

# 突破句法分析瓶颈、跨越语义鸿沟研究报告

19S103256 文荟

**摘要** 句法分析是自然语言处理中的关键技术之一，也是目前该领域研究的主要瓶颈。本文概述了众多学者在句法分析和跨越语义鸿沟相关方面的工作，主要包括五部分内容：(1) 语义鸿沟的含义，介绍了句法分析目前存在的问题和挑战和应对它的三种思路；(2) 规则和统计结合在跨越语义鸿沟的应用；(3) 深度学习在跨越语义鸿沟的应用；(4) 强化学习在跨越语义鸿沟的应用；(5) 未来研究方向，结合当前现状，引出未来的研究方向以供参考。

**关键词：** 句法分析 跨越语义鸿沟 规则统计结合 强化学习 深度学习

## 0 引言

宗成庆等学者认为，自然语言处理的最终目的是在语义理解的基础上实现相应的操作[1]。一个自然语言处理系统，如果完全没有语义分析的参与，能够获得很好的性能是不可想象的。然而，自然语言的语义计算问题十分困难，如何模拟人脑思维的过程，建立语言、知识与客观世界之间可计算的逻辑关系，并实现具有高区分能力的语义计算模型，至今仍未解决。

按照处理深度，可以将语言分析技术分为浅层分析和深层分析两类[2]。浅层分析技术主要是词汇级的处理，一般只对句子的局部进行分析，这类技术目前已经趋于成熟，如分词、命名实体识别等，已经对文本检索、文本分类、信息检索等应用产生了实质性的帮助。深层分析技术是对语言进行的语法、语义、语用级别的处理，如句法分析、语义角色标注等，这些技术需要对句子进行全局分析才能得到正确结果。目前，除开部分采用深度学习方法的应用，基于传统自然语言方法的深层分析技术距离适用实用仍有较大距离。

在深层分析技术中，句法分析处于十分关键的位置。它的基本任务是运用句法规则和其他知识将输入句子中词之间的线性次序，变成一个非线性的数据结构[3]，如短语结构树或有向无环图。具体看语言计算问题的每个阶段，以英汉翻译为例，如图1所示，步骤一般分为：原文输入、词法分析、词性标注、句法分析、语义及语用分析、生成目标形式表示等[4]。

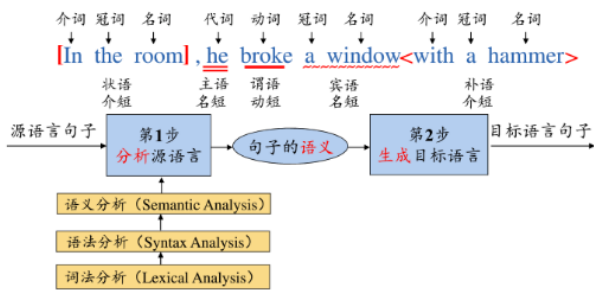


图1 英汉翻译步骤图

句法分析是语义分析(更深层次的语言理解)的必经之路，位于句法分析底层的词法分析技术已经基本成熟，之后的语义分析技术又建立在句法分析的基础上。因此，句法分析已成为当前自然语言处理的主要阻碍，如何突破瓶颈、跨越语义鸿沟是自然语言工作者们亟需解决的问题。

本文后续的篇章组织如下：第一部分介绍语义鸿沟的定义、产生背景和解决它的三种思路，第二到第四部分分别讨论这三种方法(规则和统计结合、深度学习、强化学习)在解决语义鸿沟的研究现状，第五部分介绍未来研究的方向。

## 1 语义鸿沟

### 1.1 语义鸿沟定义

语义鸿沟是指用两种不同的语言去表示两种对象的描述之间的差异。Hein 对其作了精确定义，即“不同表征系统的结构含义的差异”[5]。因此，根据定义，语义鸿沟现象不仅在自然语言处理中有所体现。事实上，它在图像识别领域也存在，而且相关的研究更甚。具体到自然语言处理领域而言，是将人类可读语言解析成机器可读语言之间的鸿

沟[6]。这就导致了计算机处理信息需要经过词法分析、句法分析等,才能进入语义和语用级别的分析。

## 1.2 语义鸿沟产生背景

句法分析除了为词义排歧、语义分析提供技术支撑外,音字转换、机器翻译、信息检索等许多本领域的问题,都需要依赖精确的句法分析才能最终得到理想的结果。然而,自20世纪50年代初机器翻译问题被提出起,句法分析一直是阻碍自然语言处理前进的巨大阻碍,它主要有三大难点:

### (1) 语法歧义多

自然语言的重要特点之一就是它存在大量的歧义现象,即一个句子对应着几种语法分析结果。人类自身可以凭借大量的先验知识有效地消除歧义,而计算机由于在知识表示和获取方面的不足,难以像人类一样进行句法分析排歧。

### (2) 搜索空间复杂

句法分析是一个较为复杂的任务,由于候选树的个数随语句长度呈现指数增长,导致搜索空间巨大。因此,在涉及句法分析模型时,人们设计的解码器往往只是在要求时间内搜索到模型定义的最优解或近似解。

### (3) 汉语句法分析的独特性

不同于英文句法分析,在句法分析之前,需要先进行分词,这无疑增加了句法分析的难度和工作量。针对汉语句法分析,朱德熙在[7][8]中指出,它具有以下几个难点:汉语没有形态、语序灵活、词类和句法成分不存在一一对应的关系、汉语句子的构造原则与词组的构造原则基本上是一致的、汉语语法形式化工作滞后等。这就导致了西方学者在英文上取得的成就难以迁移到汉语上,并且统计方法所需的树库也难以共用。

## 1.3 跨越语义鸿沟的三种思路

目前,突破句法分析瓶颈、跨越语义鸿沟主要有三种方法:统计与规则结合、深度学习、强化学习。下面我们分别对其进行探讨。

## 2 规则 and 统计结合的方法

句法分析可以分为基于规则的方法和基于统计的方法两大类。基于规则的句法分析方法的基本思路是,由人工组织语法规则,建立语法知识库,通过条件约束和检查来实现句法结构歧义的消除[9]。基于统计的句法分析方法的基本思路是,由

生成语法定义被分析的语言及其分析出的类别,在训练数据观察到的各种语言现象的分布以统计数据的方式与语法规则一起编码,如著名的概率上下文无关文法[10]。统计方法和规则方法各自存在着优点和不足,具有各自的适应场景,但在大部分的实际场景下,单独使用某一方法难以得到很高的识别率。因此,人们大多采用二者结合的形式。诸如“一只可爱的小花猫”和“一枝可爱的小花”的名词与数量词的修饰关系,以及语言中的递归现象——一只非常漂亮可爱的小花猫等,在采用规则和统计结合的方法后迎刃而解。

大体来说,该方法的尝试大多集中在上下文无关文法提出以后到深度学习出来以前,即上世纪80年代到2010年。大部分场景下,它的效果距离实用仍有差距。规则和统计结合的关键是把规则进行量化,这样语法规则的运用可以由参数 $a$ 控制,统计模型由参数 $b$ 控制。只要找到它们之间的拟合函数 $f(a, b)$ ,则混合系统就可以由拟合函数控制。

基于此,王轩等人于1999年提出了一种对规则进行量化的合成语言模型,在大词表非特定人语音识别系统中取得了理想的识别效果,将男声和女声的识别准确率分别提升到了86.6%和91.2%[11]。2003年,王晓龙等人采用规则和统计结合的方法,研究了适合于音字转换的短语搭配问题,在音字转换中取得了测试集87.7%的准确率[12]。2016年,Facebook的Mitra采用规则和统计结合的方法针对问答问题进行了探索[13],他们提出了一个新的模型,使用ASP(Answer Set Programming)作为主要的知识表示和推理语言,并且结合了标准的基于统计的自然语言处理的模型,显著提高了系统的推理能力,在十多个任务上达到了SOTA的结果。2017年,Gabbard等人将规则和统计结合的方法应用到了2016 LoReHLT评估中,并将基于统计得到的模型和基于规则的模型分开使用[14],观察到了在所用级别的数据资源下,采用规则和统计结合的方法明显优于单一方法,但在一些资源较少的数据集下,采用简单的技巧可能更加有效。

2017年,尼扎木丁等人针对维吾尔语人名,从语言的粘着性特点出发,从词干、音节、字符串三个角度对维吾尔语单词进行拆分,并把切分后的单元作为特征加入到了条件随机场中,缓解了数据稀疏的影响[15]。对于维吾尔语中的汉族人名,他们提出了基于规则的识别方法。系统的总体流程如图

2 所示,最终在人名识别的准确率、召回率和 F1 值上分别达到了 87.47%、89.12%、88.29%。

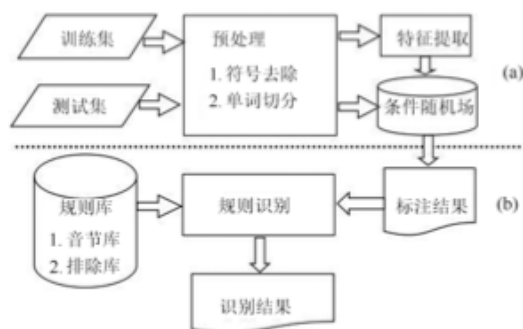


图 2 规则和统计结合方法的流程图

2017 年, Heeyoung Lee 等人针对共指消解问题设计了能够增量学习(动态改变数据集大小)的统计分析器,并对特定提及类型设计了基于规则的子模型[16],最终在 CoNLL-2012 数据集上达到了 63.2 的 F1 指数,比仅有基于规则的系统高出了 7 个百分点。

值得注意的是,深度学习的大放光彩并非意味着规则、统计的传统方法毫无用武之地, Koehn 就明确指出了神经网络在机器翻译上的六大问题[17]。2019 年, Koponen 等人在此基础上进一步定量分析了 NMT(神经网络机器翻译)、RBMT(规则机器翻译)、SMT(统计机器翻译)各自的优劣[18],他们观察到在机器翻译中,删除是 RBMT 最常见的编辑类型,插入是 SMT 最常见的编辑类型,单词形式更改和单词替换则是 NMT 中最常见的编辑类型。在跟动词形式相关的问题上, RBMT 处理得最好。

### 3 深度学习的方法

面向领域任务的深度学习研究和应用,需要解决两个主要的问题:一是应用领域的原始特征表示,二是选择合适的深度学习算法。前者实际是数据的表示问题,后者代表了深度学习结构问题,即采用何种深度学习的模型。例如,在语言处理任务中,一般选择最基本的语言单位——音素。回到面向自然语言处理的深度学习,它的研究同样需要解决上面两个问题。

#### 3.1 深度学习在自然语言处理中的特征表示

对于问题一——应用领域的原始特征表示,典型的有 Mikolov 提出的 word2vector[19], Harris 提出的词袋模型[20], Salton 提出的向量空间模型[21]

等。值得注意的是,词袋模型提出的时间是 1954 年,向量空间模型提出的时间是 1975 年。可见,改进原有的自然语言处理流程并非是在深度学习流行后才开展的,学者们很早就意识到了语义鸿沟的问题,并采取了将语言向量化的方式试图能有所突破。

在基于规则和统计的自然语言处理工作中, One-hot 是常见的表示方法,即每个词表示为一个很长的向量,只有一个维度的值为 1,代表当前词,其余绝大多数元素都为 0。One-hot 采用稀疏方式存储,形式非常简洁,结合传统的机器学习算法,如最大熵、支持向量机、条件随机场等,可以胜任自然语言处理的主流任务。但是显而易见,向量的维度就是词表的大小,这样就导致维度十分庞大。此外,它仅是孤立的表示单个词,没有表示出词之间的相关性,比如“土豆”和“马铃薯”的表示向量,不特别处理,二者的距离一般较远,无法看出是否存在关系,因此本质上还是没有解决语义鸿沟的问题。对于后者, Firth 等人提出可以用相近词表示当前词[22],通过计算不同范围的上下文相近词,从而得到当前词表示词的多种不同表示值。在 One-hot 的基础上,图灵奖获得者 Hinton 提出了用词向量表示词的方法[23],通过低维实数向量表示词,如 [0.666, 0.5, -0.107...],用 L2 距离来衡量词之间的差异。该方法不仅解决了之间的维度灾难问题,而且对于相近词的处理,可以引入合适的损失函数去学习。

显而易见,词向量主要是基于单词级别上的,而在文本基本,词袋模型是一种常见的表示方法。我们通过一个例子来简要说明,对于“Tom likes to play basketball. Mike likes too.”和“Mike also likes to play tennis.”,可以引入字典(“Tom”:1, “likes”:2, “to”:3, “play”:4, “basketball”:5, “Mike”:6, “too”:7, “also”:8, “tennis”:9)。那么,原语句就可以重新表述为向量[1, 2, 1, 1, 1, 1, 1, 0, 0]和[0, 1, 1, 1, 0, 1, 0, 1, 1]。第  $i$  个元素表示字典中第  $i$  个单词在原句中出现的次数,因此可以将词袋模型理解为词频统计。与之类似的有向量空间模型,它是将文本内容的处理简化为空间中的向量运算,通过计算向量之间的余弦距离来度量文档的相似性。目前应用广泛的 word2vector 是在上述方法基础上加以改进提出的,解决了之前维度爆炸、单纯基于词频统计、无法很好处理语料库动态更新等问题。大致思想是用神经

网络把 One-hot 形式的词向量映射为分布式的词向量，并运用了负采样等技巧加快训练速度。

### 3.2 深度学习在自然语言处理中的流程

深度学习在自然语言处理中的流程可以分为三步，一是将原始文本作为输入，自学习得到文本特征的分布表示，二是将分布式向量特征作为神经网络的输入，三是构建基本框架，需要针对处理的任务，选择合适的网络结构，构建深度学习基本模型框架。

训练阶段的流程如图 3 所示。第一步是模型检查，需要采用算法检查模型实现是否存在错误，例如梯度下降算法。第二步是模型初始化，涉及神经网络隐含层的偏置  $b$  和网络权重  $W$  的初始化。第三步是模型优化，涉及模型参数和结构的调优。第四步是模型调整，检查模型是否过拟合，是否需要运用正则化方法调整模型。

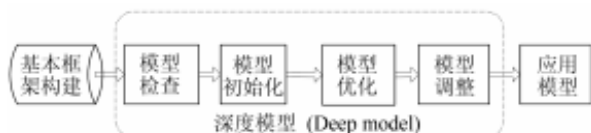


图 3 深度学习在自然语言处理中的流程图

### 3.3 深度学习在自然语言处理中的主流模型

对于问题二——如何选用合适的深度学习模型，需要根据不同的任务，自然语言的特点，选择合适的深度学习模型。下面对该领域主流的部分模型作简要介绍。

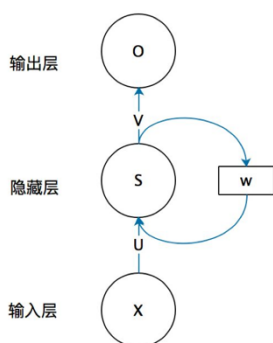


图 4 循环神经网络示意图

考虑到人类自然语言具有递归性，循环神经网络 RNN 应运而生[24]。它是隐含层和自身存在连接的一类网络，如图 4 所示。相较于前馈神经网络，循环神经网络可以将当前隐含层的计算结构用于下次隐含层的计算中，因此可以用于处理时间序列问题。Sutskever 用其解决文本生成的问题[25]，超越了之前最好的层次化非参数序列模型，取得了 SOTA 的结果。Cho 将其用于机器翻译[26]，采用

了两个 RNN 组成训练模型，一个 RNN 将符号序列编码成固定长度的向量，另一个实现表示的解码，即 Encoder-Decoder 模型，很好地实现了短语对中文语言规律的捕捉。Graves 将其用于语音识别[27]，以华尔街日报作为数据集，实现了直接对文本进行音频数据转录，即使在缺少词典的情况下，比 baseline 的错误率少了 6.7%。

一般而言，循环神经网络的反向传播只能传递 5-10 层，就会出现梯度消失的问题。为解决该问题，Hochreiter 提出了长短时记忆网络 LSTM[28]，它可以记忆不定时间长度的数值，结构如图 5 所示。图 5 中，底部是 4 个 *softmax* 函数单元，最左侧函数根据情况可能成为输入，右侧三个会经过判定决定是否输入。左侧第二个是输入门，决定该部分输入是否进入下一层。左侧第三个是遗忘门，决定是否以往区块中记住的值。第四个是输出门，决定在区块中记忆的输入是否输出。长短时记忆网络有很多变体，Greff 给出了 8 种 LSTM 之间的比较[29]。对于长短时记忆网络的应用场景，Tang 将其用于情感分析[30]，在 IMDB 影评数据集上较 baseline 提升了 5% 的准确率。Ma 用长短时记忆网络解决了命名实体识别问题[31]，取得了 91.21% 的结果，对比第二节中规则和统计结合的方法，LSTM 具有较大的优越性。

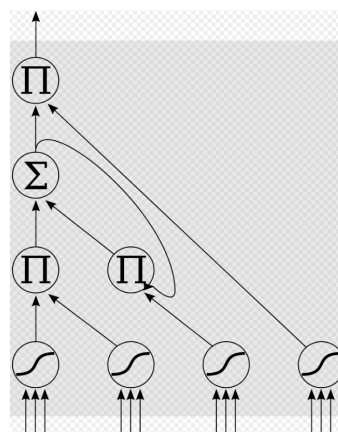


图 5 长短时记忆网络示意图

另一种解决 RNN 梯度消失的办法是 Attention 机制，严格来说它和上面的循环神经网络并不属于同一层级，更像是一种技巧和思路，它本质的网络还是循环神经网络。由于近些年来，Attention 机制在深度学习中的广泛应用，在此我们作简要介绍。

Attention 机制跟人类翻译文章的思路类似，即只将注意力集中在翻译部分对应的上下文中。做机器翻译时，在 Attention 模型中只会寻找源语句中对



应的几个词语，再去结合之前已经翻译的部分。例如，我们翻译图 6 所示的句子“Knowledge is power”，当翻译“Knowledge”时，只需将注意力放在源语句中“知识”的部分。这样，当 decoder 预测目标翻译时，就可以看到 encoder 的所有信息。基于此，Bahdanau 用其进行了神经机器翻译[30]。区别于传统的机器翻译，这里的目标是建立一个单一的神经网络，经过各种调整，最大限度地提高翻译性能。它由编码器和解码器组成，编码器用于将元语句编码成固定长度的向量，解码器负责从该向量生成翻译。通过这种方法，Bahdanau 在英译法任务中取得了 SOTA 的结果。

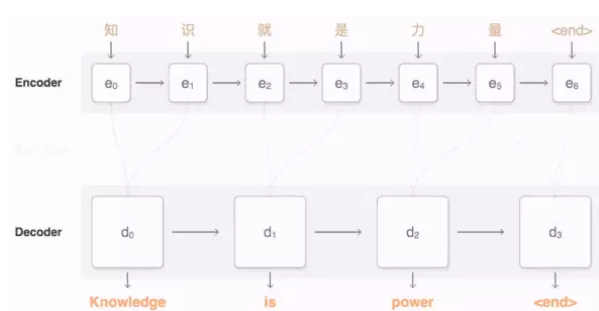


图 6 Attention 示意图

Attention 问世后，为了摆脱 RNN 只能从左向右或从右向左依次计算的牵制，Vaswani 提出了 Transformer[33]。它抛弃了传统的 CNN 和 RNN，整个网络完全由 Attention 机制组成，这也是它论文的名称“Attention is all you need”的由来。Transformer 的网络仅由 self-Attention 和前馈网络组成，通过搭建编码器和解码器各 6 层，在机器翻译中取得了 BLEU 值的 SOTA 成绩。2018 年，谷歌的 Devlin 等人基于 Transformer 提出了著名的 Bert[34]，它通过合并网络每层的信息，进行对未标记文本的双向表示预训练，在自然语言处理的 11 项任务中都取得了大幅度的提升，并在机器阅读理解任务中超越了人类。随后，大量基于 Bert 的应用开始广泛开展。Qin 用 Bert 处理了机器对话任务，主要思想是为对话模型动态提供相关的长格式文本，作为外部知识的来源[35]，在人工指标和自动化指标上均超越了之前的方法。Tripathi 基于 Bert，提出了多任务学习(MLT)网络，成功处理了自然语言理解任务[36]，将 MR(电影评论)、AG News、SST-2(斯坦福情感树库)数据集的准确率分别提高了 6%、1.4%、3.3%。

另外，递归神经网络也是深度学习在自然语言处理中常见的网络，它和循环神经网络的英文缩写

均为 RNN，需要注意区分。递归神经网络是利用树形结构递归构造而成，可用于构建句子语义信息[37]。它所用的树形结构一般为二叉树，典型的网络结构如图 7 所示，用于句法分析。其中， $x$  表示词向量， $y$  表示合并后的子树的向量。我们可以使用贪心算法，相邻的子树两两合成，计算评分，取分数最高的合并，直到形成句法树。除了用于句法分析，递归神经网络在其他自然语言处理任务中也有应用。Socher 将其用于关系分类[38]，为树的每个节点分配了一个向量和矩阵，矩阵可用于捕获相邻单词、短语的含义，实现了任意句法类型和长度的短语和句子的向量表示，并应用到情绪预测、电影分类等任务上，达到了 SOTA 的水平。随后，Socher 有将递归神经网络用于问答系统[39]，基于两个实体间的关系推理，实现了统计信息之间的共享。例如，孟加拉虎和苏门答腊虎就可以共享老虎的部分信息。最终，在 WordNet 和 FreeBase 知识库上，分别达到了 86.2% 和 90% 的准确率。

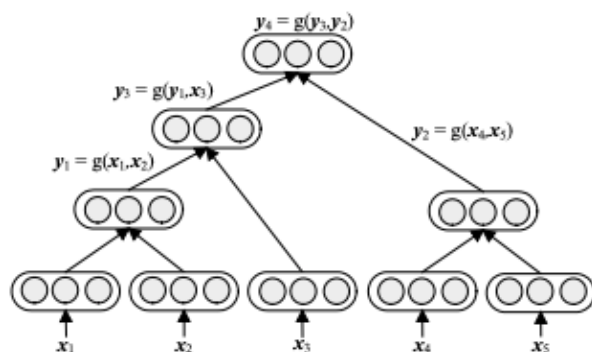


图 7 递归神经网络模型结构图

## 4 强化学习的方法

强化学习强调如何基于环境行动，以取得最大化的预期利益[40]。它往往有一个利益的函数，类似于损失函数，然后再采用不同的学习算法让 agent 去学习不同状态下的决策，如采用一阶马尔科夫决策，Q 学习就是其中典型的一类。强化学习在策略决策、文本生成上具有天然优势，因为在 Q 学习的时候，学者们尚没有发现除了强化学习之外的方法能够达到目标效果。此外，对于任务型对话系统，一般在给定一个对话之后，会有一个决策过程，即该系统下一步是要回答问题还是提出问题，或是去执行某个任务。基于此，我们可以将任务型对话问题转换为强化学习问题。2016 年，在 AlphaGo 横扫人类棋手后，更多的学意识强化学习的强大之处，

开始大量地将其应用到自然语言处理中，并在部分领域取得了成果。显而易见，强化学习可以和深度学习相辅相成，通过强化学习去指导深度学习。事实上，目前在自然语言处理领域，学者们大多也是这么做的。

目前，强化学习在自然语言处理中应用十分广泛，在一些任务上已经应用于工业界。2016 年，Covington 提出了 YoutubeDNN[41]，将信息检索任务进行了划分，首先引入了深度候选生成模型，然而再引入了一个单独的深度排序模型，在推荐任务重取得了理想的效果，并应用在了 Youtube 的推荐中，这也是它名称的由来。2017 年，Yu 等人用生成对抗网络 GAN 结合强化学习解决了文本生成任务[42]，而在此前，单纯采用 GAN 并不能解决自然语言处理中的问题。Yu 提出了一个序列生成框架 SeqGAN，在强化学习中将数据生成器建模为随机策略师，SeqGAN 就可以通过直接执行梯度策略更新来绕过生成器的微分问题。最终，在合成数据和实际任务的大量实验上，该方法比 baseline 有了显著的改进。2019 年，Peng 等人利用强化学习发展了一个情绪分析语调排歧的网络[43]，用于消除每个汉字拼音的歧义。他们讲语音特征与文本和视觉特征相融合，在五个不同的汉语情感分析数据集上取得了 SOTA 的结果。

## 5 未来研究方向

总体而言，现有的方法无论是规则和统计结合，还是深度学习、强化学习，都没有根本意义上解决语义鸿沟的问题，在机器翻译、问答系统等任务上，距离实用还有很长的距离。基于上述的研究现状，结合当前学者们开展的研究，我认为有两个方向可供突破。

一是结合周围的其他信息，不单纯依赖于语言信息进行处理。该方法在自动驾驶和语音识别上应用广泛，而且取得了一定的突破。以语音识别为例，为了提升系统的鲁棒性，不让对话系统在任何情况下都触发，学者们提出了称之为 groudred 的方法，即调用摄像头中的图像信息或是其他系统中的信息，判定用户是否有对话的意愿，来提升识别的准确度。借助 groudred 的思想，Chen 等人提出了一个新的评论员系统[44]，同时借助人类评论员的文本信息和足球比赛中的视频信息，实现了对足球事件的评论。

二是类比于规则和统计结合的方法，将基于规则的方法加入到深度学习当中，解决深度学习难以处理的特殊情况。

目前，深度学习凭借一些技巧和经验，在自然语言处理领域的一些任务上，已接近或达到实际使用水平。但是，它的不可解释性、缺乏理论基础，一直是国内外学者的痛点。个人认为，当前缺乏理论基石并非意味者全盘的不合理性，因为深度学习所使用的神经网络，的确来源于生物学领域。事实上，人脑中有 860 亿个神经元，而我们目前所使用的深度学习网络远远不及这一数量级，这也就决定了难以完全模拟大脑的运行方式。对人脑而言，就算遇到一个前所未见的事物，也可以依据一些先验知识做出判定。

那么，随着科技的进步，一方面我们可以在生物学上探索出人脑真正的思考方式，并采用计算机模拟。另一方面，也可以随着计算机硬件和海量数据的不断发展，或是量子计算机和光神经网络的突破，尽可能的接近人脑的网络规模。深度学习是骡子是马，相信终有一天，我们会揭开它的面纱。

## 参考文献

- [1] 宗成庆. 统计自然语言处理[M]. 清华大学出版社, 2013:179-275.
- [2] 刘挺, 马金山. 汉语自动句法分析的理论与方法[J]. 当代语言学, 2009 (2): 100-11
- [3] 王晓龙, 关毅. 计算机自然语言处理[M]. 清华大学出版社, 2005:114-119
- [4] Alfred V.Aho, Jeffrey D.Ullman, Ravi Sethi. 编译原理[M]. 机械工业出版社, 2003.
- [5] Hein, A.M., Identification and Bridging of Semantic Gaps in the Context of Multi-Domain Engineering, Abstracts of the 2010 Forum on Philosophy, Engineering & Technology. Colorado.
- [6] Zhang D, Wang L, Zhang L, et al. The Gap of Semantic Parsing: A Survey on Automatic Math Word Problem Solvers[J]. IEEE transactions on pattern analysis and machine intelligence, 2019.
- [7] 朱德熙. 语法讲义[M]. 商务印书馆, 1982.
- [8] 朱德熙. 语法答问[M]. 商务印书馆. 1985.
- [9] Allen J. Natural language understanding[M]. Pearson, 1995.
- [10] Booth T L, Thompson R A. Applying probability measures to abstract languages[J]. IEEE transactions on Computers, 1973, 100(5): 442-450.
- [11] 王轩, 王晓龙, 张凯. 语音识别中统计与规则结合的语言模型[J]. 自动化学报, 1999, 25(03): 309-315.
- [12] 高升, 王晓龙. 语句级汉字输入系统中语义规则研究[J]. 计算机工程与应用, 2003, 39(4): 80-82.

- [13] Mitra A, Baral C. Addressing a question answering challenge by combining statistical methods with inductive rule learning and reasoning[C]//Thirtieth AAAI Conference on Artificial Intelligence. 2016.
- [14] Gabbard R, DeYoung J, Lignos C, et al. Combining rule-based and statistical mechanisms for low-resource named entity recognition[J]. Machine Translation, 2018, 32(1-2): 31-43.
- [15] 汪昆, 帕力旦, 吐尔逊. 统计与规则相结合的维吾尔语人名识别方法[J]. 自动化学报, 2017, 43(4): 653-664.
- [16] Lee H, Surdeanu M, Jurafsky D. A scaffolding approach to coreference resolution integrating statistical and rule-based models[J]. Natural Language Engineering, 2017, 23(5): 733-762.
- [17] Koehn P, Knowles R. Six challenges for neural machine translation[J]. arXiv preprint arXiv:1706.03872, 2017.
- [18] Koponen M, Salmi L, Nikulin M. A product and process analysis of post-editor corrections on neural, statistical and rule-based machine translation output[J]. Machine Translation, 2019, 33(1-2): 61-90.
- [19] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality[C]//Advances in neural information processing systems. 2013: 3111-3119.
- [20] Harris Z S. Distributional structure[J]. Word, 1954, 10(2-3): 146-162.
- [21] Salton G, Wong A, Yang C S. A vector space model for automatic indexing[J]. Communications of the ACM, 1975, 18(11): 613-620.
- [22] Firth J R. 1968. A synopsis of linguistic theory, 1930-55[J]. Frank R. Palmer (ed.), Selected Papers of Firth JR, 1952, 59: 168-205.
- [23] Hinton G E. Learning distributed representations of concepts[C]//Proceedings of the eighth annual conference of the cognitive science society. 1986, 1: 12.
- [24] Elman J L. Finding structure in time[J]. Cognitive science, 1990, 14(2): 179-211.
- [25] Sutskever I, Martens J, Hinton G E. Generating text with recurrent neural networks[C]//Proceedings of the 28th International Conference on Machine Learning (ICML-11). 2011: 1017-1024.
- [26] Cho K, Van Merriënboer B, Gulcehre C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[J]. arXiv preprint arXiv:1406.1078, 2014.
- [27] Graves A, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks[C]//International conference on machine learning. 2014: 1764-1772.
- [28] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
- [29] Greff K, Srivastava R K, Koutník J, et al. LSTM: A search space odyssey[J]. IEEE transactions on neural networks and learning systems, 2016, 28(10): 2222-2232.
- [30] Tang D, Qin B, Liu T. Document modeling with gated recurrent neural network for sentiment classification[C]//Proceedings of the 2015 conference on empirical methods in natural language processing. 2015: 1422-1432.
- [31] Ma X, Hovy E. End-to-end sequence labeling via bi-directional lstm-cnns-crf[J]. arXiv preprint arXiv:1603.01354, 2016.
- [32] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. arXiv preprint arXiv:1409.0473, 2014.
- [33] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in neural information processing systems. 2017: 5998-6008.
- [34] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [35] Qin L, Galley M, Brockett C, et al. Conversing by Reading: Contentful Neural Conversation with On-demand Machine Reading[J]. arXiv preprint arXiv:1906.02738, 2019.
- [36] Tripathi S, Singh C, Kumar A, et al. Bidirectional Transformer Based Multi-Task Learning for Natural Language Understanding[C]//International Conference on Applications of Natural Language to Information Systems. Springer, Cham, 2019: 54-65.
- [37] Socher R, Lin C C, Manning C, et al. Parsing natural scenes and natural language with recursive neural networks[C]//Proceedings of the 28th international conference on machine learning (ICML-11). 2011: 129-136.
- [38] Socher R, Huval B, Manning C D, et al. Semantic compositionality through recursive matrix-vector spaces[C]//Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning. Association for Computational Linguistics, 2012: 1201-1211.
- [39] Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]//Advances in neural information processing systems. 2013: 926-934.
- [40] Tom Mitchell. 机器学习[M]. 机械工业出版社, 2008.
- [41] Covington P, Adams J, Sargin E. Deep neural networks for youtube recommendations[C]//Proceedings of the 10th ACM conference on recommender systems. ACM, 2016: 191-198.
- [42] Yu L, Zhang W, Wang J, et al. Seqgan: Sequence generative adversarial nets with policy gradient[C]//Thirty-First AAAI Conference on Artificial Intelligence. 2017.
- [43] Peng H, Ma Y, Poria S, et al. Phonetic-enriched Text Representation for Chinese Sentiment Analysis with Reinforcement Learning[J]. arXiv preprint arXiv:1901.07880, 2019.
- [44] Chen D L, Mooney R J. Learning to sportscast: a test of grounded language acquisition[C]//Proceedings of the 25th international conference on Machine learning. ACM, 2008: 128-135.