

人工智能在语言服务企业的应用研究

崔启亮

(对外经济贸易大学 英语学院, 北京 100029)

摘要:人工智能技术的应用促进了语言服务企业产品创新。本文分析语言服务行业报告关于机器翻译和译后编辑模式的调查数据,选择语言服务企业的典型产品,分析人工智能技术在产品中的应用现状。研究发现:自然语言处理、机器学习、计算机视觉是应用最广泛的人工智能技术,是语言服务企业产品应用最广泛的人工智能技术,各个产品都集成了多项人工智能技术。

关键词:人工智能;语言服务;企业产品;机器翻译

中图分类号:TP31

文献标志码:A

文章编号:1674-6414(2021)01-0026-08

0 引言

人工智能(AI)是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的技术。人工智能是引领未来的战略性技术,在移动互联网、大数据、超级计算、传感器、脑科学等新的理论和技术以及社会经济发展强烈需求的共同推动下,人工智能得到迅速发展,呈现出深度学习、跨学科融合、群智开放、自主操控等特征(冯志伟,2018:48)。语言服务企业是提供翻译服务、本地化服务和语言技术产品/服务的企业群体,语言服务企业是提供翻译和本地化服务的市场主体。人工智能时代技术发展速度惊人,随着人工智能、神经网络、深度学习技术不断融入,翻译技术从最初基于规则的机器翻译发展到依托云技术、大数据的神经网络机器翻译,小至手机的翻译APP,大至翻译行业乃至语言服务领域中的云翻译,翻译技术不断推陈出新,迭代更替。人工智能的迅速发展将深刻改变语言服务企业的服务模式和生态体系(王华树,2020:87)。

那么,人工智能技术在语言服务企业中应用现状如何呢?语言服务企业的产品使用了哪些人工智能技术呢?人工智能技术如何改变语言服务企业的服务模式呢?本文先介绍人工智能的基本内容,然后从语言服务行业调查报告以及人工智能在语言服务企业产品应用两个方面进行研究,以期探索人工智能对语言服务企业的影响,推动语言服务企业提升服务能力和模式变革。

1 人工智能简介

人工智能是自然科学和社会科学交叉的科学,人工智能技术涉及计算机、信息论、控制论、数学、语言学和哲学等多个学科的技术。

收稿日期:2020-10-05

基金项目:国家语委“十三五”科研规划2018年度重点项目“京津冀协同发展中语言需求及对策研究”(ZD1135-69)、对外经济贸易大学培育项目“京津冀协同发展中的语言服务研究”(19PY60)的阶段性研究成果

作者简介:崔启亮,男,对外经济贸易大学英语学院副教授,博士,硕士生导师,主要从事翻译技术、本地化、语言服务研究。

当前学者们比较认同的人工智能起源于1956年在美国召开的达特茅斯(Dartmouth)大学研讨会,会上首次提出了“人工智能”的术语。此后人工智能技术与计算机技术相互促进,得到快速发展。近十年来,机器学习、计算智能、人工神经网络和行为主义的研究成为热点。

我国的人工智能研究起步较晚。纳入国家计划的研究(“智能模拟”)始于1978年;1984年召开了智能计算机及其系统的全国学术讨论会;1998年起把智能计算机系统、智能机器人和智能信息处理(含模式识别)等重大项目列入国家高技术研究计划。2017年国务院印发《新一代人工智能发展规划》文件中,发展人工智能技术成为国家重大发展战略。

人工智能具有交叉学科的特征,产生了不同学派,其中代表性的学派包括符号主义、连接主义和行为主义。符号主义认为人工智能来自数理逻辑,连接主义认为人工智能源于仿生学,特别是对人脑模型的研究。行为主义认为人工智能源于控制论。

人工智能的研究领域包括自然语言处理、自动定理证明、智能数据检索系统、机器学习、模式识别、视觉系统、问题求解、人工智能方法、程序语言、自动程序设计(蔡自兴等,2003:9)。人工智能的研究和应用可以分为问题求解、逻辑推理与定理证明、自然语言理解、自动程序设计、专家系统、机器学习、神经网络、机器人学、模式识别、机器视觉、智能控制、智能检索、智能调度与指挥、分布式人工智能与Agent、计算智能与进化计算、数据挖掘与知识发现、人工生命、系统与语言工具。

2017年国务院印发《新一代人工智能发展规划》文件中,列出了我国重点发展的人工智能技术,包括:知识计算引擎与知识服务技术、跨媒体分析推理技术、群体智能关键技术、混合增强智能新架构与新技术、自主无人系统的智能技术、虚拟现实智能建模技术、智能计算芯片与系统、自然语言处理技术。

人工智能是复杂的学科和技术体系,融合了移动互联网、大数据、超级计算、传感网、脑科学等技术。人工智能不同学派、领域、技术和应用互相交织,构成了不断发展的人工智能全景。

2 人工智能在语言服务行业的市场数据

语言服务行业是以促进跨语言、跨文化交流为目标,提供信息转换服务和产品,以及相关研究咨询、技术研发、工具应用、资产管理、教育培训等专业化服务的现代服务业(中国翻译协会,2016:16)。本文的研究重点是人工智能在语言服务企业的应用状况,语言服务企业是提供语言信息转换服务和提供语言技术服务的企业总称,语言服务企业是语言服务行业的重要组成部分。

为了全面了解人工智能在语言服务企业中的应用,先从国内外语言服务行业的市场调查报告揭示的数据进行分析,这些行业报告以语言服务企业为调查对象,从报告数据中可以揭示语言服务企业的人工智能应用状况。我们选择了最近三年(2017—2019)国际和国内发布的语言服务行业的调查报告,国外以卡门森斯咨询公司(CSA Research)的语言服务市场报告为例,国内选择中国翻译协会(Translators Association of China, TAC)的语言服务行业发展报告进行分析。

2.1 国际语言服务企业人工智能应用调查数据

卡门森斯咨询公司是提供全球语言服务研究、调查和咨询的公司,该公司每年开展全球语言服务企业的调查,根据调查研究结果发布全球语言服务市场(The Language Services Market)调查报告,在国际语言服务行业具有广泛的影响力。表1是CSA Research发布的2017—2019年国际语言服务企业机器翻译的译后编辑市场数据。表2是CSA Research发布的2017—2019年国际语言服务企业的语言技术应用数据。

表 1 国际语言服务企业机器翻译的译后编辑数据

类型 \ 年代	2017	2018	2019
提供 MTPE 服务的企业占比	29.86%	36.44%	36.53%
MTPE 的服务占比	3.74%	3.64%	3.81%

(数据来源:CSA Research, 2017,2018,2019)

表 1 表明,全球范围内,提供机器翻译的译后编辑的语言服务企业比例不断提高,从 2017 年的 29.86% 到 2019 年的 36.53%。但是,2019 年的数据与 2018 年相差无几,说明出现了增长放缓的现象;从全球语言服务企业的业务占比情况来看,机器翻译的译后编辑业务在波动中增长,2017 年机器翻译的译后编辑业务占全部业务的比例是 3.74%,2018 年小幅下滑到 3.64%,2019 年又回升到 3.81%。

表 2 国际语言技术服务企业的语言技术应用数据

类型 \ 年代	2017	2018	2019
机器翻译 (Machine translation)	12.7%	18.04%	17.6%
音频可视化 (Audio-visual tools)	2.00%	5.53	6.5%
语音处理 (Speech processing)	1.15%	0.91%	1.20%

(数据来源:CSA Research, 2017,2018,2019)

表 2 是 2017—2019 年 CSA Research 语言服务市场报告关于人工智能技术的应用数据,包括机器翻译、音频可视化和语音处理三部分。其中,音频可视化技术在全球语言技术服务企业的产品和服务中持续增长,从 2017 年的 2.00% 到 2019 年的 6.5%;语音处理技术虽然在 2018 年小幅下降,但是 2019 年又回升到 1.20%。机器翻译技术在经过 2017—2018 年的增长后,2019 年略有下降。值得注意的是,2019 年的 CSA Research 调查报告首次出现机器口译业务,占 0.01%,这个数据虽然不大,但是意义显著,表明机器口译技术是国际语言服务的新需求,表明机器口译技术已经开始商业应用。

2.2 国内语言服务企业人工智能应用调查数据

中国翻译协会是语言服务领域的全国性社会团体,自 2012 年开始中国翻译协会发布中国语言服务行业发展报告,行业报告以全面详实的数据,完整呈现了中国语言服务行业的现状,在国内外的语言服务行业具有积极的影响。根据中国翻译协会的调查报告,分析我国语言服务企业的人工智能技术应用及发展趋势,与 CSA Research 的调查报告进行对比,可以更清晰地了解中外语言服务企业人工智能的应用现状。

由于中国翻译协会的各个调查报告结构不一致,内容详略不同,不便于按年度对同一个问题进行对比分析。因此,我们针对我国语言服务企业的人工智能应用进行研究。图 1 是根据中国翻译协会 2016 年、2018 年和 2019 年的调查报告数据。

中国翻译协会的行业报告中,人工智能的应用表现在机器翻译和译后编辑方面。从图 1 可以发现,2016 年 17.2% 的语言服务企业使用了机器翻译,2018 年 27.1% 的语言服务企业使用机器翻译,2019 年 41.0% 的语言服务企业经常使用机器翻译。对比发现,机器翻译的应用呈现连续较大增长的现象。

具有自主研发翻译工具的语言服务企业中,使用机器翻译的译后编辑完成语言服务项目的情况如何

呢? 2018 年中国翻译协会的行业调查报告显示,42.7% 的语言服务企业的 21% ~40% 的项目使用机器翻译的译后编辑;22.5% 的语言服务企业的 41% ~60% 的项目使用机器翻译的译后编辑。这些数据说明机器翻译的译后编辑方式已经成为越来越多企业的翻译方式。

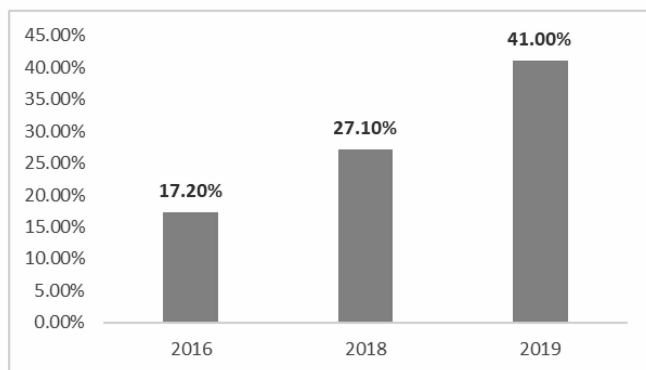


图1 我国语言服务企业机器翻译使用情况

(数据来源:中国翻译协会,2016,2018,2019)

3 人工智能技术在语言服务企业的产品应用

人工智能技术正在影响和改变翻译行业的生态面貌,改变着翻译职业能力,如翻译对象、翻译角色、翻译能力、翻译策略、翻译方式、翻译流程、翻译标准、翻译环境、翻译教育及翻译研究(孔令然等,2018:55)。传统翻译工作进行升级改造,才能提供高效专业的语言翻译服务。我国语言服务企业从数量到服务能力不断增长,越来越多的语言服务企业研发和应用基于人工智能技术的产品,这些产品在语言服务实践中应用越来越广泛,不仅提高了工作效率,而且提高了语言服务质量,体现了人工智能技术赋能语言服务的特征。

3.1 综合研究

根据企业核心业务类型分类,我国语言服务企业分为以提供翻译与本地化服务的企业(称为“翻译与本地化企业”)和提供语言服务技术和产品的企业(称为“语言技术企业”)。相比较而言,以开发语言服务技术和产品为核心业务的企业数量较少,国内还没有这方面的权威数据,而以翻译与本地化服务为主营业务的在营企业 9734 家(中国翻译协会,2019:2)。

人工智能在语言服务企业的应用是本文的研究主题,为此我们选择的语言服务企业满足三个条件:(1)公司具有知识产权的人工智能产品。(2)人工智能产品具有行业和专业领域创新特色。(3)人工智能产品已经在语言服务实践中广泛应用。

本文作者长期在语言服务企业和行业协会从事语言技术与管理工,与多家语言技术企业的高级管理者、企业技术部门保持密切沟通,多次前往这些企业进行培训、交流、研讨,对语言技术产品应用具有多年的使用体验。根据选择语言服务企业的三个条件,筛选出七家企业的产品,作为我国语言服务企业的代表产品,对其进行研究具有典型性、专业性、客观性。人工智能技术在语言服务企业产品的应用状况如表3所示。

这七家语言服务企业中,一者、优译、小牛、语帆可以归为语言技术公司,译讯、传神、精艺达可以归为翻译与本地化服务公司。从公司地理分布来看,这些公司的总部分别位于上海、成都、沈阳、北京、武汉和厦门,具有广泛的区域代表性。

表 3 人工智能技术在语言服务企业产品中的应用

公司与产品名称	产品功能	应用的人工智能技术
上海一者 YiCAT	在线翻译管理平台,翻译流程自动化、任务进度实时跟踪、集成计算机辅助翻译与机器翻译	深度学习、自然语言处理、机器翻译
成都优译译马网	在线翻译管理平台,翻译流程自动化、协同翻译、集成计算机辅助翻译与机器翻译	深度学习、自然语言处理、机器翻译智能机器人
沈阳雅译小牛快译 APP	在线实时翻译工具,支持文本翻译、语音翻译、拍照翻译、多种文档翻译	深度学习、自然语言处理、计算机视觉、人机交互、机器翻译、语音识别、图像识别、语音合成
北京语智云帆术语宝	在线术语提取、术语标注、术语分享与管理工具	自然语言处理、机器学习、机器翻译、智能决策
四川译讯译图	文档电子化的人工智能辅助格式转换与自动化排版	计算机视觉、图像识别、智能机器人、流程自动化
武汉传神云译客之 Twinslator	为译员打造个人专属机翻引擎 Twinslator(孪生译员)、译员信息采集、创建孪生翻译、自动采集译员翻译的语料、术语和翻译行为数据,训练机器翻译。	机器学习、自然语言处理、人机交互、智能机器人、机器翻译、流程自动化、智能决策
厦门精艺达九九译	在线人机交互的文档翻译、机器翻译、音视频听写、字幕翻译与视频压制	自然语言处理、机器学习、语音识别、机器翻译、语音合成、文本识别、字符识别

YiCAT 和译马网是在线翻译管理平台的代表将计算机辅助翻译、机器翻译、人工翻译相结合,将翻译技术与项目管理技术集成于一体。从应用的人工智能技术来看,深度学习的神经网络机器翻译技术、记忆库智能匹配技术、智能机器人的流程自动化技术和智能质量保证技术是在线翻译管理平台应用的主要技术。翻译记忆库和术语管理技术的底层是语言大数据技术,机器翻译技术的底层技术是自然语言处理技术。

小牛快译 APP 是快速口笔译 APP 的代表,支持文本翻译、语音翻译、拍照翻译、多种文档翻译,应用的人工智能技术包括深度学习的神经网络、机器翻译技术、语音识别、图像识别和语音合成等技术。

术语宝可以完成在线术语提取、术语标注、术语分享与管理,以良好的术语抽取效果,受到越来越多译者的欢迎。术语宝设计用到的人工智能技术包括自然语言处理、机器学习、机器翻译、信息抽取、语言大数据处理技术。

译图是语言服务的电子文档智能自动化排版产品的代表,可以完成文件格式转换和自动化排版,格式转换和排版质量和效率领先于同类产品。译图设计应用的人工智能技术包括计算机视觉的文字识别、模糊图像增强、智能机器人、流程自动化的智能排版。

云译客之 Twinslator(孪生译员)是人工智能技术在笔译翻译应用中的创新产品,该产品在线采集译员信息、创建孪生译员(Twinslator)、自动采集译员翻译的语料、术语和翻译行为数据,训练适合译员的机器翻译。Twinslator 应用到的人工智能技术包括机器学习、智能机器人、机器翻译、流程自动化和语言大数据技术。

九九译是为个人和中小语言服务需求公司提供在线人机结合翻译和本地化服务的产品,提供证件翻译、文档翻译、机器翻译、音视频语言服务。应用到的人工智能技术包括语音识别、机器翻译、语音合成、图像识别和机器学习。

以上七家公司的产品只是我国语言服务行业的代表,还有更多的人工智能语言服务产品没有在本文

中提到,例如,中译语通、文思海辉、语言桥、UTH 国际等公司的产品和服务都具有良好的应用和技术能力。

从这些产品设计应用的人工智能技术分类看,各个产品都集成了多项人工智能技术,它们共同服务于语言服务任务或项目需求。其中,自然语言处理、机器学习、计算机视觉是应用最广泛的人工智能通用技术。机器翻译、图像识别、语音识别是最常用的人工智能应用技术。

需要强调的是技术不是万能的,技术是完成任务的手段,技术、业务和人员结合是语言服务产品和人工智能技术发挥更大功用的有效形式。张爱玲等(2018:92-93)指出当前人工智能技术在笔译领域已经得到广泛应用,但是在同传口译中人工智能技术还不成熟。人工智能技术应用的深度和广度,不仅取决于技术本身的成熟度和产品质量,而且还受用户习惯和法律法规的制约,从这个意义来说,研究人工智能技术伦理可以更好地应用技术服务社会需求(任文,2019:51)。

3.2 个案研究

人类社会已进入人工智能时代,作为人工智能技术应用分支的翻译技术对翻译行业带来深刻影响(崔启亮,2019:82),前面分析的七个人工智能产品反映了这种影响。虽然七个产品各有特色,但是从人工智能技术在语言服务产品的创新应用来看,传神云译客的 Twinslator 是最具行业创新的产品,它的本质是通过深度学习和智能机器人等技术,训练和生成适合译者的机器翻译系统,Twinslator 工作原理如图 2 所示。云译客协作翻译平台可以让每一个志愿译者共同创建和维护语言和文化的多样性,共享翻译记忆软件,分享多语种平行语料,促进翻译在文化层面和技术层面上的发展(邵璐等,2020:108)。

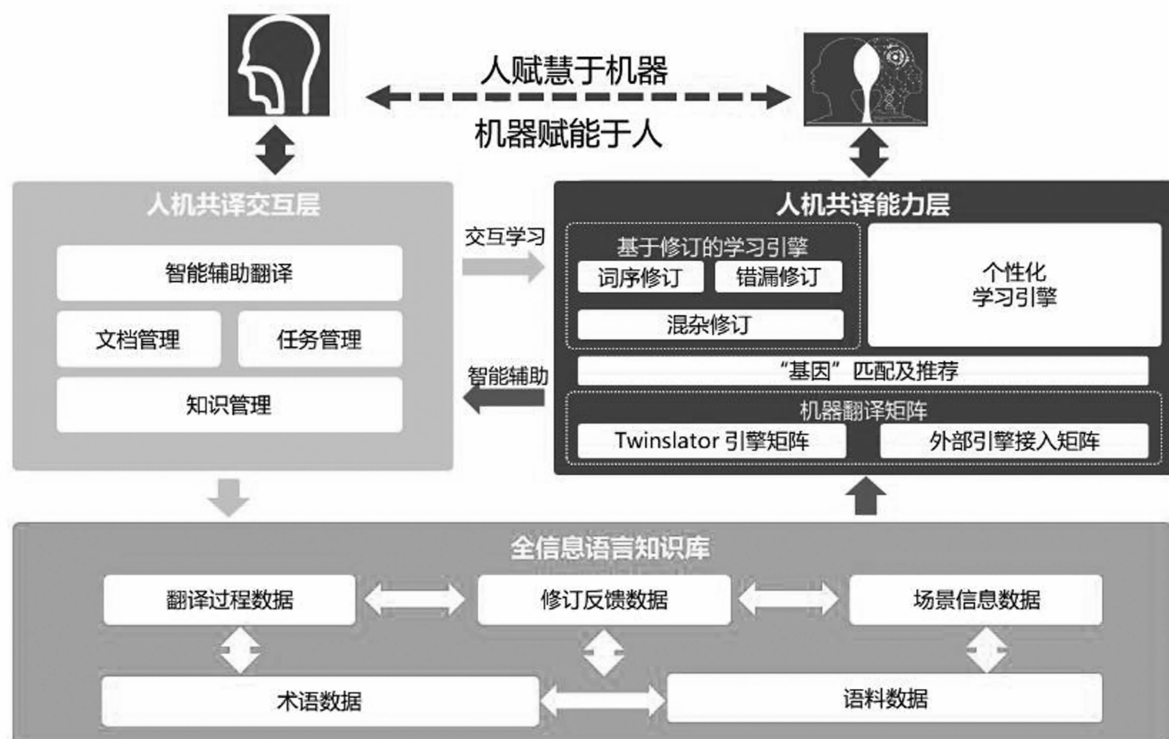


图2 传神孪生译员 (Twinslator) 工作原理

语言大数据是训练机器翻译的重要数据,早在2013年翻译自动化用户协会(TAUS,2013:41)即发布报告指出:“给予足够和适当的翻译数据,有可能在一天内为具体行业、组织乃至产品创建计算机翻译引

擎。”时至今日,这种预测正在成为现实。传神 Twinslator 以翻译过程数据和场景信息数据为语言知识库的知识来源,其中,语料数据和术语数据是语言数据的两种类型。它是以开放互联、数据共享和自定义机器翻译引擎为设计理念的创新产品,目标是实现机器翻译速度和人工翻译质量的平衡,译员加入后即可激活专属机翻引擎训练特权,拥有专属 Twinslator 机器翻译引擎。

Twinslator 可以根据译员擅长的行业和语种创建一个或多个 Twinslator,译者可以随时上传和维护翻译语料库和术语库,在线创建、编辑和保存翻译任务和译文。Twinslator 通过术语同步、语料同步以及在线翻译功能使用过程中产生的翻译特征及行为数据的自动采集实现译员 DNA 克隆。通过人机共译交互层和应用层的数据传输和处理,实现人赋慧于机器,机器赋能于人的目标。译员持续贡献的术语、语料以及翻译特征、翻译行为等数据经过数据分析和处理后,辅以云译客语言处理技术,实现 Twinslator 引擎的翻译质量从量变到质变,形成译员专属的机器翻译系统,不断提高译者翻译效率和质量。译员可以随时选择将 Twinslator 的译文与谷歌翻译、百度翻译、有道翻译和搜狗翻译等机器翻译译文进行对比。

Twinslator 是多种人工智能技术集中应用的产物,它集成了自然语言处理、深度学习、人机交互和智能机器人等技术,已经在翻译服务中取得了积极应用。至 2020 年 12 月 11 日,Twinslator 术语对总量 23764529 条,语料库 68304088 句对。其中,中英通用术语对为 97561 条,中英语料库 1195228 句对;工程行业的中英术语 86012 条,中英语料库 1277631 句对。应用领域涵盖建筑、文化、财经、电商、法律、文学、机械和信息技术等领域。Twinslator 的应用将机器处理的效率和人工翻译的创造性相互赋能,将改变传统人工翻译的方式,将译者、语料、术语、技术和管理融合,形成翻译服务新模式、新产品、新应用。

4 结语

人工智能技术在语言服务企业的产品中已经较为广泛应用。其中,自然语言处理和机器学习是应用最广泛的人工智能领域。机器翻译、图像识别、语音识别是语言服务企业产品中应用较多的具体技术。“AI 技术 + 流程 + 团队”推动语言服务模式变革,改变了语言服务的实施方式、服务能力、翻译行为和管理方式。随着智能化、数字化、自动化的快速发展,更多人工智能技术将在语言服务企业产品中得到应用,推动语言服务行业和企业不断发展进步,提供更智能、更敏捷、更专业的信息服务。

参考文献:

- CSA Research. 2017. *The Language Services Market: 2017* [R]. CSA Research.
- CSA Research. 2018. *The Language Services Market: 2018* [R]. CSA Research.
- CSA Research. 2019. *The State of the Market: 2019* [R]. CSA Research.
- TAUS. 2013. *Translation Technology Landscape Report* [R]. TAUS.
- 蔡自兴,徐光祐. 2003. 人工智能及其应用[M]. (第三版)北京:清华大学出版社.
- 崔启亮. 2019. MTI 翻译技术教学体系设计[J]. 中国翻译 (5):80-86.
- 冯志伟. 2018. 机器翻译与人工智能的平行发展[J]. 外国语 (6):35-48.
- 孔令然,崔启亮. 2018. 论信息技术对翻译工作的影响[J]. 北京第二外国语学院学报 (3):44-57.
- 任文. 2019. 机器翻译伦理的挑战与导向[J]. 上海翻译 (5):46-52,95.
- 邵璐,曹艺馨. 2020. 语篇 · 非语篇 · 语言资源:众包翻译的过程与产物[J]. 外国语 (3):102-109.
- 王华树,李智. 2020. 人工智能时代的翻译技术研究:内涵、分类与趋势[J]. 外国语言与文化 (1):86-95.

(下转第 73 页)