

# 神经网络机器翻译问题及对策研究

——以有道、DeepL、搜狗、百度翻译为中心

荣 奕, 黄成湘

广西大学外国语学院, 广西 南宁

收稿日期: 2024年4月24日; 录用日期: 2024年6月17日; 发布日期: 2024年6月27日

## 摘 要

2013年以来,神经网络机器翻译实现了机器翻译质量的大幅提升,短短几年便成为机器翻译系统的主流模式。国内外行业巨头纷纷上线神经网络机器翻译系统,推动机器翻译实用化和商品化发展。但是,神经网络机器翻译仍然面临诸多挑战。如何有效地利用在线翻译系统,方便、快速且准确地提高翻译效率和质量,是当前机器翻译研究的重点。因此,本文以有道、DeepL、百度、搜狗四个神经网络翻译软件为研究对象,概括总结了神经机器翻译译文的特点及存在问题,并据此从机器翻译软件开发、译后编辑、译者发展等不同角度,探讨在人工智能翻译飞速发展背景下的应对策略。

## 关键词

神经网络机器翻译, 机器翻译译文特征, 问题及对策

# Research on Neural Network Machine Translation Issues and Strategies

—Focusing on Youdao, DeepL, Sogou, and Baidu Translation

Yi Rong, Chengxiang Huang

College of Foreign Languages, Guangxi University, Nanning Guangxi

Received: Apr. 24<sup>th</sup>, 2024; accepted: Jun. 17<sup>th</sup>, 2024; published: Jun. 27<sup>th</sup>, 2024

## Abstract

Since 2013, neural network machine translation has significantly improved the quality of translation and has quickly become the mainstream mode of machine translation systems. Industry giants have launched neural network machine translation systems, promoting the practical and

commercial development of machine translation. However, neural network machine translation still faces many challenges. How to effectively use online translation system to improve the efficiency and quality of translation conveniently, quickly and accurately is the focus of current machine translation research. Therefore, this paper takes Youdao, DeepL, Baidu, and Sogou as the research objects of neural network translation software, summarizes the characteristics and existing problems of neural machine translation, and discusses the coping strategies under the background of rapid development of artificial intelligence translation from different perspectives such as machine translation software development, post-translation editing and translator development.

## Keywords

Neural Network Machine Translation, Characteristics of Machine Translation Text, Issues and Strategies

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

机器翻译始于 20 世纪 30 年代, 迄今为止已有近 90 年的发展历史[1]。2013 年以来, 基于人工神经网络的深度学习技术在机器翻译领域取得突破性进展, 神经网络翻译成为在线机器翻译系统的主流模式。但是, 神经网络机器翻译仍然面临诸多挑战, 如何有效地利用在线翻译系统提高翻译效率和质量成为机器翻译研究的重点。本文基于人工评测和机器评测标准, 通过分析神经机器翻译软件日译汉译文的特点及存在问题, 从不同角度探讨日汉神经网络翻译在人工智能翻译飞速发展背景下的应对策略, 以期为用户、译者、研发者提供有益参考。

## 2. 研究设计

### 2.1. 研究问题

本研究以百度翻译、谷歌翻译、有道翻译、搜狗翻译四个神经网络翻译软件的日译汉文本为对象, 主要探讨以下 2 个问题: 1) 神经网络翻译译文有何特征和问题? 2) 在人工智能翻译飞速发展背景下, 有怎样的应对策略?

### 2.2. 研究对象

本文研究的重点为在线神经网络翻译。目前国内页面浏览量、评分较高的主流在线神经网络翻译系统有百度翻译、DeepL 翻译、有道翻译和搜狗翻译。这四个在线 NMT (Neural Machine Translation, 神经网络机器翻译) 网站均采用神经机器翻译技术、可靠性高、实验数据可重复。鉴于此, 笔者选择了这四种在线神经网络翻译软件进行评估。

### 2.3. 评测语料来源

考虑到本研究的研究对象及目的, 本稿选择了翻译教材《现代日汉翻译教程》作为机器翻译评估的语料。该书用词简洁平实、准确严谨, 还涵盖各类文体、题材、风格、翻译重难点, 作为评估语料能够

涵盖机器翻译各种可能出现的误译情况。并且, 本书的对照译文以 3~5 句的语段组成, 相较于大篇段落翻译结果更优良, 相较于纯单词、纯句子文本能够更好评估翻译风格、上下文接续情况。因此, 可为机器翻译质量评估提供权威、丰富、立体的语料支撑。

## 2.4. 评价参数与评测标准

“译文质量是机器翻译评价最重要、最核心、最关键的指标” [2]。机器翻译译文评估有人工评测和自动评价两种方式。自动评价方法主要通过测量机器翻译译文与高质量人工译文的相似度, 善于处理大批量译文, 具有较强客观性, 而人工评价则会考虑翻译准确性、忠实度、流畅度等因素, 全面考察译文结果, 主观性较强。因为人工评测和自动评测各有长短, 而且这种长短又具有互补性, 因此本文将二者结合起来, 进行了全面、科学, 可信的评测。

在自动评测方面, 本文选取了 BLUE score 评估模型。BLUE score 方便、快速、结果有很高的参考价值, 本文选用的 Blue-2 值不仅能对比机器译文和人工译文的相似度, 还可以评估出译文流畅度。而人工评测方面, 本文则选取了《中华人民共和国国家标准 GB/T1968-005 翻译服务译文质量要求》, 该规范适用范围广, 且具有相当细致的评估标准, 是中国最权威的人工评测细则。

## 2.5. 研究过程

### 2.5.1. 语料筛选

为了能够较为全面地评估机器翻译, 检验机器翻译对语料的处理情况, 笔者结合先行研究中神经网络翻译系统可能出现的错误类型和翻译教材中的翻译难点分类, 抽取了 400 句中对比译文作为语料。这些语料难度适中, 无现成译本, 包括记叙文、说明文、议论文等文体, 日常问候、风俗习惯、科学技术等主题, 还包括简单句、复合句等各类句式, 是一个自然、真实、多样化的语料库。

### 2.5.2. 构建数据集

进行人工评测和机器评测还需要构建数据集。笔者将日语原文分批复制粘贴到四个神经机器翻译系统中, 在 Microsoft Excel 中记录翻译结果, 以便数据分析。实验日期为 2024 年 1 月 28 日, 为了保证翻译速度和质量, 每次的机器翻译量语段为中心, 2~3 句, 约为 200~300 字左右, 其间没有人工干预, 翻译结果概况见图 1。

人工	有道	DEEPL	百度	搜狗	原文
1 参议院决定废除用特殊符号记录议事的 2 2 3 4 5 6 7 8 9 10 11 12	参议院决定废除用特殊符号记录议事的 2 2 3 4 5 6 7 8 9 10 11 12	参议院已决定取消速记, 即用特殊符号 (速记符号) 记录会议过程。 今后, 速记员将直接在电脑上打字, 同 时观看会议的监控屏幕。 众议院今后很可能效仿参议院, 而议 会速记这一自 1890 年第一届帝国国会 以来一直延续的传统将在技术革新之前 结束其作用。	参议院决定废除用特殊符号 (速记文 字) 记录议事的速记。 今后是速记员一边看会议的监控画面 一边直接输入电脑。 众议院在未来也有可能追随参议院。 1890 年第一届帝国国会以来一直延续 的传统国会速记, 也将与技术革新前结 束其作用。	参议院确定了废除用特殊符号(速记文 字)记录议事的速记的方针。 今后的方向是速记员一边看会议的监 控画面, 一边直接输入电脑。 将来众议院也有可能追随参议院。 1890年第一届帝国国会以来持续传 统国会速记也将与技术革新前结束作 用。	参議院は特殊な符号 (速記文字) で議事を書き取る速 記を廃止する方針を固めた。 今後は速記者が会議のモニター画面を見ながら、パソ コンに直接入力する方向だ。 衆院も将来的には参院に追随する可能性が高く、1890 年の第一回帝国議会以来続く伝統ある国会速記も、技 術革新前に役割を終えることになる。

Figure 1. Translation results from 4 machine translation systems  
图 1. 4 个机器翻译系统的翻译结果

之后又对机器翻译后的文本进行预览, 并对机器翻译出现较大问题的部分进行标注。两个数据集的基本情况如表 1 所示。

**Table 1.** Basic information about experimental and control data sets  
**表 1.** 实验数据集和对照数据集的基本情况

实验数据集(机器译文)				对照数据集(人工译文)	
搜狗译文字数	百度译文字数	有道译文字数	DeepL 译文字数	译文句子数	译文字数
12,079	12,066	12,066	12,184	400	12,251

### 2.5.3. 进行人工评测

根据《中华人民共和国国家标准 GB/T1968-005 翻译服务译文质量要求》, 评估《现代日汉翻译教程》译文需要依照译文综合差错率这一指标。其计算公式如下:

$$\text{综合差错率} = K C_A \frac{c_I D_I + c_{II} D_{II} + c_{III} D_{III} + c_{IV} D_{IV}}{W} \times 100\%$$

计算这一指标最重要的是明确详细地划分译文 I、II、III、IV 类差错的类别, 按照错误分类统计错误次数(即为  $D_I$ 、 $D_{II}$ 、 $D_{III}$ 、 $D_{IV}$ )。其具体分类标准为: 第 I 类: 对原文理解和译文表述存在核心语义差错或关键词(数字)漏译、错译。第 II 类: 一般语义差错, 非关键词(数字)、句段漏译、错译、译文表述存在用词、语法错误或表述含混。第 III 类: 专业术语不准确, 不统一、不符合标准或惯例, 或专用名词错译。第 IV 类: 计量单位、符号、缩略语等未按规(约)定译法。

除此之外, 公式中的  $C_I$ 、 $C_{II}$ 、 $C_{III}$ 、 $C_{IV}$  为 I、II、III、IV 类差错的系数, 规定取值为:  $C_I = 3$ 、 $C_{II} = 1$ 、 $C_{III} = 0.5$ 、 $C_{IV} = 0.25$ 。译文目的系数  $C_A$  则按照译文文本根据不同使用目的划分 4 类, 1 类所需要的精度最高。由于本研究涵盖领域、文体丰富, 具有一定的专业性, 因而对机器翻译按照 2 类使用目的进行评估,  $C_A = 0.75$ 。本文还立足于语料文本、在线 NMT 网站的翻译质量, 结合个人翻译经验确定综合难度系数  $K$  的取值为 1。

综上所述, 可以得到四种机器翻译的译文综合差错率如表 2。

**Table 2.** Comprehensive error rate of translation  
**表 2.** 译文综合差错率

数据组	D1	D2	D3	D4	W	结果
有道	14	53	20	0	11,968	0.00658005
DeepL	17	48	27	0	12,184	0.00692507
百度	17	102	21	1	12,066	0.01017839
搜狗	23	85	16	0	12,079	0.01005878

### 2.5.4. 计算 BLEU 值

笔者在实际操作中, 在 Python 中直接调用 NLTK (Natural Language Toolkit) 的 `nltk.translate.bleu_score` 工具包进行计算。NLTK 中提供了不同的函数和选项来计算 BLEU 值, 笔者使用“`sentence_bleu`”函数来计算 BLEU 值。该函数接受两个参数, 第一个参数是一个列表, 包含了参考翻译的句子或词组, 第二个参数是待评估的机器翻译输出句子或词组, 其他各项参数均选用默认值, Bleu-2 值数据见表 3。

**Table 3.** Neural machine translation software Bleu-2 value data

**表 3.** 神经机器翻译软件 Bleu-2 值数据

数据组	有道	DeepL	百度	搜狗
有效数据	400	400	400	400
平均数	0.429	0.323	0.415	0.413
中位数	0.418	0.321	0.414	0.408
标准偏差	0.157	0.145	0.161	0.167

### 3. 机器误译问题分析

#### 3.1. 词汇翻译结果对比分析

词汇是组成句子的基本元素, 词汇翻译对译文的质量有着至关重要的影响[3]。在词汇翻译方面, 神经网络机器翻译会出现的错误一般有几种, 分别为: ① 词语直接照搬原文; ② 语义错误; ③ 词语漏译; ④ 专有名词翻译错误; ⑤ 惯用语、俗语的错译; ⑥ 多语义词汇选择语义不当; ⑦ 无法进行语义的引申。下面是一些例子:

例 1: 空き巣というと、夜間に留守宅を狙って忍び込むというイメージが強い。

译文: 说起空巢, 给人的印象是夜间瞄准不在家的房子偷偷潜入。

分析: “空き巣”在日语中是趁人不在入室盗窃的意思。中文中则是指小鸟离巢, 现在被引申为子女离开后家庭空寂。机器翻译直接将其翻译成“空巢”, 直接沿用了原文中的词汇, 无法传达出日语本来的意思。

例 2: 「あれつ。財布がない?!」頭に血がのぼって、冷や汗がたらたらと流れるところだろうが、困ったときほど落ち着いたほうがよい。

译文: “咦? 没有钱包吗?!” 虽然是头上冒着血, 流着冷汗的地方, 但越是困难的时候越要冷静下来。

分析: 頭に血がのぼって被机器译文译成了头上冒着血, 这显然是不符合原文意思的, 属于语义错误。

例 3: 彼女にとって指輪より、貯金通帳よりも子供が大切なのだ。

译文: 对她来说, 孩子比戒指更重要。

分析: 本句漏译了存折这个词, 虽然不影响理解, 但是可以看出机器翻译存在漏译问题。

例 4: クジャクのオスが美しいのは性淘汰の結果といわれる。

译文: 雄性孔雀之所以美丽, 据说是性别淘汰的结果。

分析: 日语中的“性淘汰”翻译成中文为性选择, “性别淘汰”这一译文显然是直译, 属于专业术语错误。

例 5: その「顔」という言葉、けっこう広く使われています。「顔を貸す」「顔を効かせる」とも言いますし、「顔を立てる」「顔をつぶす」「顔に泥を塗る」「顔向けができない」という言い方もありますし、「顔パス」という用語まで使われています。

译文: “脸”这个词使用得相当广泛。有“把脸借给你”、“让脸发挥作用”的说法, 也有“给你面子”、“毁容”、“往脸上抹黑”、“没脸见人”的说法, 甚至还有“脸路径”的用语。

分析: 机器对俗语、惯用语翻译需要借助于已经建好的双语对齐的俗语库, 不然很难翻译出谚语、

名言、歇后语等具有强烈历史文化内涵的语句。

例 6: 猫がある距離まで接近して、いよいよ危うくなると、ネズミははじめて逃げる。

译文: 猫接近某个距离, 终于危险了, 老鼠第一次逃跑。

分析: 日文中的“はじめて”有开始、第一次、首先等意思, 本文主要内容描述的是老鼠遇见猫后会先静止不动, 然后伺机逃跑。因此, “第一次”在文段中是不符合语境的。

例 7: 古来赤い色は魔を祓う力があると信じられていて、中国でも玄関に赤い札を貼って魔除けにする。

译文: 自古以来人们就相信红色具有辟邪的力量, 中国也会在玄关上贴红色的牌子来辟邪。

分析: 在这一句话中, 我们可以很轻易的推测出红色的牌子指的是中国的对联, 但是机器并不能识别出这一点。机器翻译在翻译隐喻或口语化表达时通常会逐字翻译, 不能符合源文本的原意。

### 3.2. 句法翻译结果的对比分析

在线翻译系统面临的主要难题是如何正确理解各句子要素, 然后根据各要素之间的关系将各要素的翻译结果进行组合。句法类错误主要表现在以下 5 个方面: ① 断句错误; ② 句子层次结构错误; ③ 时态错误; ④ 主语错误; ⑤ 语序错误; ⑥ 逻辑关系混乱。

例 8: 隔てるものは川であれ、海であれ、その向こうには何かがあるに違いないのだから、どうしても行つて見ずに入れないのが人間だ。

译文: 不管是河还是海, 把我们分隔开来, 对岸一定有什么, 是人类忍不住要去看一看。

分析: 神经机器翻译系统会将长句从中截断, 从而造成断句错误。译文将“どうしても行つて見ずに入れないのが人間だ”之间进行了切分, 出现了断句错误。

例 9: また、風呂に入るのもいい。風呂で汗をかき、汗と共にアルコールを流すのである。

译文: 另外, 洗澡也不错。在浴室里出汗, 和汗一起流酒精。

分析: 其中, 酒精会随着汗水被冲走被翻译为流出酒精, 句子层次结构十分混乱, 令人感觉不知所云。

例 10: 「イチエンじゃね。せめて五円。ご縁があるようにとね。」「値上がりですか。」

译文: 一分钱也不行。至少五日元祝你好运。要涨价了吗?

分析: 日汉两种语言采用不同的时态表现形式, 原文询问是否已经涨价, 但机器翻译把原本的过去时误译成了将来时, 出现了明显的时态错误。

例 11: 泥棒口惜しい紛れに「ああ、知つてたとも。」「知つてて、入るとはいよいよもつと間抜けな泥棒だ。おまえはここで何か盗めるかと思つたのか。」

译文: 小偷可惜地说: “啊, 你知道的。” “明明知道, 进来的竟然是更笨的小偷。你以为在这里能偷到什么吗?!”

分析: 日语中有很多省略主语的表达, 日译中时需要完善句子的语法结构。本译文虽然补充了主语, 但却错误地补充成“你”, 没有表达出原文的意思。

例 12: 犬はしつぽをふることで親愛の情を表すが、猫はしつぽを狩りに役立つ。

译文: 狗摇尾巴表示亲爱的感情, 而猫则帮助猎取尾巴。

分析: 该译文把しつぽを狩り当作一个整体, 词与词之间没有逻辑顺序, 从而导致造成整句逻辑关

系错误。

例 13: このことを、サービスの低下だと非難した声は聞かれたようです。どこの国でもやっていることですから。

译文: 似乎有人批评这是在降低服务质量。但各国都是这样做的。

分析: 日本取消了站内广播, 因此遭到了批评, 理由是每个国家都会有站内广播。但译文很明显翻译错了关联词, 将因果关系错译成了转折关系, 短句之间的逻辑关系混乱。

## 4. 日汉机器翻译的局限性分析

### 4.1. 主要发现

通过对四种在线机器翻译系统在日译汉翻译上的性能对比分析, 总结出以下问题:

(1) 机器翻译在总体翻译错误数上存在差异, 但各个错误类别在总体中的比例几乎一致, 机器翻译问题具有共性。

(2) 相较于句子的翻译, 词汇翻译错误率低, 机器翻译对于词汇翻译的表现较好。

(3) 机器翻译在句法翻译方面, 经常出现句子分割和词序错误, 导致译文比原文更加模糊且复杂, 翻译内容不成句子。

(4) 机器翻译整体来看流畅性较差, 无法联系上下文语境, 只是机械性地以句子为单位进行翻译。

### 4.2. 误译原因

#### 4.2.1. 计算而非理解

NMT 系统不了解源语和目的语的语法等信息, 也不具备理解文本所必需的常识和专业知识, 仅通过计算单词和句型的概率来完成翻译任务, 选择那些获得最高概率的单词和句型, 并将它们排列成流畅的字符串。当源句子和目标句子之间不存在一对一的对等关系时, NMT 系统很可能产生过于字面化、尴尬和不地道的翻译。

#### 4.2.2. 语料库有限

NMT 是在大型语料库上训练的, 输入的句子越类似于语料库示例, 翻译质量越高。但是, 语言是不断发展的, 世界上大多数语言对都没有双语平行语料库, 也不可能存在涵盖所有表达的语料库[4]。因此可以很容易地在 NMT 输出中检测到源句子的措辞和短语的痕迹, 例如上文提到的日语词汇直接借用到译文中的问题。

#### 4.2.3. 不可靠的 LSTM 和注意机制

NMT 系统的体系结构由 LSTM (Long Short-Term Memory, 长短期记忆网络)和注意机制组成。LSTM 存储进入 NMT 模型编码器的历史信息。注意机制则将解码器的注意力引导到源句子中相关信息最集中的位置。二者的引入提高了 NMT 系统翻译长句的能力, 然而, 在实际应用中, LSTM 和注意机制无法把握语序, 习惯了中文和英语语序, 很难理解把谓语放在句末的日语用言, 所以语序问题频出。

## 5. 机器翻译优化策略

依据上述结论, 我们可以从以下三方面对机器翻译译文所出现的问题提出对策。

### 5.1. 对机器翻译开发的建议

构建完善的语料库。扩充机器翻译系统所依赖的汉日词典; 由于大多在线翻译主要是以英语作为源

语言或目的语言的[5], 因此可以基于英语这一中间语料库提高中日对译的质量; 建立不同领域术语库, 得到准确度、匹配度更高的译文; 基于上下文对语料库进行词类扩展, 帮助机器翻译系统用少量的规则正确地分析和处理单词属性。

调整语序和句子结构。引入语言特征, 例如主谓宾的语法规则、修饰语的位置等。帮助模型更好地理解原始句子的结构, 在翻译时保持语义和语法的一致性; 引入后编辑环节, 让人工编辑者对翻译结果进行反馈, 帮助机器翻译系统更准确地捕捉语序结构, 不断迭代。

引入新技术。结合神经网络机器翻译、基于规则的机器翻译和统计机器翻译各自的独特优势, 弥补单一模型的不足, 显著提高机器翻译的质量。同时还可以引入新的人工智能技术。例如2023年初由OpenAI开发的人工智能聊天机器人ChatGPT, 与过往机器翻译相比, 具有更明了的语法逻辑和连接关系。

## 5.2. 对翻译人员的建议

未来的翻译软件将更加注重人工智能和人类智慧的结合, 以实现更好的翻译结果。译员在翻译过程中应充分发挥主观能动性, 从多方面改进翻译的质量。作为翻译专业的人员, 我们需要加强专业知识与技能, 培养跨文化沟通能力, 掌握专业翻译工具与技术, 尝试与人工智能协同合作。

## 5.3. 对译后编辑的运用

在翻译数字化和产业化时代, 基于机器翻译的译后编辑已成为翻译行业发展的新常态[6]。神经机器翻译需要人工辅助, 才能产生高质量、人性化的翻译。这就需要译后编辑。广义上译后编辑包括译前编辑、译中编辑、译后编辑等一切机器翻译行为之外的工作。

译前编辑阶段, 可以收集权威术语资源、统一术语和风格、优化原文的结构和逻辑, 消除歧义, 以便于机器翻译系统处理, 提高输出质量。译中编辑阶段则需要监控机器翻译系统的输出, 及时发现并纠正存在的错误。在译后编辑阶段, 可以使用自然语言处理工具和人工智能技术, 对翻译结果进行自动化的质量评估和修正。

## 6. 结语

从本文的研究结果来看, 目前现有的神经机器翻译系统还不能实现高质量的日译汉翻译。本研究仅涉及对单词和句子的分析, 所提到的机器翻译中出现的问题还只是冰山一角, 语篇方面的问题仍有待探讨。

通过本次的翻译实践, 笔者更加清晰地看到了机器翻译的优势和不足, 短时间内机器翻译无法取代人工翻译。但是也可以看出, 随着科技的日益进步, 人工智能翻译和人工翻译的差距在慢慢减少, 作为翻译专业的人员, 我们也需要作出一些应对策略。需要不断加强专业知识与技能, 培养跨文化沟通能力, 掌握专业翻译工具与技术, 尝试与人工智能协同合作。

## 参考文献

- [1] 胡开宝, 李翼. 机器翻译特征及其与人工翻译关系的研究[J]. 中国翻译, 2016, 37(5): 10-14.
- [2] 张政, 王贵明. 论机器翻译系统的评价体系[J]. 北京理工大学学报, 2008, 10(2): 112-117.
- [3] 罗季美, 李梅. 机器翻译译文错误分析[J]. 中国翻译, 2012, 33(5): 84-89.
- [4] 张政. 国外机器翻译理论概述[J]. 外语研究, 2003(6): 44-49.
- [5] 盛春. 汉日在线翻译方法及问题研究[J]. 科技信息, 2010(35): 315, 297.
- [6] 肖志清, 金鸣. 国内机器翻译译后编辑研究: 现状、问题与展望——基于中国知网数据库(1995-2022年)[J]. 民族翻译, 2022(4): 68-76.