

从科大讯飞造假事件看AI翻译未来的发展

张 凯

中国地质大学(武汉) 湖北省武汉市 430074

摘要:近年来,人工智能与自然语言处理领域的结合使得机器翻译在速度和准确度都在迅速提升,在语言服务业中作用日益彰显。与此同时,机器翻译将取代人工翻译的呼声日益增高。然而,最近国内龙头智能语言技术公司科大讯飞被爆出“AI同传”造假事件,再次引发了翻译界对机器翻译的思考。本文以该事件为背景分析人工智能翻译现状与不足,并结合相关领域科技发展情况,预测了人工智能翻译未来的发展方向,最后本文在此基础上对译员未来职业发展道路提出了一些建议。

关键词:机器翻译;人工智能(AI);人机耦合

2018年9月20日,一位同声传译员在社交媒体平台爆料称,科大讯飞AI会议同传技术造假,会议中的同声传译工作均由其团队完成,所谓的AI同传只不过是该公司研发的语音识别系统将译员声音转化为文字,再通过机器将文本朗读出来。以此给外界造成一种“AI实现了同声传译”的错觉。消息一经传开,科大讯飞被推上了风口浪尖,业内人士纷纷表示这样的做法在很大程度上剽窃了同传译员的工作成果。不可否认,近年来,机器翻译在人工智能技术的加持下取得了飞速的进步,以至于一些译员甚至担忧自己的“饭碗”会被机器给夺去。但是我们必须认识到,即便被赋予了“智能”,机器翻译目前的水平依然还无法胜任大部分场合的要求,而对于每一个翻译从业者来说,了解机器翻译的过去、现在和未来发展方向的重要性不言而喻。

1. 人工智能翻译及其发展史

1.1 人工智能与机器翻译

人工智能(Artificial intelligence)是指研究、开发模仿人的思维和行为的一种科学。人工智能通过模拟人的思维和行为,像人一样思考,可以在各种各样的领域中得到应用。

机器翻译,又称为自动翻译,是利用计算机将一种自然语言(源语言)转换为另一种自然语言(目标语言)的过程。以前的机器翻译速度比较慢,准确率也比较低,而且难以处理篇章的翻译。随着人工智能的发展以及其在翻译领域的运用,人工智能翻译也逐渐进入人们的生活。机器翻译在有了人工智能技术加持下,能够更加有效地分析源语言的语句和语义并将其转化为行文

更加流畅的目标语言。人工智能翻译融合了自然科学和社会人文科学中的多门学科,包括但不限于计算机科学、机器学习、语言学等。AI翻译是人工智能技术发展的一个重要方向,具有极高的科研价值和实用价值。

1.2 人工智能翻译发展简史

人工智能翻译起源于20世纪30年代,因计算机科学的兴起而飞速发展。最早提出应用计算机技术进行翻译的科学家是法国的阿尔楚尼教授。当时,主要是通过一个简单的机械装置进行词汇翻译,即通过词典来进行词汇转化。1947年,沃伦教授首次提出通过计算机来进行翻译工作,但由于当时技术条件的限制,机器翻译的质量进展缓慢。机器翻译真正进入快速发展期是2010年之后。在2011年,美国国际商用机器公司(IBM)推出了Watson系统,这个系统在进行翻译方面取得了巨大的进步和成就,随后,微软在2018年提出其公司的人工智能翻译水平已经达到与专业的翻译人员相当的水平,在同一年,谷歌推出了Google Assistant,可以完全与人进行对话,帮助人们完成一些简单的工作,这证明了人工智能的语言理解能力已经有了很大的突破。

近年来,我国许多优秀的互联网科技公司如百度、腾讯、搜狗等也推出了各有特色AI翻译平台,这些产品的出现极大地降低了基础信息交流的成本。

2. 人工智能在翻译中的应用现状

机器翻译和语音识别是人工智能在翻译中最主要的两大应用领域。机器翻译从最初的基于规则的机器翻译到基于统计的机器翻译最后进化为基于神经网络的机器翻译,也就是AI翻译。谷歌翻译(Google Translation)作为AI翻译中的佼佼者,其通过机器学习和大数据分析将机器翻译的准确度大大提升,尤其是相对规范化的文本类型,笔译工作者基本已经可以把翻译结果当作译后编

作者简介:张凯,中国地质大学(武汉)外国语学院翻译硕士,研究方向为英语笔译。

辑的初稿使用。语音识别主要是以语音作为研究对象,通过语音信号处理和识别模式,自动识别和理解不同的语言,并将其转化为文本。语音识别技术的成熟使得翻译更加方便、快捷、及时。科大讯飞在这次同声传译中被批评作假,并非是其没有使用人工智能翻译,而是其将语音识别包装成了机器翻译。

3. 人工智能翻译的局限性

3.1 难以从整体把控译文质量达到“信达雅”

谈到翻译的标准,相信每一个译者都会脱口而出反复的“信达雅”,对于人工智能翻译来说,这个标准显然也是适用的。

对信与达,近两年的神经网络机器翻译较以往有了极大的提高。最主要的进步来自“达”。传统的基于短语的统计机器模型我们可以把它理解为一个受目标语言语言模型影响的翻译模型,打一个比方,就好比一个母语是汉语,初通英语的人拿着一本词典进行汉英翻译。而神经网络机器翻译本质上是一个“受源语言限制的语言模型”,也打一个比方,我们可以把这个过程看成一个母语是英语,对汉语阅读也比较熟练的人不依赖词典进行汉英翻译。基于短语的模型往往翻出奇怪的Chinglish,而神经网络机器翻译模型的译文往往非常流畅。但是神经网络机器翻译可能因为对汉语的理解偏差而“瞎翻译”。而这种情况往往来自数据稀疏——如果这位翻译官本身不是中国通,要他翻译北京胡同里的市井俚语肯定要出大麻烦。要继续提升信和达二者,难点之一在如何提高数据覆盖(而不仅仅是增加数据量),特别是源语言的数据覆盖率。除了数据,当然翻译官本身要更加聪明。从神经网络角度出发,这一需求要求更大的模型,更合理的模型结构,更好的训练方法。同时增强模型的学习能力也需要硬件的配合。

相对信和达来说,机器翻译面临的另一个困难是“雅”。目前绝大多数机器翻译系统都是逐句翻译的,并没有利用篇章上下文信息。因此经常会出现这句话阳春白雪,下一句就是下里巴人,前言不搭后语的情况。这个问题解决起来是很不容易的。同句子相比,篇章上下文的信息量实在是非常巨大,如何能够找到合理的信息表示方法并且在翻译中应用目前还没有非常好的解决方法。推而广之,将机器翻译放置于一个更为广大的世界中,上下文不但包括文本,还包括时代背景,空间位置,知识背景等,它们对翻译结果都有影响。

3.2 难以传达语言的“人文性”

语言具有工具性和人文性的双重属性。语言是用来交流的,这是语言的工具性的体现。而语言作为人类的

交际工具,以口头和书面的各种形式创造并传承着人类的文化。语言是人类创造的自然也离不开人类文化,处处浸润着人类文化的气息。这便是语言人文性的体现。而语言这些背后的文化,恰恰是人工智能无法解码和分析的。

拿目前最成熟的人工智能翻译谷歌翻译为例,我们在使用中不难发现,在处理一些客观性较强的文本时,其译文质量尚可,但是如果使用谷歌翻译去翻译文学文本时,其译文输出质量几乎处于不可读的水平,甚至在翻译一些口语性较强的日常对话时都会出现可笑的错译。因为人工智能的“智能”的本质来源于大数据,来源于统计学。无论软件拥有多么优秀的算法,多么庞大的数据库,其能做到了本质工作还是进行源语言和目标语言之间匹配,而无法分析语言背后承载的文化,更别说是跨文化之间的转换工作了。

4. 人工智能翻译的未来发展趋势—人机耦合

人工智能翻译的质量一定会越来越高,甚至在一些相对低端的翻译领域会完全取代人工翻译,这是不可阻挡的科技发展趋势。但语言是人创造的,包含着各种各样“人”的因素,翻译是人类一项具有主观能动性的工作,这也决定了,机器翻译在可预见的未来无法完全取代人工翻译,而是会成为优秀译员的强大工具,极大提高他们的工作效率。

以这次科大讯飞AI同传造假为例,很显然,按照目前人工智能翻译的水平,实现高质量的同传显然还无法达到,所以科大讯飞才会让同传译员在后台“偷偷”翻译,这确实涉嫌虚假商业宣传,同时也是对译员辛勤工作的不尊重。但换个角度来看,虽然目前人工智能机器翻译的准确率仍然很低,但人工智能翻译应用的另一个领域—语音识别,其准确率已经高达95%以上,在同传和交传,完全可以使用这项技术将话语转换成文本,再交给译员进行翻译,这样一来原本复杂的口译工作就变成了相对简单的视译,这样不仅能提高翻译的效率和质量,也能够减轻翻译人员的工作强度。而针对一些专业性强的领域,译员可以通过根据文本建立不同领域的术语库和翻译记忆库,在以后翻译相关文本时可以借助人工智能强大的配对算法,让机器来完成枯燥的重复性工作,译员也有了更多时间来打磨译文。

5. 人工智能翻译对译员职业发展的建议

5.1 注重翻译的“人文性”

AI翻译虽然在语言转换中一定程度模拟“人的智慧”,但语言翻译中的“人文性”不可忽视。从主要AI翻译平台的译文输出质量来看,机器翻译能做到差强人

意的场景依然有限。机器学习的本质依然很大程度上依赖统计学原理，实现路径依然是标准化输出。对于源语言语境和目标语言服务目的的把握是AI无法精确识别的。

语言的背后是文化，其中包含着人类复杂的理解力和创造力，文化的不断演进决定了语言的人文性也是一个动态变化的过程。从笔译角度来看，文学性的文本翻译很大程度是译者对于语言这一属性理解的基础上进行的转移过程，每一次遣词造句都是对源语言感情色彩的细微把握。另一方面，从口译领域来看，优秀的口译员需要不仅是出色的双语能力，还有察言观色的本领和跨国文化沟通能力的要求，这些都是AI所不具备的能力。

5.2 成为机器翻译的“设计师”

回顾历史，每一次科技革命都会减少一些传统工作岗位，但也会带来大量新的工作岗位。人工智能翻译的发展势必会取代一些低端翻译工作岗位，但与此同时，相关的软件工程人员需求也急剧增加，而既掌握翻译知识，又懂得软件开发技术的人员无疑会拥有强大的就业竞争力。

大部分翻译从业者可能会因为自己是文科背景出身而畏惧学习技术方面的知识，但是目前人工智能领域相关的编程语言其实是呈现出简化的趋势，学习门槛也越来越低，像目前主流的人工智能编程语言Python、Swift

等并不要求学习者储备太多的计算机相关知识。因此，对于该领域感兴趣的翻译从业者完全可以学习人工智能编程算法相关的知识，做跨学科人才，成为机器翻译的“设计师，”以全新的身份为翻译事业做出贡献。

6. 结语

总的来说，人工智能的翻译水平正在迅速发展，它的便捷性、快捷性、创造性和实效性都让人们感受到了它的魅力。但是语言的人文性和多样性，以及不同场合对翻译质量的要求决定了人工智能翻译并无法取代人类译员，而是应该和人类协同工作实现优势互补。我们应该坚信：比人类更强大的，绝不是AI，而是掌握了AI的人类。

参考文献：

- [1] Jerry Kaplan. ARTIFICIAL INTELLIGENCE—What Everyone Needs to Know[M]. Oxford University Press, 2016
- [2] 杨宪泽. 人工智能与机器翻译[M]. 西南交大出版社, 2006.
- [3] 邵华, 杨晓峰. 人工+智能语言翻译的现状与发展[J]. 电子世界, 2018(16): 28-29.
- [4] 刘逸之. 从人工智能翻译的产生浅析人工智能在语言服务业的应用与发展[J]. 中国战略新兴产业, 2018(4): 153-154.
- [5] 华美淇. 浅析人工智能的应用现状与发展前景[J]. 计算机产品与流通, 2018(13): 18-19.