

人工智能背景下机器翻译的难点探析

王昱寒

(辽宁对外经贸学院 外国语学院, 辽宁 大连 116052)

基金项目: 本文系辽宁省国家级大学生创新创业训练项目尔雅文化翻译服务工作室阶段性研究成果, 项目编号 202010841003S。

摘要: 论文基于机器翻译理论, 分析了人工智能背景下机器翻译的难点, 阐述了人工智能对机器翻译的发展具有革命性的作用, 试探性地提出了如何全面发展该行业的新路径。

关键词: 人工智能; 机器翻译; 难点探析

1 人工智能背景下机器翻译的理论基础

人工智能被称为机器智能的计算机科学, 可以对人类进行描述、模仿, 与其他人类思维相关联, 并通过机器进行展示。伴随着计算机技术的高速发展, 人工智能自 1956 年起, 成为一门学术学科, 至今已成为囊括“机器翻译”在内的, 涉及各个领域的应用技术。

1.1 自然语言理解

自然语言理解作为人工智能的重要分支学科之一, 其研究范围包括通过计算机技术实现人类语言交流过程的模拟, 使计算机能够尽可能全面理解和运用人类社会自然语言, 如汉语、英语等, 通过实现人与计算机的自由交流令计算机帮助或代替人类进行资料查询、问题解决、整理汇编等各种脑力劳动。在翻译过程中, 需要通过计算机的辅助进行输入语的理解, 以进行翻译过程的初步阶段。随着技术的进步和自然人机交互需求面的扩大, 各个品牌的语音助手成为了自然语言理解最好的产物。为了实现人与计算机之间语言交流的自由, 计算机必须能够基于对自然语言文本含义的理解, 以自然语言的形式表达自己(包括图像, 文本, 声音等)。

但是, 自然语言理解依旧存在着难点。第一, 语法方面。语法分析只能在一个完整的句子中进行, 在一篇文章中, 被分析句子的上下文及该句所处的语境有无法忽视的影响, 所以, 省略句、歧义句、有代词存在的句子由不同的人在不同场合说出的同一句话的含义都不一样, 且尚无任何规律可查, 这一点仍需研究。第二, 知识的储存量。人类社会有不同的人种、不同的文化背景、不同的交流方式, 这就意味着, 人类在用语言交流的过程中, 理解句子含义不仅是在语法层面, 更多的是基于个人所处社会的文化背景与习惯。

1.2 自然语言处理

自然语言处理是指让计算机接受人类以自然语言输入, 并在其内部通过人类所定义的数据处理方式进行加工、计算等处理, 此过程首先需对人类自然语言的理解方式进行模拟, 随后才能生成符合人类期望的自然语言, 最终用以实现

以计算机替代人类进行自然语言信息的大批量处理。简而言之, 令计算机精准识别人类自然语言, 并与人类进行无障碍沟通交流, 其处理过程涉及步骤包括: 语音识别、自然语言理解, 自然语言生成。在翻译过程中, 需结合翻译语料库对输入语进行处理、组合、校对等步骤, 是机器翻译中最重要的一环。

对自然语言处理的技术壁垒进行归纳总结, 主要包括以下几点: 第一, 词语具有歧义性, 即词语的意义不明确, 有两种及以上的解释。这就要求技术针对语境及上下文实现对词义的恰当解析; 第二, 句子存在模糊性, 此情况易发生于句子中有代词时, 例如: 毛毛经过老师的精心辅导考取了满意的大学, 他十分开心。在此句中, “他”具体指向毛毛或是老师, 还仍待进一步考究。第三, 语言表述存在不规范性, 在人类社会这个多元文化的背景下, 不同人种的交流方式、语言、文化大相径庭, 也就意味着自然语言处理的理解过程需要非常丰富的资料库, 目前还没有任何资料库可以做到将全人类的资料进行储存。

1.3 翻译记忆

翻译记忆是针对自然语言处理的相关程序及软件所建立的数据库。其存在的意义在于对人工翻译的辅助。用户可将自己所有的原文及译文整理汇编为一个或多个翻译记忆数据库。后续过程中, 计算机系统可自动匹配相同或相似语句并输出译文以此避免用户发生重复劳动的现象。翻译记忆库的常见问题有三点, 分别是译文质量缺陷、文本类型混杂、更新维护延迟。首先, 译文质量的缺陷通常体现在错译误译、专业术语的不一致、文风不统一和乱码这四个小方面; 对原文理解不够, 语法不当会导致错译漏译, “术业有专攻”, 各个领域的专有名词在库内搜索不到会导致专业术语的不一致, 数字表达、计量单位、标点符号、缩略词未按照统一标准翻译会导致文风不统一, 疑问无法识别会导致乱码; 其次, 文本类型混杂是由于翻译记忆库未能按照专业类型分类, 如医学、法律、文学等。不同行业的专用表述方式、书面语和语言风格都会有差异。最后, 更新维护不到位主要原

因是由于原文与译文的不匹配。例如,用户更新了原始文档,而负责译文更新环节的员工只更新了译文,并未对原文做出改动,这就导致了原文和译文间的不匹配。

2 人工智能背景下机器翻译的难点

严复先生曾在《天演论》中的译例言中讲到:译事三难:信、达、雅。在书中,严老先生也并没有给“信、达、雅”下定义,这句话也只是他老人家在翻译历程中收获的心得。信指译文需要准确,不出现遗漏的情况,更不能随便更改原文意思;达是指译文不拘泥于原文形式,译文需要在准确的前提下流利通顺;雅指进行翻译工作时选用词语需要符合语境和文化背景,追求文章本身的含义。

2.1 误译

误译就是错误的翻译,例如文化背景知识的缺失、逻辑关系及词不达意现象(一词多义和一次多性)。

2.1.1 文化背景的缺失

例 1: ...while they drank tea and chewed melon seeds and chattered

机译:他们一边喝茶,一边嚼瓜子,一边聊天

突然,喝酒和用餐被认为是侮辱,是男性阴谋将我们(女性)留在我们的地方,因此我们拿出支票簿去了荷兰人。

Suddenly, drinking and dining were considered insults. It was a male conspiracy to keep us (women) in our place, so we took out our checkbook and went to the Dutch.

突然,被宴请被视为侮辱,男性阴谋的一部分,以使我们(女性)在我们的地方,所以我们我们的支票簿,去荷兰。

Suddenly, being banned was regarded as an insult, part of a male conspiracy to keep us (women) in our place, so we went to the Netherlands with our checkbook.

无法加载全部结果

重试

正在重试...

正在重试...

解析:如果在翻译前对翻译库内进行文化背景的查询就可以得知,在中国文化背景下,一般会使用嗑瓜子,而不是嚼瓜子。但在机译中直接将 chew 译为嚼,不符合中文的习惯。

译文:他们边喝茶,边嗑瓜子,边聊天。

2.1.2 逻辑关系

例 2: There was little hope of continuing my inquiries after dark, to any useful purpose in a neighborhood that was strange to me.

机译:天黑后继续我的调查几乎没有希望,在我陌生的街区有任何有用的目的。

解析:从句子形式上看,可以将句子划分为三部分,There was little hope of 是表态,continuing my inquiries after dark, to any useful purpose 和 in a neighborhood that was strange

to me 是在解释原因,按照中文的语句使用习惯,顺序应该是按照时间一件一件的说明,最后表态。其中,这句话中含有因果和假设的关系,在翻译时,需要运用增译法添加相对应的连接词。

译文:因为这一带我不熟悉,如果天黑以后继续进行调查,那么取得结果的希望是不大的

2.1.3 词不达意

例 3: The set of equipment is quite late arrivals for doctors to use.

机译:这套设备很晚才可供医生使用。

这套设备很晚才可供医生使用。

The equipment was not available to doctors until late.

该套设备是相当抵港迟到医生使用。

This set of equipment is used by doctors who arrive late to Hong Kong.

无法加载全部结果

重试

正在重试...

正在重试...

分析:该句中的 late 有“迟到的、最新的、最近的”等多种含义。该句中只能选用“最新的”,否则含义不明。

译文:这是供医生使用的一套新型设备。

例 4: All members of the party were dead against the war.

机译:该党所有成员都死于战争。

分析:此句需对单词 dead 的词性进行准确判断,如果将 dead 视为形容词,那么就必然会误译。事实上,dead 在本句中词性为副词,用以修饰 against the war 这一介词短语,译为坚决。在语法上,句子中真正的系表结构为 be against the war。

译文:全体党员都坚决反对这场战争。

2.2 局限性

2.2.1 句法的局限性

机器翻译最早的使用仅限于词对词的翻译,只起到了和电子词典相似的功能和作用,句法的使用推动了机器翻译的前进,直译法可以解决翻译的需求,以准确度最高的 Google 翻译进行举例。

原文:我是名教师。

译文:I am a teacher.

但是随着原文句子结构复杂程度的增加,机器翻译的准确率急速下降;在句法层面进行分析,机器翻译的准确度仍然是停止在结构简单的句子中,复杂句依旧需要人工进行校对和审阅。

2.2.2 语义和语用的局限性

机器翻译更倾向于对原文的词句进行粗浅对等处理,而在人工翻译过程中会根据深层意义处理从而解决浅层差异,进而实现超越不对等的翻译。通常情况下,逻辑关系混

乱极大程度上是由于句法错误造成的,即词语间不存在逻辑关系导致句子不通顺。语用方面的局限性主要表现在多义词语的意义选择上。语义的局限性需结合语义学中语素分析和翻译学中语素选择进行处理,在翻译多义词时需重点关注词语的不同意义,可通读上下文理清逻辑关系后选择最恰当的含义翻译。在机器翻译中,多义词的翻译会默认选择高频词义,这就导致了译文的错误(如例3)。尤其是在汉译英的翻译中,更容易出现语义上的错误。

2.3 语篇翻译

当使用机器翻译进行段落和语篇的翻译时,句法和语义的问题全部暴露在译文中,语篇特征无法凸显,句与句之间的逻辑关系混乱甚至错误,没有连贯的译文,非常生硬。类似这样的情况,需要制定更加详细准确翻译规则。

人工智能背景下的机器翻译最重要的一点就是语用因素。为词语或句子的同一个含义可以通过不同的叙述方法来表达,使用户根据自身需要恰当的选择句式。但是,由于语境、文化等差异,相对应的译文需要做出不同程度上归化异化的处理,这样才能使文章内容更加贴合实际。

对于任何一篇文章的翻译,都不只是用一种翻译方法,而是多种翻译方法的综合运用。这也就意味着详细的语用规则将会准确高效的解决语篇翻译中的错误。同时,机器翻译在对于情景、情感、情绪等方面很难识别,无法准确还原在原文中真实表达的感情与意图,这就形成了机器翻译中语篇翻译的障碍。

语篇翻译还需要着重考虑的一点就是潜在错误因素。尤其是在专业性较强的翻译文本中,例如药品说明书、大型机械操作指南、法律类文件等,这些内容要求达到近百分之百的准确率。在此类情况下,一旦出现翻译上的错误,可能会伤及性命,造成无法预估的损失。

3 人工智能机器翻译的发展前景

机器翻译是目前为止快速进行大数据翻译任务的最优解决方案,起初提出机器翻译这一概念,目的在于代替人工。由于机器翻译技术在可预期的将来,也就是在人工智能的技术加码上,暂时无法达到人工翻译的水平,所以目前机器翻译作为一种技术工具来使用。这也就提出当下最关键的问题,在不久的将来,机器翻译真的会取代代替人工翻译吗?

3.1 机器翻译的现状

科技的发展带来了准确度越来越高的机器翻译,但是这一技术也对翻译行业构成了一定冲击。对机器翻译的前景存在着两种不同观点,一是机器翻译在未来可以代替人工翻译,二是机器不可能代替人工翻译。其中第一个观点在目前来说占据主流。北京外国语大学高级翻译学院教授李长栓在《机器可以取代人工翻译吗?》一文中明确表示,“要想通过机器翻译,实现自然语言的处理,恐怕还有很长的路要走;也许永远没有实现的一天。”

但是,从另一角度来看,复旦大学教授蔡基刚在《AI

时代,高校翻译专业会消亡吗》一文中提出,“随着机器翻译时代的到来,高校翻译专业必然走向消亡。”他指出:机器翻译代替人工翻译可能在未来会实现,尤其是技术类文章,机器翻译译文质量也必然超过人工译者。他表示:如今,机器翻译已达到了第三代,即神经机器翻译,原理也是通过语料库处理用户输入的信息。神经机器翻译使用更加广泛的信息数据来翻译出更加准确的译文,同时,会自动将译文根据用户的需求和语法规则重新组合成流利通顺的语句。与其他翻译软件不同点在于:神经机器翻译在人工智能的支持下,不是进行逐字翻译,而是可以一次性翻译整个句子,神经机器翻译是针对多种语言间的共同点进行翻译工作,而不是以一种标准语言进行翻译。

3.2 机器翻译的发展趋势

由此可见,人工智能背景下的机器翻译还拥有很大的进步空间。首先,针对机器翻译翻译误译、局限性等难点,未来的发展的着重点应在语境的部分,挖掘机器翻译新潜能的关键正是语境,经常被使用的逐句翻译并不是最好的翻译,因为在翻译的过程中机器无法访问该句前后相关的句子。这是个真正的限制,因为在文章中(尤其是专利类文章)上下文可能包含关键信息。其次,在翻译后期过程(即输出过程),机器学习技术也是一个可发展的重点,他自动将质量或风险评估输入进机器,避免错误,提高可靠性。最后,机器的专有化,从历时的角度出发,对机器翻译的技术和翻译数据进行专有化归类,可以进一步建设某个固定领域的研究翻译,并针对某一特定语言制定统一的语法及翻译准则。

4 结语

在人工智能技术以及现当代科学技术不断发展的背景下,机器翻译在各类技术的支持下发展的更为便捷。人工智能背景下机器翻译不仅能够降低翻译成本、大幅度提高翻译效率,同时可促进各行各业发展的新路径。基于机器翻译的难点,需要注意的是:不要夸大机器翻译的力量。在教育领域,外语和翻译的教师和学生应行动起来,让技术成为教育的翅膀。多媒体、远程教育、VR技术为学生和教师创造了极为优秀的教育环境,机器翻译的诞生也提供了更大的便捷性、以及学生自主学习的途径与方法,在人机共舞的时代,现在应更多的去了解机器,不断的开拓创新,在机器翻译的帮助下,跨语言和跨文化人才会不断提高工作效率,为中国登上世界舞台,为实现中华民族的伟大复兴贡献出自己的一份力量。

参考文献:

- [1] 王华树,王鑫.人工智能时代的翻译技术研究:应用场景、现存问题与趋势展望[J].外国语文,2021(01).
- [2] 高恂.机器翻译的语言学价值与发展困境[J].山西能源学院学报,2020(06).
- [3] 吴娜,顾毅.人工智能时代译者译后编辑能力探究[J].中国轻工教育,2020(06).

- [4] 贺潇潇 . 人工智能助力翻译语料库建设 [J]. 网络传播 ,2020(12).
- [5] 杨驰 , 杨宪泽 . 西南民族大学学报 (自然科学版) [J],2020(06).
- [6] 霍小静 . 人工智能理论的机器自动翻译系统 [J]. 微型电脑应用 ,2020(11).
- [7] 高璐璐 , 赵雯 . 机器翻译的研究综述 [J]. 中国外语 ,2020(06).
- [8] 傅宇轩 . 机器翻译译后编译研究综述 [J]. 智库时代 ,2020(05).
- [9] ZHANG JiaJun, ZONG ChengQing. Neural machine translation: Challenges, progress and future[J]. Science China Technological Sciences,2020,63(10).