

2019 年 中国机器翻译行业研究报告

行业走势图



TMT 团队

卢佩珊 分析师

邮箱: cs@leadleo.com

相关热点报告

人工智能系列深度研究——科创板上市，人工智能行业借势迎来发展

人工智能系列深度研究——沙利文助力 2019 全球人工智能健康峰会成功举办，共同推动全球人工智能技术和产业发展

人工智能系列深度研究——谷歌上线“猜画小歌”，智能技术与人们生活越来越近

人工智能系列深度研究——产品智能化浪潮开启，智能家居风口已至

报告摘要

深度学习的引入变革了机器翻译领域，大幅提升机器翻译质量，使机器翻译商业化成为可能。机器翻译在深度学习作用下重获新生，随着翻译结果准确度的不断提升，机器翻译产品有望拓展更广阔的市场。

热点一：翻译市场发展潜力大，人工翻译供给不足，需机器翻译补充

中国的语言翻译服务需求仍然高涨，然而现阶段的翻译人才供给与翻译人才需求间产生较大差距，高素质的翻译人才仍缺乏，通晓小语种的翻译人员更为稀少。在翻译人才供给满足不了翻译需求的情况下，机器翻译能作为较好的补充手段。人工翻译在机器翻译的辅助下，可节省大量人工成本，提高翻译效率。

热点二：机器翻译产品逐渐实用化，拥有离线翻译等强大功能

机器翻译作为一项技术，最终将以产品的形式落地，实现商业化，产品的实用性能决定其使用价值。在旅途过程中的用户，会较为频繁地身处于没有网络覆盖或网络信号较差的地域，如飞机、地铁以及地理位置较为偏远的景区，单纯的在线翻译服务无法很好地满足旅游用户在不同场景下的翻译需求，不支持离线翻译的机器翻译产品并未带给用户良好的使用体验，因此离线翻译功能成为各大机器翻译厂商竞争的突破口。

热点三：中国机器翻译厂商多兼具软、硬一体能力

由于现阶段机器翻译软件的盈利能力较低，硬件产品是机器翻译厂商的主要收入来源，因此多数中国的机器翻译厂商集成了软件和硬件生产能力，成为软硬一体的机器翻译供应商。机器翻译行业的市场规模有限，可容纳的企业数量有限，未来行业将呈现寡头竞争局面，没有同时具备软硬件优势的企业难以抢占市场份额。

目录

1	方法论.....	5
1.1	方法论.....	5
1.2	名词解释.....	6
2	中国机器翻译行业市场综述.....	7
2.1	机器翻译定义与特征.....	7
2.2	机器翻译方法.....	7
2.3	中国机器翻译发展历程.....	9
2.4	中国机器翻译市场规模.....	12
2.5	中国机器翻译产业链分析.....	13
2.5.1	产业链上游.....	13
2.5.2	产业链中游.....	15
2.5.3	产业链下游.....	15
3	中国机器翻译行业驱动因素.....	16
3.1	技术进步提升翻译质量.....	16
3.2	翻译人才不足需机器补充.....	16
3.3	需求带动机器翻译发展.....	17
4	中国机器翻译行业制约因素.....	18
4.1	机器翻译表达不精准.....	18

4.2	机器翻译训练数据不足.....	19
4.3	翻译模型不具通用性.....	20
5	中国机器翻译行业相关政策法规.....	21
6	中国机器翻译行业发展趋势.....	23
6.1	翻译语种更加丰富.....	23
6.2	机器翻译逐渐实用化.....	23
6.3	产品逐渐支持离线翻译.....	24
7	中国机器翻译行业竞争格局.....	25
7.1	中国机器翻译行业竞争现状分析.....	25
7.2	中国机器翻译行业典型企业分析.....	26
7.2.1	新译信息科技（深圳）有限公司.....	26
7.2.2	北京分音塔科技有限公司.....	27
7.2.3	北京爱特曼科技有限公司.....	28

图表目录

图 2-1 中国机器翻译市场规模 (按营收计), 2019-2023 年预测	13
图 2-2 机器翻译产业链	13
图 5-1 中国机器翻译行业相关政策	22

1 方法论

1.1 方法论

头豹研究院布局中国市场，深入研究 10 大行业，54 个垂直行业的市场变化，已经积累了近 50 万行业研究样本，完成近 10,000 多个独立的研究咨询项目。

- ✓ 研究院依托中国活跃的经济环境，从人工智能、机器学习、自然语言处理等领域着手，研究内容覆盖整个行业的发展周期，伴随着行业中企业的创立，发展，扩张，到企业走向上市及上市后的成熟期，研究院的各行业研究员探索和评估行业中多变的产业模式，企业的商业模式和运营模式，以专业的视野解读行业的沿革。
- ✓ 研究院融合传统与新型的研究方法，采用自主研发的算法，结合行业交叉的大数据，以多元化的调研方法，挖掘定量数据背后的逻辑，分析定性内容背后的观点，客观和真实地阐述行业的现状，前瞻性地预测行业未来的发展趋势，在研究院的每一份研究报告中，完整地呈现行业的过去，现在和未来。
- ✓ 研究院秉承匠心研究，砥砺前行的宗旨，从战略的角度分析行业，从执行的层面阅读行业，为每一个行业的报告阅读者提供值得品鉴的研究报告。
- ✓ 头豹研究院本次研究于 2019 年 07 完成。

1.2 名词解释

- **API (Application Programming Interface):** 应用程序编程接口, 提供特定功能的软件。
- **BLEU (Bilingual Evaluation Understudy):** 评价翻译质量的指标, 用以标注机器译文与参考译文的相似程度。
- **颗粒度:** 指内容可显示的详细和清晰程度。颗粒度越细, 表示细节越详尽, 越有助于了解事情的全貌。颗粒度越粗, 表示细节越少, 多为抽象概括。
- **指代消解:** 系统自主在篇章中确定显性代词指向哪个名词短语的问题。
- **自适应:** 系统根据处理数据的数据特征自动调整处理方法、处理顺序、处理参数、边界条件或约束条件, 使其与所处理数据的统计分布特征、结构特征相适应, 以取得最佳的处理效果的过程。

2 中国机器翻译行业市场综述

2.1 机器翻译定义与特征

机器翻译指计算机程序将一种书写形式或声音形式的自然语言翻译成另一种书写形式或声音形式的自然语言。

机器翻译形式包括词典翻译、计算机辅助翻译和文本或语音的句子以及段落翻译。词典翻译提供基本的词语翻译服务,将一种自然语言的字词或词组翻译成另一种自然语言的字词或词组。计算机辅助翻译主要利用计算机的自动记忆功能将用户翻译过的资料转换为可重复使用的翻译内容库,协助人工翻译,为用户以后的翻译行为提供便利。文本或语音的句子以及段落翻译是计算机按照设定的规则自动将一种自然语言的文本或语音的句子及段落翻译成另一种自然语言的文本或语音的句子及段落。

机器翻译相比人工翻译,具有以下优点:

(1) 成本低

机器翻译的成本较人工翻译成本低。机器翻译过程无需人工操作,单次软件开发即可多次使用,可减少专业翻译人员的雇佣,节省翻译成本。

(2) 易于把控

机器翻译按照规则执行,流程相对固定,翻译时间能较为精准地估算。

(3) 翻译速度快

机器翻译程序的运行速度比人工翻译速度快,用户可更快速地获得翻译结果,减少翻译等待时间。

2.2 机器翻译方法

自机器翻译诞生以来,形成了两大类翻译方法,包括理性主义方法和经验主义方法。

“理性主义”以人工生成语言学为基础，表现为基于规则的机器翻译；“经验主义”以数据驱动为基础，表现为基于语料库的机器翻译。

(1) 理性主义方法

理性主义的翻译方法由语言学专家通过编撰规则的方式，将不同自然语言间的转换规律生成算法，计算机程序依据算法规则进行语言翻译。理性主义方法高度依赖专家，要求专家具备一定的语言学知识和翻译知识，深刻了解源语言和目标语言，还要求专家熟练掌握计算机的相关操作技能，而同时具备翻译知识、语言学知识和计算机知识的综合人才较少，熟悉小语种的人才更是匮乏，导致理性主义的机器翻译方法发展缓慢。

基于规则的机器翻译方法认为任何一种语言里的无限的句子均可由有限的规则推导，能够充分利用语言学家总结的语言规律为计算机翻译程序提供指引。基于规则的翻译方法直观，系统根据人为设定的不同颗粒度的规则进行翻译操作。大颗粒度的规则概括能力强，具有较好的系统适应性，不依赖于具体的机器训练语料。细颗粒度的规则具有精细的描述能力，可处理结构复杂，语义深刻的翻译内容。但基于规则的机器翻译方法主观因素较重，规则的质量和效果依赖于语言学家的知识和经验，受专家自身因素影响过大。规则一般仅从规范文本的语言现象中总结，覆盖性较差，对于弱规范或不规范的口语语言现象和网络用语处理能力较差。人为总结的规则不够客观，新规则与已有规则易发兼容性问题。规则获取的人工成本较高，规则库的建立和调试过程漫长，维护大规模规则比较困难，导致基于规则的理性主义机器翻译方法难以处理大规模真实文本的翻译工作。

(2) 经验主义方法

20 世纪 90 年代初，大量的语料库和机器学习方法开始兴起，机器翻译方法逐渐由理性主义转向经验主义。经验主义方法使用语料库作为翻译知识的来源，通过语料库训练让系统自主学习并产生翻译规则，无需人工编写规则。但经验主义方法严重依赖语料库，若语料

库数据不全面，系统总结的机器翻译规则亦不全面，通过语料库训练的方法不易获得大颗粒度的具有高概括性的翻译规则。

基于语料库的经验主义机器翻译方法包括**基于实例的翻译方法**和**基于统计模型的翻译方法**。①**基于实例的翻译方法**从语料库中检索出与源语言文本相似的实例（通常句子且可能有多个），并对实例对应的译文进行适当的替换、删除和插入等一系列变形操作，得到最终的目标译文。基于实例的机器翻译方法无需对句子进行复杂的语言分析，可直接利用已有的翻译实例库进行译文片段重组，但此情况通常要求实例库中存在与待翻译内容高度相似的实例。目前，业内普遍缺少大规模的双语对齐语料库，因此基于实例的机器翻译系统匹配率并不高。②**基于统计模型的翻译方法**从大量的语料库中获取翻译知识，为翻译内容建立统计模型，把翻译问题理解为搜索问题，认为任何一个目的语句子都有可能是任何一个源语的译文，机器翻译的任务即是从语料库中找到最大概率可匹配源语的译文句子。基于统计模型的机器翻译中的原规则结构复杂，对语料库的依赖性强，即使现在可用大规模语料库实现机器自主学习训练，但语料库数据仍显严重不足。

基于规则的理性主义翻译方法和基于语料库的经验主义方法各有优劣，目前尚未有任何一种方法能实现完美的机器翻译。近年来，机器翻译研究趋于将基于规则的翻译方法和基于语料库的翻译方法相结合，兼容并包，博采众长。

2.3 中国机器翻译发展历程

中国的机器翻译研究始于 20 世纪 50 年代，中科院语言研究所与电脑科技研究所合作开展中俄语言机器翻译研究，成功实现了九种复杂句式的中俄语言互译。但受国际外交局势以及计算机发展水平的限制，当时的机器翻译研究尤其是汉英机器翻译研究并未得到深入发展，甚至陷入停滞。直到 20 世纪 80 年代末，中国的机器翻译研究再度复苏，随着计算机技术的发展和国际交流需求的增长，机器翻译的使用需求日渐提升，中国多个科研院校与国

际研究机构开展合作，相继推出“863 智能英-汉翻译系统”、“译星”、“通译”等机器翻译软件，迅速带动机器翻译进入公众视线。

21 世纪以来，深度学习算法和神经网络架构成为机器翻译领域的翻译范式，在诸多语种及应用场景中的翻译质量超越了传统机器翻译技术的翻译质量，受到各大机器翻译研究商的追捧。**百度**的基于神经网络的机器翻译，在翻译知识获取、翻译模型、多语种翻译技术等方面有重大突破，将深度学习算法应用于多种主流翻译模型中，翻译质量有明显提升，翻译结果在国际上获得广泛认可。**腾讯**的机器翻译结合神经网络机器翻译和统计机器翻译的优点，对源语言文本进行深入理解，使翻译效果更为准确，同时支持语音翻译、图片翻译、语种识别等多种场景，大幅降低传统文本翻译的读写成本。**科大讯飞**基于深度神经网络算法优化机器翻译，并于 2014 年在国际口语翻译大赛 IWLST 上获得中英以及英中翻译的双项冠军。中国的机器翻译行业发展正逐步深入发展，未来的机器翻译应用将更具深度与广度。

前哨 2020 | 科技特训营

掌握创新武器 抓住科技红利

Insights into Tech and the Future

直播时间
每周四20:00-21:00

全年50次直播课程
+私享群互动

随报随听

王煜全

海银资本创始合伙人
得到《全球创新260讲》主理人



扫码报名

微信咨询: InnovationmapSM

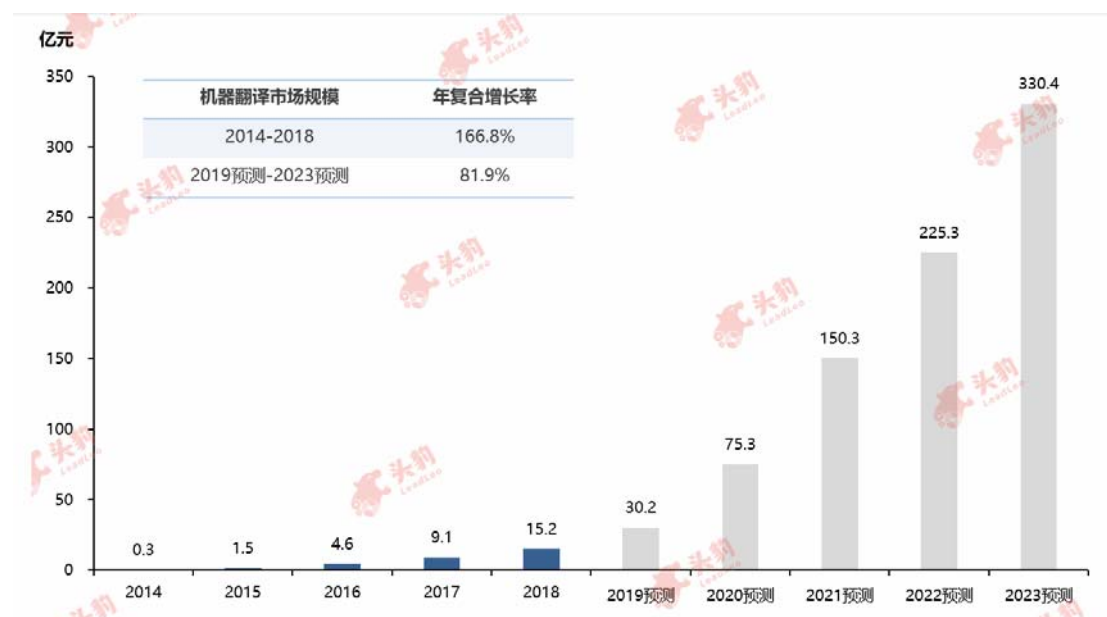
电话咨询: 157-1284-6605

2.4 中国机器翻译市场规模

机器翻译涉及软件（如在线翻译软件、计算机辅助翻译软件和翻译 API、SDK 等）和硬件（翻译机、蓝牙翻译耳机和翻译手机等）产品。目前机器翻译软件产品在中国市场上尚未产生明显的收入，众多翻译软件免费供用户使用，部分翻译 API、SDK 在特定情况下向用户收取少量的费用（如百度通用翻译 API 高级版和尊享版在翻译字符超 200 万字的情况下，按 49 元/百万字符向用户收取相应的使用费），但此类软件收入占比甚少。

中国机器翻译行业的营收主要由机器翻译硬件产品产生，目前主要有翻译机、蓝牙翻译耳机以及翻译手机三种硬件产品。在机器翻译硬件产品中，中国的蓝牙翻译耳机以及翻译手机并不普及，产生的营收有限，而翻译机产生的营收额占市场营收总额的比例超 85%。翻译机产品成为各大机器翻译厂商抢占市场红利的最佳消费级电子硬件产品，如科大讯飞、腾讯、搜狗、网易等厂商纷纷推出智能翻译机产品，当前的智能翻译机产品功能不断增强，支持多种语言甚至是方言的即时互译以及拍摄翻译等。随着智能翻译机功能的增强以及用户体验提升，智能翻译机的应用有望更进一步普及，带动机器翻译市场规模的增长。2018 年，中国翻译机销售量近 100 万台，产生的营收额为 15.2 亿元人民币，预计 2023 年，翻译机的销量将超 2,000 万台，届时将带动中国机器翻译市场规模大幅提升，实现中国机器翻译市场 2019-2023 年 81.9% 的年均复合增长。

图 2-1 中国机器翻译市场规模 (按营收计), 2014-2023 年预测

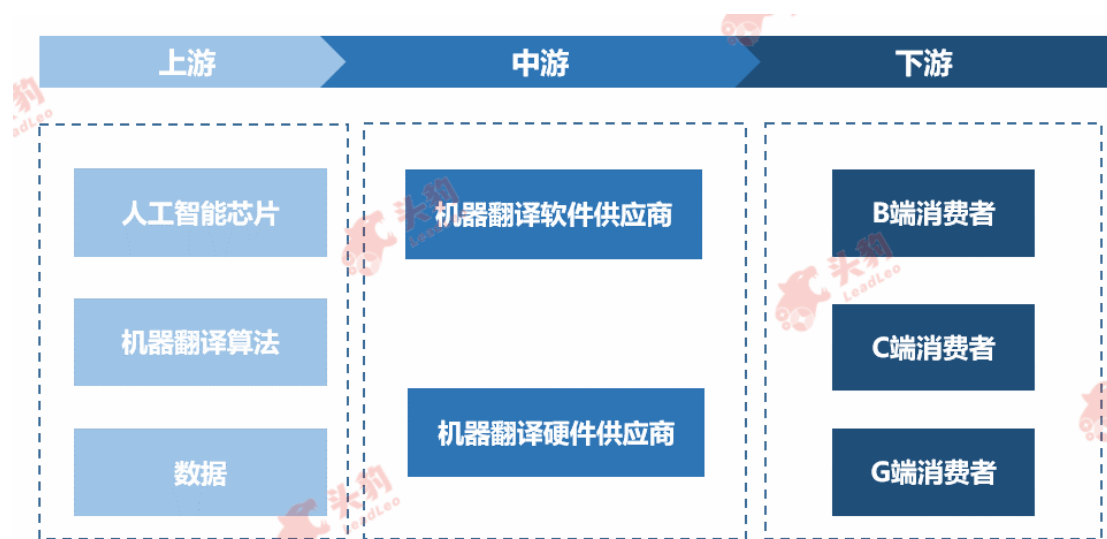


来源: 头豹研究院编辑整理

2.5 中国机器翻译产业链分析

机器翻译产业链由上游的人工智能芯片供应商、机器翻译算法供应商和数据供应商, 中游的机器翻译软件供应商和硬件供应商以及下游的 B 端、C 端和 G 端消费者组成。

图 2-2 机器翻译产业链



来源: 头豹研究院编辑整理

2.5.1 产业链上游

机器翻译产业链上游由人工智能芯片供应商、机器翻译算法供应商以及数据供应商组成。现阶段, 中国机器翻译行业并未诞生专门的数据供应商, 数据并不对外出售, 多数机器翻译

算法供应商以及数据供应商与机器翻译软件供应商集成一体,厂商通过自主搜集到的数据运用于训练自主研发的机器翻译算法,并将算法内嵌于自主研发的机器翻译软件中。部分缺乏机器翻译算法能力的厂商,通过调用百度的机器翻译 API 来提供机器翻译软件和硬件生产基础。

人工智能芯片涉及 GPU、FPGA 和 ASIC,目前的人工智能芯片市场主要由非定制化芯片 GPU 和 FPGA 主导,但非定制化芯片在计算性能、功耗、成本等方面存在明显不足,市场需求必将推动人工智能芯片从通用向专用发展,由通用的 GPU,半通用的 FPGA,向专用化的 ASIC 发展。

通用型 CPU 与 GPU 在集成电路产业中最早发展成熟并被广泛应用,但在人工智能时代,CPU 的串行逻辑架构无法适应大数据量运算,单独的 CPU 解决方案不再出现于人工智能应用中。GPU 天然的并行逻辑架构可以轻易地适应大数据量运算环境,因此被最早应用于人工智能市场。但 GPU 高度缺乏灵活性,难以为各种特定的人工智能场景定制性能与算力。

半通用的 FPGA 可与硬件结合紧密,在编程、底层配置和构建层面具有较高灵活度,在专用芯片发展尚未成熟的阶段能较好缓解 GPU 在人工智能领域中的发展不适,赋予特定人工智能应用场景所需的芯片可塑性。短期内,半通用 FPGA 的市场前景可观,市场潜力有望被挖掘。如深鉴科技的“深度学习处理单元”芯片(DPU)基于 FPGA 平台,利用指令集与编译器,实现了以 ASIC 级别的功耗提供优于 GPU 的性能。微软也把重心放在 FPGA 人工智能芯片上,推出了基于 FPGA 的视觉芯片 A-eye,可应用于无人机、机器人等多种人工智能产品。由于 FPGA 的单片成本较 ASIC 低,在芯片需求未成规模、深度学习算法未稳定的情况下,运用 FPGA 芯片来实现半定制化的人工智能芯片成为了众多企业的最佳选择。

2.5.2 产业链中游

机器翻译产业链中游包括机器翻译软件供应商和硬件供应商。由于现阶段机器翻译软件的盈利能力较低，硬件产品是机器翻译厂商的主要收入来源，因此多数中国的机器翻译厂商集成了软件和硬件生产能力，成为软硬一体的机器翻译供应商，如百度翻译包括百度翻译 APP 和百度翻译机；科大讯飞推出讯飞翻译 APP 和晓译翻译；搜狗推出搜狗翻译 APP 和搜狗翻译机、搜狗旅行翻译宝等。

2.5.3 产业链下游

机器翻译产业链下游市场主体为 B 端、G 端和 C 端消费者，其中 B 端消费者主要为行业应用，如金融行业的机器翻译应用、医疗行业的机器翻译应用和文学领域的机器翻译应用等，但由于行业应用对机器翻译结果的准确度要求高，而目前的机器翻译水平尚未能满足行业应用的需求，因此 B 端的行业应用难以普及。

G 端消费者主要为政府机构，政府机构将机器翻译产品多应用于跨国交流场景，如 2019 年博鳌亚洲论坛启用讯飞人工智能翻译机，2018 年博鳌亚洲论坛将腾讯同传作为官方指定的独家人工智能同声传译合作伙伴，但会议论坛上呈现的机器翻译结果不佳。

C 端消费者为个人用户，可应用于个人境外旅游、个人商务交流合作、个人语言学习等场景。现阶段，C 端市场是各大机器翻译厂商的主要发力点，消费级机器翻译硬件产品是机器翻译厂商的主要盈利来源。

3 中国机器翻译行业驱动因素

3.1 技术进步提升翻译质量

深度学习的引入变革了机器翻译领域，大幅提升机器翻译质量，使机器翻译商业化成为可能。传统机器翻译模型（基于规则的机器翻译方法和基于统计模型的机器翻译方法）需要人工设计特征，使机器根据人工标记特征进行翻译操作，而端到端神经机器翻译是一种全新的机器翻译方法，通过非线性的神经网络直接将源语言文本映射成目标语言文本，直接完成自然语言间的转换，不再需要由人工设计特征以及词语对齐、短语切分、句法树等隐结构。端到端神经网络机器翻译技术相较于传统机器翻译技术是一种革命性的改变，将机器翻译带到一个全新的纪元，大幅提升机器翻译质量。如谷歌上线的神经网络机器翻译系统产出的译文接近自然语言，用词更为准确，语序更为自然合理。曾有研究以维基百科和新闻作为语料，对谷歌的神经网络机器翻译系统进行译文质量测试，结果显示谷歌翻译在多个样本的翻译中，将错误率降低了 55%-80%。百度推出的基于注意力机制的神经网络机器翻译的汉英翻译结果在 BLUE 指标上提升了 4.9 点。科大讯飞的神经网络机器翻译系统的中译英翻译结果的 BLUE 值得分达到 0.43（BLUE 值达到 0.2-0.4 分较为适合人的理解和阅读）。

翻译准确度是机器翻译产品的核心竞争要素，过往的机器翻译的结果多受人诟病，由于翻译不准确而闹出各种争议，导致机器翻译产品的商业应用一直受限。只有在机器翻译准确度有所保证的前提下，机器翻译的商业化才能提上日程。深度学习的深入发展使机器翻译重获新生，随着翻译结果的不断提升，机器翻译产品有望拓展更广阔的市场。

3.2 翻译人才不足需机器补充

根据中国翻译协会发布的《2018 中国语言服务行业发展报告》数据，语言服务需求方对笔译和口译服务的需求较大，对英语、日语、法语、德语、俄语、西班牙语、葡萄牙语、

阿拉伯语、朝鲜语等语种的翻译服务需求较高，截至 2018 年 6 月，65.8%的语言服务需求方需要寻求笔译服务，62.8%的语言服务需求方需要口译服务，其中语言需求方对英语翻译需求比例达 96.3%，对日语、法语和德语的翻译服务需求分别为 58.4%、45.0%和 40.5%，而对其他语言的翻译需求亦有 2.6%。

中国的语言翻译服务需求仍然高涨，然现阶段的翻译人才供给与翻译人才需求间产生较大差距，高素质的翻译人才仍缺乏，通晓小语种的翻译人员更为稀少。翻译人才培养体系不够完善，多数院校的翻译教师没有翻译行业的从业经历，对教学内容是否属于翻译素质的培养无从判别，课堂教学内容多为单纯地迁移书本知识，仅对外语学科解决语言问题进行延伸，学生并未接受到真正意义上的职业型翻译教学培养，真正的翻译人才培养仍需时日。

在翻译人才供给满足不了翻译需求的情况下，机器翻译能作为较好的补充手段。人工翻译在机器翻译的辅助下，可节省大量人工成本，提高翻译效率，如美国国防预先研究计划局 (DARPA) 在多语种政治、经济、军事等领域情报信息日益激增的情况下，积极投入机器翻译研究以缓解单靠人工翻译与阅读带来的不足，提高翻译处理效率。

3.3 需求带动机器翻译发展

境外旅游及跨国商务场景的翻译需求催生机器翻译硬件产品。境外旅游和跨国商贸场景对语言翻译服务存在刚需，便携式的机器翻译产品能较好地满足此类场景的翻译需求。翻译硬件产品主要应用于境外旅游、国际企业交流、商务会谈、外语教育学习、国际贸易、文学翻译、国际会展交流会等场景，其中境外旅游是翻译机发展潜力最大的市场。随着中国居民消费水平提升，传统的固定路线跟团旅游方式无法满足消费者的旅游需求，个性化定制的深度境外游成为年轻一代消费者追求的旅游新方式。个性化定制旅游方式下，用户依据自主制定的路线完成旅程，期间没有传统跟团游的导游，旅途中需要用户自主解决一切问题，克服语言沟通难题，而机器翻译硬件产品成为境外旅游用户快速解决语言沟通难题的优质选择。

随着中国境外旅游人数增长，机器翻译硬件产品的市场空间有望提升。2018 年中国出境旅游人数达 14,972 万人次，同比增长 14.7%，机器翻译产品在境外旅游场景应用中消费潜力巨大。

4 中国机器翻译行业制约因素

4.1 机器翻译表达不精准

机器翻译结果不准确，限制机器翻译应用推广，制约机器翻译行业发展。机器翻译是一种在人工智能、大数据等技术指引下的机器行为，由于机器难以拥有生活常识、文化背景和人类风俗习惯知识储备，对于作为文化载体的语言，只能通过系统设定的规则机械地理解，因此机器翻译凸显自动化、机械化以及无意识等特征，翻译过程难以结合具体的语言环境、风俗习惯等因素分析源语，仅能够机械地分析源语的语法结构，在复杂句子结构或层层句式嵌套的情况下，则将复杂句子机械地拆解为多个子句单独翻译，并将子句译文根据源语排列组合获得最终翻译结果，容易导致翻译结果缺乏逻辑关联、生硬乏味。

机器翻译在篇章翻译及文学翻译领域呈现的结果更是让读者难以理解。机器翻译无法理解文学作品所处时代的经济、宗教等人文因素以及体会字里行间的情感流露，无法使译文表露同样的情感，不具备人工翻译“信、雅、达”的优势。在篇章翻译中，机器翻译无法处理指代消解问题和主语省略等现象，无法在译文中将源文所指代的内容自动恰当地补全，导致篇章翻译内容逻辑不通，句子结构不全，语法问题通篇存在。当前的人工智能技术在自然语言理解方面尚未有质的突破，机器翻译的应用领域仍然较为狭窄（较多运用于对翻译质量要求不高的个人休闲消费领域），机器翻译质量仍需较长时间提升。

4.2 机器翻译训练数据不足

现阶段，机器翻译以经验主义方法为主导，利用语料库数据资源驱动机器翻译质量的提升，但可运用于机器翻译训练的语料库数据资源仍然较为缺乏，导致机器翻译质量难有大的提升。基于语料库的机器翻译是一种数据驱动的方法，翻译结果的准确度主要受语料库数据的影响。实现不同语言间的互译，需要使用不同语言的语料库训练机器，而在实际应用中，众多语言缺少语料库积累，语言互译资源更是稀缺，如多数中国的少数民族语言缺少电子文本和知识库，少数民族语言间及少数民族语言与汉语间的互译文本资料甚少，造成机器缺乏训练数据，无法自主学习总结及修正该类语言间的翻译规则，导致翻译结果不佳。

目前机器翻译获取的翻译数据多来自于政府文档，如联合国官方文件配备了多语种版本，语料资源获取方便，但政府文档资源覆盖面狭窄，多为官方领域语料，缺少体育、娱乐等领域的语料数据。此外，政府文档全是正式化语体材料，缺少口语及网络用语语料，而口语及网络用语是弱规范甚至是不规范的语言，与官方用语具有明显区别，单纯的官方用语语料库训练的机器翻译规则难以适应口语及网络用语的翻译场景，难以使机器自主训练学会翻译生活化用语。而实际的语言交流场景中，口语及网络用语使用更为频繁，机器翻译的应用更应覆盖口语及网络用语场景。仅面向规范文本提供的机器翻译服务难以适应市场需求，提高非

规范文本的处理能力和翻译效果，仍是机器翻译行业面临的难题。

4.3 翻译模型不具通用性

现阶段的机器翻译系统具有明显的缺陷，基于某一领域语料库建立的机器翻译模型不具有通用性，应用于其他领域时翻译结果明显较差，系统性能显著下降。机器翻译模型不具通用性限制了机器翻译产品在不同应用领域的推广，大幅减缓了机器翻译行业的发展进程，机器翻译的行业应用至今仍未有明显发展。虽然近年来研究者们针对翻译模型、语言模型、短语调序模型和不同领域新词翻译等问题对系统进行了众多领域自适应的调试，但每种调试方法只能解决单方面的自适应问题。机器翻译行业内仍缺少解决领域自适应问题统一的有效框架，距离系统全方位适应不同领域的翻译场景仍较为遥远。

5 中国机器翻译行业相关政策法规

2015 年以来，中央及各地方政府陆续出台相关政策，积极支持人工智能产业的发展，并推进人工智能在各个细分领域的渗透和壮大。2015 年 7 月，国务院发布《关于积极推进“互联网+”行动的指导意见》，将“互联网+人工智能”列为其中 11 项重点行动之一。2017 年 7 月，国务院印发《新一代人工智能发展规划》，将人工智能上升至国家战略，并确立了人工智能发展三步走的目标，为中国人工智能产业发力奠定制度基础。2017 年 10 月，十九大将“人工智能”写进党代会报告，并提出加快建设制造强国，发展先进制造业，推动互联网、大数据、人工智能和实体经济深度融合等意见。截至 2018 年，中央及各地方政府在短短 3 年内出台了 60 余项与人工智能产业发展相关的政策，足以看出国家对人工智能产业化的重视。机器翻译是人工智能的应用体现，同样受到政策的鼓励发展。

2016 年 5 月，发改委、科技部、工信部及中央网信办联合发布《“互联网”+人工智能三年行动实施方案》要求加强产学研用合作，明确支持国家工程实验室、国家工程（技术）研究中心等创新平台建设，布局国家级创新中心，共同推动人工智能基础理论、共性技术、应用技术研究。加快基于人工智能的计算机视听觉、生物特征识别、复杂环境识别、新型人机交互、自然语言理解、机器翻译、智能决策控制、网络安全等应用技术研发和产业化，加强前沿技术布局，构造未来融合创新技术基础。

2017 年 7 月，国务院发布《新一代人工智能发展规划》，明确提出大力发展人工智能新兴产业，开发面向人工智能的操作系统、数据库、中间件、开发工具等关键基础软件，突破图形处理器等核心硬件，研究图像识别、语音识别、机器翻译、智能交互、知识处理等智能系统解决方案，培育壮大面向人工智能应用的基础软硬件产业。

2017 年 12 月，工信部印发《促进新一代人工智能产业发展三年行动计划（2018-2020

年)》，明确指出培育智能翻译行业，推动高精度智能翻译系统应用，围绕多语言互译、同声传译等典型场景，利用机器学习技术提升准确度和实用性。到 2020 年，多语种智能互译取得明显突破，中译英、英译中场景下产品的翻译准确率超过 85%，少数民族语言与汉语的智能互译准确率显著提升。

2018 年 11 月，工信部发布《新一代人工智能产业创新重点任务揭榜工作方案》，加快推进中国新一代人工智能产业创新发展。重点突破一批技术先进、性能优秀、应用效果好的人工智能标志性产品、平台和服务，为产业界创新发展树立标杆和方向；要求大力发展智能翻译系统，到 2020 年，明显突破多语种智能互译，显著提升少数民族语言与汉语的智能互译准确率，产品达到国际先进水平。

图 5-1 中国机器翻译行业相关政策

政策名称	颁布日期	颁布主体	主要内容及影响
《新一代人工智能产业创新重点任务揭榜工作方案》	2018-11	工信部	要求大力发展智能翻译系统，到2020年，明显突破多语种智能互译，显著提升少数民族语言与汉语的智能互译准确率，产品达到国际先进水平
《促进新一代人工智能产业发展三年行动计划（2018-2020年）》	2017-12	工信部	明确指出培育智能翻译行业，推动高精度智能翻译系统应用，围绕多语言互译、同声传译等典型场景，利用机器学习技术提升准确度和实用性
《新一代人工智能发展规划》	2017-07	国务院	大力发展人工智能新兴产业，开发面向人工智能的操作系统、数据库、中间件、开发工具等关键基础软件，突破图形处理器等核心硬件，研究图像识别、语音识别、机器翻译、智能交互、知识处理等智能系统解决方案，培育壮大面向人工智能应用的基础软硬件产业
《“互联网”+人工智能三年行动实施方案》	2016-05	发改委 科技部 工信部 中央网信办	加快基于人工智能的计算机视觉、生物特征识别、复杂环境识别、新型人机交互、自然语言理解、机器翻译、智能决策控制、网络安全等应用技术研发和产业化

来源：头豹研究院编辑整理

6 中国机器翻译行业发展趋势

6.1 翻译语种更加丰富

机器翻译模型将从资源丰富的语言向资源匮乏语言渗透,机器翻译的覆盖语种将更丰富。以往的机器翻译服务主要集中在应用范围较广、使用人数较多的语种中,对使用频率不高、使用范围不广的语种关注度较低,不利于机器翻译产品的推广。近年来,各大机器翻译厂商逐渐扩大翻译语言的覆盖范围,开始向小语种倾斜,提升翻译语言的丰富度。如科大讯飞翻译机支持汉语与 50 种语言的即时互译,支持日、英、俄、维、藏等重点语种以及 23 种方言与汉语的互译。腾讯翻译君支持汉语、英语、日语、韩语、法语、德语、意大利语、土耳其语等 15 种语言和 80 个语种的互译。百度翻译目前支持全球 28 种语言互译,包括汉语、日语、韩语、泰语、德语、法语等。随着各机器翻译厂商对机器翻译语言覆盖范围的拓展,机器翻译产品有望覆盖更多元的语言,帮助用户实现全世界语言间的无障碍交流。

6.2 机器翻译逐渐实用化

机器翻译作为一项技术,最终将以产品的形式落地,实现商业化,产品的实用性能决定其使用价值。由于行业应用领域(如医疗、金融、能源、文学等)对翻译结果质量要求高,机器翻译尚未能实现高水平的语义理解翻译,翻译结果无法满足行业应用需求。对于生活化场景,现阶段的机器翻译机具有较大的市场空间,翻译机将更多地与生活场景结合,帮助用户真正实现语言的无障碍转换,如翻译机能帮助用户翻译菜单、店名、商品信息等,还附带语音秘书、全球景区导览等功能,帮助出国旅游用户无障碍地完成旅游生活规划以及享受出国旅游生活。机器翻译产品还可帮助用户翻译外语剧集和国外新闻播报等,拓展用户的资讯接收方式。机器翻译产品能更多地结合生活场景,为用户提供更多元的服务。

6.3 产品逐渐支持离线翻译

机器翻译产品在出国旅游场景中的需求较大，而在出国旅游应用场景中，存在部分出国旅游用户由于流量资费较贵，不愿开通国际漫游服务，亦不愿意购买当地的上网方案，在线的机器翻译产品无法满足未能接入网络服务的用户需求。此外，在旅途过程中的用户，会较为频繁地身处于没有网络覆盖或网络信号较差的地域，如飞机、地铁以及地理位置较为偏远的景区，单纯的在线翻译服务无法很好地满足旅游用户在不同场景下的翻译需求，不支持离线翻译的机器翻译产品并未带给用户良好的使用体验，因此离线翻译功能成为各大机器翻译厂商竞争的突破口。

机器翻译离线化包括软件层面的离线和硬件层面的离线，其中硬件层面的离线是把原有的神经网络模型离线部署于硬件端中，软件层面的离线则需要调整原有的机器翻译算法模型，软硬件层面的离线均是机器翻译离线化的关键。实现软硬件层面的离线翻译要求机器翻译厂商具备强大的平台计算性能以及离线状态下的数据存储和识别能力，对厂商的要求较高，因此离线翻译的进入门槛较高，能较好地展现机器翻译厂商的竞争优势。优质离线机器翻译将成为众多机器翻译厂商追求的目标，如科大讯飞已推出汉语与英语、日语、俄语、韩语四种语言间的离线翻译功能，百度翻译机、搜狗翻译机以及网易翻译王均支持离线翻译。未来，离线翻译将成为机器翻译产品的标配，而拥有优质离线翻译能力的产品更容易抢占 C 端用户市场。

7 中国机器翻译行业竞争格局

7.1 中国机器翻译行业竞争现状分析

由于机器翻译软件的盈利能力较弱,机器翻译厂商主要依靠机器翻译硬件产品实现盈利,因此机器翻译硬件产品成为中国机器翻译厂商的竞争焦点,截至 2018 年,中国机器翻译硬件产品领域的参与厂商超 50 家,仅 2018 年推出的机器翻译硬件产品多达 10 款。当前机器翻译硬件市场的参与厂商可分为技术派和应用派:

- (1) 技术派:强悍的人工智能底层技术为技术派厂商进军机器翻译领域保驾护航,如科大讯飞在语音识别领域优势明显,其以语音识别技术切入语音翻译领域则顺理成章;百度、搜狗以搜索引擎起家,在自然语言处理技术上积累深厚,其语义理解、语音处理、以及数据积累为其开展机器翻译研发提供有利支撑。
- (2) 应用派:传统网页端翻译领域积累的大量数据有助于应用派厂商拓展机器翻译硬件领域,如网易有道深耕在线翻译多年,从传统的网页端和移动端翻译,延伸至翻译硬件领域是合理的企业扩张路径。

尽管机器翻译硬件市场参与厂商数量较多,但市场份额相对集中,在中国机器翻译硬件市场中,科大讯飞占据 50% 市场份额,分音塔占 25%,其余市场份额被其他厂商瓜分,其中包括百度、腾讯、阿里巴巴、网易、搜狗和众多创业企业。根据从事机器翻译行业商务拓展长达 3 年时间的专家分析,虽然 BAT 的人工智能技术积累深厚,但并未在机器翻译硬件领域大力投入。在机器翻译行业较为知名的厂商较多是由原先的技术积累潜移拓展至机器翻译领域,顺带研发投产机器翻译产品。鉴于机器翻译行业商业模式未被开发,仍以机器硬件产品为主要收入来源,专门开展机器翻译业务的创业企业没有软件和硬件优势,需要较长时间进行研发投入以及产业链构建,耗费时间及成本较大,竞争优势并不明显。机器翻译行业

的市场规模有限，可容纳的企业数量有限，未来行业将呈现寡头竞争局面，没有同时具备软硬件优势的企业难以抢占市场份额。

7.2 中国机器翻译行业典型企业分析

7.2.1 新译信息科技（深圳）有限公司

7.2.1.1 企业简介

新译科技成立于 2014 年 12 月，是一家专注于人工智能技术的科技公司，以机器翻译和语音识别等技术为基础，向用户提供在线机器翻译、在线辅助翻译平台和互联网语言服务平台等一系列关于翻译领域的产品和服务。新译科技以软、硬件智能翻译为承载体，服务于国家“一带一路”多语传播平台建设，为金融、专利、法律、医学等垂直领域提供全球化基础性多语沟通服务。

7.2.1.2 产品研发

(1) B 端和 G 端产品：新译科技的 B 端和 G 端产品包括软件和硬件产品，主要面向大型央企、军工企业、互联网媒体及政府机关等企业和政府机构提供服务。

①定制化机器翻译是一个混合神经网络翻译引擎，该服务能统一管理企业语料，给企业特定的数据和管理账号，满足个性化、精细化、专业性的需求，达到保护企业数据隐私和数据的效果。

②智能翻译插件包括办公翻译插件和网页翻译插件，办公翻译插件可以在 WPS 和 Office 两大办公系统上使用，满足多语言文档的快速转换和阅读；网页翻译插件支持各大浏览器网页翻译，一键转化母语阅读。

(2) C 端产品：新译科技主要向个人消费者提供在线文档翻译、PC 端翻译、智译 APP 服务以及可穿戴式翻译产品。智译 APP 可实现语音和文本的多种语言智能翻译，适用于旅游、社交、住宿、商务等领域。在线文档翻译支持 10 种常用文档格式输入，可应

用于通用或专业领域的文档翻译，翻译质量较高。

7.2.1.3 竞争优势

新译科技拥有较强的技术优势，同时拥有澳门大学自然语言处理与葡中智能翻译 (NLP2CT) 实验室、葡萄牙里斯本 L2F 语音实验室、美国卡耐基梅隆大学 LTI 实验室以及新译-澳大-清华人工智能研究院四所优质自然语言研究实验室提供的技术人才和原创技术支持，公司所研发的智能翻译产品于 2017 年通过国家科技成果认定，并获得深圳青年技术奖以及吴文俊人工智能科学技术奖。新译科技目前拥有近百人的团队，核心技术研发人员主要在清华大学自然语言处理研究中心，澳门团队主要开发机器翻译的各个衍生品，深圳团队做自然语言处理底层设计，北京团队负责前端、产品和销售。新译科技未来将会着力提升限定性神经网络翻译技术、增强交互式辅助翻译水平、提高语义翻译的技术、改进自动修正技术提高后编译能力水平。

7.2.2 北京分音塔科技有限公司

7.2.2.1 企业简介

分音塔科技成立于 2016 年 7 月，是一家专注于人工智能领域的高科技公司，开发不同国家和民族之间跨语言同声传译产品，为用户提供可穿戴智能设备、人工智能 API 综合解决方案和虚拟运营服务。

7.2.2.2 产品研发

准儿翻译机系列是分音塔的主要产品，包括准儿 Pro、准儿 Go、准儿 Fly 以及准儿 First。准儿翻译机是由分音塔科技采用世界领先的人工智能技术和独有的口语翻译技术，针对泛旅游场景潜心打磨而就，支持中、英、日、韩、泰等 50 多种语言的实时对译语聊。在旅游场景下，中英、中日翻译准确率分别可达 96% 和 90%。准儿翻译机可轻松应对一些出国热门目的地的常用的语言，如英语、日语、韩语、泰语、法语、西班牙语、俄语等。

7.2.2.3 竞争优势

分音塔科技精准卡位泛旅游类应用场景，受众明确，旗下的准儿翻译机在“紫禁城巅峰旅游论坛暨 2017 首届旅游投资峰会”中荣获最具投资潜力奖；在环球旅讯峰会上荣获“2017 旅游产品创新大赛”全场大奖双料冠军以及在中国劲旅奖颁奖获“最佳旅行社服务商”荣誉称号。目前，准儿翻译机和科大讯飞的晓译翻译机，二者的中英互译准确率不相上下，准儿翻译机的语种更丰富，且首创带屏幕，可以同步显示中外文字，实时校验内容，尤其是应对汉语大量的同音字、口音差异时，有屏幕的翻译机能增强内容识别正确率，使用户感知的准确率更高。

7.2.3 北京爱特曼科技有限公司

7.2.3.1 企业简介

爱特曼科技成立于 2016 年 1 月，是专注于研发医学领域的人工智能技术和产品的企业，主要为用户提供基于深度学习的神经网络机器翻译服务。Atman 于 2017 年 9 月获得由长江会基金、北极光创投参投的 560 万美元的 A 轮融资。

7.2.3.2 产品研发

Atman 已落地机器翻译服务，推出 TransGod 翻译工具，并为医药客户提供基于 TransGod 的翻译服务，客户包括强生、太美医疗等。强生使用了 Atman 专属定制的开发医药领域机器翻译系统 J-SmarT 后，提升了 20%的工作效率。

- (1) Atman 机器翻译 API: 专注于医学领域机器翻译，依托领先的 Atman 神经网络机器翻译引擎和亿级的医学数据优势，为开发者带来即时、精准且便捷的医学领域翻译服务，开发者可使用该机器翻译 API 构建丰富多样的功能或应用。
- (2) 医学翻译: Atman 医学翻译团队由长期从事医学翻译工作的专业人员组成，为用户提供医药、医疗器械、医学研究、诊疗诊断等文件的翻译服务。

7.2.3.3 竞争优势

海量的医学数据基础为 Atman 的机器翻译提供明显的竞争优势，Atman 拥有 2 亿基础语料以及众多客户资源，可通过基础语料以及客户的机器翻译模型和算法不断优化机器翻译引擎。此外，Atman 有严谨的翻译流程，符合 ISO 17100 (际翻译服务企业标准) 和 TAC 001-2016 笔译服务要求 (中国翻译协会发布的翻译服务企业标准) 的规范，制定了医学翻译行业风格指南对医学翻译行业中的常见问题进行抽丝剥茧，对产品名、术语和缩略语以及公司名称、地址、书写方式等事项进行了严格规定，不但确保了行文风格、术语及格式等事项的规范化，更增强了客户对高质量服务的预期。

头豹研究院简介

- 头豹研究院是中国大陆地区首家 B2B 模式人工智能技术的互联网商业咨询平台，已形成集行业研究、政企咨询、产业规划、会展会议行业服务等业务为一体的一站式行业服务体系，整合多方资源，致力于为用户提供最专业、最完整、最省时的行业和企业数据库服务，帮助用户实现知识共建，产权共享
- 公司致力于以优质商业资源共享为基础，利用大数据、区块链和人工智能等技术，围绕产业焦点、热点问题，基于丰富案例和海量数据，通过开放合作的研究平台，汇集各界智慧，推动产业健康、有序、可持续发展



四大核心服务：

企业服务

为企业提供定制化报告服务、管理咨询、战略调整等服务

云研究院服务

提供行业分析师外派驻场服务，平台数据库、报告库及内部研究团队提供技术支持服务

行业排名、展会宣传

行业峰会策划、奖项评选、行业白皮书等服务

园区规划、产业规划

地方产业规划，园区企业孵化服务



报告阅读渠道

头豹科技创新网 —— www.leadleo.com PC端阅读全行业、千本研报



头豹小程序 —— 微信小程序搜索“头豹”、手机扫右侧二维码阅读研报



图说



表说



专家说



数说

详情请咨询



客服电话

400-072-5588



上海

王先生：13611634866

李女士：13061967127



南京

杨先生：13120628075

唐先生：18014813521



深圳

李先生：18916233114

李女士：18049912451