

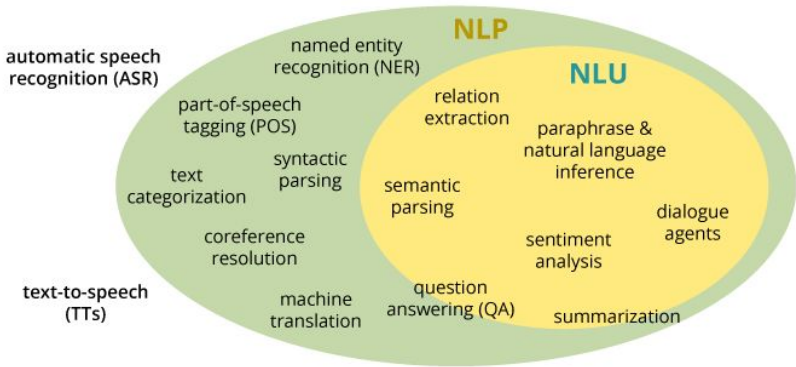
- 4G
- wifi
- 出门问路
- TechCrunch
- Square
- 智能音箱
- Spotify
- 华硕
- 奥迪
- HUD
- Arduino

随着人工智能的进步，相关技术变得越来越复杂，我们希望现有的概念能够包容这种变化 - 或者改变自己。同理，在自然语言处理领域中，自然语言处理（NLP）的概念是否会让位于自然语言理解（NLU）？或者两个概念之间的关系是否变得更微妙，更复杂，抑或只是技术的发展？

在这篇文章中，我们将仔细研究NLP和NLU的概念以及它们在AI相关技术中的优势。

值得注意的是，NLP和NLU尽管有时可以互相交换使用，但它们实际上是两个有一些重叠的不同概念。首先，他们都用来处理自然语言和人工智能之间的关系。他们都试图理解类似于语言之类的非结构化数据，而不是像统计，行为等结构化数据。然而，NLP和NLU是许多其他数据挖掘技术的对立面。

Terminology: NLU vs. NLP vs. ASR



来源: <https://nlp.stanford.edu/~wcmac/papers/20140716-UNLU.pdf>

自然语言处理

NLP是一个已经很成熟，已有数十年历史的领域。在计算机科学、人工智能、数据挖掘等领域进行交叉。NLP的最终目的是通过机器读取、解读、理解和感知人类语言，将某些任务用机器来代替人类来处理，包括在线聊天机器人，文本摘要生成器，自动生成的关键字选项卡，以及分析文本情绪的工具。

NLP的作用

NLP在广义上来说可以指广泛的工具，例如语音识别、自然语言识别和自然语言生成。然而，在历史上，NLP常用于：

- 符号化
- 解析
- 信息提取
- 相似度
- 语音识别
- 自然语言和语音生成等等

在现实生活中，NLP用于文本摘要、情感分析、主题提取、命名实体识别、词性标注、关系提取、词干提取、文本挖掘、机器翻译、自动问答、本体论、语言建模和我们可以想到的所有与语言相关的任务。

NLP技术

NLP的两个支柱是句法分析和语义分析。

总结：NLP依靠机器学习通过分析文本语义和语法从人类语言中获得意义。

自然语言理解（NLU）

NLP可以追溯到20世纪50年代计算机程序员开始尝试简单的语言输入。NLU在20世纪60年代开始发展，希望让计算机能够理解更复杂的语言输入。NLU被认为是NLP的一个子方向，主要侧重于机器阅读理解：让计算机理解文本的真正含义。

NLU实际上做了什么

与NLP类似，NLU使用算法将人类语音转化为结构化本体。然后使用AI算法检测意图，时间，位置和情绪等。但是，当我们查看NLU任务时，我们会惊讶地发现这建立了多少NLP概念：



自然语言理解是许多过程的第一步，例如分类文本，收集新闻，归档单个文本，以及更大规模地分析内容。NLU的实际例子包括从基于理解文本发布短命令到小程度的小任务，例如基于基本语法和适当大小的词典将电子邮件重新发送到合适的人。更为复杂的行为可能是完全理解诗歌或小说中的新闻文章或隐含意义。

总而言之：最好将NLU视为实现NLP的第一步：在机器处理语言之前，必须首先理解它。

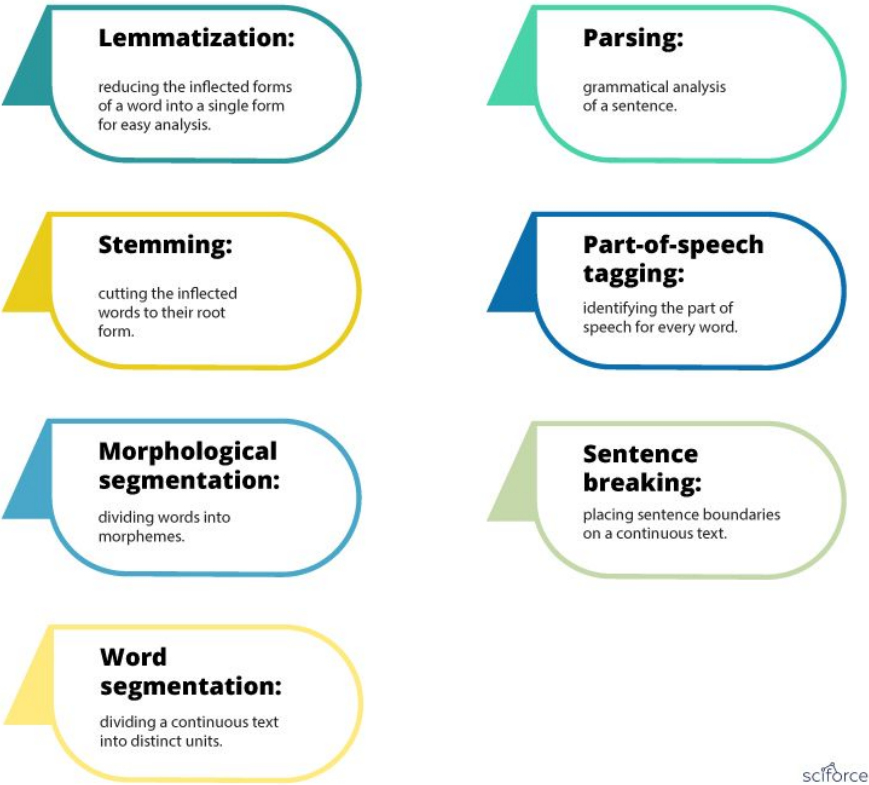
NLP和NLU的相关性

从其任务可以看出，NLU是NLP的组成部分，它负责人类理解某个文本所呈现的含义。与NLP最大的区别之一是NLU超越了解单词，因为它试图解释和处理常见的人类错误，如错误发音或字母或单词的颠倒。

推动NLP的理论是Noam Chomsky在1957年的“句法结构”中所设定的假设：“语言L的语言分析的基本目标是将L的句子的语法序列与不符合语法的序列分开。这不是研究语言L的句子，而是研究语法序列的结构。”

句法分析确实用于多个任务，通过将语法规则应用于一组单词并通过多种技术从中获得意义来评估语言如何与语法规则保持一致：

- 词形还原：将单词的变形形式简化为单一形式，以便于分析。
- 词干：将变形的词语切割成它们的根形式。
- 形态分割：将单词划分为语素。
- 分词：将连续文本分成不同的单元。
- 解析：句子的语法分析。
- 词性标注：识别每个单词的词性。
- 句子破坏：将句子边界放在连续文本上。



句法分析技术介绍

然而，语法正确性或不正确性并不总是与短语的有效性相关。想想一个无意义而又语法化的句子“colorless green ideas sleep furiously”的经典例子。更重要的是，在现实生活中，有意义的句子通常包含轻微的错误，并且可以被归类为不符合语法的。人工交互允许产生的文本和语音中的错误通过优秀的模式识别和从上下文中添加附加信息来补偿它们。这显示了以语法为中心的分析的不平衡性以及需要更加关注多级语义。

语义分析是NLU的核心，涉及应用计算机算法来理解单词的含义和解释，尚未完全解决。

以下是语义分析中的一些技巧，仅举几例：

- 命名实体识别（NER）：确定可以识别并分类为预设组的文本部分。
- 词义消歧：根据语境赋予词语意义。
- 自然语言生成：使用数据库导出语义意图并将其转换为人类语言。

然而，为了完全理解自然语言，机器不仅需要考虑到语义提供的字面意义，还需要考虑到预期的信息，或理解文本试图表达的内容。这个级别称为语用分析（pragmatic analysis），它刚刚开始引入NLU / NLP技术。目前，我们可以在一定程度上看到情绪分析：评估文本中包含的消极/积极/中立的感受。

NLP的未来

追求创建一个能够以人类方式与人类互动的聊天机器人的目标 - 并最终通过图灵测试，企业和学术界正在研发更多的NLP和NLU技术。他们想到产品实现的目标应该轻松，无监督学习，并能够以适当的方式直接与人们互动。

为实现这一目标，该研究分三个层次进行：

- 语法 - 理解文本的语法
- 语义 - 理解文本的字面意义
- 语用学 - 理解文本试图表达的意思

不幸的是，理解和处理自然语言并不是提供足够大的词汇量和训练机器那么简单。如果要取得成功，NLP必须融合来自各个领域的技术：语言，语言学，认知科学，数据科学，计算机科学等。只有结合所有可能的观点，我们才能揭开人类语言的神秘面纱。雷锋网(公众号：雷锋网)雷锋网雷锋网

想要继续查看该篇文章相关链接和参考文献？