

自然语言处理

2020 年 3 月 12 日

Springer

目录

- 1 自然语言处理 1
 - 1.1 简介 1
 - 1.2 语言模型 2
 - 1.3 词法分析 3
 - 1.4 语法分析 6
 - 1.5 主题模型 8
 - 1.6 对话理解 9
 - 1.7 智能客服 11
 - 1.7.1 检索型对话系统 11
 - 1.7.2 多轮任务型对话系统 12
 - 1.7.3 生成型对话系统 13
 - 1.8 自然语言处理技术展望 13

Chapter 1

自然语言处理

1.1 简介

自然语言处理（Natural Language Processing，简称NLP），属于计算机科学与语言学的交叉学科，所以又称计算语言学；它是用计算机来理解、处理、运用人类语言的学科。人类通过自然语言进行沟通协作，可以说如果没有语言人类的智能将无从谈起，它是人类区别于动物的重要标志。也可以说，只有当计算机具备了准确的自然语言的理解处理能力时，才算真正实现了人工智能。

- **研究内容：**NLP研究内容主要包括词法分析、句法分析、语义分析、篇章理解、机器翻译等。
- **应用场景：**NLP广泛应用信息系统方方面面。例如：手写体识别、光学符号识别、语音识别、语音合成、信息检索、机器翻译、对话系统等。
- **关联学科：**NLP紧密相关的研究领域包括机器学习、数据挖掘、知识图谱等；紧密相关的学科包括信息论、语言学、计算机科学等。

NLP面临的两大难题是歧义消解、未知语言现象。

- **歧义消解：**在自然语言的词法、句法、语义等各个层次中存在大量的歧义现象。比如“什么是一个词”，这就是NLP面临的一个难题。因为不同的人对词语粒度、标准有不同的理解。再比如，语言中存在大量一词多义的现象，在上下文语境中如何准确找到对应的词义，这些都是NLP研究面临的实实在在的难题。

- 未知语言现象：未知语言现象主要由两个方面的原因导致。第一点，人类语言一直处于不断演化过程中，同一个语言表达，在新的时空环境下，可能已经不再是以前的含义；而且，由于信息网络的发达，人们构造、传播新语言现象的能力大大增强，比如互联网上每天都在涌现新的语言词汇。第二点，在NLP研究中，由于整理收录能力、知识表达能力等现实因素的制约，实际中并没有一种可以准确、全面表达人类语言知识工程或工具。

自然语言处理研究的内容非常广泛，应用范围也非常广泛，如下举例一些常见的应用场景：

- 机器翻译：利用计算机实现自然语言（英语、汉语等）之间的自动翻译。
- 自动摘要：利用计算机自动地从原始文档中提取全面准确地反映该文档中心内容的简单连贯的短文。
- 文本分类：在预定义分类体系下，根据文本的特征，将给定文本于一个或多个类别相关联的过程。
- 情感分类：根据文本所表达的含义和情感将文本划分为褒扬或者贬义的两种或几种类型，是对作者倾向性、观点、态度的划分，因此也称倾向性分析。
- 信息抽取：从非结构化或半结构的自然语言文本中提取出于某个主题相关的实体、关系、事件等事实信息，并且形成结构化信息输出。
- 信息检索：用户输入一个表述需求信息的查询字段，系统回复一个包含所需要信息的文档列表。其核心技术在于索引构建和相关性计算。
- 问答系统：接受用户自然语言形式描述的问题，从大量异构数据中查找或者推断用户问题答案的信息检索系统。

1.2 语言模型

语言模型（Language Model）用于计算语言序列 w_1, w_2, \dots, w_n 的概率，数学表示为 $P(w_1, w_2, \dots, w_n)$ ，它是对语句的概率分布的建模。其最直接的应用就是判断一句话来自于人生成的语句的概率，例如在我们自然语言中，句子“我去吃饭”相比于“吃饭去我”的出现的概率更高，因此 P (“我去

吃饭”) $> P(\text{“吃饭去我”})$ 。讲到这里，最直接的一个问题就是，如何计算 $P(w_1, w_2, \dots, w_n)$ 呢？我们下面介绍一种最基本的语言模型：**n-gram**语言模型。

n-gram语言模型是一种最基础的语言模型。根据链式法则(Chain Rule)，公式 $P(w_1, w_2, \dots, w_n)$ 可以得到：

$$P(w_1, w_2, \dots, w_n) = P(w_1)P(w_2|w_1) \cdots P(w_n|w_1, \dots, w_{n-1})$$

其中的每一项 $P(w_i|w_1, \dots, w_{i-1})$ ，可以用以下公式来估计，即：

$$P(w_i|w_1, \dots, w_{i-1}) = \frac{C(w_1, \dots, w_{i-1}, w_i)}{C(w_1, \dots, w_{i-1})}$$

其中， $C(\cdot)$ 表示该序列在训练语料中出现的次数。但是，当序列长度很长时候，计算 $P(w_i|w_1, \dots, w_{i-1})$ 比较困难，一种常见的处理方式是引入马尔可夫假设(Markov Assumption)，即假设当前词出现的概率只依赖于前 $n-1$ 个词，也就是：

$$P(w_i|w_1, \dots, w_{i-1}) = P(w_i|w_{i-n+1}, \dots, w_{i-1})$$

根据 n 取值的不同，我们可以得到不同的**n-gram**语言模型：

- Unigram: $P(w_1, \dots, w_{i-n}) = \prod_{i=1}^n P(w_i)$
- Bigram: $P(w_1, \dots, w_{i-n}) = \prod_{i=1}^n P(w_i|w_{i-1})$
- Trigram: $P(w_1, \dots, w_{i-n}) = \prod_{i=1}^n P(w_i|w_{i-1}, w_{i-2})$

在计算**n-gram**时候，一个很重要的问题就是测试集中出现了训练集中未出现过的词而导致语言模型计算出的概率为零，我们称这些词为未登录词(OOV)。平滑(Smoothing)技术就是为了缓解这类问题，常见的平滑技术有：拉普拉斯平滑(Laplace Smoothing)、古德图灵法 (good-turing)、线性减值法 (Linear Discounting) 等，感兴趣的读者可以深入阅读相关论文。

1.3 词法分析

词法分析是自然语言处理的第一步，是做NLP深层次分析的先决条件。词法分析的核心是，将自然语言解析为词序列，并判断每个词的词

性、专名信息，为后续分析做好准备。词法分析主要由分词、词性标注、命名实体识别3个子任务组成。

- 自动分词（Word Segmentation）：是将连续的自然语言文本，切分成具有语义合理性和完整性的词汇序列的过程。
- 词性标注（Part-of-Speech Tagging）：是指为自然语言文本中的每个词汇赋予一个词性的过程。
- 命名实体识别（Named Entity Recognition，也称专名识别）：是指识别自然语言文本中具有特定意义的实体，主要包括人名、地名、机构名、时间等。



图 1-1 词法分析示例

如图1-1所示，将输入句子切分成词序列，然后给每个词汇标记出名词、动词、介词等词性；并且识别出“2003年10月15日”是时间，识别“杨利伟”是人名等。因为词法分析的自动分词、词性标注、专名识别本质上是类似的，所以它们的发展历史方也是类似的。总结起来，都大致经历词典匹配、机器学习、深度学习这3个发展阶段。基于词典匹配的优点是实现简单、运行速度快；缺点是面临词典收录更新困难、未登录词难处理；同时由于消除歧义策略过于简单，通常效果不甚理想。基于机器学习的序列标注模型中，条件随机场（Conditional Random Field, CRF）[?][?][?]是集大成者。相比于隐马尔科夫模型（Hidden Markov Model, HMM）[?][?][?][?]，CRF去除了输出独立性要求，对于整个序列内部的信息和外部观测信息都可以有效利用，可以更加有效建模上下文。其优点是通过人工设计的特征工程，充分地挖掘了序列的上下文信息。模型的歧义消解胜过词典匹配的方法；同时具有很强的泛化能力，能够很好地处理未登录词问题。其缺点实特征工程不经需要耗费大量人力，而且需要大量语言学知识，设计和寻找有效特征存在较高门槛；另外，这些浅层模型，通常使用离散的binary特征，无法表达复杂先验，比

如没法利用词向量。随着大数据、神经网络、深度学习的快速发展，很多研究提出利用前馈神经网络来解决词法分析[?][?]任务。总的来说，与前面基于机器学习的方法类似，也是把词法分析任务作为序列标注问题进行求解，只是把人工设计特征函数，改成了使用多层前馈神经网络进行自动特征抽取。后续有很多研究，深度学习框架基础上，对前馈神经网络进行改进，进一步探索了卷积神经网络[?]、循环神经网络[?]、递归神经网络[?]等复杂结构对词法分析任务的影响。总结起来，基于深度学习方法的优点是深度神经网络自动学习多层特征抽象，避免了复杂的特征工程；模型的歧义消解、泛化能力通常都很好。只是，深度学习模型需要训练充分，通常对数据量、计算量都有较高的要求。

通常NLP中深层次的语法语义分析通常都是以词作为基本单位，所以词法分析都是这些深层次分析的基础和先决条件。词法分析作为自然语言处理的第一步，是下游所有分析任务的基础。这些下游应用小句法分析、语义分析，大到对话系统、文本分类、自动摘要、机器翻译[?]、信息检索、搜索引擎、语音合成等等。几乎只要有利用到自然语言处理技术的地方，词法分析都是不可或缺的基础技术。比如在搜索引擎中，用户输入查询，系统回复包含所需要信息的文档列表。其核心技术在于索引构建和相关性计算。索引构建：需要对文档进行，分词、词性标注、专名识别，为所有文档建立倒排索引。相关性计算：需要对query进行词法分析，然后以词为粒度计算query和文档的相似度。

1.4 语法分析

短语结构语法（Constituency Structure Grammar）和依存关系语法（Dependency Grammar）是现在常见的两种语法关系。短语结构语法又叫上下文无关文法（Context-Free Grammars, CFGs），它从一个特殊的初始符号出发，不断的应用一些产生式规则，从而生成出一个字串的集合（如句子）。产生式规则指定了某些符号组合如何被另外一些符号组合替换。它呈现一个树分类关系，句法根据一定的规则进行转换分析。每一个词的转换都是需要按照设定的树值规则进行目的性的转换。

依存语法(从属关系语法)是由法国的语言家Lucien Tesnière提出的[?]，它将句子各个词语之前的搭配关系描述成预先定义好的依存关系。

它基于一个基本假设：句法结构本质上包含词和词之间的依存（修饰）关系。一个依存关系连接两个词，分别是核心词（head）和依存词（dependent）。依存关系可以细分为不同的类型，表示两个词之间的具体句法关系。比如主语依赖于谓语（SBV），宾语也依赖于谓语（VOB）以及定语依赖于名词性短语（ATT）等。依存句法认为“谓语”中的动词是一个句子的中心，其他成分与动词直接或间接地产生联系。除了以上介绍的两种句法体系外，国内外都开展了对句法分析的研究。不论是国外的链语法(Link Grammar)、组合范畴语法（Combinatory categorial grammar, CCG）等，还是国内黄曾阳教授提出的HNC理论(Hierarchical Network of Concept)[?]都是目前行业内常用的语法，只是由于设定区域的不同，所以使用有一定的局限性。

短语结构分析的语法集是由固定的语法集组成，较为固定和呆板，依存语法则更加的自由。另外依存语法树标注简单且parser准确率更高，再加之通用依存数据集（Universal Dependencies Treebanks）的发展，依存语法分析受到专家学者普遍的青睐，得到越来越多的应用。这里也将着重介绍依存语法分析。

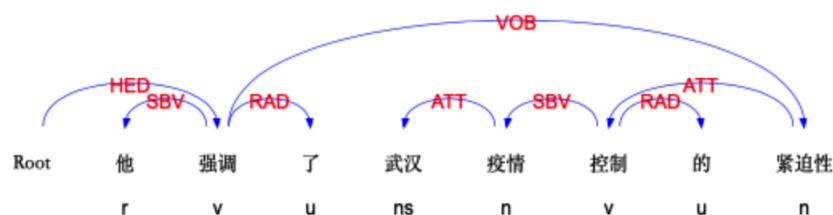


图 1-2 LTP依存句法分析样例

例如，从上述例子中我们可以看到，句子的核心谓词为“强调”，主语（SBV）是“他”，强调的宾语（VOB）是“紧迫性”，“紧迫性”的修饰语（ATT）是“武汉疫情控制”。

依存句法分析方法主要可以分为两种，一种是基于图的方法（Graph Based），一种是基于转移（Transition Based）的方法。基于图的方法先建立句子中所有词语的全连接图，然后求图中的最大生成树。两种方法中更主流的算法是基于转换的依存句法分析，基于转移的方法将依存树的构成过程建模为一个动作序列，将依存分析问题转化为寻找最优动作序

列的问题。基于转移的解析过程是线性的，动作步骤随句子长度线性增长，而基于图的方法需要在全图上做搜索，所以时间复杂度上基于转移的方法会有优势。但是基于转移的方法在解析的每一步都只是利用局部信息，会导致错误传播，性能比基于图的略差。近几年，分别出现了针对这两种不同方法的神经网络模型 [?????]。

依存语法分析在信息检索，评价抽取和情感分析等NLP任务上都有很多应用。比如“谢霆锋的儿子是谁”和“谢霆锋是谁的儿子”两句话如果不使用依存语法分析，很有可能就返回了一样的结果。依存句法分析能够更直接地通过语法结构的规则约束筛选出可能正确结果，提升相应任务的准确性。又比如“深圳大学非常漂亮，学生都很聪明”，这里“漂亮”形容的是“深圳大学”，“聪明”形容的是“学生”，通过依存句法分析，就可以抽取对应的搭配。再比如“我家音响声音很大”和“我家洗衣机声音很大”，两者在情感上前者是正评价，后者是负评价，需要使用依存句法分析来识别“声音很大”的修饰对象。

1.5 主题模型

主题模型 (Topic Models) 是近些年来非常重要的一项NLP技术，它被广泛应用于工业和学术界。在主题模型中，我们一般用 d 来表示要分析的文档，例如一篇文章或者一个网页等，而一个文档 d 通常由一系列词 (w_1, w_2, \dots, w_n) 组成，其中 w_n 是文档中的第 n 个词。多份文档共同构成了我们要分析的语料集，我们用 \mathcal{D} 来表示， $\mathcal{D} = (d_1, d_2, \dots, d_m)$ 组成，其中 d_m 是语料库中的第 m 个文档。主题一般用 z 来表示，它由一些词组成，同时也有该词在这个主题下的概率。主题模型泛指由一类可以从语料库中抽取主题并利用这些主题表示文档的模型，常见的主题模型有PLSA, LDA，以及各种LDA的变种，例如SentenceLDA等。在熟悉了这些基本概念之后，我们通过一种常见的主题模型Latent Dirichlet Allocation (LDA)来认识主题模型。

2003年Blei等人在《Latent Dirichlet Allocation》[?]一文中提出了LDA模型。如图1-3所示，其中空心节点表示隐藏变量，实心变量表示客观观测变量，整个模型具有 K 个主题， M 个文档和 N 个词。LDA将文档的主题分布 $P(z|d)$ 看做随机变量 θ ，并且假设 θ 从一个狄利克雷先验中产生。同

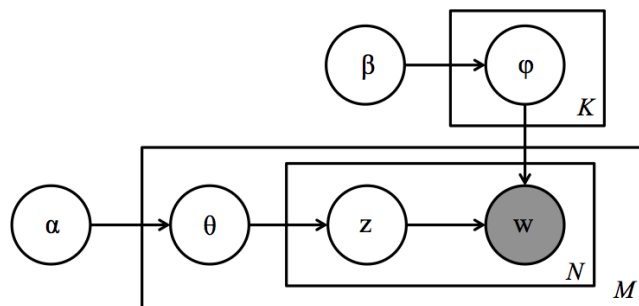


图 1-3 LDA图模型

时，由于训练数据之外的文档对应的主题分布 θ 可以从上述狄利克雷分布中产生，训练数据之外的文档的 θ 可以更自然地进行计算。读者可以使用常见的开源主题模型工具如LightLDA[?]、Familia[?]来尝试训练LDA模型。近年来，出现了各种对主题模型的扩展工作，例如联邦学习场景下的主题模型[?]工作等。

1.6 对话理解

对话理解是指希望计算机跟人一样，具备自然语言理解的能力，从对话内容中挖掘对话意图，理解对话意图，用户情绪识别等。例如，在客服与用户交互的对话中，用户询问“今天的天气如何”，这里就是一个“询问天气”的意图。这里对话可以包含语音对话和文本对话，如果是语音对话，我们一般可以利用语音自动识别技术将语音转为文本。后续我们要讨论的内容是文本的对话理解。

一般而言，文本的对话理解从技术角度上可以分为两类：文本匹配和文本分类。

文本匹配 文本匹配的目标是得到 $f(text_1, text_2)$ 的语义匹配得分，其中 $text_1$ 和 $text_2$ 是输入的文本， f 是文本匹配模型。

文本分类 文本分类的目标则是得到 $g(text)$ 的类别标签，其中 $text$ 是输入的文本， g 是文本分类模型。在任务型对话中，类别标签就是意图。

下面以智能质检为例，阐述对话理解相关技术是怎么应用的。智能质检使用人工智能算法，分析坐席呼叫场景下人工客服与客户的对话，实现全量质量检查，提高人工客服的服务质量和客户的满意度。智能质检系统的输入是一通人工客服和客户对话的录音，输出是质检报表，显示该录音在不同质检项的合格情况。质检项的重要性通过质检项的分数来决定。智能质检无需人工介入，节省质检人力，覆盖率高（100%），提升质检效率，降低漏检错检率。

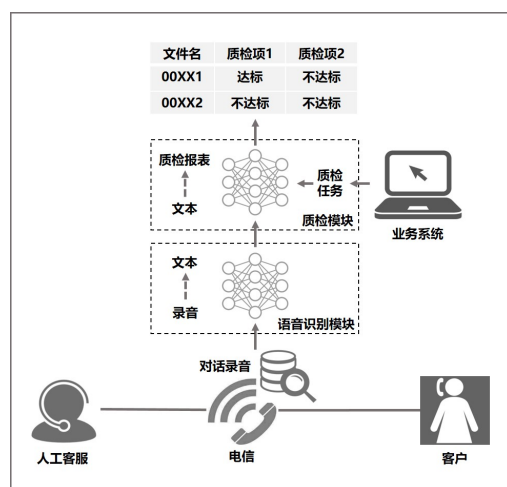


图 1-4 智能质检基本流程图

从智能质检的基本流程图中我们可以发现，对话录音经过语音识别模块之后，我们得到了客户和人工客服之间的对话文本。质检员配置了质检项之后，我们将对话文本输入质检模型，最后得到了质检报表。

假设我们定义了质检项要求客服在对话中“核实用户的工作地址”，比如“你的公司地址在哪里”。在冷启动时，一种简单的实现方案是通过规则配置“算子+逻辑操作符”或者正则表达式，如果录音文本中满足匹配条件，则命中该质检项。如果我们有标注数据，就可以使用文本匹配和文本分类的方法。计算 $f(\text{录音文本}, \text{质检例句})$ 的语义匹配得分或者 $g(\text{录音文本})$ 的质检项标签。我们通过数据驱动的方法让模型越来越聪明，业务方只需要提供标注数据就能进行质检，不需要人工定义规则，模型具有一定的泛化能力。但遇到bad case没有基于规则的方法容易修复，另外需要标注数据积累到一定规模才能发挥模型的优势。

当前智能质检的应用可以包含离线质检和坐席实时质检。离线质检是指结合语音识别和自然语言处理技术，对海量录音数据进行批量的智能化分析。离线质检在质检过程无需人工介入，可以提供内容质检，敏感词识别等质检结果。坐席实时质检是指在人工客服和客户通话过程中，提供实时质检功能，辅助人工客服判断客户情绪和实时分析对话过程的信息，及时提醒人工客服从而使客户获得更好的服务。现在智能质检的产品形态包含SaaS云服务和私有化部署。SaaS 级产品部署，让中小企业也能够享受智能质检带来的高效与便捷，克服了采购费用高部署周期长的问题。未来随着多方业务的使用，可以基于联邦学习进行智能质检，在满足数据安全和私隐保护的前提下，通过模型的参数梯度共享，获得了把所有数据放在一起训练的效果，使得不同的业务方合作共赢，建立更准确的数据模型。

1.7 智能客服

现在越来越多的交易和交流都可以通过网上平台完成，如果顾客在使用产品中碰到问题，他们可以在随时在网上提问，跟客服交流。如果传统公司雇佣一大批客服专员24小时值班，那导致的人工成本会非常高昂。智能客服系统使用人工智能技术进行问题语义分析，答案查找，可以自动完成问答，能提高顾客的用户体验，大大降低公司所需客服专员数量。客服系统假设客人和客服专员通过文字进行交流。智能客服系统会根据用户输入的问题和当前聊天语境，输出对应的文字解答。根据客服系统的功能不同，一般分为检索型对话系统，任务型对话系统[?]和生成型对话系统[????]。

1.7.1 检索型对话系统

顾名思义，检索式模型就是查找问题库中跟当前问题最相似的问答对，输出答案。检索型对话系统适用于能一句话就能描述清楚的问题，而且答案比较固定不会经常变动的情况。这样的例子包括“你们公司的地址在哪里？”和“你们的咨询电话是多少？”这类简单问题。检索式对话系

统一般包括问题库和问题匹配模型两个部分。问题库记录了常见问题和标准的回答，而匹配模型会计算用户的输入问句跟问题库中的所有问题之间的匹配度，系统输出最高的问题的标准答案。检索式对话系统的核心在于问题匹配模型，就是如何计算用户输入问题跟标准问题库之间的相似度。读者可能很容易就可以想到的一种方法就是对所有的输入问题，经过分词，提取TF-IDF特征，然后使用两个问题的特征向量来计算问题之间的相似度。由于自然语言中，问题本身的表述形式非常多样，难以穷尽。这种人工特征的性能有很大改进空间，经常对同一种意思的几种不同说法不是太鲁棒。基于深度学习的句子语义匹配，可以在训练集模型上学习把意思相近，可是说法不一样的一对句子映射到相似的语义空间中去。基于深度学习的句子匹配模型复杂，一般来说需要大量的训练数据才能训练好。有一些工作提出了利用大数据上预训练的NLP模型来对深度语义模型进行初始化，取得了良好的效果。检索性对话系统的缺点包括问题库更新困难，标准答案不能实时变动等。

1.7.2 多轮任务型对话系统

多轮任务型对话系统[?]，又称填槽型对话系统，会通过提问，来引导用户提供完成当前任务所需的多个关键信息“槽位”，进而帮用户完成复杂任务。多轮任务型对话系统适用于不能用一句话描述清楚的，需要多轮交流才能完成的问题，这样的例子包括购买机票，预定酒店等。

多轮任务型对话系统跟单轮系统不一样的是，任务型系统需要考虑对话的状态。对于一个任务，设计者需要指定完成要获取哪些槽位信息，而在对话过程中，每一个信息的获取情况，就可以认为是对话的状态。例如在定机票任务中，需要获取的槽位包括“起飞时间”，“起飞城市”，“到达城市”等，而对话状态就是这些槽位的获取状态。

多轮任务型对话系统包括四个模块，自然语言理解模块，对话状态跟踪模块，对话策略模块和自然语言生成模块。口语理解模块自然语言理解模块会根据对话的每一轮输入，从用户的问题中识别出用户的意图和槽位值。对话状态跟踪模块会根据当前的对话记录，计算最新的对话状态。对话策略模块会根据最新的对话状态计算系统的回复动作和槽位值。自然语言生成模块会根据回复动作和槽位值生成回复的句子。

1.7.3 生成型对话系统

生成型对话系统[?]，会根据用户的提问逐词地生成合适的回复。生成型系统最灵活，多用于对用户的问题进行闲聊式回复。目前比较常见的方法是编码器解码器模型，还有很多新的方法使用了记忆模块，注意力模块等技术来改进生成型模型。生成型对话很灵活，不过由于模型参数多，一般需要比较多的训练数据来进行训练。后面有一些使用预训练模型来减少模型所需训练数据的工作。除此之外，生成型对话系统容易产生不包含任何信息的“安全回复”，例如“我不知道”等。

1.8 自然语言处理技术展望

随着算法的持续改进以及数据和算力的持续增长，自然语言处理技术在最近十年间取得了长足的进步，基于深度神经网络的模型在多个自然语言处理任务中刷新了历史记录。这些新兴的自然语言处理技术在改善了传统自然语言处理任务的效果的同时，也极大地促进了一些相关领域的发展。例如，在信息检索领域，自然语言理解技术能够帮助搜索引擎更准确地识别用户的潜在需求；文本摘要技术能够帮助搜索引擎更好地刻画文本中的重要信息；自然语言对话技术有极大的潜力带来搜索引擎和用户交互方式的深刻变革。在语音识别领域，最新的语言模型技术(例如BERT、XLNet等[?])能够帮助语音识别系统更好地控制识别结果在语法和语义层面的合理性；通过自然语言理解技术和语音识别系统的无缝结合，智能音箱等设备正在逐步走进人们的日常生活，也重构了AIOT行业的生态版图。

然而，当前的自然语言处理技术还面临着诸多挑战:

- 基于深度神经网络的自然语言处理模型往往是Model-blind的，其效果严重依赖大量的数据和超强的算力，这通常会导致极高的模型成本，也极大地限制了自然语言处理技术的公平性和普惠性。因此，以贝叶斯程序学习 (Bayesian Program Learning) [?]为代表的Model-based小样本学习正在逐步引起研究人员的关注。通过引入领域相关的先验知识，贝叶斯程序学习等技术可以有效降低模型训练中对于大量训练数据的依赖，节省数据标注成本和算力开销。

- 当前的自然语言处理技术通常只能发现数据之间的关联(Association)而不能发现数据之间的因果关系(Causality)。例如，当前的对话系统技术虽然可以生成一些似是而非的语句，但是从多轮的语境来看，这些机器生成的对话往往存在非常严重的逻辑问题。导致这种现象的重要原因是神经网络模型通常只是对训练数据做曲线拟合，缺乏对数据背后的复杂关系的深刻理解。最近日益受到科研人员重视的因果推理(Causal Inference) [?] 技术有助于人们对数据进行因果关系分析，该技术和传统自然语言处理技术的深度结合有潜力使自然语言处理在认知层面更上一层楼。
- 随着诸多类似欧盟General Data Protection Regulation (GDPR) 法规的陆续出台，数据监管正在变得日益严格，单一研发机构获取大量的训练语料变得越发困难。如何在合规前提下统筹利用行业内的数据助力自然语言处理的研发任务是当前需要研究的重要课题。以联邦学习(Federated Learning) [?] 为代表的技术可以在保护各个研发机构数据隐私的前提下，提供一种挖掘跨机构数据背后的价值的计算范式。通过应用联邦学习这种计算范式，整个行业可以有效解决数据孤岛问题，打造更强大的自然语言处理模型，实现一个互利共赢的业界生态。

综上所述，自然语言处理技术的发展同时面临着机遇与挑战。我们在体验着近年来自然语言处理的飞速发展带来的技术红利的同时，也期待在可见的将来该领域能在上述三个方面取得可喜的进展，从而带来自然语言处理相关行业的又一次腾飞。

