

# 自然语言处理的现状研究与未来发展初探

柴海伦

(杭州师范大学, 浙江 杭州 310000)

**摘要:** 随着科技领域的不断发展, 自然语言处理仍旧是目前科研以及人工智能领域需要研究的课题。作为最热门的课题之一, 研究其发展历程和方法、了解其技术原理, 并深入思考和探究其在各大领域中的现状以及应用程度也是非常有必要的。从而可以推测和观察出未来的发展方向和发展趋势。对我国的科技发展具有十分重要的意义。

**关键词:** 自然语言处理 自然语言处理现状 自然语言处理发展方向

**中图分类号:** TN941.1 **文献标识码:** A **文章编号:** 1003-9082 (2022) 09-0140-03

## 一、自然语言处理技术的概念、应用及任务

### 1. 什么是自然语言处理

自然语言处理是研究“如何使人和计算机二者之间采取自然语言并进行合理有效沟通”的集语言学以及计算机科学为一体的数据科学, 简而言之就是: 机器语言能和自然语言之间有效切换交流、沟通的一座桥梁, 用来实现“人机交流”的目的。它并不是一般的自然语言, 而是针对人工智能以及计算机科学领域的重要发展方向。它是数据科学领域最受欢迎, 也是最热门的课题之一。

### 2. 自然语言处理的应用

自然语言处理主要应用于: 信息搜索、机翻、问答自动化、语音识别、文本自动摘要以及情感分析等, 随着不断地发展, 将会有更多要应用的领域<sup>[1]</sup>。

### 3. 自然语言处理的核心任务

自然语言处理的核心任务就是: 期望人工智能机器如同人一样, 与人类能产生正常的语言和理解能力。将非人类语言的数据转换成人类能理解的语言格式。

## 二、自然语言处理的发展“轨迹”

### 1. 萌芽期

美国工程师韦弗最先提出自然语言处理技术中的“机器翻译”的方案。按当时社会环境而言, 自然语言处理最先产生于语言翻译领域, 在计算机未被发明出来前, 翻译工作都是由人工承担的, 但随着对自然语言处理的客观需求, 促使了人们对计算机语言翻译工作的提出和改进。

### 2. 发展期

20世纪中期, 法国数学家沃古瓦在原有的基础上, 创建出了一套全新完整的计算机翻译步骤, 并将其运用到了实际的法语和俄语的翻译工作中, 获得了较好的成果。但由于人类自然语言的不同, 以及即便是相同词语, 其表达出的意思也可能不同, 从而导致翻译出来的语句原意差异性

非常大。1974年, 英国人工智能专家Y.A.威尔克斯设计出的另一套翻译系统, 可读性较高最具代表性, 在当时的环境以及对未来的发展中具有非常大的突破性。

上述工作主要是自然语言处理技术的“机器翻译”。1972年, 维诺格拉德的SHRDLU系统将语言分析和知识推理结合在了一起, 这对于自然语言处理的研究方向迈出了一大步。同年, 伍兹提出扩充转移网络(ATN), 并建成了LUNAR系统。ATN还成了现今自然语言处理研究中广泛采用的方法。

### 3. 繁荣期

20世纪末, 自然语言处理技术终于从诞生走向繁荣, 并具有两大特征: 规模性大以及真实可用性强。规模性大意味对该技术的发展和改革需要有更深层次的要求; 真实可用性强说明对于文本处理内容需要更加的丰富。由此可见, 规模性和真实可用性二者缺一不可, 相辅相成。而该技术之所以能够不断改革和发展进入繁荣期, 也说明了人们的生活已离不开科技, 进而促使了网络技术的不断发展和壮大。而且计算机技术也必将会随着国际互联网的日益发展逐渐走向成熟。

## 三、“自然语言处理”的现状及其影响

### 1. 自然语言处理的发展现状

现阶段, 自然语言处理技术的主要发展趋向是: 人工智能、语言工程、数据处理以及科学认知。从目前来看, 除数据处理之外, 其他三类主要受实验室的限制, 而数据处理则有可能是未来应用方向发展最多的技术。

如今专业领域上的文本翻译、搜索引擎、文字录入等研究成果已经很大程度为人类提供了可靠的辅助性帮助。但随着日渐增长的迫切需求, 如信息服务、情报信息处理、国家安全和通信、网站内容管理, 语义表示与计算等, 针对这些方面的研究工作仍任重道远。目前正在进行的部

分研究也还缺少坚实的理论基础，从而暴露出很多问题。要想发展出真正更加实用的技术还需要很长时间，在此基础上也需要针对这些理论知识进行更深入的探讨。

### 2. 自然语言处理技术面临的现实问题

第一，跨模态语言理解（语言智能）存在局限性。机器人和人类对于常识性问题无法产生“共鸣”。人可以直接对手机发出具体指令，如“查找附近餐饮店”，手机助手就会标注出附近餐饮的全部位置。但若人对手机发出“饿了”的指令，手机助手则会无动于衷。就是因为机器人无法拥有和人类相同的常识，除非将所有常识性问题搬到系统中。但设计者不可能做到把所有人类常识问题总结出来并输入到AI系统中，这对目前来说是不现实的。换言之，目前的系统可能连三四岁孩童的语言和理解能力都达不到。

第二，低资源问题。对于标注数据资源缺乏的问题，自然语言处理还没有办法能够解决。针对这些不能被解决的资源问题，除却专业知识的加入整合，人工标注数据也不失为一个好方法，也能对数据加强有一定的帮助。这点是如何让自然语言处理技术变得更“广”的问题。

### 3. 自然语言处理对现实的影响

目测从现在以及未来很长一段时间，人类都将处于大数据时代的阶段。而大数据想要体现出其本身的价值，就肯定离不开人工智能技术以及机器的学习；同样人工智能想要体现出优势也必须以大数据为基础，二者相辅相成。很多大企业虽然都拥有属于自己的海量数据库，并运用人工智能技术进而展现出它的价值。但若单从数据量本身来看，或许目前最大的数据量应该是互联网上的茫茫网页，目前针对这些网页的利用率还是比较浅的，还具备很大的挖掘价值，每个网页几乎都有定量的文本内容，这些网页绝大部分得靠文本来展现其核心内容，这些文本内容都是自然语言。那么研究自然语言处理的价值就相当大了，只有更好的自然语言处理方法才能深度挖掘网页价值并创造出更大的价值。

根据目前情况来看，人工智能技术的发展情况很不错，最火的应用应属于图像、语言方向等，在文本处理方面相对还比较欠缺。因此，对于图像和语言方向无论是技术还是人才都将进入相对平缓的阶段，文本处理即将登上舞台。

机器理解自然语言的含义对人工智能来说是非常重要的，机器要实现智能化，若连人类语言都无法理解，如何才能体现出其智能之处，难道让两个机器之间交流信息便称这就是智能？总而言之，对于自然语言处理的研究会让大家的生活越来越方便，金融、电商或医疗等各大领域也

让自然语言处理技术得到了广泛的应用。

## 四、自然语言处理技术的研究方向

针对自然语言处理的研究方向以及涉及范围非常广。如信息提取、系统问答、机翻、文字校对或编辑、语音合成或识别等。

第一，提取信息。从特定的文本中提取出如人物、时间、地点、原因以及结果等重要信息。换言之就是要让机器了解某人在某时因某原因做了某事，以及产生了何种结果。

第二，文本生成。让机器人如同人类一般能够使用自然语言进行表达和写作。

第三，问答系统。人为通过对计算机提出问题，由计算机的问答系统利用自动搜索等方法作出精准的答案。这需要计算机对自然语言查询语句进行语义分析，从而在众多候选答案中找出最佳答案。

第四，对话系统。可通过与用户进行问答聊天完成任务。为了能更智能化需要系统具备多轮对话的能力。

第五，机器翻译。将源语言文本输入到系统内，从而自动获得另一种想要的语言文本。机翻的方法截止到目前已逐渐形成了比较严谨的一套体系。

第六，舆情分析。将收集到的海量信息，通过系统自动化对互联网上的舆论导向加以分析，以此来实现能及时应对舆情的目的。

第七，信息过滤。通过系统将符合条件的信息进行自动识别和过滤。如将互联网有害信息的自动化过滤，保护信息安全做好防护工作。

## 五、自然语言处理的技术应用

### 1. 采用NLP技术支持的AI系统追踪客户的“反馈信息”

为了了解顾客对自己店面的真实反馈并与顾客进行真实互动，美国一家连锁酒店使用NLP技术支持的人工智能系统进行网络跟踪，并根据顾客在互联网上留下的评论等反馈信息进行相应的整改。到目前为止，AI系统已经做出了95%的准确判断，帮助门店实现了良好的“沟通”效果。

### 2. 帮客户节省“一万小时”的工作时间

美国某IT集团每个月都会收到近10万个人站请求，出于这个原因，他们创建了“数字工作者”的智能机器，用来阅读这些电子邮件，读取并单独回复。对于更复杂的问题，则交给人类工程师解决。自该数字工人“上岗”以来，每月为工程师节省了近1万小时的工作时间，大大提高了工作效率。

### 3. 将NLP技术纳入医疗领域的应用开发

NLP语音识别技术正变得越来越成熟，因此被应用于医疗临床。这项技术与医学领域的深入结合使医生从繁重的

电子健康记录中解放出来，在护理病人方面更有效。

#### 六、自然语言处理的发展趋势

站在新时代的互联网“风口”，自然语言处理技术在未来几年将产生整体动态影响，发展趋势将不可限量。对于跨模态的融合以及对各专业领域的需求及解决方案，人机智能互动也将有突破性的变化。

##### 1. “符号表示”转变为“分布式表示”

自然语言处理是由词汇和符号体现的，因此当出现两个词性接近，但词形不同的词语时，计算机就会判定它们是不同的词语。这给现实中的应用带来了很大的不便。因此，如果在一个语义当中，改变传统的思路，采用组合词语相结合的方法，就能计算出不同级别的语言单元间的相似性。运用这种新型方法再进行深度学习也会带来很大的转变。

##### 2. 从浅到深的学习模式

浅层学习是按步骤走，可能还仅停留在低级认知的层次；直接的深层学习则是一步到位的端到端（end-to-end），而从浅到深的学习基于对浅层模型的学习。大部分语言信息用“稀疏”表示，从而会导致“维数灾难”类的问题；而“密集向量”表示则取得了较好的效果。这一大趋势是由词嵌入和深度学习模式的成功引发的。

##### 3. 从封闭到开放的NLP（自然语言处理）技术平台化

NLP技术非常复杂，不仅要处理与之相关的专业性问题，还要考虑到和其他领域相结合可能产生的问题，所以就显得更加琐碎。NLP技术领域的研究数据和程序从以前的封闭，程序员不愿分享成果，到如今的开放状态，都揭示了其使用门槛越来越低，无论是大型企业还是各大高校也都愿意提供更多的平台。从另一角度来看，NLP技术的发展也将会越来越好。

##### 4. 从人工创建到自动化构成的语言知识

美国某公司预做金融预测，结果只招聘与计算机和数学方面的人才。这恰好说明一个问题：计算机是运用现有算法解决存在的问题，并非是人类高手进行对决学习。从人工创建到自动化构建NLP技术领域，以前需要的大量显性知识，如今可采用自动化方法来构建，比如自动发现词汇与词汇之间的关系，像人类身上的血管一般融入各个方面。

##### 5. 从通用到场景化，将机器人与特定场景相结合

自从助手Siri“出道”后，国内也开始跟随潮流做语音助手，但因为“只能听得到，但却听不懂”所以很快就下马了，这也导致后续服务跟不上，实用性也不够强。如今是将特定场景和机器人结合在一起，进行人机对话的任务，非

常具有趣味性和实用价值。

##### 6. “浅层分析”过渡到“深度理解”的文本推理

谷歌推出的自动化测试机器人，已经识别并报告出广泛使用的项目中存在的漏洞。该技术的工作原理是：通过让其阅读并查找软件应用程序中的大量随机数据并进行分析其输出异常的问题，从而测试它是否能解决出现的BUG，反过来也能为开发人员提供可能存在的错误应用程序代码。这非常具有难度，但也证明了该技术的重要性。

##### 7. 从事实性文本分析到情感文本分析

目前文本情感分析已涵盖了文本挖掘、抽取信息、机器学习等多个领域，而且这一技术已经得到了较为广泛的应用，如商业和舆情方面。相较于事实性文本，文本情感分析更受群众欢迎。

##### 8. 从传统媒体“转型”到社交媒体

从传统媒体过渡到社交媒体说明互联网技术的不断发展，如今人们还会用社交媒体做股票以及票房的预测，它与日常生活更加息息相关。从长远角度来看，人文社会和互联网技术相结合更具有历史意义。自然语言处理作为最基础也最为重要的技术，其未来的发展前景自然是无限与广阔的。

##### 9. 从规范文本到自由文本

近几年文本生成作为NLP领域的另一大重要技术，从利用范本构建文本转变成如今的自由文本，研究价值和其意义都是非常大的。

##### 10. NLP与行业领域深度结合从而创造更大价值

随着自然语言处理技术的不断发展，目前已被广泛应用到各个行业中。因这些专业领域对NLP技术的需求非常大，所以NLP技术势必会发展得越来越好。

#### 结语

在如今21世纪的信息科技化时代，随着互联网的不断进步和发展，自然语言处理技术也会被不断的拓展到不同领域及应用，更会成为引领科技领域发展的焦点。它对我国科技乃至世界科技的进步和发展都具有极大的意义并具有深远影响。从人类日常生活到其他专业领域都离不开它。上述对于自然语言处理技术的现状研究以及未来发展均展开了深入探析，从中可以得知：对于自然语言处理的研究，未来更着重于与多领域的结合以及未来的实用性，此研究对未来的发展极具重要意义。

#### 参考文献

[1]包胜,谢为时,秦现德,等.基于城市信息模型的城市管理理体系架构[J].施工技术.2021(13):128,133.