

# 直面隐忧 中国人工智能企业机会何在



本报记者 华凌

今年人工智能领域的发展迎来新一波高潮,犹如枝繁叶茂的大树渗透到各行业的蓝天之中,跃跃欲试服务于众多领域。有人欢呼,人工智能商业应用元年已经到来。

2018年,人工智能领域的另一趋势是大额融资频发。

清华大学近日发布的《中国AI发展报告2018》显示,自2013年以来,全球和中国人工智能行业投融资规模都呈上涨趋势。2017年全球人工智能投融资总规模达395亿美元,融资事件1208笔,其中中国的投融资总额达到277.1亿美元,融资事件369笔。中国AI企业融资总额占全球融资总额的70%,融资笔数达31%。

## 隐忧一:发展结构“头重脚轻”

重点突破基础领域,建立自己的生态体系

早在2015年,谷歌开放其内部使用的机器学习软件TensorFlow源代码,脸书、亚马逊和微软也纷纷发布其工程师用于机器学习的开源软件。似乎AI进入了“免费原材料”时代,人人都可以顺手取材。但是,“国外的开源布局对于我国AI行业发展而言,埋藏着巨大隐患。”远望智库人工智能事业部部长、图灵机器人首席战略官谭芝洲指出。

谭芝洲告诉记者:“开源模式会引导技术方向、路线图,形成开源生态,创造商业模式,这些由发起开源项目的核心利益者掌控,不仅控制行业上层的应用,还控制底层的生态,构建了整个帝国,掌控极大的权利。因此,开源是开放的资源,但现在免费并不代表未来不会收费和控制。如安卓系统是一种开源手机操作系统及应用开发平台,而谷歌实际上主导着整个生态的发展。”

谭芝洲认为,若我国企业今后过度依赖目前的AI开源平台,采用大量现成的源代码,仿佛在起跑线上丧失优势,创新及工艺再精深,也是在人家的体系中做零部件的更新改造。“如同温水煮青蛙,今后可能会给行业带来很大影响。这将是最大的隐忧。”他说。

在业界看来,投融资的热情不减,主要是看中人工智能与各行业结合的广阔前景。

然而,有业内人士近日指出,目前国内跟人工智能有关的公司大概有四千多家,但是能够得到投资人青睐或关注,并且愿意投资的,大概不到三分之一。如果没有后续资金投入,很多初创企业有可能难以生存下去。由于人工智能产生收益的时间存在不确定性,巨大繁荣的背后存在隐忧。

那么,什么才是人工智能企业的核心竞争力?对于初创企业来说,如何才能站稳脚跟而不被市场淘汰?直面隐忧,中国人工智能企业的机会何在?

赛迪研究院公布的《2018中国人工智能产业展望》提出,由于我国人工智能产业应用层技术、轻基础理论,底层技术积累薄弱,存在“头重脚轻”的结构不均衡问题,使我国人工智能产业犹如建立在沙滩上的城堡,根基不稳。基层技术积累薄弱使人工智能核心环节受制于人,阻碍重大科技创新,不利于国内企业参与国际竞争。

那么,建立我国自己的AI生态体系,还有机会吗?当然,“谭芝洲斩钉截铁地答道,“在时间上还来得及,因为国外也才刚刚发展。从国家层面洞悉AI发展态势,重点突破基础领域,针对人工智能底层技术,加强对以深度学习为代表的底层算法模型的深入研究,并积极布局影响人工智能未来发展的前沿基础理论研究。现在国内也有一些小团队在做相关开发项目,有一定潜质,而且我们拥有全世界最多的应用开发者,非常多的应用场景、大体量的市场、蓬勃的创新创业环境等,这些都是国外比不了的。”

据了解,科技部指导下的新一代人工智能产业技术创新战略联盟,已联合深圳鹏城实验室于7月在深圳启动了中国的“启智开源开放平台(Open)”的建设。

## 隐忧二:商业应用路径不明确

瞄准市场需求,实现落地是关键

据亿欧智库《2018中国智能商业落地研究报告》统计,2017年中国人工智能创业公司获得累计融资超过500亿元,但商业落地百强创业公司累计收入不足100亿元,90%以上人工智能企业亏损。不少业内人士担心,国内人工智能领域存在巨大泡沫,或将迎来一波倒闸。

《2018中国人工智能产业展望》提出,我国人工智能产业处于早期发展阶段,商业化应用路径尚不明确,商业落地痛点突出,致使近期实际商业价值变现难度较大。

谭芝洲指出,“对初创企业而言,人工智能有门槛,创业成本较高。因此,建议企业不要盲目,要尽快找准发力方向,而AI项目商业应用场景的落地是其成败与否的关键,快速积累核心技术优势,打造商业模式,才能做出真正有市场需求的产品,产生现金流。这也有助于人工智能行业回归理性。”

“未来产品形态应能把智能交互和后面的服务及产品联系在一起。”新一代人工智能产业技术创新战略联盟副秘书长、科大讯飞副总

裁兼AI研究院联席院长李世鹏分析,亚马逊成功通过智能音箱将人工智能引入美国家庭的方式值得借鉴,我们需要有亚马逊这样既卖服务又卖产品和内容的企业。

据《2017年中美人工智能创投现状与趋势研究报告》显示,中国智能机器人与无人相关技术创业最为火爆;其次为语义分析、语音识别、聊天机器人等自然语言系列技术;然后是人脸识别、视频/监控、自动驾驶、图像识别等计算机视觉系列技术;另外,情感计算包含心理学、语义、视觉、环境感知等多种复杂应用的技术也在慢慢成长。

李世鹏表示,人工智能包括算法、数据和处理能力。从投资角度首要看数据,BAT、微软、苹果、脸书在很多领域已占先机,想去撼动它们经过十几年积累的数据并不容易。所以,对于初创公司,没有多少资源去做范围太大、体量太大的事情,其成败的关键在于是否有渠道获得海量独特的数据,并通过这些数据为用户提供新的价值,比如大幅提高传统行业的生产力。

## 隐忧三:专业人才成稀缺资源

加快AI及相关学科布局,培养跨学科人才

“目前,人工智能最大痛点之一是人才难得,AI被炒得很热,稍微懂点算法的人一出来就能收到很多Offer,身价水涨船高。”李世鹏表示。

《2017年中美人工智能创投现状与趋势研究报告》指出,目前中国人工智能的人才培养已成为一个关键问题,人才缺失可能会对AI产业发展产生牵制作用。美国国家科技委员会发布的2017年人工智能全球大学排名中,前50名均位于欧美地区,我国大学无一上榜。此外,国内缺乏人工智能与传统行业的跨界人才,不利于AI在各垂直行业应用推广。

据业内对中美AI人才分析显示,截至2017年6月,中国共有592家人工智能公司,拥有员工约39200名。相比之下,美国人才数量是我国两倍。据领英数据显示,我国从业经验10年以上的AI人才比例不足40%,而美国这一比例超

过70%;美国人工智能基础层、技术层和应用层的人才数量占比分别为22.7%、37.4%和39.9%,而中国为3.3%、34.9%和61.8%。

李世鹏建议,我国需加快人工智能及相关学科布局,高校加强学科建设,依托现有人工智能相关学科,培养跨学科人才,并鼓励高校、科研院所加大与人工智能企业、国外高校及相关机构的合作力度,打造多种形式人才培养平台;针对人工智能芯片、基础算法模型等重点领域,充分利用现有各类人才计划,并设立专门通道和定向优惠政策,加大对国际顶级科学家和高层次人才吸引力,加快人才引进效率,扩大人才引进规模;重视培养贯通人工智能基础理论、软硬件技术、市场产品及垂直领域应用的纵向跨界人才,以及兼顾人工智能与经济、社会和法律等横向跨界人才。

## 产业界

### 血液科实验室 来了位AI实习生

本报记者 雍黎 通讯员 曾理

近日,在陆军军医大学新桥医院血液科实验室,主管技师彭贤贵有了一名新的实习生。特别之处在于,“它”是个机器人——AI智能影像诊断系统。它将通过在临床实习,成为一名诊断高手。

从外形上看,它是个灰白色、方方正正的箱子,与一台曲面显示器相连,医生将做好的骨髓涂片放入它的大脑中,不到一分钟,显示屏上就标出了识别出的各种细胞名称,并对发生病变的细胞进行重点标注,为医生出具病理检验报告提供了依据。彭贤贵介绍,他从事临床工作30年,每天都在显微镜下看一个个这样的骨髓涂片,根据肉眼观察到的不同细胞形态为患者出具检验报告,现在,这台设备可以帮助医生来识别患者骨髓涂片中的各个细胞。

新桥医院血液科主任张曦教授介绍,骨髓各系细胞形态学检查是对人体的骨髓细胞进行分型检查,用于诊断和辅助诊断某些造血系统或非造血系统疾病。其检查结果会提供给临床医生提供诊断和治疗的最终依据,所以往往这项检查又被称为血液系统疾病患者最终的“判决书”。目前,这位“实习生”还只能根据骨髓细胞中各个细胞的大小、颗粒及细胞核的复杂程度等测量值来初步判断、认定细胞特性,但由于骨髓细胞中存在粒细胞系、红细胞系和巨核细胞系三大系统,各种细胞间有些大小等数据区别较小,所以其最初的识别准确率尚不及经验丰富的临床医生。但是,它拥有强大的“大脑”,有过目不忘和海量储存信息的能力,所以,在最初的学习阶段,可以由它先扫描识别最初的骨髓细胞、辨认信息,再由经验丰富的临床医生逐一辨别其对细胞识别的准确率,然后对正确的识别进行确认存储,对有误的识别进行纠错校正。

“经过我们的培训,它的辨识能力正在逐渐提高,目前对骨髓细胞的识别准确率已接近90%。”彭贤贵说,将来,随着它看的片子越来越多,“临床经验”会越来越丰富,准确率还会逐步提高,将极大减轻医生的工作量,提高血液病理检验效率,缩短病理报告出具的时间。

据了解,这套AI智能影像诊断系统由新桥医院与杭州一高科技公司合作研发,通过临床医生的参与,将有助于更好地将医生的经验复制推广。

## 新鲜事

### 计算机博弈大赛 一场机器之间的棋艺较量

本报记者 吴长锋

当AlphaGo一路厮杀,当卡内基梅隆大学冷扑大师机器人击败人类德州扑克顶尖玩家,当人机大战已没有悬念……人机较量成为新时期的关注热点。计算机博弈究竟是怎么回事,透过它又能看到人工智能在哪些领域的应用?近日,2018年中国大学生计算机博弈大赛暨第十二届中国计算机博弈锦标赛在安徽大学举办,科技日报记者对此进行了探访。

“计算机博弈,就是人工智能棋艺比赛。许多人都有过跟电脑玩家下棋或打牌的经历,其实电脑玩家就是一种AI。”安徽大学计算机学院副院长李学俊介绍说,“就像孟德尔将果蝇用于基因研究一样,计算机博弈是人工智能领域的果蝇,是一个挑战无穷、生机勃勃的研究领域。”计算机博弈大赛以喜闻乐见的、高对抗性的棋牌游戏为研究载体,以理论与实际系统构建相结合为原则,锻炼了大学生的科研思维、创新意识、团队合作能力和沟通交流能力,调动了大学生的学习与研究热情,激发了大学生的创新潜能,培养了大学生的科学素养和专业实践能力。

AI下棋对战,规则和人类大同小异。以大赛中的幻影围棋为例,这种棋赛采用9×9棋盘,规则和围棋基本相似,但加入了信息不完全的限制,有点像现实中的盲棋。“今年大赛部分选手采用Google开源的深度神经网络,将之应用到斗地主、桥牌等非完全信息博弈,即机器人和人都不知道对方的牌,选手们取得了优异战绩。”李学俊介绍说,大赛还保存对弈双方的棋谱,建立国内计算机博弈的棋谱大数据。今后将建立开源棋谱大数据平台,为人工智能的科研人员和博弈爱好者提供开放数据接口。值得一提的是,今年的大赛当中,安徽大学研发了NAO机器人与人对弈的人机对弈原型系统,而不仅仅是计算机博弈,扩展了新的外延,将有利于计算机博弈应用到人工智能中小学教育领域。

“去年国务院发布了《新一代人工智能发展规划》,使得今年的计算机博弈大赛更具有时代意义,通过竞赛促进了我国计算机博弈技术的发展,加快了我国迈进智能化时代的新步伐。”李学俊说。

此次大赛共设置了五子棋、六子棋、点格棋、苏拉卡尔塔棋、亚马逊棋、幻影围棋、不围棋、爱恩斯坦棋、军棋、海克斯棋、中国象棋、围棋、九路围棋、国际跳棋(100)、国际跳棋(64)、二打一扑克牌、桥牌、德州扑克等18个比赛项目,其中德州扑克项目是第一次出现在大赛中。

(本版图片来源于视觉中国)

# 借助机器学习,对垃圾信息说“不”

## 好机友

实习记者 于紫月

七夕将至,你是否收到过电商店铺推送的打折促销短信?平时是否经常收到理财、贷款、购房等信息?如果你是苹果用户,那你是否有过半夜收到iMessage推送赌场、假货、股票以及色情信息的经历?

商家给安卓用户推送广告信息需要通过移

动、联通等运营商,运营商会拦截掉一些非法信息,而针对苹果用户,商家甚至不法分子则可利用iMessage渠道通过互联网直接向用户推送垃圾信息。目前苹果公司尚未采取有效的筛选、拦截等手段。

近日,苹果官方终于发声,表示正在探索更先进的机器学习模型识别,过滤垃圾信息。机器学习模型如何过滤垃圾信息?又面临哪些技术难点呢?科技日报记者采访了相关专家。

作为人工智能的核心技术,机器学习是计算

机模拟人类思考方式的一种学习行为,它让计算机变得“聪明伶俐”。“基于机器学习的垃圾信息过滤技术实际上是一个二元分类过程,机器学习模型需回答‘是’或‘不是’,以便将垃圾文本从大量信息中分离出来。”北京语言大学大数据与语言教育研究所所长荀东教授在接受科技日报采访时说,“首先应准备人工标注的数据,进而构建机器学习参数化模型,最后对其进行训练、测试,直至模型应用,解决实际问题。”

“目前市场上识别垃圾邮件、短信的机器学习模型绝大部分采用的是针对文本式特征的分析与提取。”荀东东解释,所谓“显式”特征,是指垃圾信息的关键词、表达形式、特殊符号、异体字、敏感词语表达方式等“外在”特征。将这些多元、离散的特征元素汇总,便可构成显式特征列表,进而构建模型对垃圾信息进行甄别。

“这种分类方式效率较高、成本较低,所依赖数据较少,但也存在适应性差,识别精度不高等不足之处。”荀东东指出,如果垃圾信息发送方掌握了用户拦截系统的显式特征列表,便可对垃圾信息的敏感词汇作出相应调整,变换表达形式,从而有效规避拦截系统。因此分类器需要实时动态更新显式特征列表,即便如此,该列表元素数量也十分有限,导致分类器过滤效果不佳,精确度不高。

基于显式特征的分类方式“先天不足”,苹果

公司有可能将目光投向隐式特征分类,即深度学习模型。荀东东表示,深度学习模型可对海量数据的信息进行深度挖掘,从信息的语义和内容上对垃圾信息进行甄别。也就是说,同样一条推销短信,深度学习模型基于庞大的数据库,可分析出它的多种表达方式,从而做出更加准确的判断,仅靠改变文字形式无法逃脱深度学习模型的“火眼金睛”。深度学习模型省去了前期建立大量显示特征列表的工作,只需标注垃圾、非垃圾信息即可,不仅效率提高,识别精度也得到很大提升。

基于深度学习的垃圾信息过滤技术目前也面临诸多技术难点。复旦大学计算机科学技术学院张军平教授表示,虽然有大数据支撑,深度学习模型在很多方面还是无法像人一样有效分析和理解信息中的内容。例如,个别汉字的顺序颠倒不一定能影响阅读,然而深度学习模型并没有人脑这样“聪明”。另外,组合爆炸问题也是难点之一。垃圾信息涉及领域广泛,形式变化多端,大数据深度学习模型可能对已有的垃圾信息有效分类,但面对新出现的垃圾信息可能就会“蒙圈”。

“这需要借助类似短时记忆网络或更新的一些技术,对深度学习模型进行定期更新。”张军平认为,还可以考虑自然语言处理中的一些句与句的关系,进一步完善深度学习算法,从而改善其预测、分类能力。



扫一扫  
欢迎关注  
AI瞭望站  
微信公众号

