自然语言处理的发展概述

#### **摘要：自然语言是指汉语、英语、日语等人们日常使用的语言，通常情况下是指一种随社会自然演化的语言，而不是人造的语言，可进一步解释为以语音为介质，由词汇和语法两部分组成的符号系统。自然语言对人类思维与交流的影响话意义往往是决定性的。随着人工智能的快速发展，自然语言处理在计算机领域的应用取得了很大的进步，对当代人类的日常生活产生了及其深远的影响。本文将从背景、应用方向和展望未来三个方面浅析自然语言处理。**

#### **关键词：自然语言处理，背景；应用方向;展望未来.**

### 背景

“自然语言处理可以定义为研究在人与人交际中以及在人与计算机交际中的语言问题的一门学科。自然语言处理要研制表示语言能力和语言应用(linguistic performance)的模型,建立计算框架来实现这样的语言模型,提出相应的方法来不断地完善这样的语言模型根据这样的语言模型设计各种实用系统，并探讨这些实用系统的评测技术。【1】

从20世纪40年代开始，人们开始研究自然语言处理，随着对机器翻译需求的增加，围绕自然语言处理这个命题，研究者进行了许多实践。1948年，香农在论文中概述了一种准确测量消息中信息量的方法，为定义数字时代的信息论奠定了基础。在实验过程中，研究人员不仅建立了语言的统计模型，还尝试利用现有的统计规则来生成全文。在离散马尔可夫概率模型的基础上，香农将该理论应用于语言描述并取得了成功。与此同时，美国数学家克莱因发展了有限自动机和正则表达式。1956年，美国语言学家乔姆斯基将他的上下文无关语法应用于自然语言处理。这一时期的科学家将自然语言处理分为两个不同的方向：基于规则和基于概率。

Chomsky在《自然语言形式分析导论》一文中,从数学的角度给语言提出了新的定义,指出:“这个定义既适用于自然语言,又适用于逻辑和计算机程序设计理论中的人造语言”。【2】  
 20世纪60年代，人们将自然语言处理作为人工智能的研究范畴。这一时期的自然语言处理研究分为两大阵营，即基于规则的符号学派和基于概率的随机学派。在此期间，两种不同的研究思路都取得了长足的发展。符号学派在这一时期进行了形式语言理论和高级句法的研究，并在60年代后期对形式逻辑进行了深刻研究。随机学派在这一时期也取得了很大的进步。他们将贝叶斯概率模型作为统计研究方法，进行了大量的研究。

### 进入21世纪，基于人工神经网络的深度学习成为人工智能领域最热门的研究方向。许多科技公司已经通过使用深度学习模型进行了图像和语音识别的研究。2013年，谷歌开源了一款词向量工具，可以在数百万词典和数亿数据集上高效训练自然语言处理。这一成就在自然语言处理史上具有里程碑的意义。随着深度学习技术的发展和对图形处理器硬件的不断突破，研究人员开始利用图形处理器更强大的计算能力来训练大型人工神经网络，在深度学习框架和图形处理器技术中，语音识别技术的精确度已达到90%以上。

2.应用方向

自然语言处理旨在设计算法使计算机像人一样理解和处理自然语言，是互联网和大数据时代的必然，涉及许多领域，包括聊天机器人、语音助手，语言翻译、情感分析、语法检查、搜索引擎等。随着通信和计算机相关技术的发展，自然语言处理的应用需求面会越来越广大。  
 聊天机器人是一种人工智能形式，经过编程可以使它们像人类一样与人进行语言互动。根据聊天机器人的复杂性，它们可以只响应特定的关键字，或者进行完整的对话，这使得人们很难将它们与人类区分开来。聊天机器人是使用自然语言处理和机器学创建的，这要求它们能够理解自然语言的复杂性和句子的实际含义，并且它们还能在与人类的对话中学习并随着时间的推移变得更好。聊天机器人主要通过两个步骤进行工作，首先，机器人能够确定所问问题的含义，并从用户那里收集回答问题可能需要的所有数据。然后，它们给予适当的回答。  
 近年来，语音助手风靡全球。无论是苹果的Siri、微软小冰，还是百度小度，人们可以方便地使用它们来拨打电话、发出提醒等。这些语音助手让生活变得更加轻松。它们使用语音识别、自然语言理解和自然语言处理的复杂组合来理解人类在说什么，然后采取行动。语音助手的长期目标是成为人类与互联网之间的桥梁，并提供基于语音交互的各种服务【3】。  
 谷歌翻译、百度翻译、有道翻译等都已经进入了人们的工作生活。虽然它们不能达到100%准确，但仍然可以实现文本从一种语言转换为另一种语言。谷歌翻译和其他翻译工具都使用序列到序列建模，这是自然语言处理中的一种技术。这种模型将一系列单词从一种语言转换为另一种语言，即翻译。早些时候，语言翻译人员使用统计机器翻译分析了数百万已经从一种语言翻译成另一种语言的文档，然后寻找该语言的常见模式和基本词汇。然而，与序列到序列建模相比，这种方法并不那么准确。  
 电商平台可以使用情感分析来了解特定类型的用户对特定主题、产品等的感受。很多互联网公司通过使用自然语言处理、计算语言学、文本分析等来了解用户对其产品的总体情感和服务，并找出情绪是好是坏还是中性。这样可以通过多种方式使用情绪分析，找出目标的情绪，了解评论，衡量他们的情绪等。除了商业领域，政府部门也可以使用情绪分析来寻找流行观点并发现对国家安全的任何威胁【4】。]

### 3.展望未来

在NLP领域模型层面，语言模型预训练方法在很多NLP任务中取得了突破，受到了各界的广泛关注。目前，神经网络的训练基本都是基于反向传播算法，通过随机梯度下降法等优化算法随机初始化网络模型参数和优化模型参数。预训练的思想是参数不再随机初始化，而是训练一个任务得到一组模型参数，然后用这些参数初始化模型进行训练。即通过在大量语料库上预训练语言模型，将预训练好的模型迁移到特定的下游NLP任务中，从而提升模型的能力。在应用层面，NLP的相关研究虽然比较抽象，但其最基本的研究仍然是语法、句法和语义的研究，重点是语言和文本。自然语言处理的难点在于理解语言不仅要靠逻辑，还要靠强大的知识库，需要相互支持才能更好地处理数据并对文本做进一步的理解和分析。未来，自然语言的发展趋势可能是从人工构建走向自动化构建，人们可以使用一些显性知识来构建一种方法来探索语言组件之间的关系，从而避免烦琐耗时的手动构建。在对自然语言的理解和推理层次上可以由浅入深，完成对语言的深层次理解。

参考文献：

[1] A.Bakushinsky&A.Goncharsky.III-posed Problems:Theory and Application[M].Dordrecht/Boston/London:Kluwer Academic Publishers, 1994

[2] 冯志伟.自然语言处理的历史与现状[J].中国外语,2008,5(1):14-22.

[3] 人工智能时代背景下自然语言处理技术的发展 林莉 长春工业大学人文信息学院

[4] 自然语言处理在情感分析领域应用综述 王颖洁1朱久祺1汪祖民1白凤波2,3弓箭3