

# 能中英双语实时复述的智能读唇模型

■ 金木

## 帮助听力障碍者读懂唇语

随着人工智能系统被赋予越来越多的功能,它帮助我们解决更多问题。AI读唇正成为听力障碍患者的耳朵,帮助他们“听到”原本消失在耳朵里的声音。

据统计,目前全世界约有4.66亿人不幸患有听力障碍,大约占到了世界人口的5%。根据世界卫生组织的预估,到2050年,这个数字将会攀升至9亿。大多数患有听力障碍的人交流时会存在困难,手语是其中的一种交流方式。如何帮助听力障碍者和聋哑人与外界交流,是很多人都在关注的问题。交流首先要解决的是“听懂”别人的话,手语和唇语都可以作为一种相互理解的方法。听障患者通过读唇语得以获取信息,但是这个技能需要长时间的练习,并且即使掌握了识别率也很低。

将“读唇术”交给AI实际上已经不是什么新鲜事了,早在2016年,谷歌deepmind和牛津大学的研究人员就开发了名为Lipreading视频标注系统,准确率达95.2%,远远优于受训的读唇者。2018年哥伦比亚大学研究团队开发了大型视觉语音识别系统。为了追求更高性能的系统,近日,阿里巴巴、浙江大学和斯蒂文斯理工学院的研究人员共同开发出名为LIBS(Lip by Speech)的算法,在将从语音识别中提取出来的特征当作补充材料,其准确率达业界领先水平,同时还针对目前最大的普通话唇语语料库CMLR进行了训练,读懂中文也不在话下。

## 人工智能如何学会读唇

机器读唇很困难,因为它需要从视频中提取时空特征(位置和运动都很重要)。现大多数机器学习系统只能进行单词分类,而不进行句子级的序列预测。LIBS的研究者表示,这个系统在两个基准上管理着行业领先的准确性,在字符错误率上分别比基准高出7.66%和2.75%,能帮助有一定听力障碍的人群观看无字幕的视频。

LIBS可以从视频中的多个层次提取有用的音频信息,包括在序列层、语境层和帧数层。将提取的数据与视频数据对齐,识别对应部分(由于样本数量和

近年来,随着深度学习和大数据应用的发展,AI读唇的研究也取得了前所未有的进步。AI读唇可用于帮助深受听力障碍困扰的患者,改善嘈杂环境中的语音识别等,具有巨大的潜力。

缺失值不同,可能存在开头或结尾部分的视频和音频序列长度不一致的情况),并采用特定的筛选方法对有用特征数据进行了提纯。

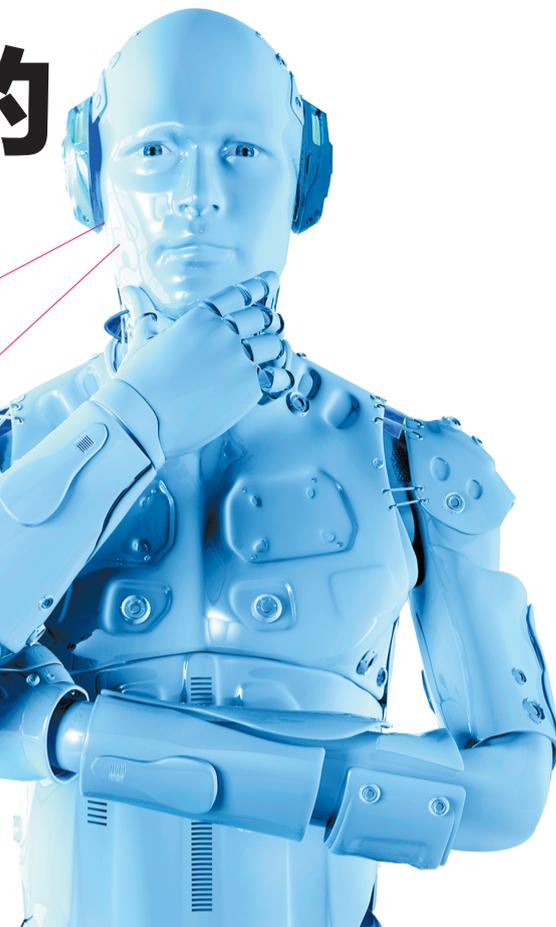
LIBS的语音识别和唇语识别都是基于注意力机制的“序列到序列”模型,这是一项机器翻译方法,将序列(例如音频或视频序列)输入转化为标签和注意值的输出。值得注意的是,序列到序列模型在自动语音识别(ASR)领域越来越受欢迎,因为它将传统ASR系统的独立组件折叠成了单一的神经网络。总的来说,LIBS通过引入一种新的过滤策略来从语音识别器中提取特征,并通过采用基于交叉模式对齐的方法,来进行帧级知识提取,从而解决两个序列之间采样率不一致的问题,以实现准确的唇语识别。

研究人员分别使用LRS2和CMLR数据集在上述模型中训练,LRS2数据集中包含45000条来自BBC的句子音频,而CMLR来源于中国网络电视网,是包含100000条以上自然语言句子的最庞大的普通话唇语语料库(包括3000个以上的汉字和20000条以上的词组)。

## 句子越长读得越准确

LIBS的研究团队发现,模型在使用过短句子(如LRS2数据集)进行预训练时得到的结果不理想,因为解码器从少于14个字母的句子中提取有效信息的难度较大。然而,一旦模型使用最大长度为16个单词的句子进行预训练,解码器由于获得了语境层的知识,对LRS2数据集的句末解码质量有了显著提高。

早在1982年,就有Easton和Basala的研究表明,



人的唇读能力会随着长单词的出现而增强,这表明了在模糊的沟通渠道中,上下文之间的特征词捕捉时间的重要性。2016年的LipNet模型也是针对长句在读唇上的优势构建的。该模式利用时空卷积、递归网络和连接主义时间分类损失,将一个可变长度的视频帧序列映射到文本。LipNet系统通过6个不同的电视节目、共超过10万个句子进行5000小时的训练。最终这个AI系统通过只看每个说话人的嘴唇,就能准确地破译整个短语。

近年来,随着深度学习的发展和训练大数据的可用性,人工智能系统学习读唇取得了前所未有的进步,表现也有了很大的提高。读唇对于人类来说也是一项困难的工作,当人们看到说话人的嘴巴时,通常会被细微的、容易混淆的唇形变化所困惑。所以说,AI读唇可以用于改善助听器、公共空间中的默写、嘈杂环境中的语音识别、生物识别和无声电影处理。AI读唇不久后将可作为应用程序整合到手机中,这使得听力障碍患者无论走到哪里都能随身携带一只“耳朵”。这样的系统还可以帮助那些因为声带受损而不能说话的人“发出声音”。另外,LIBS团队表示在未来的工作中,他们期待着将同样的框架应用到其他模式中,比如语音和手语系统。

## 我国首家人工智能机器人博物馆 落户远东理工学院

本报讯(通讯员 陆伟)近日,我国首家人工智能机器人博物馆落户哈尔滨远东理工学院。此博物馆依托上海太敬集团与其创办的哈尔滨远东理工学院共同打造的具有国内一流水准的人工智能与机器人科创中心。场馆共有五层,由教育博物馆、科技体验馆、竞技场、研发与考试中心、大数据中心、学术报告厅六大部分组成,集科学研究、科技体验、技术推广、工程与艺术设计等诸多职能与声光电效果于一体,实物展示与声光电效果相得益彰,

图文并茂地向您展示人工智能与机器人科创的跨界融合之作。

哈尔滨远东理工学院作为全国第一所建立“机器人学院”的高校,建成了首家专业的机器人博物馆,随着科技的迅猛发展,云计算、物联网、大数据、人工智能等足以改变世界的新技术不断涌现,越来越多新技术、新应用走进了人们的日常生活,人工智能机器人博物馆将在不久后对外开放,接待团队参观。

## 我国获得智能动态防雷国际标准制定权

本报讯(通讯员 李伟)近年来,随着全球气候变化,雷电灾害愈加难以防治,这也是一个世界性难题。近日由国家电力公司主导,国网江苏电力和苏州供电承担自主研发的智能动态防雷科技成果,已通过国际大电网会议(CIGRE)39个以上成员国论证并立项,正式承担“动态防雷”国际标准制定,今后将在世

界范围内进行推广应用。这是我国在这个领域首次承担国际标准制定,标志着试点应用的电网防雷技术达到国际领先水平。

据了解,国际大电网会议(CIGRE)是电力系统三大国际标准组织之一,其下属的国际大电网工作组、报告及技术规范,是电力行业国际标准体系的重要组成部分。

## 英国科学家研究出 可穿戴式脑扫描仪

英国科学家团队报告一项神经科学最新成果——一种自行车头盔改造的可穿戴式脑扫描系统。该设备可以让儿童的脑扫描更简单、更可靠。

脑磁图是一种通过大脑产生的弱磁场测量脑部活动的技术,受检者需要将头部置于特别敏感的测定器中,仪器可测出颅脑的极微弱的脑磁波,再用记录装置把这种脑磁波记录下来,形成图形。这一类脑扫描技术可以提供关于大脑功能的有用信息。英国诺丁汉大学研究人员马修·布鲁克斯及其同事,在自行车头盔里安装了一个可穿戴脑磁图设备,并在头盔上固定了一些小型轻便的传感器,用于追踪头部运动,让扫描质量不会受到患者运动的影响。

测试中,他们用这个设备记录了2岁和5岁幼童的大脑对外部接触的反应。儿童可以在家或骑车时佩戴一个同款头盔,先习惯一下,从而减少实际扫描时的焦虑感。研究人员用该头盔记录了一名青少年打电子游戏和一名24岁年轻人弹尤克里里时的脑活动,证明该头盔可以调节适应各种头型和尺寸。研究人员表示,除了研究自闭症和癫痫这类神经发育性疾病外,该脑扫描仪还能用来研究健康大脑的发育。(本报综合)