

# 文本自动生成研究进展与趋势

CCF 中文信息技术专业委员会

万小军 冯岩松 孙薇薇

北京大学计算机科学技术研究所，北京

## 摘要

我们期待未来有一天计算机能够像人类一样会写作，能够撰写出高质量的自然语言文本。文本自动生成就是实现这一目的的关键技术。按照不同的输入划分，文本自动生成可包括文本到文本的生成、意义到文本的生成、数据到文本的生成以及图像到文本的生成等。上述每项技术均极具挑战性，在自然语言处理与人工智能领域均有相当多的前沿研究，近几年业界也产生了若干具有国际影响力的成果与应用。本文对上述前沿技术的国内外研究现状进行了全面总结，并对发展趋势进行了展望。

关键词：自然语言生成、文本到文本的生成、意义到文本的生成、数据到文本的生成、图像到文本的生成

## Abstract

We expect that computers can write high-quality natural language texts like human beings in the near future. Automatic text generation is the key technique for achieving this goal. According to different data types of inputs, automatic text generation techniques include text-to-text generation, meaning-to-text generation, data-to-text generation and image-to-text generation. All the above text generation techniques are very challenging, and they are the frontier research topics in the natural language processing and artificial intelligence fields. In recent years, a few internationally influential achievements and applications have been yielded in academia and industry. In this article, we conduct a comprehensive survey of recent advances of automatic text generation at home and abroad. We also discuss the research and development trends.

Keywords: natural language generation, text-to-text generation, meaning-to-text generation, data-to-text generation, image-to-text generation

# 1. 引言

文本自动生成是自然语言处理领域的一个重要研究方向,实现文本自动生成也是人工智能走向成熟的一个重要标志。简单来说,我们期待未来有一天计算机能够像人类一样会写作,能够撰写出高质量的自然语言文本。文本自动生成技术极具应用前景。例如,文本自动生成技术可以应用于智能问答与对话、机器翻译等系统,实现更加智能和自然的人机交互;我们也可以通过文本自动生成系统替代编辑实现新闻的自动撰写与发布,最终将有可能颠覆新闻出版行业;该项技术甚至可以用来帮助学者进行学术论文撰写,进而改变科研创作模式。

按照不同的输入划分,文本自动生成可包括文本到文本的生成(text-to-text generation)、意义到文本的生成(meaning-to-text generation)、数据到文本的生成(data-to-text generation)以及图像到文本的生成(image-to-text generation)等。上述每项技术均极具挑战性,在自然语言处理与人工智能领域均有相当多的前沿研究,近几年业界已产生了若干具有国际影响力的成果与应用。最值得一提的是,美联社自 2014 年 7 月开始已采用新闻写作软件自动撰写新闻稿件来报道公司业绩,这大大减少了记者的工作量。美国洛杉矶时报也有一种用来撰写突发新闻的应用软件。美国已有多家公司能够提供新闻写作软件与服务,比如美国“自动洞察”公司(Automated Insights)已采用“语言专家”软件撰写了 3 亿篇报道,包括橄榄球、财经报道。这些进展标志着文本自动生成不再属于纸上谈兵的技术,而是已经对人类工作和生活产生了重大影响。

目前国内学界与工业界对文本自动生成技术的重视程度并不够,普遍缺乏对该方向前沿技术与进展的了解。因此,本技术报告将首次对文本自动生成前沿技术进行综合全面的调研、分析与总结,为国内同行提供一个全面了解文本自动生成技术的重要参考。同时,期望学界和工业界一起努力,尽早实现中文文本自动生成系统,抢占中文文本自动生成技术的制高点。

需要指出的是,自然语言处理领域的自然语言生成技术专指从机器可读数据生成自然语言文本的技术,而本文所介绍的文本自动生成技术的范畴则更加广泛,还包括了文本到文本的生成技术、以及图像到文本的生成技术。

## 2. 文本到文本的生成

### 2.1 国际研究现状

文本到文本的生成技术主要指对给定文本进行变换和处理从而获得新文本的技术,具体说来包括文本摘要(Document Summarization)、句子压缩(Sentence Compression)、句子融合

(Sentence Fusion)、文本复述(Paraphrase Generation)等。国际上对上述不同技术均进行了多年的研究,相关研究成果主要发表在自然语言处理相关学术会议与期刊上,例如 ACL、EMNLP、NAACL、COLING、AAAI、IJCAI、SIGIR、INLG、ENLG 等。国际上几个主要的研究单位包括密歇根大学、南加州大学、哥伦比亚大学、北德克萨斯大学、爱丁堡大学等。需要指出的是,机器翻译从某种程度上也可看作是一种从源语言到目标语言的文本生成技术,但由于机器翻译自身是相对独立的一个研究领域,因此本文的内容不再涵盖机器翻译技术。

### 2.1.1 文本摘要

文本摘要技术通过自动分析给定的文档或文档集,摘取其中的要点信息,最终输出一篇短小的摘要(通常包含几句话或上百字),该摘要中的句子可直接出自原文,也可重新撰写所得。摘要的目的是通过对原文本进行压缩、提炼,为用户提供简明扼要的内容描述。

根据不同的划分标准,文档摘要可以主要分为以下几种不同类型:

根据处理的文档数量,摘要可以分为单文档摘要和多文档摘要。单文档摘要只对单篇文档生成摘要,而多文档摘要则对一个文档集生成摘要。

根据是否提供上下文环境,摘要可以分为为主题或查询无关的摘要和主题或查询相关的摘要。主题或查询相关的摘要在给定的某个主题或查询下,能够诠释该主题或回答该查询;而主题或查询无关的摘要则指不给定主题和查询的情况下对文档或文档集生成的摘要。

根据摘要所采用的方法,摘要可以分为生成式和抽取式。生成式方法通常需要利用自然语言理解技术对文本进行语法、语义分析,对信息进行融合,利用自然语言生成技术生成新的摘要句子。而抽取式方法则相对比较简单,通常利用不同方法对文档结构单元(句子、段落等)进行评价,对每个结构单元赋予一定权重,然后选择最重要的结构单元组成摘要。抽取式方法应用较为广泛,通常采用的结构单元为句子。

根据摘要的应用类型,摘要可以分为标题摘要、传记摘要、电影摘要等。这些摘要通常为满足特定的应用需求,例如传记摘要的目的是为某个人生成一个概括性的描述,通常包含该人的各种属性,例如姓名、性别、地址、出生、兴趣爱好等。用户通过浏览某个人的传记摘要就能对这个人有一个总体的了解。

文档自动摘要的研究在图书馆领域和自然语言处理领域一直都很活跃,最早的应用需求来自于图书馆。图书馆需要为大量文献书籍生成摘要,而人工摘要的方式效率很低,因此亟需自动摘要方法取代人工高效地完成文献摘要任务。随着信息检索技术的发展,文档自动摘要在信息检索系统中的重要性越来越大,逐渐成为研究热点之一。文档自动摘要技术的第一

篇论文来自 Luhn (1958)<sup>[1]</sup>，经过数十年的发展，同时在 DUC<sup>1</sup>与 TAC<sup>2</sup>组织的自动摘要国际评测的推动下，文本摘要技术已经取得长足的进步。值得一提的是，由南加州大学 Chin-Yew Lin 博士(现就职于微软亚洲研究院)开发的摘要质量自动评估工具 ROUGE<sup>3</sup>的广泛使用也是自动摘要技术快速发展的一个推动力。国际上文档自动摘要方面比较著名的几个系统包括 ISI 的 NeATS 系统<sup>[2]</sup>，哥伦比亚大学的 NewsBlaster 系统<sup>4 [3]</sup>，密歇根大学的 NewsInEssence 系统<sup>5 [4]</sup>等。2013 年雅虎耗资 3000 万美元收购了一项自动新闻摘要应用 Summly，标志着新闻摘要技术走向成熟。

目前的文本摘要方法主要基于句子抽取，也就是以原文中的句子作为单位进行评估与抽取。这类方法的好处是易于实现，能保证摘要句子具有良好的可读性。该类方法主要包括两个步骤：一是对文档中的句子进行重要性计算或排序，二是选择重要的句子组合成最终摘要。第一个步骤可采用基于规则的方法，利用句子位置或所包含的线索词来判定句子的重要性；也可采用各种机器学习方法(包括深度学习方法)，综合考虑句子的多种特征进行句子重要性的分类、回归或排序，例如 CRF<sup>[5]</sup>，HMM<sup>[6]</sup>，SVM<sup>[7][8]</sup>，RNN<sup>[9]</sup>等。第二个步骤则基于上一步结果，需要考虑句子之间的相似性，避免选择重复的句子(如 MMR 算法<sup>[10]</sup>)，并进一步对所选择的摘要句子进行连贯性排列(如自底向上法<sup>[11]</sup>)，从而获得最终的摘要。近几年学界进一步提出了基于整数线性规划的方法<sup>[12][13][14]</sup>以及次模函数最大化的方法<sup>[15][16]</sup>，可以在句子选择的过程中同时考虑句子冗余性。

不同于上述方法，压缩式文本摘要方法则考虑对句子进行压缩，以在较短长度限制下让摘要涵盖更多的内容。最有代表性的做法为同时进行句子选择与句子压缩<sup>[17][19][19]</sup>，能够取得更优的 ROUGE 性能。除了压缩之外，部分工作还利用句子融合等技术来对已有句子进行变换，得到新的摘要句子<sup>[20][21]</sup>。

国际上还有部分研究者研究真正意义上的生成式摘要，也就是通过对原文档进行语义理解，将原文档表示为深层语义形式(例如深层语义图)，然后分析获得摘要的深层语义表示(例如深层语义子图)，最后由摘要的深层语义表示生成摘要文本。最近的一个尝试为基于抽象意义表示( Abstract Meaning Representation, AMR)进行生成式摘要<sup>[22]</sup>。这类方法所得到的摘要句子并不是基于原文句子所得，而是利用自然语言生成技术从语义表达直接生成而

---

<sup>1</sup> <http://duc.nist.gov/>

<sup>2</sup> <http://www.nist.gov/tac/>

<sup>3</sup> <http://www.berouge.com>

<sup>4</sup> <http://www1.cs.columbia.edu/nlp/newsblaster/>

<sup>5</sup> <http://lada.si.umich.edu:8080/clair/nie1/nie.cgi>

得。这类方法相对比较复杂，而且由于自然语言理解与自然语言生成本身都没有得到很好的解决，因此目前生成式摘要方法仍属于探索阶段，其性能还不尽如人意。

上述摘要方法均面向新闻摘要，而近年来针对学术文献的摘要越来越受到大家的重视。一方面，可以利用学术文献之间的引用关系以及引文来帮助进行学术文献摘要<sup>[23]</sup>；另一方面，对学术文献进行自动综述也是一个很有意思的研究问题<sup>[24]</sup>。更多的有关文本摘要技术的内容可参考综述<sup>[25]</sup>。

### 2.1.2 句子压缩与融合

句子压缩与句子融合技术一般用于文本摘要系统中，用于生成信息更加紧凑的摘要，获得更好的摘要效果。

句子压缩技术基于一个长句子生成一个短句子，要求该短句保留长句中的重要信息，也就是重要信息基本不损失，同时要求该短句是通顺的。下面给出一个句子压缩的例子：

原句：But they are still continuing to search the area to try and see if there were, in fact, any further shooting incidents.

压缩后的句子：They are continuing to search the area to see if there were any further incidents.

学界尝试了多种方法实现句子压缩，包括从句子中删除词语<sup>[26]</sup>，或对句子中的词语进行替换、重排序或插入<sup>[27]</sup>。其中，从句子中直接删除词语的做法因其复杂程度较低而成为主流方法。研究人员提出多种方法用于实现基于词语删除的句子压缩，包括噪声信道模型<sup>[28]</sup>，结构化辨别模型<sup>[29]</sup>，树到树的转换<sup>[30]</sup>，整数线性规划<sup>[31]</sup>，等等。但就总体效果而言，对于大部分句子所删除的词语一般较少，压缩效果体现并不明显。

句子融合技术则是合并两个或多个包含重叠内容的相关句子得到一个句子。根据目的的不同，一类句子融合只保留多个句子中的共同信息，而过滤无关的细节信息（类似于集合运算中的取交集运算），另一类句子融合则只过滤掉多个句子之间的重复内容（类似于集合运算中的取并集运算）。下面给出两个相关的句子以及人工合并后得到的句子：

句子 1：In 2003, his nomination to the U.S. Court of Appeals for the District of Columbia sailed through the Senate Judiciary Committee on a 16-3 vote.

句子 2：He was nominated to the U.S. Court of Appeals for the District of Columbia Circuit in 1992 by the first President Bush and again by the president in 2001.

合并后的句子(取交集)：He was nominated to the U.S. Court of Appeals for the District of Columbia Circuit.

合并后的句子(取并集): In 2003, his nomination by the first President Bush, and again by the second Bush in 2001 to the U.S. Court of Appeals for the District of Columbia sailed through the Senate Judiciary Committee on a 16-3 vote.

针对句子融合问题, MIT 的 Regina Barzilay 和哥伦比亚大学的 Kathleen McKeown 提出一条流水线算法, 包括共同信息识别(Identification of Common Information)、融合网格计算(Fusion lattice computation)、网格线性化(Lattice linearization)三个步骤<sup>[20]</sup>。研究人员针对句子融合问题提出的其它代表性方法包括基于结构化辨别学习的方法<sup>[32]</sup>, 基于整数线性规划的方法<sup>[33]</sup>, 基于图最短路径的方法<sup>[34]</sup>等。

上述研究均面向英文, 少数研究者在网上公开了所使用的数据集, 但这些数据集的规模相对较小, 覆盖面较窄, 业界也没有组织过句子压缩或融合相关的评测。近些年, 与句子压缩与句子融合技术相关的学术论文比较少见, 与上述因素不无关系。

### 2.1.3 文本复述

文本复述生成技术通过对给定文本进行改写, 生成全新的复述文本, 一般要求输出文本与输入文本在表达上有所不同, 但所表达的意思基本一样。文本复述生成技术应用相当广泛, 例如, 在机器翻译系统中可利用文本复述技术对复杂输入文本进行简化从而方便翻译, 在信息检索系统中可利用文本复述技术对用户查询进行改写, 在儿童教学系统中可利用文本复述技术将难以理解的文本简化为儿童容易理解的文本。

根据实际的需求, 通过复述生成技术得到的输出文本与原文本相比, 可以只是一两个词发生了改变(如例 1), 也可以是整段文本面目全非(如例 2)。

例 1: all the members of → all members of

例2: He said there will be major cuts in the salaries of high-level civil servants. =>

He claimed to implement huge salary cut to senior civil servants.

简单的文本复述生成可以通过同义词替换来实现, 也可以通过人工或自动构建的复述规则来实现<sup>[35]</sup>, 例如根据变换状语位置的一条规则, 可以获得下面句子的简单复述句子:

输入: He booked a single room in Beijing *yesterday*.

输出: *Yesterday*, he booked a single room in Beijing.

为了实现复杂的文本复述生成, 研究人员提出了基于自然语言生成的方法<sup>[36]</sup>、基于机器翻译的方法<sup>[37]</sup>与基于支点(Pivot)的方法<sup>[38][39]</sup>等。基于自然语言生成的方法模拟人类的思维方式, 首先对输入句子进行语义理解, 获得该句子的语义表示, 然后基于得到的语义表示生成新的句子。基于机器翻译的方法则将文本复述生成问题看作是单语言机器翻译问题, 从



而利用现有机翻译模型(例如噪声信道模型)来为给定文本生成复述文本。基于支点的方法则将当前语言中的输入文本翻译到另一种语言(支点),然后将翻译得到的文本再次翻译回当前语言。由于每次翻译过程均要求源语言和目标语言中文本的语义保持一致,因此可以预期最后得到的文本在语义上能跟输入文本保持一致。注意支点语言可以只采用一种语言,也可采用多种语言。例如,下面的例子中采用意大利语作为支点语言,通过两次翻译为输入的英文句子生成复述文本:

输入英文句子: What toxins are English most **hazardous** to *expectant mothers*?

翻译后的意大利文句子: Che tossine sono più pericolose alle donne incinte?

再次翻译后的英文句子: What toxins are more **dangerous** to *pregnant women*?

总体而言,现有方法能够为给定文本生成具有较小差别的复述文本,但是难以有效生成高质量的具有很大大差别的复述文本,原因在于对于改写甚多的复述文本而言,一方面难以保证其与原文本的语义一致性,另一方面则难以保证该文本的可读性。近几年已经极少在自然语言处理重要会议上看到文本复述生成相关的学术论文,表明针对该问题的研究已经遇到了瓶颈。

需要指出的是,句子简化(Sentence Simplification)可以看作是一类特殊的复述生成问题,其目的是将复杂的长句改写成简单、可读性更好、易于理解的多个短句,方便用户快速阅读。在实现上仍可采用上述各类方法,例如基于单语言机器翻译的方法<sup>[40]</sup>,基于树转换的方法<sup>[41]</sup>等。针对句子简化问题的很多研究都采用维基百科<sup>6</sup>以及对应的简单维基百科<sup>7</sup>数据来进行学习和测试。简单维基百科面向的阅读对象为儿童以及正在学习英语的成人,简单维基百科的作者要求使用简单的词汇和简短的句子来撰写文章。一个简单维基百科文章一般对应一个普通维基百科文章,因此通过这种文本之间的对齐关系能够获取大量的有用语料。爱丁堡大学的 Kristian Woodsend 与 Mirella Lapata 则提出基于准同步文法(Quasi-synchronous grammar)与整数线性规划模型将普通维基百科文章简化为简单维基百科文章<sup>[42]</sup>。

## 2.2 国内研究现状

### 2.2.1 文本摘要

相比机器翻译、自动问答、知识图谱、情感分析等热门领域,文本摘要在国内并没有受到足够的重视。在文本摘要方面从事过研究的单位包括北京大学计算机科学技术研究所、北京大学计算语言所、哈工大信息检索实验室、清华大学智能技术与系统国家重点实验室等。

---

<sup>6</sup> <http://en.wikipedia.org>

<sup>7</sup> <http://simple.wikipedia.org>

其中,北京大学计算机科学技术研究所在文本摘要方面进行了长期深入的研究,提出了多种基于图排序的自动摘要方法<sup>[43][44][45][46]</sup>与压缩式摘要方法<sup>[47]</sup>,并且探索了跨语言摘要、比较式摘要、演化式摘要等多种新颖的摘要任务<sup>[48][49][50]</sup>。在学术文献摘要方面,则分别提出基于有监督学习和整数线性规划模型的演示幻灯片的自动生成方法<sup>[51]</sup>与学术论文相关工作章节的自动生成方法<sup>[52]</sup>。

国内早期的基础资源与评测<sup>8</sup>举办过单文档摘要的评测任务,但测试集规模比较小,而且没有提供自动化评价工具。2015年CCF中文信息技术专委会组织了NLPCC评测<sup>9</sup>,其中包括了面向微博的新闻摘要任务,提供了规模相对较大的样例数据和测试数据,并采用自动评价方法,吸引了多支队伍参加评测,目前这些数据可以公开获得。但上述中文摘要评测任务均针对单文档摘要任务,目前还没有业界认可的中文多文档摘要数据,这在事实上阻碍了中文自动摘要技术的发展。

近些年,市面上出现了一些文本挖掘产品,能够提供文档摘要功能(尤其是单文档摘要),例如方正智思、拓尔思(TRS),海量科技等公司的产品。百度等搜索引擎都能为检索到的文档提供简单的单文档摘要。这些文档摘要功能均被看作是系统的附属功能,其实现方法均比较简单。由于这些模块均未参加公开评测,因此其性能不得而知。

### 2.2.2 句子压缩与融合

国内有少数单位与学者对句子压缩问题进行了研究,例如北京大学语言计算与互联网挖掘研究室提出基于对偶分解的句子压缩方法<sup>[53]</sup>,清华大学智能信息获取研究小组提出基于马尔科夫逻辑网的句子压缩方法<sup>[54]</sup>,等等。而对于句子融合问题的研究,国内单位和学者基本没有涉猎。

国内学者的上述研究仍面向英文数据,主要原因在于缺少相关的中文评测数据,而构建一个高质量的中文句子压缩或融合评测数据集并不简单,需要依靠对语言有深刻理解的标注者。

### 2.2.3 文本复述

国内有少数单位和学者对文本复述生成进行了一些研究,例如哈工大信息检索中心与微软亚洲研究院、百度等单位合作,提出利用多种资源(包括多种词典、平行语料等在内)改进基于机器翻译的复述生成方法<sup>[55]</sup>、利用多种机器翻译引擎的复述生成方法<sup>[56]</sup>,以及面向不同应用的复述生成方法<sup>[57]</sup>。

---

<sup>8</sup> <http://www.863data.org.cn>

<sup>9</sup> [http://tcci.ccf.org.cn/conference/2015/pages/page05\\_evadata.html](http://tcci.ccf.org.cn/conference/2015/pages/page05_evadata.html)



上述研究仍面向英文领域，采用英文数据进行评测，而中文复述生成技术则极少有人涉足，这是一件很令人遗憾的事情。

## 2.3 发展趋势与展望

文本到文本的生成包括多项任务，这些任务之间具有紧密的联系，很多方法也都对不同任务具有通用性。在未来几年，随着深层语义分析技术的发展，研究者可以在研究过程中充分利用深层语义分析结果，此外，深度学习技术的成熟则为我们的研究打开了另外一扇门，但是大家需要认真思考如何才能用好深层语义分析技术与深度学习技术。而随着社交媒体的广泛使用，我们也可充分利用社交媒体数据为我们的研究服务。

为了更好的推动文本到文本的生成技术的发展，业界可从以下几个方面着手：

其一，构建大规模评测数据集。数据是研究的基石，大规模、高质量的评测数据集对于研究工作至关重要，而目前上述多个任务均缺少大规模评测数据集，尤其是中文评测数据集。数据集的构建需要耗费大量人力物力，因此一个可行的途径就是采用众包的方式。

其二，构建开源平台。尽管针对上述各项任务业界均提出了多种解决方法，但很多方法并不易实现。业界需要为每个任务构建一个开源平台，将主流算法集成到该平台中，将会极大方便后来者的研究，推动研究的发展。

## 3. 意义到文本的生成

### 3.1 国际研究现状

不同于文本到文本的生成，意义到文本的生成这一任务的输入在学界并没有达成一致，其根本在于不论是哲学家还是语言学家对何为自然语言的语义都未能形成较为一致的定义。在计算语言学领域，研究人员普遍遵循的语义研究原则建立在“真值条件（**Truth Condition**）”的基础上，认为寻找到了能够使自然语言语句成真的条件，即是在某种程度上刻画了自然语言的语义。在真值条件假设基础上，学者普遍采用逻辑的方法来对语义进行表征，并分别从模型论（**Model Theory**）和证明论（**Proof Theory**）两个角度来展开研究，很多学者也常常称这类型的语义为逻辑语义。目前已有的意义到文本的生成研究，普遍假设使用逻辑语义表征——以逻辑表达式为代表——作为输入，而以自然语言语句作为输出，本文也围绕这些研究展开介绍。图 3.1 给出了一个基于类型  $\lambda$  演算进行语义表征的实例，在该例子中，问题的输入是一个  $\lambda$  表达式，而输出是一个英语句子。

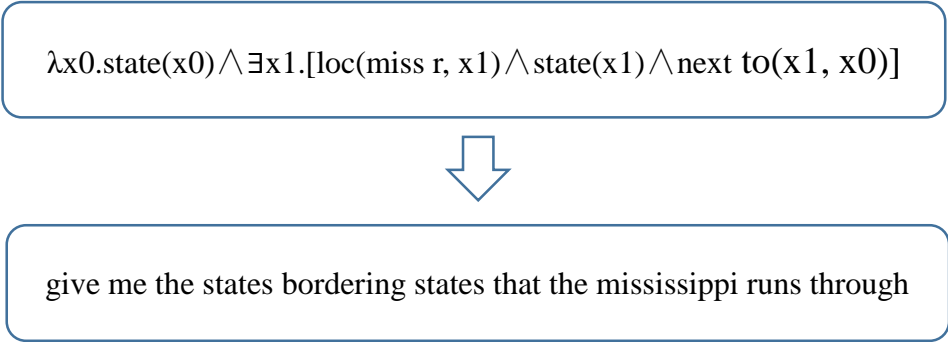


图 3.1:  $\lambda$  表达式到文本的生成实例<sup>[58]</sup>

意义到文本的生成和组合语义分析（Compositional Semantic Parsing）密切相关，语义分析旨在对线性的词序列进行自动句法语义解析并得到其真值条件。因为在分析过程中遵循了弗雷格所提之组合原则（Principle of Compositionality），因而称为组合语义分析，以与分布式语义（Distributional Semantics）相区别。组合语义分析是自然语言处理的一项核心技术，是迈向深度语义理解的一座重要桥梁，在多个自然语言处理核心任务中有着潜在应用，如智能问答、机器翻译等。从问题自身的定义来看，意义到文本的生成与组合语义分析是一对互逆的自然语言处理任务。在当前的国际研究中，仅专注于意义到文本的生成这一任务的学者并不多，部分以句法语义分析研究为主的学者会兼顾这方面的研究。

### 3.1.1 基于深层语法的文本生成

在早期的自然语言处理研究中，计算语言学发挥了很大的作用，计算语言学家从形式化、可计算的角度对自然语言进行建模，提出一系列的旨在解释语言运作机理的句法语义模型，并根据这些模型构建自然语言处理系统。相关研究在上个世纪八九十年代取得了丰硕的研究成果，一系列兼具语言本体解释力和可计算性的语法范式（Grammar Formalism）被提出，如组合范畴语法（Combinatory Categorical Grammar；简称 CCG）<sup>[59]</sup>和中心语驱动的短语结构语法（Head-driven Phrase-Structure Grammar；简称 HPSG）<sup>[60]</sup>等。不同于目前句法分析所主要使用的上下文无关文法（Context-Free Grammar；简称 CFG），上述语法范式具有超越上下文无关的表达能力，其语法推导过程往往更复杂，蕴含更多的信息，而这些信息可以用来做更透明的语义分析，简单而言，这些深层语法范式能够更好地支持句法语义同步的语言分析。在深层语法的支撑下，通过句法语义的协同推导可以获取自然语言的组合语义；而当以语义表征作为输入，通过其逆过程可以完成意义到文本的生成。

Shieber <sup>[61]</sup>提出了一个统一的框架用于进行句法语义分析与生成。在这一框架中，Shieber 将语言处理统一理解为逻辑推演（Deduction）过程，其差别在于推演的始点——公理——与推演的终点——目标——不同。在这一视角下，传统的句法分析（Parsing）技术可

以移植到文本生成上来，如线图分析法（Chart Parsing）技术可以转化为线图生成（Chart Generation）技术<sup>[62]</sup>。Shieber 后续又同其他学者合作，将推演的思想细化，利用合一语法来表达句法语义接口（Syntax-Semantics Interface），提出了语义中心驱动的生成<sup>[63]</sup>。

深层语法复杂度较高，如何构造对错综复杂的语言现象具有高覆盖度（Broad Coverage）的语法规则本身是一个极大的难题。以上研究主要是对原型算法进行讨论，而因为真实可用的大型深层语法当时没有得到很好的开发，以上研究并没有呈现极具代表性意义的经验结果。经过十余年的漫长开发，研究人员在 HPSG 理论的基础上开发了英语资源语法（English Resource Grammar；简称 ERG）<sup>10</sup> <sup>[64]</sup>，它是一个比较成功的具有较高覆盖率的深层语法规则系统，而围绕的 ERG 所展开的文本生成研究也取得了有益的进展。Carroll 和 Oepen <sup>[65]</sup>基于 ERG 和真实测试数据重新讨论了基于线图的生成技术，给出了极具参考意义的经验评估；另外，他们也提出了两项新的技术来改进基于合一语法的可行解紧致表示（Compact Representation）及其相关解码算法——Selective Unpacking，尤其后者，有效地利用了判别式学习模型来改进文本生成过程中所遇到的歧义消解。

组合范畴语法是一个广受自然语言处理领域学者关注的语法范式，其设计遵循了类型透明（Type Transparency）的原则，具有精简的语法语义接口，常常被语义分析<sup>[66]</sup>和文本生成<sup>[67]</sup>模型所采用。White 和 Baldridge <sup>[67]</sup>讨论了如何将线图生成法与组合范畴语法结合，并开发了开源的基于组合范畴语法的句子实现（Realization）工具——OpenCCG<sup>11</sup>。White 又同其他学者联合提出了一些进一步改进文本生成的算法<sup>[68][69][70]</sup>。

### 3.1.2 基于同步文法的文本生成

在过去的二十年间，统计句法分析与统计机器翻译是公认的两个取得长足进步的自然语言处理任务。除了从成熟的统计句法分析中借鉴成功经验——如判别式消歧——之外，不少学者也尝试复用成功的机器翻译模型来完成文本生成。机器翻译的目标是将某种自然语言语句翻译成另外一种自然语言的语句，并尽量保持意义不变；而文本生成则可以视为将某种形式语言语句翻译成一种自然语言语句，二者具有极强的可比性。

Chiang <sup>[71]</sup>提出了层级基于短语的翻译模型（Hierarchical Phrase-based Model），其核心是利用同步上下文无关文法（Synchronous Context-Free Grammar）来协同源语言语句的解析和目标语言语句的生成。目前同步文法也已经被借鉴到文本生成的研究中<sup>[72][58]</sup>。Wong 与 Mooney <sup>[72]</sup>两位作者讨论了两种形式语言用于表征意义：第一种是用于指挥机器人动作的形

---

<sup>10</sup> <http://www.delph-in.net/erg/>

<sup>11</sup> <https://github.com/OpenCCG/openccg>

式语言，第二种是一种无变量的数据库检索语言；而 Lu 与 Ng<sup>[58]</sup> 则针对表达能力极强的类型  $\lambda$  表达式（Typed  $\lambda$ -expression）展开研究。两项研究的共同点是构建形式语言的基于树的结构，在将相关结构与待生成的自然语言的树结构建立一致性对应，从而完成文本生成任务；另一个共同点则是广泛地使用了现有的机器翻译技术（包括开源软件等）来进行文法抽取、解码等。

## 3.2 国内研究现状

国内语言学界与计算语言学界针对自然语言语义的形式化研究较少，针对汉语进行全方面组合语义刻画的研究目前尚属空白。另一方面，从事自然语言处理的研究人员也较少涉猎深层语言结构处理问题，而对意义到文本的生成研究则更是鲜有，很少能见到相关学术成果发表在重要学术会议和期刊上。

## 3.3 发展趋势与展望

随着深层自然语言理解的发展，研究者将越来越多的目光投向了意义到文本的生成这一自然语言生成核心任务上。意义到文本的生成这一任务随着意义表征体系的不同问题的复杂度也会随之变化，传统的基于深层语法分析的生成方法面临的解码效率差、语法鲁棒性不够等问题仍需要更好的技术解决方案。近些年来，有零星的一些工作尝试将较为成熟的组合优化技术应用到在句法分析和机器翻译，如拉格朗日松弛<sup>[73][74]</sup>，尝试去求解一些所涉及到的 NP 难问题。应对意义到文本的生成这一复杂度高的问题，我们也可以尝试应用相关技术。而针对深层语法鲁棒性不够的问题，基于数据驱动的语法近似（Grammar Approximation）<sup>[75]</sup>取得了不错的结果，结果显示低阶语法近似能够有效改进深层语法分析的鲁棒性，如何应用相关思想来解决文本生成中所遇到的问题也是一个非常值得研究的方向。

而就针对汉语的文本生成研究来说，需要国内外学界做出更大的努力。首先，在语言本体分析方面，需要学者们建立相关的语义表征体系及针对汉语的特殊语言现象的分析，以支持汉语的深层处理。其次，在文本生成算法方面，也需要我们投入更多的科研精力设计适合汉语自动生成的模型算法等。

# 4. 数据到文本的生成

## 4.1 国际研究现状

数据到文本的生成技术指根据给定的数值数据生成相关文本，例如基于数值数据生成天

气预报文本、体育新闻、财经报道、医疗报告等。数据到文本的生成技术具有极强的应用前景，目前该领域已经取得了很大的研究进展，业界已经研制出面向不同领域和应用的多个生成系统。针对数据到文本的生成技术的研究单位主要集中少数几个单位，例如英国阿伯丁大学、英国布莱顿大学、爱丁堡大学等，相关研究成果主要发表在 INLG、ENLG 这几个专业学术会议上。

英国阿伯丁大学的 Ehud Reiter 在三阶段流水线模型<sup>[76]</sup>的基础上提出了数据到文本的生成系统的一般框架，见下图：

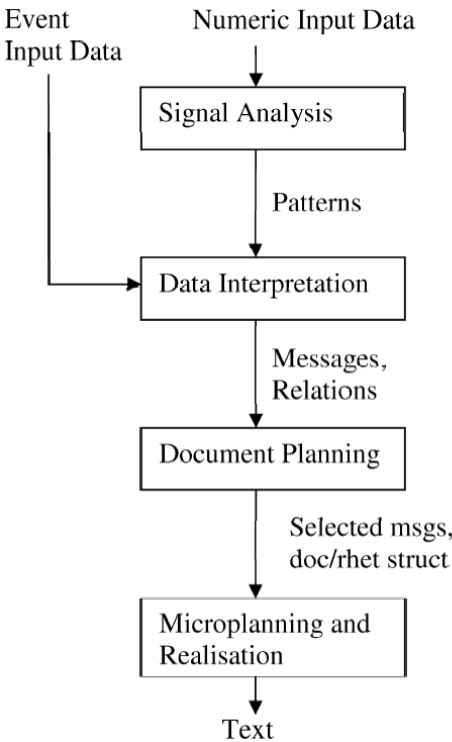


图 4.1：数据到文本的生成系统的一般框架<sup>[77]</sup>

其中：

信号分析模块(Signal Analysis)的输入为数值数据，通过利用各种数据分析方法检测数据中的基本模式，输出离散数据模式。例如股票数据中的峰值，较长期的增长趋势等。该模块与具体应用领域和数据类型相关，针对不同的应用领域与数据类型所输出的数据模式是不同的。

数据阐释模块(Data Interpretation)的输入为基本模式与事件，通过对基本模式和输入事件进行分析，推断出更加复杂和抽象的消息，同时推断出它们之间的关系，最后输出高层消息以及消息之间的关系。例如针对股票数据，如果跌幅超过某个值则可以创建一条消息。还需要检测消息之间的关系，例如因果关系、时序关系等。值得说明的是，数据阐释模块并不

是在所有文本生成系统中都需要，例如，在天气预报文本生成系统中，基本的模式足以满足要求，因此并不需要采用数据阐释模块。

文档规划模块(Document Planning)的输入为消息及关系，分析决定哪些消息和关系需要在文本中提及，同时要确定文本的结构，最后输出需要提及的消息以及文档结构。从更高的层次来说，信号分析与数据阐释模块会产生大量的消息、模式和事件，但文本通常长度受限，只能描述其中的一部分，因此文档规划模块必须确定文本中需要说明的消息。一般可根据专家知识、消息的重要性、新颖性等进行选择和确定。当然，该模块与领域也很相关，不同领域中对消息的选择所考虑的因素不一样，文档的结构也会不一样。

微规划与实现(Microplanning and Realisation)模块的输入为选中的消息及结构，通过自然语言生成技术输出最终的文本。该模块主要涉及到对句子进行规划以及句子实现，要求最终实现的句子具有正确的语法、形态和拼写，同时采用准确的指代表达。所采用的技术在学术界有相当多的研究，具体可参考本文第 3 节“意义到文本的生成”。

目前，业界已经研制了面向多个领域的数据到文本的生成系统，这些系统的框架与上述一般框架并无大的差别，部分系统将上述框架中的两个模块合并为一个模块，或者省去了其中一个模块。

数据到文本的生成技术在天气预报领域应用最为成功，业界研制了多个系统对天气预报数据进行总结，生成天气预报文本。例如，FoG 系统<sup>[78]</sup>能够从用户操作过的数据中生成双语天气预报文本；SumTime 系统<sup>[79]</sup>能够生成海洋天气预报文本，实验评测表明用户有时候更倾向于阅读 SumTime 所生成的天气预报，而非专家撰写的天气预报<sup>[80]</sup>。此外，英国阿伯丁大学的 Anja Belz 提出概率生成模型进行天气语言文本的生成<sup>[81]</sup>。Anja Belz 和 Eric Kow 进一步基于天气预报数据分析对比了多种数据到文本的生成系统，结果表明采用自动化程度较高的方法并不会降低文本生成质量，同时文本质量的自动评价方法会低估基于手工规则构建的系统，而高估自动化系统<sup>[82]</sup>。

业界面向其他领域也研制多个文本生成系统，例如针对空气质量的文本生成系统<sup>[83]</sup>，针对财经数据的文本生成系统<sup>[84]</sup>，面向医疗诊断数据的文本生成系统 TOPAZ<sup>[85]</sup>、Suregen<sup>[86]</sup>、BT-45<sup>[87]</sup>等。其中 BT-45 能够为新生儿重症监护病房(NICU)的监控数据生成文本摘要，帮助医生进行决策。下面两张图分别给出了 BT-45 系统的输入样例与生成得到的文本。



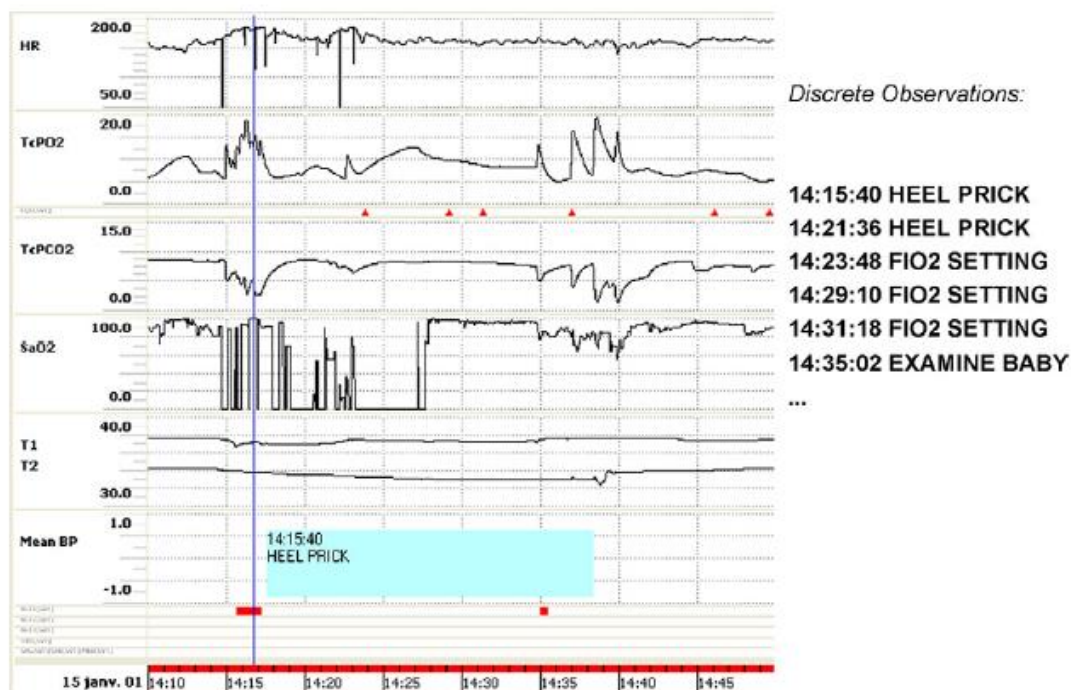


图 4.2: NICU 数据样例, 从上到下分别表示 HR, TcPO<sub>2</sub>, TcPCO<sub>2</sub>, SaO<sub>2</sub>, T<sub>1</sub> & T<sub>2</sub>, and Mean BP

[Portet et al., 2009]<sup>[87]</sup>

You saw the baby between 14:10 and 14:50. Heart Rate (HR) = 159. Core Temperature (T<sub>1</sub>) = 37.7. Peripheral Temperature (T<sub>2</sub>) = 34.3. Transcutaneous Oxygen (TcPO<sub>2</sub>) = 5.8. Transcutaneous CO<sub>2</sub> (TcPCO<sub>2</sub>) = 8.5. Oxygen Saturation (SaO<sub>2</sub>) = 89.

Over the next 30 minutes T<sub>1</sub> gradually increased to 37.3.

By 14:27 there had been 2 successive desaturations down to 56. As a result, Fraction of Inspired Oxygen (FIO<sub>2</sub>) was set to 45%. Over the next 20 minutes T<sub>2</sub> decreased to 32.9. A heel prick was taken. Previously the spo<sub>2</sub> sensor had been re-sited.

At 14:31 FIO<sub>2</sub> was lowered to 25%. Previously TcPO<sub>2</sub> had decreased to 8.4. Over the next 20 minutes HR decreased to 153.

By 14:40 there had been 2 successive desaturations down to 68. Previously FIO<sub>2</sub> had been raised to 32%. TcPO<sub>2</sub> decreased to 5.0. T<sub>2</sub> had suddenly increased to 33.9. Previously the spo<sub>2</sub> sensor had been re-sited. The temperature sensor was re-sited.

图 4.3: BT-45 系统生成的对应文本 [Portet et al., 2009]<sup>[87]</sup>

由于数据到文本的生成技术的巨大应用价值, 工业界成立了多家从事文本生成的公司, 能够为多个行业基于行业数据生成行业报告或新闻报道, 从而节省大量的人力。比较知名的公司有 ARRIA<sup>12</sup>、AI<sup>13</sup>、NarrativeScience<sup>14</sup>等。其中 ARRIA 是一家总部设在欧洲的公司, 其前称为 Data2Text, 由来自阿伯丁大学的两名教授 Ehud Reiter 与 Yaji Sripada 创办, 后来自然语言生成领域的另一位科学家 Robert Dale 也加入了该公司, 该公司的核心技术为 ARRIA NLG 引擎。AI (Automated Insights) 则是一家美国人工智能公司, 由一名思科的前工程师 Robbie Allen 所创办, 最早基于体育数据生成文本摘要, 目前能为包括金融、个人健身、商业智能、网站分析等在内的多个领域内的数据生成文本报告, 其核心技术为 WordSmith NLG 引擎。目

<sup>12</sup> <https://www.arria.com/>

<sup>13</sup> <http://automatedinsights.com>

<sup>14</sup> <http://www.narrativescience.com>

前，AI 公司已经为美联社等多家单位生成数亿篇新闻报道，造成了巨大的影响力。NarrativeScience 则是根据美国西北大学的一个研究项目 StatsMonkey 发展而来，其核心技术为 Quill NLG 引擎。Forbes 是 NarrativeScience 的一个典型客户，在网站上有个 NarrativeScience 专页<sup>15</sup>，全部文章都是由 NarrativeScience 自动生成。下面给出一篇自动生成的样例新闻：

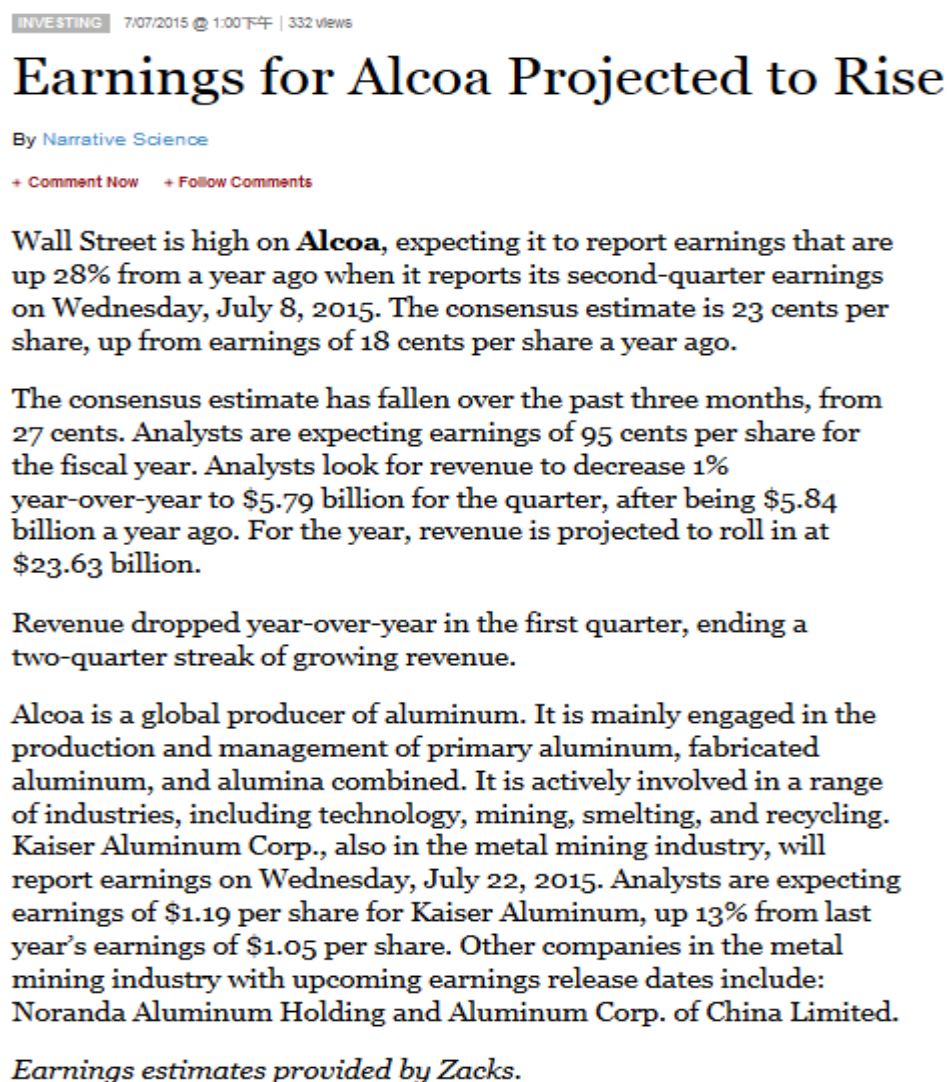


图 4.4：NarrativeScience 自动生成的样例新闻

## 4.2 国内研究现状

国内学术界对数据到文本的生成鲜有研究，也很少见到相关学术成果发表在重要学术会议和期刊上。国内工业界则有部分单位研制了基于模板的文本生成系统。例如新华社已开发了从财报数据生成企业财报年报的系统，该系统基于人工模板，将需要的数据填入写好的模板中，从而生成财报年报。由于采用的模板比较固定，所以为不同企业生成的财报年报都比

<sup>15</sup> <http://www.forbes.com/sites/narrativescience>

较类似，而不够生动。

### 4.3 发展趋势与展望

从数据到中文文本的生成技术很有研究意义，同时实用性很强。如果能实现从数据到中文新闻的生成，那么将极大缓解编辑和记者的负担，实现媒体、出版行业的变革。而实现这样的系统，必须依靠科研院所和新闻出版机构的合作，新闻出版机构能够提供大量的数据和专家知识，而科研院所则擅长自然语言理解与生成的理论与方法。

此外，要研制一套通用的面向不同领域的数据到文本的生成系统相当复杂和困难，因此一个更好的做法是先选择一两个领域(如财经、体育)进行系统研制，待系统成熟后再考虑将系统迁移到其他领域。

## 5. 图像到文本的生成

### 5.1 国际研究现状

图像到文本的生成技术是指根据给定的图像生成描述该图像内容的自然语言文本，例如新闻图像附带的标题、医学图像附属的说明、儿童教育中常见的看图说话、以及用户在微博等互联网应用中上传图片时提供的说明文字。依据所生成自然语言文本的详细程度及长度的不同，这项任务又可以分为图像标题自动生成和图像说明自动生成。前者需要根据应用场景突出图像的核心内容，例如，为新闻图片生成的标题需要突出与图像内容密切关联的新闻事件，并在表达方式上求新以吸引读者的眼球；而后者通常需要详细描述图像的主要内容，例如，为有视力障碍的人提供简洁详实的图片说明，力求将图片的内容全面且有条理的陈述出来，而在具体表达方式上并没有具体的要求。

对于图像到文本的自动生成这一任务，人类可以毫不费力地理解图像内容，并按具体需求以自然语言句子的形式表述出来；然而对于计算机而言，则需要综合运用图像处理，计算机视觉和自然语言处理等几大领域的研究成果。作为一项标志性的交叉领域研究任务，图像到文本的自动生成吸引着来自不同领域研究者的关注。自 2010 年起，自然语言处理界的知名国际会议和期刊 ACL、TACL 和 EMNLP 中都有相关论文的发表；而自 2013 年起，模式识别与人工智能领域顶级国际期刊 IEEE TPAMI 以及计算机视觉领域顶级国际期刊 IJCV 也开始刊登相关工作的研究进展，至 2015 年，计算机视觉领域的知名国际会议 CVPR 中，更有近 10 篇相关工作的论文发表，同时机器学习领域知名国际会议 ICML 中也有 2 篇相关论文发表。图像到文本的自动生成任务已被认为是人工智能领域中的一项基本挑战。

与一般的文本生成问题类似,解决图像到文本的自动生成问题也需要遵循三阶段流水线模型<sup>[76]</sup>,同时又需要根据图像内容理解的特点,做出一些调整:

在内容抽取方面,需要从图像中抽取物体、方位、动作、场景等概念,其中物体可以具体定位到图像中的某一具体区域,而其他概念则需要进行语义标引。这部分主要依靠模式识别和计算机视觉技术。

在句子内容选择方面,需要依据应用场景,选择最重要(如图像画面中最突出的,或与应用场景最相关的),且意义表述连贯的概念。这部分需要综合运用计算机视觉与自然语言处理技术。

最后,在句子实现部分,根据实际应用特点选取适当的表述方式将所选择的概念梳理为合乎语法习惯的自然语言句子。这部分主要依靠自然语言处理技术。

早期工作主要依照上述三阶段的流水线模式来实现。例如,在 Yao 等人的工作<sup>[88]</sup>中,图像被细致的分割并标注为物体及其组成部分,以及图像所表现的场景,并在此基础上选择与场景相关的描述模板,将物体识别的结果填充入模板得到图像的描述文字。而 Feng 与 Lapata<sup>[89][90]</sup>则采用概率图模型对文本信息和图像信息同时建模,并从新闻图片所在的文字报道中挑选合适的关键词作为体现图像内容的关键词,并进而利用语言模型将所选取的内容关键词、及必要的功能词汇链接为基本合乎语法规则的图像标题。还有一些工作<sup>[91][92][93][94]</sup>则依靠计算机视觉领域现有的物体识别技术从图像中抽取物体(包括人物、动物、花草、车、桌子等常见的物体类型),并对其定位以获得物体之间的上下位关系,进而依赖概率图模型和语言模型选取适当的描述顺序将这些物体概念、介词短语块串联成完整的句子。Hodosh 等人<sup>[96]</sup>则利用基于核函数的典型关联分析(Kernel Canonical Correlation Analysis, KCCA)来寻找文本与图像之间的关联,并依据图像信息对候选句子排序,从而获得最佳描述句子。值得说明的是,Hodosh 等人的工作<sup>[96]</sup>和 Feng 与 Lapata 的工作<sup>[90][91]</sup>均没有依靠现有的物体识别技术。

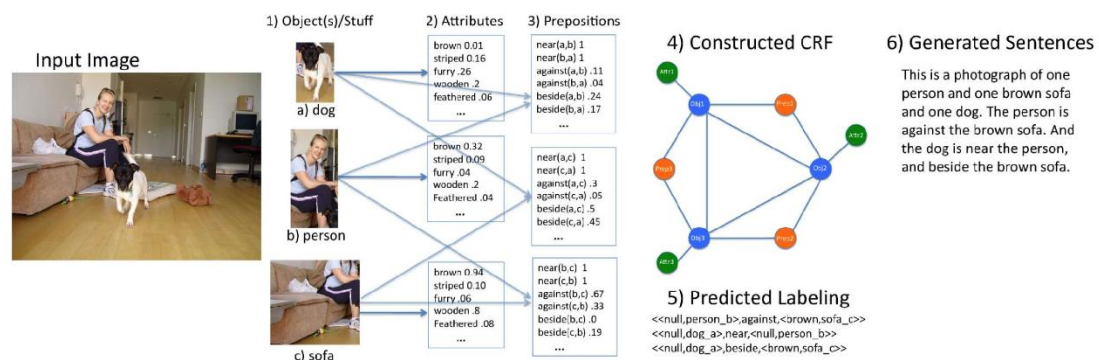


图 5.1: 一种典型的流水线模型<sup>[93]</sup>



随着深度学习方法在模式识别、计算机视觉及自然语言处理领域的广泛应用，基于海量数据的大规模图像分类、语义标注技术得到了快速发展；同时，统计机器翻译等与自然语言生成相关的技术也有了显著的提高。这也催生了将图像语义标注及自然语言句子生成进行联合建模的一系列工作，一方面在图像端采用多层深度卷积神经网络（Deep Convolution Neural Network, DCNN）对图像中的物体概念进行建模，另一方面在文本端采用循环神经网络（Recurrent Neural Network, RNN）或递归神经网络（Recursive Neural Network）对自然语言句子的生成过程进行建模<sup>[97]</sup>。传统图像语义标注工作主要关注具体某个物体的识别以及物体之间的相对位置关系，而对动作等抽象概念的关注较少。Socher 等人<sup>[98]</sup>提出利用递归神经网络对句子建模，并利用句法解析树突出对于动作（动词）的建模，进而将图像端与文本端进行联合优化，较好的刻画了物体与动作之间的关系。为了将两种不同模态的数据统一在一个框架下，Chen 与 Zitnick<sup>[99]</sup>将文本信息与图像信息融合在同一个循环神经网络中，利用图像信息作为记忆模块，从而指导文本句子的生成，同时又借助于一个重构图像信息层，实现了图像到文本、文本到图像的双方向表示。而 Mao 等人<sup>[100]</sup>则通过 DCNN 得到的图像信息与文本信息融合到同一个循环神经网络（m-RNN）中，将图像信息融入到了自然语言句子生成的序列过程中，取得了不错的结果。类似的想法也被 Donahue 等人<sup>[101]</sup>应用于动作识别和视频描述生成过程中。但在 m-RNN 的句子生成过程中，在图像端并没有显著的约束，例如在下图中，当生成单词“man”的时候，并没有与图像信息中的任务标注发生直接或间接的关联。

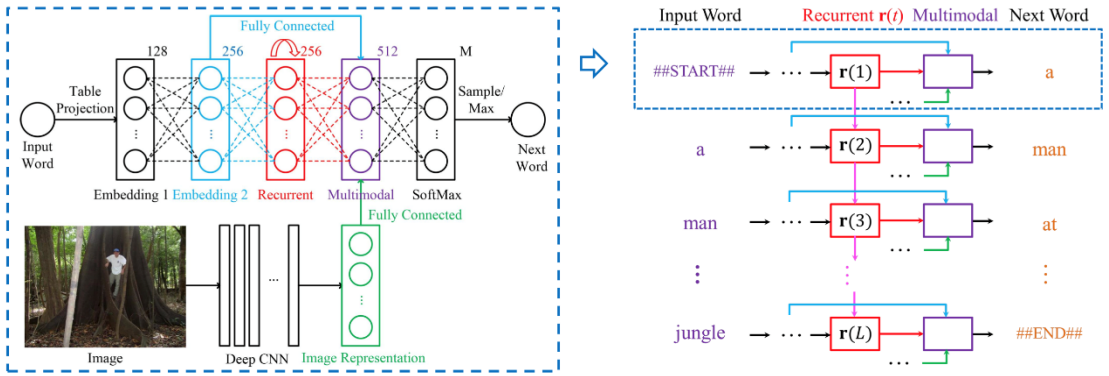


图 5.2：多模态 m-RNN 模型<sup>[100]</sup>

谷歌和加拿大蒙特利尔大学和多伦多大学的研究人员则分别借鉴了统计机器翻译领域的最新研究进展来推进图像到文本自动生成的联合建模<sup>[102] [103]</sup>。前者利用深层卷积神经网络 DCNN 对图像建模，将图像信息“编码”（encoding）后，直接由另一个与之相连接的 LSTM 神经网络（Long-Short Term Memory Network, LSTM）“解码”（decoding）成自然语言句子，

无需进行图像-词对齐、调序等传统模型的子步骤。而后者则在基于神经网络的机器翻译框架下，提出利用计算机视觉领域中的“注意”（Attention）机制来促进词语和图像块之间的对齐，从而在句子生成过程中，模拟人视觉的“注意”转移过程能够与词语序列的生成过程相互促进，使生成的句子更符合人的表述习惯。

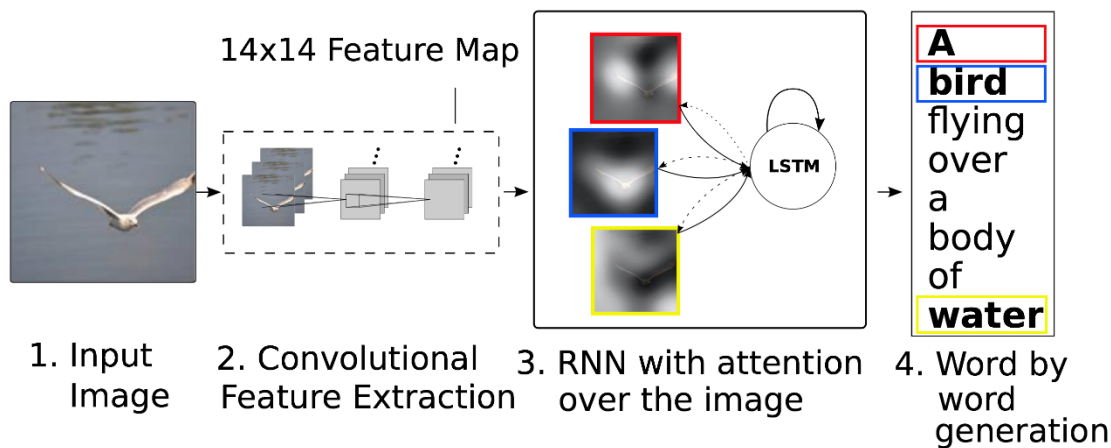


图 5.3：视觉“注意”引导的图像标题生成过程<sup>[103]</sup>

此外，微软的研究人员<sup>[104]</sup>利用卷积神经网络 CNN 和多示例学习（Multiple Instance Learning, MIL）对图像建模，并利用判别式语言模型生成候选句子，并采用统计机器翻译研究中经典的最小误差率训练 (Minimum Error Rate Training, MERT) 来发掘文本和图像层面的特征对候选句子进行排序。

虽然图像到文本的生成技术还处在探索阶段，距离实际产业应用还有一定的距离，但工业界已经开始注意到这一技术的理论研究价值和潜在应用前景，积极与学术界合作拓展研究方向。在 2015 年的计算机视觉知名国际会议 CVPR 2015 上举办的 LSUN Challenge（Large-scale Scene Understanding）挑战活动中也进行了图像标题自动生成的评测任务，最终谷歌公司<sup>[102]</sup>和微软研究院<sup>[104]</sup>取得了总成绩并列第一名，蒙特利尔-多伦多联队<sup>[103]</sup>和另一只微软研究院队伍<sup>[105]</sup>总成绩并列第三名，加州伯克利分校<sup>[101]</sup>获得第五名。

## 5.2 国内研究现状

国内学术界对图像到文本的生成技术研究开展较晚，大部分科研单位专注于跨媒体数据的语义标注和检索等任务上，只有人大、清华、北大、北航和中科院等科研单位开展了相关研究，如人民大学与腾讯合作在 2015 年欧盟组织的 ImageCLEF 评测中，在图像句子生成（Image Sentence Generation）任务中取得了第一名。

在工业界方面，百度和腾讯等科研机构也依靠自身在跨媒体语义标注、分类和检索等方



面的研究优势,逐步开展相关方向的研究工作,如百度与UCLA合作的m-RNN系统在CVPR 2015 LSUN 评测的图像标题自动生成任务中也取得了不错的成绩。

### 5.3 发展趋势与展望

从图像到文本的生成技术需要集成模式识别与机器学习、计算机视觉、自然语言处理,甚至认知科学领域的研究成果,具有极高的理论研究价值和实用前景。从一定程度上讲,这一技术同图像语义标注等任务一道,已成为各大顶尖科研机构在人工智能领域综合研究实力的较量方式,必将促进其快速发展。

而对于这一任务本身而言,更大的挑战仍然在于如何正确的抽取图像的内容,同时根据人类的语言习惯选择适当的表述方式将图像内容转换为自然语言句子。需要指出的是,目前的研究仍然聚焦在是否将图像中的物体概念抽取完全,是否选择了正确的词语,所生成的句子是否符合语法习惯等;可以预见在不久的将来,实际应用场景和上下文语境等约束将进一步推进相关技术的进步,必将广泛应用于新闻传播、在线教育、智能家居等多个领域。

## 6. 总结与展望

本文对文本自动生成技术进行了全面的介绍,包括文本到文本的生成、意义到文本的生成、数据到文本的生成、图像到文本的生成等。由于上述每项技术均有众多的研究者在研究,相关的学术成果也层出不穷,因此本文的总结难免有遗漏之处。希望本文的内容能够对相关研究人员和从业者有所帮助。

对文本自动生成技术的国际研究现状和国内研究现状进行比较可以看到,国内对该领域的研究投入和产出均远远不够,原创性的方法、资源、系统以及产品都相对比较匮乏,而且该领域没有受到业界足够的关注。我们必须奋起直追,建设相关中文资源,提出原创性文本生成方法,搭建中文文本生成系统并开发相关产品,才能占领中文文本生成的制高点。我们期待第一个中文文本生成系统由国内单位研制而成。

### 参考文献

- [1] Luhn, H. P. (1958). The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2), 159-165.
- [2] Lin, C. Y., & Hovy, E. (2002, July). From single to multi-document summarization: A prototype system and its evaluation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* (pp. 457-464). Association for Computational Linguistics.

- [3] Evans, D. K., Klavans, J. L., & McKeown, K. R. (2004, May). Columbia newsblaster: multilingual news summarization on the Web. In *Demonstration Papers at HLT-NAACL 2004* (pp. 1-4). Association for Computational Linguistics.
- [4] Radev, D., Otterbacher, J., Winkel, A., & Blair-Goldensohn, S. (2005). NewsInEssence: summarizing online news topics. *Communications of the ACM*, 48(10), 95-98.
- [5] Shen, D., Sun, J. T., Li, H., Yang, Q., & Chen, Z. (2007, January). Document Summarization Using Conditional Random Fields. In *IJCAI* (Vol. 7, pp. 2862-2867).
- [6] Conroy, J. M., & O'leary, D. P. (2001, September). Text summarization via hidden markov models. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 406-407). ACM.
- [7] Schilder, F., & Kondadadi, R. (2008, June). FastSum: fast and accurate query-based multi-document summarization. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers* (pp. 205-208). Association for Computational Linguistics.
- [8] Ouyang, Y., Li, W., Li, S., & Lu, Q. (2011). Applying regression models to query-focused multi-document summarization. *Information Processing & Management*, 47(2), 227-237.
- [9] Cao, Z., Wei, F., Dong, L., Li, S., & Zhou, M. (2015, February). Ranking with recursive neural networks and its application to multi-document summarization. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- [10] Carbonell, J., & Goldstein, J. (1998, August). The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 335-336). ACM.
- [11] Bollegala, D., Okazaki, N., & Ishizuka, M. (2010). A bottom-up approach to sentence ordering for multi-document summarization. *Information processing & management*, 46(1), 89-109.
- [12] McDonald, R. (2007). *A study of global inference algorithms in multi-document summarization* (pp. 557-564). Springer Berlin Heidelberg.
- [13] Gillick, D., & Favre, B. (2009, June). A scalable global model for summarization. In *Proceedings of the Workshop on Integer Linear Programming for Natural Language Processing* (pp. 10-18). Association for Computational Linguistics.
- [14] Li, C., Qian, X., & Liu, Y. (2013, August). Using Supervised Bigram-based ILP for Extractive Summarization. In *ACL (1)* (pp. 1004-1013).
- [15] Lin, H., & Bilmes, J. (2010, June). Multi-document summarization via budgeted maximization of submodular functions. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 912-920). Association for Computational Linguistics.
- [16] Lin, H., & Bilmes, J. (2011, June). A class of submodular functions for document summarization. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 510-520). Association for Computational Linguistics.
- [17] Qian, X., & Liu, Y. (2013). Fast Joint Compression and Summarization via Graph Cuts. In *EMNLP* (pp. 1492-1502).
- [18] Li, C., Liu, Y., Liu, F., Zhao, L. & Weng, F. (2014). Improving Multi-documents Summarization by Sentence Compression based on Expanded Constituent Parse Trees. In *EMNLP*.

- [19] Berg-Kirkpatrick, T., Gillick, D., & Klein, D. (2011, June). Jointly learning to extract and compress. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 481-490). Association for Computational Linguistics.
- [20] Barzilay, R., & McKeown, K. R. (2005). Sentence fusion for multidocument news summarization. *Computational Linguistics*, 31(3), 297-328.
- [21] Bing L., Li P., Liao Y., Lam W., Guo W., & Passonneau R. J. (2015). Abstractive Multi-Document Summarization via Phrase Selection and Merging. In *ACL*.
- [22] Liu, F., Flanigan, J., Thomson, S., Sadeh, N., & Smith, N. A. (2015). Toward Abstractive Summarization Using Semantic Representations. In *NAACL*.
- [23] Abu-Jbara, A., & Radev, D. (2011, June). Coherent citation-based summarization of scientific papers. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 500-509). Association for Computational Linguistics.
- [24] Saif Mohammad, Bonnie Dorr, Melissa Egan, Ahmed Hassan, Pradeep Muthukrishnan, Vahed Qazvinian, Dragomir Radev, and David Zajic. 2009. Using citations to generate surveys of scientific paradigms. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 584-592. Association for Computational Linguistics.
- [25] Nenkova, A., & McKeown, K. (2012). A survey of text summarization techniques. In *Mining Text Data* (pp. 43-76). Springer US.
- [26] Knight, K., & Marcu, D. (2002). Summarization beyond sentence extraction: A probabilistic approach to sentence compression. *Artificial Intelligence*, 139(1), 91-107.
- [27] Cohn, T., & Lapata, M. (2008, August). Sentence compression beyond word deletion. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1* (pp. 137-144). Association for Computational Linguistics.
- [28] Knight, K., & Marcu, D. (2000, August). Statistics-based summarization-step one: Sentence compression. In *AAAI/IAAI* (pp. 703-710).
- [29] McDonald, R. T. (2006, April). Discriminative Sentence Compression with Soft Syntactic Evidence. In *EACL*.
- [30] Cohn, T. A., & Lapata, M. (2009). Sentence compression as tree transduction. *Journal of Artificial Intelligence Research*, 637-674.
- [31] Clarke, J., & Lapata, M. (2008). Global inference for sentence compression: An integer linear programming approach. *Journal of Artificial Intelligence Research*, 399-429.
- [32] Thadani, K., & McKeown, K. (2013). Supervised sentence fusion with single-stage inference. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing* (pp. 1410-1418).
- [33] Elsner, M., & Santhanam, D. (2011, June). Learning to fuse disparate sentences. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation* (pp. 54-63). Association for Computational Linguistics.
- [34] Filippova, K. (2010, August). Multi-sentence compression: finding shortest paths in word graphs. In *Proceedings of the 23rd International Conference on Computational Linguistics* (pp. 322-330). Association for Computational Linguistics.
- [35] Barzilay, R., & Lee, L. (2003, May). Learning to paraphrase: an unsupervised approach using multiple-sequence alignment. In *Proceedings of the 2003 Conference of the North American Chapter of*

- the Association for Computational Linguistics on Human Language Technology-Volume 1* (pp. 16-23). Association for Computational Linguistics.
- [36] Fujita, A., Inui, K., & Matsumoto, Y. (2005). Exploiting lexical conceptual structure for paraphrase generation. *IJCNLP 2005, LNAI 3651*, pp. 908-919.
  - [37] Quirk, C., Brockett, C., & Dolan, W. B. (2004, July). Monolingual Machine Translation for Paraphrase Generation. In *EMNLP* (pp. 142-149).
  - [38] Duboue, P. A., & Chu-Carroll, J. (2006, June). Answering the question you wish they had asked: The impact of paraphrasing for question answering. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers* (pp. 33-36). Association for Computational Linguistics.
  - [39] Max, A. (2009, August). Sub-sentential paraphrasing by contextual pivot translation. In *Proceedings of the 2009 Workshop on Applied Textual Inference* (pp. 18-26). Association for Computational Linguistics.
  - [40] Wubben, S., Van Den Bosch, A., & Krahmer, E. (2012, July). Sentence simplification by monolingual machine translation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1* (pp. 1015-1024). Association for Computational Linguistics.
  - [41] Zhu, Z., Bernhard, D., & Gurevych, I. (2010, August). A monolingual tree-based translation model for sentence simplification. In *Proceedings of the 23rd international conference on computational linguistics* (pp. 1353-1361). Association for Computational Linguistics.
  - [42] Woodsend, K., & Lapata, M. (2011, April). WikiSimple: Automatic Simplification of Wikipedia Articles. In *AAAI*.
  - [43] Wan, X., Yang, J., & Xiao, J. (2007, January). Manifold-Ranking Based Topic-Focused Multi-Document Summarization. In *IJCAI* (Vol. 7, pp. 2903-2908).
  - [44] Wan, X., & Yang, J. (2008, July). Multi-document summarization using cluster-based link analysis. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 299-306). ACM.
  - [45] Wan, X., & Zhang, J. (2014, July). CTSUM: extracting more certain summaries for news articles. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 787-796). ACM.
  - [46] Yan, S., & Wan, X. (2014). SRRank: leveraging semantic roles for extractive multi-document summarization. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 22(12), 2048-2058.
  - [47] Jin-ge Yao, Xiaojun Wan, Jianguo Xiao. (2015). *Compressive Document Summarization via Sparse Optimization*. In *IJCAI*.
  - [48] Yan, R., Wan, X., Otterbacher, J., Kong, L., Li, X., & Zhang, Y. (2011, July). Evolutionary timeline summarization: a balanced optimization framework via iterative substitution. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 745-754). ACM.
  - [49] Wan, X. (2011, June). Using bilingual information for cross-language document summarization. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 1546-1555). Association for Computational Linguistics.
  - [50] Wan, X., Jia, H., Huang, S., & Xiao, J. (2011, July). Summarizing the differences in multilingual news. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 735-744). ACM.

- [51] Hu, Y., & Wan, X. (2015). PPSGen: Learning-Based Presentation Slides Generation for Academic Papers. *Knowledge and Data Engineering, IEEE Transactions on*, 27(4), 1085-1097.
- [52] Hu, Y., & Wan, X. (2014). Automatic Generation of Related Work Sections in Scientific Papers: An Optimization Approach. In EMNLP.
- [53] Yao, J. G., Wan, X., & Xiao, J. (2014). Joint Decoding of Tree Transduction Models for Sentence Compression. In EMNLP.
- [54] Huang, M., Shi, X., Jin, F., & Zhu, X. (2012, July). Using first-order logic to compress sentences. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.
- [55] Shiqi Zhao, Cheng Niu, Ming Zhou, Ting Liu, and Sheng Li. 2008. Combining Multiple Resources to Improve SMT-based Paraphrasing Model. In Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-08: HLT), pages 1021-1029.
- [56] Shiqi Zhao, Haifeng Wang, Xiang Lan, and Ting Liu. 2010. Leveraging Multiple MT Engines for Paraphrase Generation. In Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010), pages 1326-1334.
- [57] Shiqi Zhao, Xiang Lan, Ting Liu, Sheng Li. 2009. Application-driven Statistical Paraphrase Generation. In Proceedings of Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2009), pages 834-842.
- [58] Wei Lu; Hwee Tou Ng. 2011. A Probabilistic Forest-to-String Model for Language Generation from Typed Lambda Calculus Expressions. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*.
- [59] Mark Steedman. 2000. *The Syntactic Process*. MIT Press.
- [60] Carl Pollard, Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press.
- [61] Stuart M. Shieber. 1988. A uniform architecture for parsing and generation. In *Proceedings of the 12th International Conference on Computational Linguistics*.
- [62] Martin Kay. 1996. Chart Generation. In *Proceedings of the 34th annual meeting on Association for Computational Linguistics*.
- [63] Stuart M. Shieber, Gertjan van Noord, Fernando C. N. Pereira, and Robert C. Moore. 1990. Semantic-head-driven generation. *Computational Linguistics*.
- [64] Dan Flickinger. 2002. On building a more efficient grammar by exploiting types. *Collaborative Language Engineering*.
- [65] Carroll, J., & Oepen, S. (2005). High efficiency realization for a wide-coverage unification grammar. In *Natural Language Processing-IJCNLP 2005* (pp. 165-176). Springer Berlin Heidelberg.
- [66] Luke S. Zettlemoyer and Michael Collins. 2005. Learning to Map Sentences to Logical Form: Structured Classification with Probabilistic Categorical Grammars. In *Proceedings of UAI*.
- [67] Michael White and Jason Baldridge. 2003. Adapting Chart Realization to CCG. In *Proc. of the 9th European Workshop on Natural Language Generation*.
- [68] Michael White. 2004. Reining in CCG Chart Realization. In *Proc. of the 3rd International Conference on Natural Language Generation*.
- [69] Michael White. 2006. CCG Chart Realization from Disjunctive Inputs. In *Proc. of the 4th International Conference on Natural Language Generation (INLG-06)*.
- [70] Michael White, Rajakrishnan Rajkumar and Scott Martin. 2007. Towards Broad Coverage Surface Realization with CCG. In *Proc. of the 2007 Workshop on Using Corpora for NLG: Language Generation and Machine Translation*.

- [71] David Chiang. 2005. A Hierarchical Phrase-Based Model for Statistical Machine Translation. In *Proceedings of the 43rd annual meeting on Association for Computational Linguistics*.
- [72] Yuk Wah Wong; Raymond Mooney. 2007. Generation by Inverting a Semantic Parser that Uses Statistical Machine Translation. In *Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics*.
- [73] Terry Koo, Alexander M. Rush, Michael Collins, Tommi Jaakkola, and David Sontag. 2010. Dual Decomposition for Parsing with Non-Projective Head Automata. In *Proceedings of EMNLP 2010*.
- [74] Alexander M. Rush and Michael Collins. 2011. Exact Decoding of Syntactic Translation Models through Lagrangian Relaxation. In *Proceedings of ACL 2011*.
- [75] Yi Zhang, Hans-Ulrich Krieger. 2011. Large-Scale Corpus-Driven PCFG Approximation of an HPSG. In *Proceedings of 12th International Conference on Parsing Technologies*.
- [76] Reiter, E. & Dale, R. (2000). *Building natural language generation systems* (Vol. 33). Cambridge: Cambridge university press.
- [77] Reiter, E. (2007, June). An architecture for data-to-text systems. In *Proceedings of the Eleventh European Workshop on Natural Language Generation* (pp. 97-104). Association for Computational Linguistics.
- [78] Goldberg, E., Driedger, N., & Kittredge, R. (1994). Using natural-language processing to produce weather forecasts. *IEEE Expert*, 9(2), 45-53.
- [79] Sripada, S., Reiter, E., & Davy, I. (2003). SumTime-Mousam: Configurable marine weather forecast generator. *Expert Update*, 6(3), 4-10.
- [80] Reiter, E., Sripada, S., Hunter, J., Yu, J., & Davy, I. (2005). Choosing words in computer-generated weather forecasts. *Artificial Intelligence*, 167(1), 137-169.
- [81] Belz, A. (2008). Automatic generation of weather forecast texts using comprehensive probabilistic generation-space models. *Natural Language Engineering*, 14(04), 431-455.
- [82] Belz, A., & Kow, E. (2009, March). System building cost vs. output quality in data-to-text generation. In *Proceedings of the 12th European Workshop on Natural Language Generation* (pp. 16-24). Association for Computational Linguistics.
- [83] Bohnet, B., Lareau, F., & Wanner, L. (2007). Automatic production of multilingual environmental information. In *Proceedings of the 21st Conference on Informatics for Environmental Protection (EnviroInfo-07), Warsaw, Poland*.
- [84] Kukich, K. (1983, June). Design of a knowledge-based report generator. In *Proceedings of the 21st annual meeting on Association for Computational Linguistics* (pp. 145-150). Association for Computational Linguistics.
- [85] Kahn, M. G., Fagan, L. M., & Sheiner, L. B. (1991). Combining physiologic models and symbolic methods to interpret time-varying patient data. *Methods of information in medicine*, 30(3), 167-178.
- [86] Hüske-Kraus, D. (2003, April). Suregen-2: A shell system for the generation of clinical documents. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics-Volume 2* (pp. 215-218). Association for Computational Linguistics.
- [87] Portet, F., Reiter, E., Gatt, A., Hunter, J., Sripada, S., Freer, Y., & Sykes, C. (2009). Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, 173(7), 789-816.
- [88] B. Yao, X. Yang, L. Lin, M. W. Lee, and S.-C. Zhu. 2010. I2t:image parsing to text description. *IEEE Xplore*.



- [89] Y. Feng and M. Lapata, “How Many Words Is a Picture Worth? Automatic Caption Generation for News Images,” *Proc. Assoc. for Computational Linguistics*, pp. 1239-1249, 2010.
- [90] Y. Feng and M. Lapata. 2013. Automatic caption generation for news images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35.
- [91] Y. Yang, C. L. Teo, H. Daumé III, and Y. Aloimonos. Corpus-guided sentence generation of natural images. In *EMNLP*, 2011
- [92] G. Kulkarni, V. Premraj, S. Dhar, S. Li, Y. Choi, A. C. Berg, and T. L. Berg. 2011. Baby talk: Understanding and generating image descriptions. In *CVPR*.
- [93] Kulkarni, Girish, Premraj, Visruth, Ordonez, Vicente, Dhar, Sagnik, Li, Siming, Choi, Yejin, Berg, Alexander C, and Berg, Tamara L. Babytalk: Understanding and generating simple image descriptions. *PAMI, IEEE Transactions on*, 35(12):2891–2903, 2013.
- [94] Mitchell, Margaret, Han, Xufeng, Dodge, Jesse, Mensch, Alyssa, Goyal, Amit, Berg, Alex, Yamaguchi, Kota, Berg, Tamara, Stratos, Karl, and Daumé III, Hal. Midge: Generating image descriptions from computer vision detections. In *European Chapter of the Association for Computational Linguistics*, pp. 747–756. *ACL*, 2012.
- [95] Elliott, Desmond and Keller, Frank. Image description using visual dependency representations. In *EMNLP*, pp. 1292–1302, 2013.
- [96] Hodosh, Micah, Young, Peter, and Hockenmaier, Julia. Framing image description as a ranking task: Data, models and evaluation metrics. *Journal of Artificial Intelligence Research*, pp. 853–899, 2013.
- [97] A. Karpathy and L. Fei-Fei. Deep visual-semantic alignments for generating image descriptions. *CVPR*, 2015.
- [98] R. Socher, A. Karpathy, Q. V. Le, C. D. Manning, and A. Y. Ng. Grounded compositional semantics for finding and describing images with sentences. *TACL*, 2014.
- [99] X. Chen and C. L. Zitnick. Mind’s Eye: A Recurrent Visual Representation for Image Caption Generation. *CVPR*, 2015
- [100] Junhua Mao, Wei Xu, Yi Yang, Jiang Wang, Zhiheng Huang and Alan L. Yuille, Deep Captioning with Multimodal Recurrent Neural Networks (m-RNN), *ICLR* 2015
- [101] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell. Long-term recurrent convolutional networks for visual recognition and description. *CVPR*, 2015.
- [102] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: A neural image caption generator. *CVPR*, 2015.
- [103] Xu, K., Ba, J., Kiros, R., Courville, A., Salakhutdinov, R., Zemel, R., & Bengio, Y. (2015). Show, attend and tell: Neural image caption generation with visual attention. In *ICML*.
- [104] H. Fang, S. Gupta, F. Iandola, R. Srivastava, L. Deng, P. Dollár, J. Gao, X. He, M. Mitchell, J. Platt, C. L. Zitnick, and G. Zweig. From captions to visual concepts and back. *CVPR*, 2015
- [105] Jacob Devlin, Hao Cheng, Hao Fang, Saurabh Gupta, Li Deng, Xiaodong He, Geoffrey Zweig, Margaret Mitchell, Language Models for Image Captioning: The Quirks and What Works, *arXiv* 2015

## 作者简介

**万小军** 博士，北京大学计算机科学技术研究所研究员，博士生导师。

主要研究方向为自然语言处理、文本挖掘。Email: wanxiaojun@pku.edu.cn



**冯岩松** 博士，北京大学计算机科学技术研究所讲师。主要研究方向为自然语言处理。Email: fengyansong@pku.edu.cn



**孙薇薇** 博士，北京大学计算机科学技术研究所讲师。主要研究方向为计算语言学。Email: ws@pku.edu.cn



注：本文第 1，2，4，6 节由万小军撰写、第 3 节由孙薇薇撰写、第 5 节由冯岩松撰写。博士生姚金戈参与了校对工作。