运动后三方面融入运动健身,针对不同群体提供个性化服务。在这方面国外起步较早,而我国现已开始研发,通过机器学习算法构建运动训练模型,预测运动员训练的体能消耗情况,针对性制订合适运动方案。”



在算法技术上,AI如何实现为用户精准制订个性化方案?李浩解释:“关键在于数据积累,需对大量的健身人群的运动过程进行记录。此外,每个人的特征数据也要尽可能全面,包括形态数据、机能数据、素质数据、运动数据等,训练结果作为标签数据也要记录,由此可训练出好的机器学习算法模型,针对不同类型的用户,输入用户的特征数据,系统就可制订相应方案。”他举例说,比如这家健身房,在第一次运动健身之前,对每个人的身体状况做一遍检测,系统基于这些数据,结合已有模型,制订出智能健身方案;在运动过程中,通过佩戴设备实时检测运动员的心率、呼吸频率等运动数据,对运动状态进行记录和分析,指导健身者按照预定方案进行锻炼,以达到目标效果。



在这里,除了首次需要教练讲解,之后完全可自行操作整个流程,无需私教指导。“未来人工智能健身房一定会让健身教练转型,从目前的销售角色转向服务角色,促进健身行业的健康发展。”朱睿说。

那么,私教被人工智能替代的是哪些方面,哪些又是不能替代的?“在制订健身方案方面,随着数据量的积累和机器学习算法模型的优化,人工智能完全可以替代私教。但是,私教在指导学员锻炼的过程中,可以保护学员,避免出现运动损伤,这一点却是机器无法替代的。”李浩说。

相比传统健身房,人工智能健身房未来有哪些明显优势?李浩分析说:“AI健身房场地空间要求不大,快餐式运动流程,可针对不同用户提供个性化方案。特别适合大城市快节奏人群,可以智能、就近、快速完成一次健身。未来健身房的AI元素也许会越来越多。”

(本版图片除标注外来源于网络)

取代人?别逗了!

和人类PK写作, AI“道行”还不够

本报记者 杨仑

写作是语言的精髓、文化的灵魂,与1和0组成的数据、单词和符号组成的代码看上去格格不入,灵感、直觉和纯熟的语言训练诞生的产物,怎能由数据和代码组成呢?在AI席卷各大行业的浪潮下,写作作为艺术形式的一种,仿佛成为了人们心目中的“孤岛”。

然而,这座“孤岛”也被AI浪潮“破门踏户”了。

近日,某电商平台推出了名为“莎士比亚”的AI写作系统,一秒钟可以“写”出千条文案。业内人士戏言,终于轮到编辑“下岗”了。

AI+写作真能发挥出这么大的魔力吗?

别慌!“莎士比亚”是款“文案处理器”

据该电商平台发布的新闻稿称,这套系统在借鉴传统NLG和语言模型方法的基础上,基于该平台自身在商品标签和搜索数据库层面积累的大数据,从句子层面做结构解析、训练模型和语言生成,从而能够一秒钟“吐”出千条文案,并根据用户不同的需求自主选择各类文风格。据介绍,该系统还能根据用户对文本的选择实现机器算法的优化。

然而该系统一面世就遭到了网友的吐槽:“既然能智能写文案,能不能智能刷好评呢?”“人工智能+文案,是千人千面还是千人一面?”

对此,该系统负责人表示:“传统人工创作费时费力,尤其是电商平台‘大促’期间,时间紧、任务重,很多商家修改页面时往往捉襟见肘。这套系统就是为了极大地赋能商家,提高效率、降低人力成本,逐

渐实现机器写文案、系统排版页面。”

显然,这套系统的主要应用场景是商品文案写作,离写诗、做文章这些人们想象中AI写作还有很远的距离。记者找到了一篇由人工智能写作的文章,虽然看起来逻辑通顺,但其主要内容均来自于新闻媒体的一篇专访稿件,机器所做的工作,仅仅是对稿件进行了删减和截取。

“事实上,AI技术应用于写作仍有相当大的局限性。”拓璞产业研究院分析师林贞好告诉科技日报记者,“AI写作需要输入大量某特定类型的文章数据,进行机器学习训练,付出包含时间和金钱的训练成本,才能让AI学会某类文章的写作模式。”从某种程度上看,该系统更像是丰富的文案数据库,可以根据不同的风格,向用户推荐更适合该商品的文本。

有局限:只能在特定领域“展神威”

AI对艺术领域的“破门踏户”可不是第一次了。谷歌曾利用AI创作一幅油画作品,最后竟然卖到了8500美元。在一

项测试中,AI创作的画作成功“欺骗”了超过半数观众的眼睛,使他们误以为该作品是由人完成的。

杨强则表示,人工智能驱动的未来企业是一个全新的题目。“新的技术变化和旧的商业理论需要磨合,实现机器的自主学习自主决策,达到‘最大化效益、最小化费用’的多目标多任务。在这方面,机器学习要做到‘短、频、快’。”短”指从数据到模型再到数据的回路一定要短,最好其中不要有太多人为参与和决策。“频”指数据和反馈的获得要频繁。“快”指模型更新一定要快。“短、频、快”是未来企业成功的必要条件,但不是充分条件,还有很多其他的条件,但是这三条是必不可缺的。”

在AI写作的道路上也不乏探索者。2016年,一款AI机器人曾参与了当年高考作文的创作,但成绩只能用惨不忍睹来形容。两篇文章行文风格磕磕绊绊,句子之间逻辑也欠通顺,用词单调重复,仅有几个句子被认为有点“文艺范”,可惜在上下文结合的过程中,也失去了韵味。事实上,日本也曾经使用AI参与过升学考试,积累了多次失败经验后,已经能够达到大学入学的水准,然而其擅长的也是数学、历史等学科,但在语言处理能力方面非常一般。

但在消息新闻写作、报告类数据采集等方面,AI写作则获得了不小的成

功。2007年,美国一家公司就推出了新闻编写软件,用于撰写财经类和体育类新闻稿件。只要导入最新的数据,每分钟最快可生成2000篇报道。近年来,我国媒体也陆续开始使用人工智能技术进行新闻写作。

林贞好表示,要让AI学会写作,必须要锁定某个特定应用目标,进行大量资料搜集和模型训练;训练范围拉得越大,则训练的成效与精确度往往会越差。“因此,目前AI技术只能在风格类似、词汇量使用范围较小的领域完成文本生成,比如天气预报、财经新闻等内容,AI可以很好地输出相关报道和消息。但对于那些风格具有创新要求、情感描述丰富的文章,AI就无能为力了。”

待进步:现阶段只是“不懂装懂”

“现阶段,我们使用较多的还是NLG(自然语言生成)技术,以该平台人工智能系统为例,它套用了大量数据积累下来的模板,而且最终还需要人工参与对其进行校正。”清华大学苏研院大数据中心主任林辉告诉科技日报记者。

更重要的是,人工智能现阶段只是“不懂装懂”。这和我们的语音处理方式一样,机器通过大数据分析,学会了某些词汇、字句可以连在一起使用,人工智能写作出来的内容,只是让人看上去认为机器已经了解了其中的含义,但事实上机器并不知道。”林辉说。举例来说,就像2013年上映的人工智能科幻电影《Her》一样,当女主角说出“我爱你”的时候,她

只是发出了音节,却完全理解不了背后的情感。

林辉介绍,受技术的限制,当下的人工智能+写作整体还比较初级,一方面需要人力的参与;另一方面,机器在自然语言处理、理解方面都还存在着障碍。从目前来看,AI写作的主要目的也并非替代人力,而是作为人的助手,协助人们处理一些繁琐、重复的劳动。有新闻学系教授也撰文表示,被人工智能取代的将是明确规则的智力劳动,但在新闻领域,深度报道、新闻评论等需要创造性思维的部分,将是人类进一步深耕的领域,这是机器所无法取代的。

商业决策这件事,未来或被AI承包?

产业界

实习记者 崔爽

在日前召开的品友互动人工智能大会上,麻省理工大学人工智能实验室主任丹妮拉·鲁斯教授表示,人工智能正在改变商业模式,也改变了“工作”本身。香港科技大学教授杨强则表示:“现代商业以人为中心,未来商业以机器为中心。人和机器各有各的位置,人来设计目标、提供数据、清洗和标注数据,但人工智能的模型才是中心。AI经营的企业就像无人驾驶汽车一样。”

“原来大家都是靠经验和想法来做决策,这是传统商业决策的第一特点。”品友互动联合创始人谢鹏表示。决策和实施过程需要消耗大量人工劳动,每次营销的成果也无法积累和沉淀。“往往是广告投完,受众遗忘。”谢鹏说。

而用数据和技术驱动的商业实践不同,既可

以通过数据模型预测营销结果,又可以在过程中不断反馈不断优化。谢鹏举了打靶的例子:“以前只能盲打,不知道射出去怎么样。现在一边打靶一边可以知道打了几环,随时调整策略。”

丹妮拉·鲁斯表示:“任何有数据的企业、商业活动都能从AI当中受益。比如企业可以去分析历史的数据,在历史数据的基础上解释过去发生了什么、预测将来发生的事情,决定如何做商业决策。”

杨强则表示,人工智能驱动的未来企业是一个全新的题目。“新的技术变化和旧的商业理论需要磨合,实现机器的自主学习自主决策,达到‘最大化效益、最小化费用’的多目标多任务。在这方面,机器学习要做到‘短、频、快’。”短”指从数据到模型再到数据的回路一定要短,最好其中不要有太多人为参与和决策。“频”指数据和反馈的获得要频繁。“快”指模型更新一定要快。“短、频、快”是未来企业成功的必要条件,但不是充分条件,还有很多其他的条件,但是这三条是必不可缺的。”

杨强表示。

此外,杨强还强调:“AI可以进行机制设计。”传统的博弈论是给定一个游戏,最大化它的收益。机制设计反过来,即确定最大化的收益,再考虑如何设计这个游戏。在商业决策中,利用AI进行机制设计可以被应用到信息推荐、广告设计等多个领域。

但要实现人工智能对商业的赋能,障碍还很多。“商业和企业的金钱、利益直接相关,一个决策可能直接影响企业的生死。”丹妮拉·鲁斯表示。真实的商业环境不像AlphaGo学习围棋棋谱那样算法明确、边界清晰,除了数据表现,环境、社会甚至天气等很多因素都有可能施加影响,AI的加持只是针对计算量和算法,如何结合客观现实因素作出准确判断才是商业决策的重点。

另一个制约AI决策的现实因素是,机器学习远未达到充分发展的程度。丹妮拉·鲁斯现场展示了一段视频:一个抱着书的孩子,故意向外撞击紧闭的门,门不开,他走开。一旁围观全程

的18个月大的小孩看完后,走到门边,向内拉动把手成功开门。“解决问题需要深度推理,现在机器学习还没有能力做到这一点。这其实还是一个谜题,我们自己也在努力理解大脑的工作机制到底是怎么样的。”丹妮拉·鲁斯表示。

在她看来,机器学习面临一些挑战,第一个挑战是数据标注。机器学习需要上千万的数据点,这些数据需要得到标注。“对于文字、语言、图片,我们已经有了工具和能力去进行标注,然后来创造大的数据团,但在其他领域如机器人方面,我们还没有能力得到数据团。需求和成本问题都值得考量。”丹妮拉·鲁斯说。第二个挑战是,虽然通过机器学习处理了大量数据,但处理数据不等于创造知识,复杂计算也不能带来自主性。

品友互动CEO黄晓南则强调,数据安全和隐私的问题。商业数据需要高保密性,在AI提供服务的过程中,采集的所有数据都是非实名数据,没有任何个人信息,这对技术提出更高要求。

扫一扫
欢迎关注
AI瞭望站
微信公众号

