

机器翻译近十年发展综述

——基于CiteSpace可视化分析

张 玥

上海海事大学外国语学院, 上海

收稿日期: 2023年2月21日; 录用日期: 2023年3月21日; 发布日期: 2023年3月30日

摘 要

机器翻译是利用计算机将源语言转换为目标语言的过程, 是计算语言学的一个分支, 是人工智能在翻译领域的一大进步, 具有重要的科学研究价值。随着科技的迅猛发展, 机器翻译从“机械大脑”到基于规则的机器翻译和基于统计的机器翻译, 再到基于深度学习的神经网络机器翻译, 技术上不断推陈出新, 算法上不断升级优化。同时, 在经济全球化的大背景下, 机器翻译技术在促进政治、经济、文化交流等方面起到越来越重要的作用。本文将借助CiteSpace可视化分析来探讨2012年至2022年机器翻译的发展, 发现围绕机器翻译主题所引出的研究热点有译后编辑、人工智能等方面, 并且机器翻译逐渐与其他学科相结合, 形成跨学科研究模式。因此, 在未来的发展中, 我们要多领域地培养人才, 以此促进机器翻译技术创新。

关键词

机器翻译, 研究热点, 发展趋势

Overview of the Development of Machine Translation in the Past Ten Years

—Based on CiteSpace Visual Analysis

Yue Zhang

College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: Feb. 21st, 2023; accepted: Mar. 21st, 2023; published: Mar. 30th, 2023

Abstract

Machine translation is the process of converting the source language into the target language by using computers. It is a branch of computational linguistics and a great progress of artificial intel-

ligence in the field of translation. It has important scientific research value. With the rapid development of science and technology, machine translation has evolved from “mechanical brain” to rule-based machine translation and statistic-based machine translation, and then to neural network machine translation based on deep learning. In terms of technology, it continues to bring forth new ideas, and in terms of algorithms, it continues to upgrade and optimize. At the same time, in the background of economic globalization, machine translation technology plays an increasingly important role in promoting political, economic and cultural exchanges. This paper will explore the development of machine translation from 2012 to 2022 with the help of CiteSpace visual analysis. It is found that the research hotspots around the topic of machine translation include post-translation editing, artificial intelligence and other aspects, and machine translation is gradually combined with other disciplines to form an interdisciplinary research model. Therefore, in the future development, we should cultivate talents in multiple fields to promote the technological innovation of machine translation.

Keywords

Machine Translation, Research Hotspots, Development Trend

Copyright © 2023 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

机器翻译也被称作自动翻译,指利用计算机将一种语言文本(源语言: Source Language)转换为另一种语言文本(目标语言: Target Language)的翻译过程[1]。随着科技的不断发展创新,机器翻译经历了基于规则的机器翻译到基于统计的机器翻译,再到基于深度学习的神经网络机器翻译,翻译技术不断革新,算法不断优化,硕果累累。机器翻译这股不可忽视的新兴力量必将在翻译行业的未来中占据重要的席位,相比于人工翻译,机器翻译节省了许多人力和财力,但是,当下的机器翻译技术不能满足正式和专业的场合以及一些领域的翻译工作,其翻译结果仍然存在很多不恰当的地方,其仅仅能在日常生活等基础层面为人所用。本文将通过 CiteSpace 可视化分析来观察近十年来机器翻译的研究热点,并探讨其发展趋势。

2. 研究回顾

2.1. 国内外机器翻译发展

2.1.1. 国外机器翻译发展

Weaver 在 1949 年首次提出了使用计算机进行翻译的思想,并提出了避免出现“字对字”翻译的四条具体原则。此后,机器翻译研究得以兴起并迅速发展,麻省理工学院的博士后研究者,以色列著名哲学家、语言学家和数学家 Bar-Hillel 于 1952 年组织召开了第一次机器翻译大会,之后美国的乔治敦大学、华盛顿大学、国际商业机器学习也积极展开有关机器翻译的研究[2]。早期机器翻译研究得以快速发展是由于研究领域的新颖与跨学科性以及不同国家对机器翻译研究的实际需求。

1950 年到 1960 年间机器翻译主要针对理论语言学中的句法分析,建立起多种基于句法分析的机器翻译模型。研究者们发现,机器翻译产出的译文质量良莠不齐,难以达到预期要求,因而对其可实施性提出了质疑。其中,Bar-Hillel 认为,机器翻译最大的屏障就是语义歧义问题。语言中语义的多样性要求机器翻译必须拥有一套能够充分对语言结构与语义问题进行分析的语法、句法系统。而当时应用的转换

法则与句法分析模型显然不能解决各种复杂的语言问题。因此，他建议将自动机器翻译修改为机器辅助翻译，在译前和译后阶段为译者提供必要的辅助[3]。对机器翻译的质疑催生了语言自动处理咨询委员会(The Automatic Language Processing Advisory Committee, ALPAC)的建立。ALPAC 成立的主要目的就是为机器翻译发展及资助情况做出评估。1966 年，ALPAC 发表的报告指出，机器翻译由于需要建立规则算法及模型设置，译文质量的不尽如人意与可观的经济投入显示出产出与投入的明显失衡。报告的发表使得机器翻译研究相继停止，但加拿大、法国等地进行的机器翻译研究在具体应用领域却表现突出。

从 1990 年开始，大量的双语、多语语料应运而生。基于语料库的统计机器翻译通过在平行文本中词对齐、短语对齐、句子对齐的方式自动构造机器翻译模型。近些年，基于“深度学习”的神经机器翻译系统产生并快速发展。深度学习(Deep Learning)由 Hinton 等人在 2006 年提出，现已成为机器学习领域发展最快的主流技术之一。由于深度学习是机器进行的一种多层次非线性处理，与传统的浅层次线性处理相比，在处理模型分析和分类问题上更准确、性能更高[4]。模拟神经网络的深度学习技术能够使机器拥有自动学习抽象特征表达的能力，并且能够将学习结果灵活地应用到其他任务中，因而基于深度学习的机器翻译可以尽可能省去人工调配，从而实现由机器自动推断最佳翻译结果[5]。

目前云计算、大数据和移动终端的普及给机器翻译带来了更加广泛的应用，如口语翻译、照相翻译、文字扫描翻译等。

2.1.2. 国内机器翻译发展

1956 年，我国就曾经把机器翻译研发列入了科学技术工作之中并小有成就，1959 年国庆十周年前夕，在我国大型通用电子计算机上成功地进行了实验。词典包括 2030 个词条，语法规则系统由 29 个线路图组成。这次实验翻译了九个不同类型的较繁杂的句子。在经过了一段停滞期之后，20 世纪八十年代以后，我国机器翻译迎来了高速的发展阶段，第三次科技革命的浪潮席卷而来，经济与科技高速发展中的中国迈着改革开放的步伐，在机器翻译研究领域也有了质的飞跃。

在经历了众多探索研究之后，我国的机器翻译也逐渐走向了应用化、大众化和商业化。中国软件技术公司在 1988 年推出了“译星 I 号”，标志着我国机器翻译系统正式走向市场，并迅速拥有了大量用户且走向海外。继“译星”之后，北京的高立系统、天津的通译系、陕西的朗威系统等一系列机器翻译系统也进入了大众视野。在 21 世纪，金山词霸、有道翻译、百度翻译等一系列 APP 的开发，更是极大地满足了普通用户对于翻译的需求[6]。

2.2. 机器翻译种类

2.2.1. 基于规则的机器翻译

基于语言规则的机器翻译系统类型大致可分为直接翻译系统、转换系统和中间语言系统。源语和目标语之间的词汇单元的对应关系是直接翻译系统设计的依据。转换系统的源语言分析独立于目标语言，双语转换时不仅需一部双语对应词典，还需要考虑双语结构差异，进行结构调整。中间语言系统的源语言不与目标语言直接接触，它把源语的文本用中间语表达出来，然后再把中间语言所表达的意义用目标语表达出来[7]。

2.2.2. 基于统计的机器翻译

统计机器翻译的原理使基于语料库的，语料库是他们获得翻译知识的来源。统计机器翻译采用数据驱动的方式，主要依靠双语或多语言并行语料库，通过对词对齐和翻译规则提取实现翻译建模，然后利用所学知识和翻译规则进行自动翻译。统计机器翻译的任务是将观察到的源语言还原为最有可能的翻译，即同一源语句段可能对应多个候选翻译句段，呈现概率最大的便是译文[8]。

2.2.3. 基于神经网络机器翻译

2013年, Kalchbrenner 和 Blunsom 提出了一种基于神经网络的翻译方法。端到端序列生成模型是神经机器翻译的基本建模框架。它是一种将输入序列转化为输出序列的框架和方法。其基本思想是通过神经网络直接实现自然语言之间的自动翻译, 因此, 神经机器翻译通常使用编码器-解码器框架实现序列到序列的转换[9]。神经网络法在模型训练完毕后无需再使用语料库, 只运用神经网络获取源语言到目标语言的转换。但是, 它需借助注意力机制、长短时记忆网络和门限循环单元等[10]。

3. 研究方法

3.1. 研究问题

通过对知网核心期刊中的数据进行可视化分析, 本研究将围绕以下问题展开: ① 近十年机器翻译研究主要围绕哪些热点议题展开?② 近十年机器翻译研究呈现怎样的发展趋势, 未来走向呈现怎样的特点?

3.2. 数据来源

本研究数据来源于知网核心期刊, 同时设置高级检索主题为“机器翻译”, 时间设置为 2012 年至 2022 年, 文献检索类型为北大核心和 CSSCI, 经筛选后得到有效文献 200 篇。

3.3. 分析工具

本文以 CiteSpace 软件作为分析工具, 其中 CiteSpace 可利用名词性术语找出文献中的突现词, 发现机器翻译在一段时间内呈现的变化和动态趋势, 探索研究领域的研究热点与未来发展趋势。

4. 研究热点分析

从下图 1 中发现, 近 10 年机器翻译研究主要聚焦在以下三个议题。

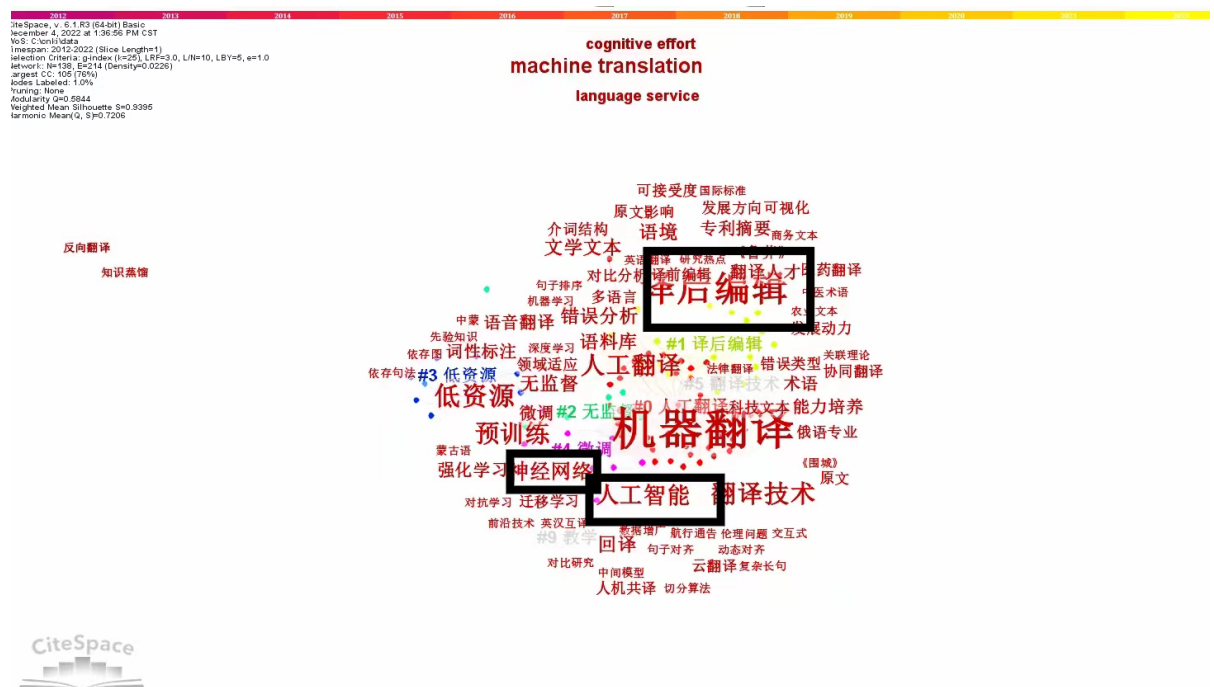


Figure 1. Machine translation research keyword co-occurrence network map (2012~2022)
图 1. 机器翻译研究关键词共现网络图谱(2012~2022)

4.1. 译后编辑

为克服机器翻译的局限性,提升译文质量,译后编辑由此应运而生,与之相关的译者也因此要具备译后编辑的相关能力。机器翻译译后编辑能力指“根据一定目的与要求对机器翻译的原始产出进行加工与修改所需的知识体系与认知素养”[11]。相比于翻译能力,译后编辑能力是一个新兴概念。Kontinen等人总结出专门针对译后编辑能力的子能力要素,包括策略能力、人际交往能力、态度或生理心理能力以及工具能力[12]。译后编辑能力培养已成为翻译人才培养中的重要议题,需要相关方携手努力,培养出语言服务行业真正需要的专业人才。

4.2. 人工智能

机器翻译平台、语料加工平台、语音转写等人工智能技术已经在翻译领域广泛应用。人工智能引入翻译行业后,可以有效地提高翻译质量,在某些方面甚至可以代替人工翻译,使用极其方便。但与此同时,人工智能翻译也有一些局限性,它始终不能像人类大脑一样具有共情能力,不能根据文本类型选择恰当的翻译策略,因此人工智能翻译本身尚存在大量的技术进展空间。

4.3. 神经网络

神经机器翻译是近年来一种新兴的机器翻译技术,作为一种全新机器翻译模型,神经机器翻译通过使用深度学习神经网络来获取自然语言之间的映射关系,实现源语言直接翻译成目标语言,减少了传统统计机器翻译中繁琐的转化过程,这一翻译技术因其简洁、高效得到了学界广泛认可。

从上图1中也可以看到,神经网络机器翻译周围出现的高频词如强化学习、低资源、译后编辑和翻译技术等,说明神经网络机器翻译近年来侧重于研究如何提升翻译质量、人机交互尤其是译后编辑的交替应用等方面。随着神经网络机器翻译在高资源语言翻译任务(如英法翻译、英汉翻译)取得的突破性进展,学界还将关注点聚焦在如何提升低资源语言情况下机器翻译的任务,例如加泰隆语和汉语之间的翻译实践质量。

4.4. 发展趋势

由上述可知,可以从宏观角度和微观角度来分析机器翻译的研究热点,以此来看机器翻译的发展趋势。宏观来说,机器翻译的研究热点主要聚焦于人工智能、翻译技术、能力培养和教学等方面。其中,翻译技术的提升包括切分算法、数据增广等,能力培养包括译前编辑和译后编辑能力等,体现了机器翻译和其他学科形成跨学科的研究趋势。微观来说,机器翻译的研究热点主要聚焦在语言句法处理上、文本分类及术语处理等方面。其中,语言句法处理包括多种语言的拓展,如俄语、蒙古语,以及句法上涉及到的句子排序和介词结构,文本分类包括商务文本、文学文本以及农业文本等更多新领域,体现了机辅翻译的发展将更加具体化和专业化。

5. 结论

本论文借助CiteSpace对核心数据库中近十年来以“机器翻译”为主题的文献进行可视化分析,研究发现:首先,机器翻译随着科学技术的发展,其翻译技术得到了不断升级,翻译质量也随之得到提升;其次,机器翻译在译前编辑和译后编辑等方面展开了拓展研究,并积累了相关经验;再次,机器翻译与其他学科,如计算机、语言学等学科知识相融合,形成跨学科研究趋势,并且逐渐应用于更多文本类型,如低资源语种文本;最后,由于机器翻译有智能但缺智慧,因此学界还关注到了神经网络机器翻译和人机共译,以达到更好的翻译效果。

在未来更具科技化的时代里，机器翻译在语言服务行业占据重要地位，应该更加关注新兴研发人才的培养，从而促进科技创新并拓宽机器翻译的研究。

参考文献

- [1] Hutchins, W.J. (1986) *Machine Translation: Past, Present, and Future*. Ellis Horwood Limited, Chichester.
- [2] Bar-Hillel, Y. (1958) Some Linguistic Obstacles to Machine Translation. *Proceedings of the Second International Congress on Cybernetics*, Belgium, 3-10 September 1958, 197-207.
- [3] Bar-Hillel, Y. (1959) Report on the State of Machine Translation in the United States and Great Britain.
- [4] Hutchins, J. (2010) Machine Translation: A Concise History. *Journal of Translation Studies*, **13**, 29-70.
- [5] Poibeau, T. (2017) *Machine Translation*. The MIT Press, Boston. <https://doi.org/10.7551/mitpress/11043.001.0001>
- [6] 徐一灿, 刘继斌. 机器翻译的现状和前景[J]. 海外英语, 2017(21): 117-118+148.
- [7] 冯志伟. 机器翻译与人工智能的平行发展[J]. 外国语(上海外国语大学学报), 2018, 41(6): 35-48.
- [8] 侯强, 侯瑞丽. 神经机器翻译研究——洞见与前景[J]. 外语学刊, 2021(5): 54-59.
- [9] Bahdanau, D., Cho, K. and Bengio, Y. (2014) Neural Machine Translation by Jointly Learning to Align and Translate. ArXiv: 1409.0473.
- [10] Chung, J., Gulechre, C., Cho, K. and Bengio, Y. (2014) Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. ArXiv: 1412.3555.
- [11] 冯全功, 刘明. 译后编辑能力三维模型构建[J]. 外语界, 2018(3): 55-61.
- [12] Kontinen, K., Salmi, L. and Koponen, M. (2020) Revision and Post-Editing Competences in Translator Education. In Koponen, M., Mossop, B., Robert, I.S. and Scocchera, G., Eds., *Translation Revision and Post-Editing: Industry Practices and Cognitive Processes*, Routledge, London & New York, 187-202. <https://doi.org/10.4324/9781003096962-15>