



视觉中国供图

自然语言处理与视觉处理,都重在对不同模态数据所包含的语义信息进行识别和理解,但是两种数据的语义表现形式和处理方法不同,导致存在所谓的“语义壁垒”,现在这种壁垒正在被AI打破。

# 把文本变成画作, AI对艺术“下手”了

◎本报记者 马爱平

1月初,美国人工智能公司OpenAI推出两个跨越文本与图像次元的模型:DALL·E和CLIP,前者可以基于文本生成图像,后者则可以基于文本对图片进行分类。这个突破说明通过文字语言来操纵视觉概念现在已经触手可及。自然语言处理和视觉处理的边界已经被打破,多模态AI系统正在逐步建立。

“数据的来源或者形式是多种多样的,每一种都可以称为一种模态。例如图像、视频、声音、

文字、红外、深度等都是不同模态的数据。单模态AI系统只能处理单个模态的数据。例如对于人脸识别系统或者语音识别系统来说,它们各自只能处理图像和声音数据。”中国科学院自动化研究所副研究员黄岩在接受科技日报记者采访时表示。

相对而言,多模态AI系统可以同时处理不止一种模态的数据,而且能够结合多种模态数据进行综合分析。“例如服务机器人系统或者无人驾驶系统就是典型的多模态系统,它们在导航的过程中会实时采集视频、深度、红外等多种模态的数据,进行综合分析后选择合适的行驶路线。”黄岩说。

据判断,这个数据集的语义类别。”黄岩说。

“语言数据最常见的就是句子,是由不同的词语序列化构成的一维结构。不同于图像像素,文本中每个词语已经包含了非常明确的语义类别信息。而自然语言处理则是在词语的基础上,进行更加高级的语义理解。”黄岩说,例如相同词语排列的顺序不同将产生不同的语义、多个句子联合形成段落则可以推理出隐含语义信息。

可以说,自然语言处理主要研究实现人与计算机直接用自然语言进行有效信息交流,这个过程包括自然语言理解和自然语言生成。自然语言理解是指计算机能够理解人类语言的意义,读懂人类语言的潜在含义;自然语言生成则是指计算机能以自然语言文本来表达它想要达到的意图。

由此可以看出,自然语言处理要解决的问题的层次深度超过了计算机视觉,自然语言处理是以理解人类的世界为目标,而计算机视觉所完成的就是所见即所得。这是两个不同层次的任务。目前来说,自然语言处理在语义分析层面说要高于视觉处理,二者是不对等的。如果强行将两者进行语义关联的话,则会产生“语义壁垒”。

## 不同层次任务强行关联会产生“壁垒”

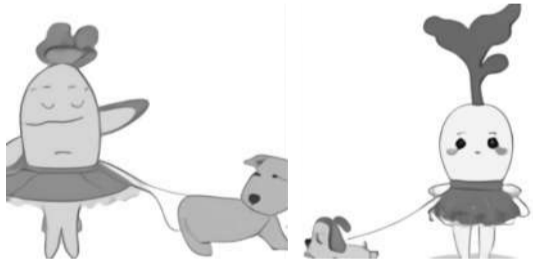
就像人类有视觉、嗅觉、听觉一样,AI也有自己的“眼鼻嘴”,而为了研究的针对性和深入,科学家们通常会将其分为计算机视觉、自然语言处理、语音识别等研究领域,分门别类地解决不同的实际问题。

自然语言处理与视觉处理分别是怎样的过程,二者之间为什么会有壁垒?

语义是指文字、图像或符号之间的构成关系及意义。“自然语言处理与视觉处理,都重在对不同模态数据所包含的语义信息进行识别和理解,但是两种数据的语义表现形式和处理方法不同,导致存在所谓的“语义壁垒”。”黄岩说。

视觉处理中最常见的数据就是图像,每个图像是由不同像素点排列而成的二维结构。像素点本身不具有任何语义类别信息,即无法仅凭一个像素点将其定义为图像数据,因为像素点本身只包含0到255之间的一个像素值。

“例如对于一张人脸图像来说,如果我们只看其中某些像素点是无法识别人脸图像这一语义类别信息的。因此,目前计算机视觉领域的研究人员更多研究的是如何让人工智能整合像素点数



图为人工智能系统DALL·E根据文本“穿着芭蕾舞裙遛狗的小萝卜”绘制的图像  
图片来源:OpenAI官网

# 介绍对象这件事,算法和朋友谁更懂你

人工智能算法正在渗入人们生活的方方面面,不仅衣食住行,夹杂情感需求的恋爱、婚配问题也开始被人算法寻求解决方案。

数据显示,如今全球线上交友产业市值已接近40亿美元,未来还将有更多资本参与其中。这些社交应用产品可利用人工智能算法,通过分析用户的出生日期、性格、爱好等指标为其进行对象匹配,还能通过自然语言处理算法和用户进行口语化交流,并根据用户信息对他们进行“约会指导”,甚至能够通过后期回访来调整推荐给用户的配对人选。

在信息技术尚未发展之时,媒人这一角色主导着婚恋市场的信息流通,那么算法匹配和朋友推荐,这两种介绍对象的方式,哪一种更能够被当代年轻人接受呢?

1月9日,由腾讯研究院举办的“腾讯科技向善暨数字未来大会2021”召开,来自不同行业的专业人士针对“算法推荐对象和朋友介绍对象,哪个更靠谱”展开了激烈交锋。

## 婚恋App靠数据算法做红娘真的靠谱吗

腾讯研究院副院长李刚认为,在婚恋交友过程中,算法推荐和朋友介绍实质上是两种不同的渠道,比较这两种渠道时应当关注如下标准——样本量、可持续性、是否偏见等,还有渠道能否在反馈中不断学习进步,匹配对象越来越靠谱。“姻缘可能是最古老的社会关系,姻缘确定

## 算法应该更好地服务人而非束缚人

中国社会科学院新闻与传播研究所助理研究员孙萍表示,根据2020年的相关数据显示,中国目前有近7700万独居青年,其中七成人没能通过社交途径找到合适的对象。

“算法是不是可以帮帮忙?当然可以。首先算法数据量足够大,并且精确度很高。”孙萍认为,

## AI打破自然语言处理和视觉处理的边界

此前,OpenAI斥巨资打造的自然语言处理模型GPT-3,拥有1750亿超大量数据,是自然语言处理领域最强AI模型。人们发现GPT-3不仅能够答题、写文章、做翻译,还能生成代码、做数学推理、数据分析、画图、制作简历。自2020年5月首次推出以来,GPT-3凭借惊人的文本生成能力受到广泛关注。

与GPT-3一样,DALL·E也是一个具有120亿参数的基于Transformer架构的语言模型,不同的是,GPT-3生成的是文本,DALL·E生成的是图像。

在互联网上,OpenAI大秀了一把DALL·E的“超强想象力”,随意输入一句话,DALL·E就能生成相应图片,这个图片内容可能是现实世界已经存在的,也可能是根据自己的理解创造出来的。

此前,关于视觉领域的深度学习方法一直存在三大挑战——训练所需大量数据集的采集和标注,会导致成本攀升;训练好的视觉模型一般只擅长一类任务,迁移到其他任务需要花费巨大成本;即使在基准测试中表现良好,在实际应用中可能也不如人意。

## 多模态交互方式会带来全新的应用

随着人工智能技术发展,科学家也正在不断突破不同研究领域之间的界限,自然语言处理和视觉处理的交叉融合并不是个别。

“语音识别事实上已经加入其中,最近业内出现很多研究视觉+语音的新任务,例如基于一段语音生成人脸图像或者跳舞视频。”黄岩说,但是要注意到,语音其实与语言本身在内容上可能具有较大的重合性。在现在语音识别技术非常成熟的前提下,完全可以先对语音进行识别将其转换为语言,进而把任务转换为语言与图像交互的常规问题。

无论是DALL·E还是CLIP,都采用不同的方法在多模态学习领域跨出了令人惊喜的一步。今后,文本和图像的界限是否会被进一步打破,能否顺畅地用文字“控制”图像的分类和生成,将会给

现实生活带来怎样的改变,都值得期待。

对于多模态交互方式可能会带来哪些全新应用?黄岩举了两个具有代表性的例子。

第一个是手机的多模态语音助手。该技术可以丰富目前手机语音智能助手的功能和应用范围。目前的手机助手只能进行语音单模态交互,未来可以结合手机相册等视觉数据,以及网络空间中的语言数据来进行更加多样化的推荐、查询、问答等操作。

第二个是机器人的多模态导航。该技术可以提升服务机器人与人在视觉和语音(或语言)方面的交互能力,例如未来可以告诉机器人“去会议室看看有没有电脑”,机器人在理解语言指令的情况下,就能够结合视觉、深度等信息进行导航和查找。

“人工智能对于情感的分析正不断走向深层,未来用机器模仿人的情感可能会达到比今天更高的水平,甚至表现出共情的能力。”彭兰指出,人与人的沟通中,有一样难以被机器复刻,那就是人的生命历程。机器没有人的生命历程,哪怕它可以描绘模拟情感,但它未必有真正跟人对话的可能性。

算法有大量数据作为基础,在某些方面的判断选择可能比人类靠谱。但算法会形成模型和套路,有时候机器的套路可以打破人类思维局限,但同时我们又需要警惕彻底被套路。

人类的情感到底容不容得下智能的结构?数据算法的进化可能性有多大?在这个时代,是人类面对智能,面对大数据不得不去思考的话题。

“如果真有所谓算法推荐的合适伴侣,那么它会不会再出一个如何和与对方相处的指南呢?”米果文化传媒公司CEO胡渐彪认为,利用算法干预人类情感,在现阶段其启示意义远大于实用意义。算法能否成功帮助人类解决婚恋问题,技术的发展最终会给出答案,“可能10年之后我们要思考的就是这个问题的第二篇章”。

“内容来源于微信公众号“腾讯研究院”)

## 好机友

◎本报记者 陈曦  
通讯员 刘晓艳

一只灵巧的“虫子”,牢牢地抓住圆管形,一伸一缩中蜿蜒前行。其实,这只灵活的“虫子”是一款采用了天津大学科研人员左思洋、刘建彬课题组研发的新型模块化柔性驱动方法3D“打印”出来的软性机器人,可应用于人工肌肉和管道爬行机器人中。该成果今年1月初在线发表在《美国电气电子工程师学会机器人和自动化快报》上。

软性机器人因其较高的柔性、对人体安全等优点,近年来得到广泛关注。3D打印的优势在于制造复杂形体、复杂结构可一次成型,不需要后续加工。打印免组装机结构,是3D打印技术制造软性机器人比较典型的应用。

左思洋、刘建彬课题组提出了一种基于薄膜气缸的新型模块化柔性驱动方法,可根据具体应用改变排列组合方式以及合理布置连接方案,将其应用于人工肌肉和管道爬行机器人中。

“每一个薄膜气缸就好比人体的一小块肌肉,或者爬虫的一个‘节’,只不过是用热塑性聚氨酯材料做的。”刘建彬解释说,如果把这个新型薄膜气缸结构比喻成一个基本的肌肉单元,根据不同的应用需求对这些单元的连接方式进行组合,就像是把一个个肌肉单元连接起来形成一整块肌肉,然后再应用于不同场景。

整块“肌肉”的制造过程采用了3D打印技术,一次成型,省掉了传统机电设备加工制造中的装配流程,大幅降低了驱动模块的制造成本和周期,且具备耗气量小、动态响应高、可靠性高、对应用场景适应性等特点。

基于此创意,课题组首先提出了一种新型气动人工肌肉,可应用于柔性外骨骼等人机交互装备的驱动中。气动即以压缩空气为动力源,带动机械完成伸缩或旋转动作。与传统气动人工肌肉相比,该设计最突出的特点是不会产生厚度方向的膨胀,从而避免了对人体的挤压。

此外,课题组还提出了一种新型气动管道爬行机器人,可应用于工业管道设施的检查和实时监控。该管道爬行机器人采用仿生尺蠖原理,通过巧妙布置薄膜气缸单元之间的连接,实现机器人在管道内、外壁面爬行。柔性驱动方式的应用使该机器人能够适应大范围管道直径的变化,并可应对直管、弯管、竖管、水平管以及各种角度倾斜管的应用场景,同时机器人可承受自重80倍以上负载。

因为采用了气动方法驱动,软性机器人只能拖着长长的气管尾巴工作。如果将传感器集成到设备中,就能去掉这些气管尾巴,使机器人更独立精致。

## 用AI打造应急管理平台 为疫情防控提供有力保障

科技日报讯(通讯员王筱骄 记者王春)1月24日,中国科学院上海高等研究院(以下简称高研院)科研人员的手机上弹出一则好消息:沈阳此次疫情的新冠肺炎患者全部出院。自去年12月23日以来,沈阳市新增多例新冠肺炎确诊病例。高研院迅速调集百人科研及开发团队,两周内其研发打造的“新冠疫情防控应急管理平台”紧急上线,在全市推广应用,为沈阳市抗击疫情提供了有力的技术保障。

据高研院张坤林博士介绍,该应急管理平台依托人工智能及区块链技术,面向疾控中心以及密切接触者管理相关单位,对密切接触者追踪、访查、管理的工作需求,实现了多项目标功能。

据介绍,这套“新冠疫情防控应急管理平台”的技术亮点在于通过人工智能及区块链等技术对接人员的行程轨迹、接触地点、交通工具等进行分析,实现密接人员关系图谱构建、区域风险提示、重点区域及重点人员预警等功能。

在前期调查期间,此平台还提供引导式的密接者信息录入,建立统一规范的密接者档案。张坤林解释道,流调有着规范的框架,但因为具体工作人员的方式不同会产生不同的结果,所谓引导式录入即情景式的提示,如在流调中的某个环节提示密接者手中是否有接触物,在引导式过程中提醒工作人员注意此点,做到不落项,确保密接调查信息的准确性、高效性。

人工智能技术主要体现在建立密接者数据库,从时间、空间、人物关系等维度对密接者进行基本信息、轨迹信息、以及接触人的充分描述。通过摄像头人脸入侵检测,人体手环等技术手段对隔离人员进行集中管控,以便于各级实施掌握密接者动态,快速上报、下发相关情况。

此平台通过区块链智能分析,通过密接者的关系图谱、行程轨迹,进行大数据分析,提供密接线索的精准推送,以及疑似密接人员、或范围的重点区域进行精准预测,对下一个可能集中爆发的点实施管控,减少传播量。张坤林谈道:“正常情况下是事后排查,也就是人员确诊之后排查,但现在利用此平台可在密接潜伏期对其接触人员彻底排查,在确诊前事前管控。”

3D打印软性机器人可应用于人工肌肉