

# 基于“人工智能+”的语言服务行业创新技术研究\*

汤丽拿

(浙江长征职业技术学院, 浙江 杭州 310023)

**摘要:** 人工智能技术的兴起给语言服务行业的发展带来了新挑战。旨在分析人工智能背景下的语言服务行业发展趋势, 在结合国内外语言服务需求的基础上, 对语言服务中的创新技术应用和策略进行研究, 寻找有效的互联网翻译技术及语言服务解决方案, 推进大数据、人工智能与语言服务的协同创新发展。

**关键词:** 人工智能; 语言服务; 翻译技术; 语音识别

**中图分类号:** H059

**文献标志码:** A

**DOI:** 10.15913/j.cnki.kjycx.2020.20.053

## 1 语言服务行业发展趋势

随着社会语言服务需求的不断增长, 人们对语言服务行业发展的研究不仅包含国家层面上的总体战略规划, 并且逐渐开始向社会层面的各专业和业务领域深度发展研究延伸。语言服务行业涉及翻译服务、本地化服务、语言技术工具开发与应用(机器翻译 MT 与计算机辅助翻译 CAT)、语言教学与培训等众多业务领域。

人工智能语言服务的出现, 突显了全球化大数据背景下的语言信息特点以及翻译智能技术平台时代的来临, 未来即将是“一带一路”的个性化语言服务和智能化翻译需求激增的时代。目前国内语言服务行业的发展趋势和现状是: 语言服务人才市场供求失衡, 语言服务人才培养体系不完善, 翻译技术和软件运用不足, 语言类服务企业和高校的创新协同合作有待深入, 网络翻译平台的开发与应用研究需要新技术和新实践等。

## 2 人工智能技术背景下的语言服务需求

知识增长和技术发展往往呈现交替加速发展的特点, 当今尤其以云计算、大数据、人工智能、物联网、5G 等为代表的新技术正逐渐改变着人类生活的各个领域。人工智能技术正改变着智能制造、金融贸易、教育培训、公共事业、医疗应用等各个行业, 并成为变革社会的重要推动力。世界各国都高度重视发展人工智能, 加强人工智能领域的技术研发和创新。2017 年国务院发布的《新一代人工智能发展规划》将发展人工智能作为国家重点发展战略之一。2018 年教育部印发的《高等学校人工智能创新行动计划》中指出要加快人工智能在教育领域的创新应用。

人工智能的应用开始涉及各行各业。目前, 中国的人工智能在金融、安防、客服等行业领域已得到广泛运用。人工

智能在语言服务行业的应用主要是自然语言处理技术(NLP), 它是计算机科学、人工智能、语言学关注计算机和人类(自然)语言之间的相互作用的领域, 主要包括机器翻译和语音识别技术。机器学习、语音识别、语义识别、图像识别技术等不同形态的语言产品将不同程度地融入新的智能技术平台, 未来语言服务业的发展无疑会进入“语言智能+”的时代。语言智能的目标是实现自然语言与机器语言的连续交互, 是人工智能的高级阶段, 在未来人工智能的应用上有着非常重要的作用。

## 3 语言服务行业技术应用与创新

### 3.1 翻译技术的应用

人工智能翻译技术的应用在语言服务业中发挥着越来越重要的作用。传统的简单机械的人工翻译模式已经无法适应快速发展的市场需求, 今后会逐渐被机器翻译、语音识别技术和同步多语字幕翻译技术、神经网络机器翻译技术所取代。

在新一代技术的影响下, 翻译产品也发生着巨大的变化。2018 年底由腾讯 AILab 研制出的人工智能辅助翻译产品 TranSmart 在业内引起轰动。腾讯交互翻译 TranSmart 是一款面向人工翻译场景的机器翻译产品, 也是业内首个落地的交互式机器翻译产品。该产品结合了腾讯 AILab 自主研发的神经网络机器翻译、交互式机器翻译、翻译输入法、翻译记忆片断复用技术, 配合亿级双语数据, 帮助用户更快、更好地完成翻译任务。针对人工翻译市场现状, 特别加强了常见专业领域的翻译性能, 对于教育、合同、法律和专利等领域的文本效果尤为出色。

另外, 针对 2020 年初的突发性全球新冠肺炎疫情, 很多语言服务行业人士运用自己的语言技能, 服务于抗疫战

\* [基金项目] 2018 年上海外语教育出版社全国高校外语教学科研项目“人工智能背景下语言服务行业发展策略研究”(编号: 2018ZJ0038B) 课题成果

役,将中国宝贵的抗疫经验通过多国文字及时地传递给全世界,让更多的人能够积极参与防疫抗疫。

一个机器翻译矩阵 Tmxmall 为了给更多正在为抗疫出力的翻译团队及相关人士提供技术便利,及时上线了聚焦于医疗的垂直领域机器翻译引擎——阿里医疗机器翻译。阿里医疗引擎是阿里云达摩院机器翻译产品,能为译者提供更加专业和准确的医疗翻译结果。该系统支持中英、英中、中日、日中及多语种互译,并且支持 YICAT 企业版/团队版、Tmxmall MT Plugin、SDL Trados、memoQ 中应用。本次疫情期间,阿里也在不间断地使用优质的疫情双语语料来训练医疗机器翻译引擎,帮助大家更加轻松高效地完成新冠肺炎疫情及医疗相关的翻译工作。未来,翻译服务也将从通用领域的机器翻译:谷歌翻译、有道翻译、云译、腾讯翻译君、搜狗翻译、aws、阿里翻译、小牛翻译、DeepL、百度翻译、Bling、Yandex 等,拓展到更精准更专业化方向的垂直领域机器翻译。未来几年,智能化的翻译技术必将加速推进语言服务行业的发展。人工智能背景下的语言服务必将是专业领域和业务领域深度融合发展的语言服务。

### 3.2 语音识别技术的运用

语音识别作为自然语言处理的关键技术,被逐渐运用于移动终端,各种语音识别产品层出不穷。语音识别是以语音为研究对象,通过语音信号处理和模式识别让机器自动识别人类口述的语言。

本次研究以大数据采集、抽取样本的研究方法研究语音识别技术的创新应用。以杭州好云教育科技有限公司提供的 MyET 语音识别软件在浙江省高校互联网听说比赛系统中的运用为实证研究,以大数据评价学生学习习惯的养成和在实践中口语能力的变化情况。MyET 自动语音分析诊断技术(ASAS©, Automatic Speech Analysis System),拥有中国、英国、法国、德国、日本等九个国家及地区的发明专利(中华人民共和国专利第 ZL02160031.7)。系统会根据语言学习四个要素——发音、语调、流利度、音量去分析学习者在口语上的问题,并提出有效的口语改善建议。比赛收集的数据和报表包括以下几种。

详细成绩报表(Score Details)数据内容:在比赛期间中,系统会完整详实记录选手历次进行比赛刷分的信息,包含每次比赛刷分成绩、比赛刷分次数、登录比赛刷分时间点及每次完成比赛所花的时间。数据意义为教师可以从详尽的成绩数据中,了解每一位学生对大赛的投入程度,并看出学生在大赛实践过程中口语能力(成绩)的变化。

统计报表(Statistics 1)比赛总时长(Contest Duration)

数据内容:在比赛期间中,系统会自动侦测每位选手每次进入比赛与完成比赛当中实际操作所花的时间,计算出选手实际投入在比赛模块的总在线时间。数据意义为教师可以了解每位学生总共花了多少时间在进行比赛,看出通过大赛是否能够有效激发学生的学习热情,启发学习的积极主动性。

课程学习总时长(Accumulated Learning Time)数据内容:每位参赛选手将获得一组 MyET 课程包,选手可在比赛期间进行课程练习,系统会自动侦测并计算每位选手使用课程内各学习模块(包含单句练习、角色扮演、强化练习、自我检定、跟读、复诵等)进行自主学习的总体时间。数据意义为教师可以了解每位学生在课堂外,花了多少时间在进行口语课程的自主学习,有效增加了多少口语学时。

以学校为单位的分析报表(Analysis by school)数据内容:在比赛期间中,同一场比赛里,报表会呈现每所参赛学校的总参赛人数、每所学校参赛选手总刷分次数、每所学校参赛选手总刷分时间。数据意义为教师可以了解在比赛中,各学校对比赛投入程度及自己学校的学生在比赛的总体表现程度。

刷分时段分析报表(Analysis by Time Slots)数据内容:在比赛期间中,报表会呈现全体选手刷分时段分布统计。数据意义为教师可以了解学生上线时间分布,看出学生习惯在哪些时间段内学习,体现通过大赛是否有效地帮助学生养成学习习惯及妥善运用碎片化时间学习。

在人工智能语言服务改变人类经济社会生活方式的新形势下,抓住机遇,努力进行技术创新变革,逐步建立语言服务行业规范及标准,培养新一代创新技术应用人才是当前语言服务行业发展的当务之急。

### 参考文献:

- [1] 蔡自兴.人工智能及其应用[M].北京:清华大学出版社,2016.
- [2] 罗华珍,潘正芹,易永忠.人工智能翻译的发展现状与前景分析[J].电子世界,2017(21):21-23.
- [3] 王华树.语言服务的协同创新与规范发展——2016 中国语言服务业大会暨中国译协年会综述[J].中国翻译,2017(1):85-88.
- [4] 2016 中国语言服务行业发展报告[R].中国翻译协会,2016.

作者简介:汤丽拿(1979—),女,浙江杭州人,硕士研究生,副教授,研究方向为课程与教学论。

[编辑:严丽琴]