

大数据技术具有军民通用的显著特点,军方应学习借鉴政府、企业在建设、应用和技术方面的成功经验,积极对接国家相关资源建设计划,在资源共享利用、技术协同创新、配套手段建设和人才合作培养等方面加强军民深度合作,推动形成优势互补、协同创新、互惠共赢的军事大数据发展生态环境。

军民融合 互惠共赢

大数据构筑最强“军事大脑”

■本报记者 李惠钰 通讯员 刘雪涛



军事大数据成为世界各国科技竞争的战略制高点。

随着大数据时代到来,数据已经成为继物质、能源之后,各国竞相争夺的第三大战略资源。特别是在军事领域,大数据的应用正在前所未有地激发创新活力、助推转型发展,成为主要国家构筑军事优势的“造血增智”工程。

7月5日~6日,由军事科学院主办,中国科学院、清华大学、中国指挥与控制学会等共同协办的第一届军事大数据论坛在京召开。论坛按照分组“主题报告+现场互动”的形式,由杨学军、梅宏、张斌、杨小牛、陈左宁、何友、陈志杰等两院院士就军事大数据相关战略、理论、技术、应用等作大会报告及主题交流。

与会专家一致认为,大数据技术具有军民通用的显著特点,军方应学习借鉴政府、企业在建设、应用和技术方面的成功经验,积极对接国家相关资源建设计划,在资源共享利用、技术协同创新、配套手段建设和人才合作培养等方面加强军民深度合作,推动形成优势互补、协同创新、互惠共赢的军事大数据发展生态环境。

从电商到国防

当前,大数据热潮几乎席卷了人类社会的各个领域,由其电商阿里巴巴就是典型代表。

“阿里巴巴万物互联、云计算构建的数据智能商业基础设施,相对以人为核心的传统商业基础设施,有巨大的优势。”阿里云计算有限公司副总裁宋杰在会上表示,在基础设施落后、个人素质素质差的前提下,阿里巴巴靠数据和计算高效调动微商、快递员、车辆等,仅用13年就超越了沃尔玛用54年实现的规模,成为世界第一零售平台。

在宋杰看来,从支付、售后、物流、商品到营销,数据智能几乎重塑了商业的各个环节。但重塑不止于此。对于城市治理体系,大数据也发挥着巨大作用。比如,实现对城市运营的监测预警、应急指挥、多网格化管理、智能决策、事件管理、城市仪表盘、协同联动等综合服务,为新时代智慧社会的可持续发展提供有力保障。

如今,军事大数据也成为世界各国科技竞争的战略制高点。中国工程院院士、海军航空大学信息融合研究所所长何友表示,当前,国防领域的竞争正围绕认知优势和决策优势而展开,大数据则是确立竞争优势的基础,为掌握未来战争的主动权,各军事强国纷纷布局国

防大数据研究,着力打造先进的数据汇聚和处理能力。

北京理工大学计算机学院院长黄河燕也表示,通过互联网搜集挖掘情报大数据信息,已成为各国军事情报数据分析的重要手段。

为增强学术交流的广度和深度,从今年4月份开始,军事科学院便面向社会公开征集军事大数据相关学术论文,最终收录论文121篇结集成册。从收录的论文中不难看出,其内容涵盖了大数据理论方法、标准规范、技术应用、建设实践、安全治理等多个方面,涉及作战训练、装备建设、后勤保障、军民融合和人才队伍等多个领域军事应用。

面临技术挑战

黄河燕表示,大数据时代最重要的三个理念就是,采集数据不是样本而是全体;获取数据不是精确而是混杂;分析数据不是因果关系

而是相关关系。相比于本就复杂的民用大数据,军事大数据则要复杂得多。

“除了军事数据,国防大数据还涉及天文、地理、社会、政治、经济、科技、文化、民族特性等方面。”何友表示,由于关乎国防安全,因此其需要更全面、更可靠、更有效的安全保障,保密性要求也远超民用大数据。

何友还表示,占国防数据总量不到20%的小数据蕴含着80%的价值,而体量超过80%的大数据也蕴含着极其宝贵的20%的价值。随着科学技术的飞速发展,小数据的价值已经发挥出60%~80%,但大数据价值的挖掘却刚刚起步。

“在未来的开发和应用中,小数据和大数据发挥的作用将互为补充,并且将长期共存,分别发挥不同的作用。”何友说。

由于数据量巨大,军事大数据也面临着非常多的技术挑战。分析情报是社会重大决策规划和实施中

的“耳目和尖兵”。黄河燕就以大数据情报分析为例指出,军事情报领域随着空间、空中、地面各种多维度、多角度、近程、远程信息获取手段逐步增多,各种监测手段不断丰富,各种装备与设备的日常检测与测试手段的完备,各种新闻媒体、互联网上产生的新闻信息都会产生越来越多的海量数据,数据来源复杂多样,聚合在一起将形成一个巨大的数据流。

“虽然当前获取数据的渠道极大地拓宽了,但同时也带来了数据数量庞大、冗余增大、分析不足等新问题。”黄河燕说,如何集成并融合多源多样化的数据,如何从海量数据中提炼出有价值的情报并预知变化,如何让机器像人一样掌握知识并深入理解情报数据,如何存储和管理以及有效分析和理解等,都是目前军事大数据应用面临的挑战。

期待以“智”取胜

军为民用,民为军需。在军民融合的大环境下,军需和民用成为拉动创新的两驾马车,军事大数据也期望从民用中寻找突破口。

凭借多年商业大数据的处理经验,宋杰认为,数据智能的迭代需要相应的数据智能生产装备,需要敏捷的开发前台、服务化数据中台、分布式全功能计算基础设施,组建数据部队,并选择自主可控、自主领先、有稳定商业模式、长期技术供应。还应采用“各业务域智能化+全域服务化”协同的发展路径。

另外,国防大数据中含有大量不公开非合作数据,这些数据真实性对决策过程和结果影响巨大,需要研究数据伪装、数据欺骗的机理与方法,识别不公开非合作数据中的数据伪装与数据欺骗,从伪装数据中探测获取真实数据。

黄河燕表示,大数据方法与“事实数据+工具方法+专家智慧”情报方法的有效聚合,是大数据情报分析的新需求。另外,对于信息挖掘,要根据应用需求和数据基础,构建应用模型。而对于情报信息挖掘的共性问题分析,也应减少人工干预,将情报信息挖掘方法与语义技术相结合,提升挖掘深度和准确度。

黄河燕认为,军事大数据还应重视认知计算的应用。“情报学的分析方法将会从原来的计算机辅助分析为主体转变为计算机认知为主体的智能分析,形成类似于IBM Watson的大数据情报认知计算及分析平台。”

前沿点击

近日,来自巴塞罗那自治大学的科学家们通过研究,利用基因疗法成功治愈了小鼠的肥胖和II型糖尿病,相关研究刊登于国际杂志EMBO Molecular Medicine上。研究者表示,利用了一种名为腺相关病毒载体(AAV)来携带成纤维细胞生长因子21(FGF21)基因的疗法(AAV-FGF21疗法)进行研究,单次使用这种载体就能操纵肝脏、脂肪组织和骨骼肌,使其持续产生FGF21蛋白。这种蛋白是由机体多个器官自然分泌的一种特殊激素,其能在很多组织中发挥作用,维护正常的能量代谢,通过基因疗法来诱导动物机体产生FGF21蛋白就能够帮助减肥,并且降低机体对胰岛素的耐受性,从而治疗肥胖和II型糖尿病。

目前研究人员已经在两种不同的肥胖小鼠模型中成功进行了试验,即饮食或遗传突变所诱导的肥胖。此外,研究者还观察到,当特殊的基因疗法用于健康小鼠,就能促进小鼠健康,并且抑制年龄相关的特种增长和胰岛素耐受性的产生。利用AAV-FGF21疗法治疗后,小鼠体重就会下降,其脂肪积累的水平 and 脂肪组织的炎症水平也会降低,同时肝脏中的脂肪含量、炎症及纤维化也会被逆转,而且机体胰岛素的敏感性会增加。

当研究者对三种不同的组织(肝脏、脂肪组织和骨骼肌)进行遗传操作使其产生FGF21蛋白后也能够再现上述结果,这或许就提高了这种新型疗法的灵活性,其不仅能够选择最合适的组织来进行作用,而且还能抑制疾病并发病的发生。当一种组织产生FGF21蛋白时,其会将这种蛋白分泌到血液中,最后分散到全身。

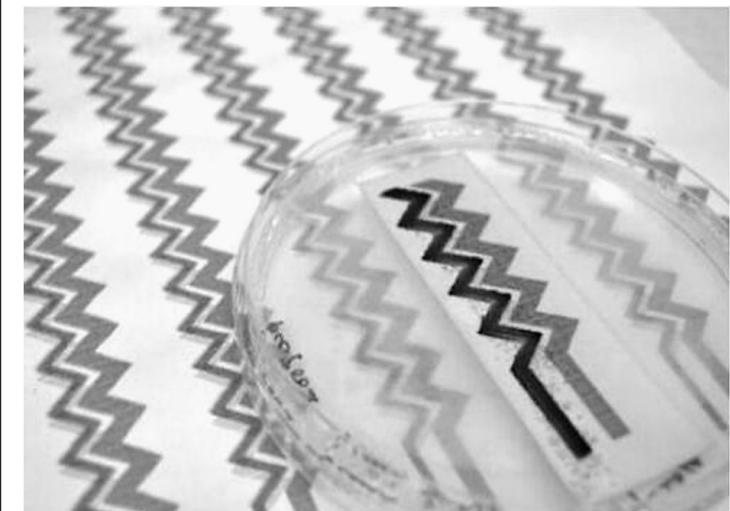
这项研究中,研究人员首次发现,利用AAV-FGF21基因疗法能够长期逆转机体的肥胖状态和对胰岛素的耐受性,未来有望应用于人体临床试验。相关研究结果也表明,AAV-FGF21疗法是一种安全有效的疗法。

同时,研究人员还指出,这种基因疗法能帮助机体有效抵御长时间高热量饮食所诱发的肝脏肿瘤产生的风险。当用于常规治疗时,FGF21蛋白的寿命较短,因为目前制药工业中开发出了FGF21的类似物,而且这些类似物已经进入临床试验。然而,科学家们需要定期给予FGF21类似物来调节临床效益,但同时也会产生一些与外源性蛋白相关的免疫性问题的风险;研究人员所开发的基因疗法载体就能够诱导小鼠产生与人类机体相同的FGF21激素,而且在使用FGF21后并不会给机体带来任何副作用。

下一步在进行患者临床试验之前,研究人员还需要进行大规模的动物研究来检测这种基因疗法的安全性,AAV所介导的基因疗法目前已经在欧洲和美国获批用来治疗多种类型的基因,而且研究人员在利用AAV介导的肝脏和骨骼肌基因转移上也具有广泛的临床经验,因此未来研究人员有望利用基于FGF21的基因疗法来治疗II型糖尿病、肥胖症及其相关的疾病。(陶朵朵整理)

基因疗法有望治愈人类II型糖尿病

酷技术



阴天也能发电的“细菌太阳能电池”

据加拿大不列颠哥伦比亚大学官网近日消息,该校研究人员开发了一种便宜且可持续的方法,利用细菌将光转化为能量来制造太阳能电池,这种新电池产生的电流密度比此前同类设备更强,且在昏暗光线下的工作效率与在明亮光线下一样。

为了解决上述问题,研究人员在不列颠哥伦比亚省这样阴雨天气比较多的地方广泛采用太阳能电池,这项创新迈出了重要一步。以前建造源于生物的电池时,采用的方法是提取细菌光合作用所用的天然色素,但这种方法成本高且过程复杂,需要用到有毒溶剂,且可能导致色素降解。

为解决上述问题,研究人员将色素留在细菌中。他们通过基因工程改造大肠杆菌,生成了大量番茄红素。番茄红素是一种赋予番茄红色的色素,对于吸收光线并转化为能量特别有效。研究人员为细菌涂上

了一种可以充当半导体的矿物质,然后将这种混合物涂在玻璃表面。他们采用涂膜玻璃作为电池阳极,生成的电流密度达0.689毫安/平方厘米,而该领域其他研究人员实现的电流密度仅为0.362毫安/平方厘米。

项目负责人、哥伦比亚大学化学和生物工程学教授维克拉姆帝亚·亚达夫表示:“我们记录了源自生物的太阳能电池的最高电流密度。我们正在开发的这些混合材料,使其可通过经济且可持续的方法制造,且最终效率能与传统太阳能电池相媲美。”

亚达夫相信,这一工艺会将色素的生产成本降低10%。他们的终极梦想是找到一种不会杀死细菌的方法,从而无限地制造色素。此外,这种源于生物的材料还可广泛应用于采矿、深海勘探以及其他低光环境等领域。(陶朵朵整理)

人工智能如何拯救方言

■祖满清

当前,随着经济、文化互动的全球化,主流或通用语言更加强势,弱势语言正濒临消亡。目前世界上大约有6000~10000多种语言,而据语言学家预测大部分将于本世纪末消失。

濒危语言保护(下简称为“语保”)已经成为一项重要而迫切的工作。在中国,普通话的强势地位已经造成一些少数民族语言、方言等弱势语言的使用人数明显减少,如不及时对弱势语言采取保护措施,我们将失去对人类文化遗产完整记录的机会。

当前方法手段不能满足语保进程

我国对语保工作早就有所重视,并有着深厚的方言研究基础。2005年我国启动国家语言资源保护工程(以下简称“语保工程”),我国学者对方言保护的主要研究方法是田野调查,研究内容包括中国语言资源有声数据库、方言词典、方言地图等。

国际语言学专家也对濒危语言进行了语言资源记录。2017年,美国科学家提出语音罗塞塔计划,旨在通过“未知”语言的语音和“已知”语言的文本的平行关系记录没有文字的“未知”语言(即濒危语言)。

归纳、确定被研究语言的基本音位是语言记录的基本工作之一,但目前这项工作很大程度上依赖于调查者对语音的主观感知和“口耳”工作。由于依赖人工,分析语料局限于孤立字、词,导致研究进度受限,很难将研究内容扩大到连续语音,从音位归纳上升到句法、语义层面的分析。并且,很多中国方言,特别是南方方言中,孤立音节的声韵调在连续话语中表现多变,在复杂的连续话语中,去除语境、韵律结构、情感等诸多因素的干扰,归纳完整的语音变化单靠人力是力所不及的。

同时,随着社会发展的日新月异,每隔数年语言会发生明显变化。因此,语言记录和分析需要高效的解决方案。

利用AI技术实现“语言复制”迫在眉睫

利用人工智能技术系统地研究濒危语言、方言的语音结构、语言结构,实现对一种语言的完整“复制”迫在眉睫。

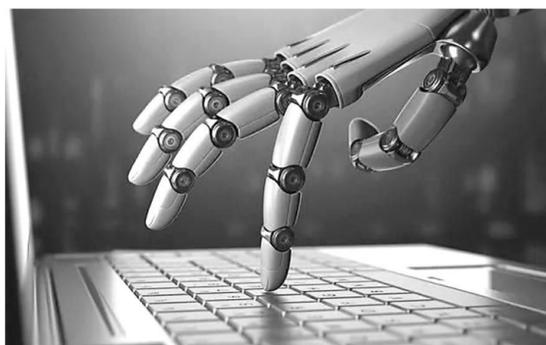
“语言复制”的概念是通过智能语音技术对一种语言实现完整记录。记录内容包括确定该语言的语音结构(例如音节语言的声母、韵母、声调等)、完整分析该语言的句法结构、连续语音的音变和连续变调分析、基本意义单位和主流语言的对应关系以及这个语言的任意文本或语音和主流语言之间的互译关系。

基于主流语言语音系统,完成语音复制需要建立被研究语言的语音合成系统(文语转换系统)、语音识别系统以及和主流语音之间的翻译系统。科大讯飞智能语音技术的发展和多年来的语言积累,可以助力语保工程。一些核心技术的突破和语言积累,使得不同语种之间互译成为可能。

科大讯飞人工智能(AI)研究院有着丰厚的智能语音研究基础,到目前为止实现了中文、英文以外的30多种语言(包含多种少数民族语言)的语音合成、语音识别、翻译,其中许多语音系统属拓荒性系统。研究院基于深度学习技术,采用全球文本、声学解决方案,在除中文普通话、英语等强势语言以外的许多语音合成系统上突破了语音合成MOS4.0的门槛,目前正尝试在部分濒危语言和方言上进行语言复制。

需要更多热爱母语的人参与

不同的研究目的会产生不同的语言分类。从人工智能的角度出发,我们将语言分为主流语言和非主流语言。中文普通话就是主流语言。非主流语言又分为三个类别。第一类是文字、口语都被广泛使用的语言,例如维吾尔语、藏语等。在这类语言的使用区域,虽然文字被广泛使用,但是往往缺乏正字规范。第二类是有文字但较少使用、口语仍被正常使用的语言,例如彝语、锡伯语等,语言群体内大多数成员仅在家乡口语交流时使用,多数群体成员不能熟练使用文字或基本



不识字。第三类为濒危语言及没有文字的语言,包括只有少数老人还在使用、群体内几乎所有其他的成员都已放弃使用的语言以及没有文字的语言。对这类语言进行完整记录比较困难,实现语言复制也有相当难度。

对于文字、口语都被广泛使用的语言实现语言复制是可行的;对于有文字但较少使用、口语仍被正常使用的语言,实现语言复制也是可能的。对于没有文字的语言可以收集被研究语言的语音,并在有条件的情况下转写为主流语言的文字,使用这样的平行数据,利用人工智能领域的另一端技术实现被研究语言语音到主流语言文本之间的转换,即美国科学家正在实施的“语音罗塞塔方案”,这在逻辑上是可行的。但被研究语言的采集、文本转写缺乏规范并存在许多具体困难。

在可能的情况下尽可能多地收集自然语音和文本的平行数据是十分有意义的。有了足够的数据,即使目前处理不了,今后仍有机会可利用。利用人工智能技术进行语言记录是一个研究方法的问题,在具体工作中仍然需要采用正确的技术路线进行操作,即使使用了人工智能技术,语言数据的处理仍然脱离不了人力支持。语言是全人类的共同财富,每种语言背后都有精彩的文化。语保工程不应该只是少数人的事业,应该有更多热爱自己母语的人群参与。

(作者系科大讯飞多语种高级研究员,本报记者赵广立编辑整理)