



AI翻译离无障碍交流有多远

本报记者 姜靖

世界这么大,我想去看看!十一长假临近,梦想中的你背起行囊,自由走在异国的大街小巷。然而现实的画风很有可能是这样的:走出国门没几天,你就发现期待已久的旅行,因为看不懂听不懂,而身在囧途。此时的你捶胸顿足地感慨道:世界上最遥远的距离,不是天涯海

角,是你站在我面前,我却不知道你在说什么。9月21日,百度智能翻译机在日本发布。名古屋有哪些好玩的景点?怎么乘车最便捷?翻译机流畅精准的中日互译令人惊叹不已。AI翻译服务通过硬件、软件连接千千万万个应用场景,将会打破语言不通的尴尬局面。甚至有人认为,AI翻译是人工翻译的终结者。现实真的如此吗?

从 path to path no man in sight。”这不仅符合近代著名翻译家严复提出的“信、达、雅”翻译原则,而且其中的意境与精髓只有熟练汉语的人才能领会。

“文学艺术翻译要体现情感、色彩、语调、温度、韵致,字里行间的意味种种微妙之处。要传达出这些,译者水准、敏感、境界和心灵力量尤为重要。”中国社会科学院外国文学研究所编审高兴说。

缘何能称霸棋坛的人工智能AlphaGo,却在翻译领域不尽完美?何恩培告诉记者:“机器翻译一直被公认为人工智能领域最难的课题之一。而且语言和围棋不同,语言背后的多元文化和复杂社会属性,注定了语言规则不可能规律化。”“文学作品,以及有引申含义或逻辑关系复杂的句子,AI翻译很难搞定,因为翻译它们不仅是基于理解,而且要使某种意义上待翻译的语言和目标语言对齐。”华为诺亚方舟实验室资深研究员、自然语言处理专家吕正东说。

此外,AI翻译要想达到“信、达、雅”的高

度,还需克服口语化的两大难题。一方面要听得清,能准确判断出指令发出者的语音、停顿,并在极短时间内进行“语音断句”。另一方面要克服口语交谈中的语法问题,以及句子不连贯、没有明确句子边界等问题。

机器靠什么解决上述问题?在清华大学计算机系研究员刘洋看来,AI机器翻译的难点其实也是整个人工智能的难点——如何让机器真正像人一样有智能行为。他认为,机器翻译采用数据驱动的方法,其准确程度取决于给计算机提供哪些数据。目前提供最多的翻译数据来自于政府文档,比如联合国多种官方语言,基本每份文件都有多种语种版本,但基本没有体育、娱乐等领域的数据库。再加上各国的口语中都存在较多随意性口语或网络用语,生活化气息浓厚,多为非正式的语体材料,因此智能翻译需要庞大的基础词汇库支撑。“其实所有上述问题,理论上来说都可以靠更多的数据来弥补,但是我们数据是有限的。”吕正东说。

现状:产品虽多,准确率有待提高

2017年可谓是AI大行其道的一年,作为现代科技的流行趋势,不管是谷歌、苹果,还是百度、阿里,都在相继发布自己的人工智能产品,向智能行业进军。而AI翻译,几乎是所有互联网巨头都想涉足的领域。

机器翻译已出现了70年,但过去一年里涌现出的AI翻译软硬件产品,已超过最近30年的总和。

在国外,微软发布了翻译工具Microsoft Translator,可以实现9种语言实时语音转为文本翻译。国内的互联网巨头不甘示弱。百度技术委员会联席主席、自然语言处理部技术负责人吴华表示,百度翻译支持全球28种语言互译,756个翻译方向,每日响应过亿次的翻译请求。除百度外,科大讯飞也走在语音交互前列,推出了讯飞听见升级版产品,加入多语种翻译功能,五种语言同步翻译成汉语,还发布了智能翻译产品“晓译翻译机”。就连网络游戏开发大户腾讯公司,也悄然上线了一款实时语音翻译APP“翻译君”,这款以AI内核驱动的

翻译产品,运用了语音识别+NMT(神经网络机器翻译)等技术,可以实现“同声传译”功能,对着屏幕说话时,边说边翻,就像身边带了一位私人翻译官。此外,电商大咖京东也开始涉足AI翻译,推出了一个可以进行多国语言实时翻译的人工智能翻译机——准儿翻译机的众筹项目,目前已经有5008名支持者,轻松筹得320万元的项目基金。

尽管AI翻译进步神速,却不被传统的翻译行业所看好。传神语联网科技股份有限公司董事长何恩培表示:“无论是翻译机还是APP,中文翻译成英文或其他语种时准确率还可以,但外国人说英文或其他语言时,翻译成中文的准确率就低,在实际交流时几乎没法用。”

今年2月,谷歌全新NMT系统在韩国世宗大学与人类译员进行的翻译对战就是一例佐证。在规定的50分钟内,人类和机器同时翻译两段随机文本。结果,人类以巨大优势战胜了机器。赛后,评委表示,NMT系统翻译出的文本90%都有语法问题。

难点:数据有限,且语言规则不规律

全球范围内,服务于各大跨国会议、发布会等的30多种翻译机能帮人们解决一些场景中语言交流的问题,但目前的现状就如科大讯飞声明的那样:虽然机器翻译已经取得非常大的进步,但距离高水平翻译所研究的

“信、达、雅”还存在很大差距。以“千山鸟飞绝,万径人踪灭”这句人们耳熟能详的诗为例,国际翻译界最高奖项之一“北极光”杰出文学翻译奖得主许渊冲教授将其译为“from hill to hill no bird in flight,

展望:前路漫漫,发展需数据推动

近两年,神经网络机器翻译(Neural Machine Translation, NMT)技术异军突起使翻译应用进一步革新。

NMT模仿人脑的神经思考模式,产出媲美人工翻译的高质量译文,并将误差降低了55%—85%。目前,谷歌公司已将该技术应用到网页翻译与手机应用,译文质量明显提升。此外,腾讯、百度、阿里巴巴等国内互联网公司也将深度学习理念应用到机器翻译。同时,语言处理、语音交互等技术的进步又加速了翻译产品的场景升级,促使机器辅助翻译(CAT, Computer Aided Translation)过渡到“人工智能交互翻译(AI Interactive Translation)。”

“语言是形式,而不是实质。”现代语言学理论奠基者费尔迪南·德·索绪尔说。同样,对各种花哨的技术概念抽丝剥茧,其核心仍然是AI处理人类语言的三种方法:即基于规则的机器翻译、基于实例的机器翻译和统计的翻译方

法。目前谷歌、搜狗等公司基本采用的是最后一种方法。

伴随着翻译方法的完善,机器将完成未来大部分的简单翻译需求,而那些细腻、多元、充满人文特质的复杂沟通以及专业化翻译,人工智能究竟能不能实现?还需要哪些改进呢?

清华大学计算机系教授孙茂松说:“世界语言好几千种,几千种之间的互译,其中绝大多数语言之间都没有足够语料。所以这方面的翻译还需要去做,包括一些模型的探索。”

“AI翻译是一个交叉学科,取决于数学、语言学、计算机科学、神经认知科学等很多方向的进展,因此AI翻译应该和更多的‘知识’结合在一起,让数据推动机器翻译发展。”刘洋说。

此外,在吕正东看来,如何真正让计算机去理解语意中那些委婉的对应关系和不规范的表述,可能需要相当长的时间才能实现。

产业界

专家共话人工智能算法未来

金婉霞 本报记者 王春

作为全国双创周的重要活动之一,近日,“人工智能学术暨产业化高峰论坛”在中科院上海微系统所展开。“算法就是一切吗?”会上,业界大咖纷纷亮相观点。

小i机器人研究院院长陈成才在讲话中提出:“深度学习是可能使人工智能产生质变的技术吗?”除了深度学习之外,深度随机森林、深度神经网络、深度贝叶斯,或者其他算法有没有可能实现突破?陈成才说:“要实现人工智能,我们还有很长的路要走。”此前,谷歌首席科学家李飞飞也曾表示,最重要的事情就是继续进行基础研究,“若基础研究跟不上,AI发展很快会枯竭。”

对于我国来说,除算法以外,传感器、芯片都是亟须突破的关键点。“没有传感器的人工智能都是空的。”上海微系统所副所长袁晓兵介绍人工智能如何帮助森林防火:在林场中“安插”含水率探测传感器、手持暗火探测仪等传感器,由它们将数据传输至“移动基站”。若仅有分析后采取对应的消防应急措施,而算法不与传感器数据结合,何谈人工智能?“人工智能最重要的内容有两点,一是基础工业,二是前端传感器。”袁晓兵说。

“基于深度学习的创新刚刚开始,单点的突破就可能爆发出巨头企业。”亚马逊孵化器CEO赵晔直言,“中国拥有海量的应用市场。”

新鲜事

逼真人造肌肉问世 让机器人更灵活



据英国每日邮报报道,美国哥伦比亚大学的工程师研发出一种可用于人形机器人的合成肌肉。这种3D打印的合成软体肌肉抛弃了之前模型所用的外部压缩机或高压设备,能够推、拉、弯曲和扭曲,并能举起比自身重千倍的物体。

研究团队表示,目前没有任何材料能够真正作为一种软体肌肉。这是由于所有材料都无法满足高强度应力和应变的需求。为了解决这个问题,团队首席研究者Aslan Miriyev博士使用了微泡中遍布乙醇的硅橡胶材料。这种解决方案将其他材料的弹性特性和极端体积变化属性组合在一起,而且具有易制造、低成本和对环境无污染的好处。

在用3D打印打造出需要的形状后,研究人员使用一根细电线和8伏低电压对人造肌肉进行电刺激。当它表现出重要的膨胀和收缩特性之后,研究人员进行了一系列机器人应用研究。团队负责人Hod Lipson教授称:“软体材料机器人可用于需要与人类接触互动的领域,如制造和医疗行业。”研究团队计划用导电性材料取代嵌入式的电线,以加快肌肉的反应速度,延长使用寿命。Miriyev称,未来他们将在研究中探索人工智能能操控这种肌肉的方法,那将是让机器人完成自然动作的“最后一座里程碑”。

AI诊断新算法 可提前十年查出老年痴呆症



近日,意大利巴里大学的研究人员研究出一种新算法,可以在阿尔兹海默症出现前10年,发现由疾病引起的大脑微小的结构变化,以达到尽早检测出阿尔兹海默症的目的。

研究小组通过对67个核磁共振成像扫描来训练他们的人工智能,其中有38个来自老年痴呆症患者,29个来自健康人的控制。研究人员将扫描结果分成小区域,并让他们的人工智能分析神经元之间的连接。在训练完成后,他们通过对148个实验对象进行脑部扫描来测试这种算法。在这些测试之外,还有48次对患有此症的病人进行扫描,48次对患有轻度认知障碍的人进行扫描,最终,人工智能得以全面地检测阿尔兹海默症。

据报道,该研究团队的人工智能算法能够诊断出85%的老年痴呆症,检测出轻微认知损伤的正确率能达到84%,这使其成为早期诊断潜在老年痴呆症的有效工具。这项研究的意义在于,未来有可能为阿尔兹海默症患者争取10年宝贵的治疗和心理准备时间。不过,研究人员现在的测试数据仅限于在南加州大学阿尔兹海默症的神经成像项目数据库,算法暂时只能对该数据库进行扫描学习。

(本版图片除标注外来源于网络)

智能信号灯:与老大难路口“死磕”

好车友

本报记者 李伟

“堵车”这一世界难题,到底有解没解?尤其是一些以堵车闻名的十字路口,简直成了出行者每天绕不过去的梦魇。近日,北京市交管局宣布正式启用百度地图为其量身定制的“智慧信号灯研判平台”,用以实现对城市路口路况的实时监测。北京交管局高级工程师陈大夫教授在接受科技日报记者采访时透露,该平台初期主要覆盖二环内以及上地、望京主要商务出行区域400多个路口的路况监测,将原来的人工监测转变为远

程平台化监控。有了“人工智能”的加持,会为“首堵”带来哪些改变呢?高峰期出行过的人都知道,有信号灯的路口很容易发生拥堵甚至堵车的情况,如果不能被及时发现和处理,很可能升级为整个区域的交通瘫痪。而传统条件下,依靠人工的信号灯路口监测,相对被动和滞后。

“现在我们借助人工智能技术和大数据资源,就能实现对主城区及商务出行频繁区域交通拥堵的分钟级报警和规律性分析,包括为每一个路口定制一个‘病历表’,定位易发多异常拥堵路口。”陈大夫认为,从被动的处理报警,到主动发现问题并及时调整信号策略,避免拥堵升级,这是大数据平台和人工

智能技术在交通领域应用后带来的升级和改变。“过去,信号灯路口的交通监测主要依靠人工调度和经验判断。付费请一个临时人员帮助交警统计路口的车流量,我们觉得贵,别人还不愿意来。”陈大夫说,关键是人工统计会受到各种因素影响,“这样的数据可靠性不高。”

事实上,2016年9月9日,广州市交警便对外宣布,引入阿里云人工智能技术ET(工业大脑)搭建“互联网+信号灯”控制优化平台。这款号称“城市数据大脑”的平台可对路口车辆运行情况进行实时分析,并输出对红绿灯时间的调整建议。同年10月,阿里云人工智能ET接管杭州交通治理,为这座拥有2200多年历史的城市也安装一个人工智能中枢。杭州市萧山区部分路段的初步试验结果显示,通过城市大脑智能调节红绿灯,车辆通行速度最高提升了11%。

那么,如果不借助智能信号灯这类黑科技,在体量堪称“巨无霸”的首都,交警和民众将面对怎样的出行现状?百度地图官方数据显示,2016年交通运输公路里程较之2015年增长2.6%,而汽车保有量的增幅为12.8%左右,“一除就知道,2016年汽车保有量的增幅是公路里程增幅的5倍。”百度地图开放平台总经理李志堂在接受采访时表示,这表明在北京,交通需求和供给之间存在很大矛盾,“出行资源相对而言是有限的,我们要在这样的背景下讨论交通话题。”

科技日报记者在采访中了解到,早在2016年11月,北京市交管局便联合百度地图围绕“互联网大数据优化交通信号配时”进行了创新和探

索。比如,通过优化治理,把海淀区后厂村路的拥堵指数峰值从4.6降至3.96,让上地小营西路到上地三街路段的每辆车平均上下班时间分别减少了3.35分钟和4.89分钟,使朝阳北路和朝阳路的平均通行速度分别提升了3.25%和3.99%。

今年7月,百度地图智慧信号灯研判平台刚上线的时候,北京市交管局根据系统发出的2600多次异常报警定位了8处易发生拥堵的交叉路口,并为每个路口量身定制了“病历表”。经过系统化调整和治理后,这8个路口从每月每天平均死锁时间13分钟下降到9月的1.4分钟。这让从业多年的高级工程师陈大夫“眼界大开”,“这充分论证了海量路况数据在信号灯配时优化方向的潜力。”更关键的是,陈大夫笑着补充说,“还是免费的。”

“北京交警经常向我们提出需求,比如‘能不能不看实时数据,通过历史数据就能更精准地定位某个路口的问题?’”百度地图智慧交通总经理严孙荣告诉记者,除了引入路口数据,双方讨论更多的还有要不要引入大数据来统计信号灯路口每个方向的流量数据,“通过抽样分析可以得出,某个路口每天左转有多少车辆,右转有多少车辆,左转时段内的平均延误时间是多少,每个车辆通过时的平均停车次数大致是多少,结合整体历史数据回溯,就可以发现路口的问题。”

而这些正是双方希望未来能与信号机打通的重要数据,被称为“信号机感知的神经终端”,通过更大量、更广泛的路口数据,做信号灯配时评估、绿波带分析及优化评估报告等,甚至更进一步地反向优化控制信号灯。



近日,首批34处带有“行人过街按钮”的信号杆在天津完成安装。没有行人过街的时候,信号灯延时让机动车顺利通过;行人需要过街时,按下按钮,系统就会根据实时的道路车流量情况、行人过街情况“预约”绿灯。视觉中国