

机器翻译研究进展与趋势

CCF 中文信息技术专委会

摘要

机器翻译实现自然语言之间的自动互译,是自然语言处理和人工智能领域的经典难题之一。近年来,机器翻译受到越来越多的关注。一方面,机器翻译技术拥有迫切的社会需求,并不完美的机器翻译技术在产业界的应用越来越广泛,实时语音翻译系统和在线翻译服务已进入日常生活;另一方面,机器翻译一直是学术界研究的一个热点,关于机器翻译技术的研究论文在数量上长期占据计算语言学(或自然语言处理)领域各大学术会议的头名。本文首先简要介绍机器翻译研究的概况,着重阐述近年来国内外的相关研究现状和技术进展,然后对机器翻译的发展趋势进行展望。

Abstract

Machine translation (MT) performs automatic translation between two natural languages and it is one of the classic problems in the area of natural language processing and artificial intelligence. In the recent years, MT has drawn more and more attention. On one hand, there are imperative application demands and more and more imperfect MT systems appear in the industry, such as real-time speech translation and on-line translation. On the other hand, MT is a hot research topic and research articles focusing on MT account for the largest part in most top-tier conferences in the area of computational linguistics or natural language processing. This article first briefly introduces the history and status of MT, and then presents recent research work and progress on MT in detail. Finally, we propose some future research directions for MT research and development.

1. 引言

人们通常习惯于感知(听、看和读)自己母语的声音和文字,在与说外语的人交流或阅读外语文本时,翻译(将外语翻译成母语)便成为一种自然的需求。打破语言障碍,实现准确的机器自动翻译成为人类的一个美好梦想。计算机的出现为机器翻译思想的提出奠定了基础。受到英国在第二次世界大战期间使用计算机成功破解德国英格玛(Enigma)密码的启发,Warren Weaver于1949年正式提出了机器翻译的问题:外语只是母语的一种编码,我们只需利用破解密码的方式将外语翻译成母语。由于这一比喻是如此形象,加之破解密码的成功,人们好像看到了希望,对机器翻译的研究充满了期待。然而,60多年的研究历程表明,机器翻译的发展道路布满了曲折和艰辛。

基于双语词典的词汇替换方法和基于简单分析、转换与生成的规则方法是机器翻译早期

研究阶段的两种主流方法。由于当时这些方法相对粗糙,无法实现令人信服的自动翻译应用,因此,在1966年美国国家科学院发布的ALPAC报告的影响下,美国政府和军方对机器翻译研究所投入的资助几乎都终止了,机器翻译研究在世界范围内陷入了低谷。

但是,并不是所有的机器翻译研究人员都放弃了希望,他们一直在基于规则的(rule-based)理性主义思维的研究道路上不断探索。20世纪70年代和80年代,出现了一批商用的翻译系统,譬如加拿大蒙特利尔大学开发的天气预报信息翻译系统TAUM-METEO、美国SYSTRAN¹公司开发的“俄-英”、“法-英”翻译系统以及日本计算机公司研制的“日-英”翻译系统等。然而,基于规则的方法存在一些难以突破的瓶颈问题,例如,规则获取成本高、大规模规则维护困难,以及新规则与已有规则易发冲突等问题。

自20世纪80年代末90年代初,随着计算机硬件设备和计算能力的飞速提升,以及互联网的快速发展和普及,大规模双语平行语料库的建立与发展,以及机器学习技术的应用,使得基于经验主义思维的研究方法(数据驱动方法)逐步取代了理性主义方法,成为机器翻译研究的主流方法。基于数据驱动的(data-driven)翻译方法可进一步划分为基于实例的(example-based)翻译方法和统计机器翻译(statistical machine translation, SMT)方法。统计机器翻译方法的出现是一大突破,2000年左右随着一批开源工具(词对齐工具GIZA++²、语言模型工具SRILM³和评价测度BLEU⁴等工具)的发布,逐渐成为机器翻译研究领域的主流。

现在的机器翻译技术已全面进入人们的日常生活。在很多应用领域,机器翻译即将成为领域革新的关键技术。在体育、文化和教育领域,人们接触到越来越多的外文资料,越来越频繁地与持各种语言的人通信和交流,因而对机器翻译的需求越来越强烈。在旅游、商贸和娱乐领域,机器翻译将在旅游资讯的翻译、各种商业产品与游戏的本地化方面发挥越来越大的作用;在国家信息安全领域和军事情报领域,机器翻译技术将扮演重要角色。我们看到大数据时代已经来临,互联网上每天都在产生海量的各种语言的数据,离开机器翻译,基于大数据的多语言信息挖掘、发现和决策将成为空中楼阁。

当然,机器翻译技术还不完美,但随着机器翻译研究的不断深入、翻译质量的不断提升,其实用性和应用性也将得到实质性的提升。因此,从各方面讲,机器翻译的研究在现阶段和未来一段时间将非常重要。

本文下面的部分将对机器翻译的方法、最新进展和技术现状等,做详细阐述。

2. 国内外研究现状

自机器翻译诞生以来,其研究路线围绕理性主义方法和经验主义方法两种思潮经过了两个

¹ <http://www.systranet.com/translate>

² <http://www.fjoch.com/GIZA++.html>

³ <http://www.speech.sri.com/projects/srilm>

⁴ <http://www.itl.nist.gov/iad/mig//tools/>

次转变。所谓“理性主义”，就是指以生成语言学为基础的方法，在机器翻译领域表现为基于规则的研究方法；所谓“经验主义”，就是指以数据驱动为基础的方法，在机器翻译领域表现为基于语料库的研究方法[1]。

很多研究方向在发展道路上都经历过辉煌和低谷，机器翻译的发展则更加曲折。享受过追捧也遭受过嘲讽，但学者们从未放弃，研究也从未停止。

2.1 基于规则的机器翻译

从 20 世纪 50 年代机器翻译问题的正式提出到 90 年代，机器翻译研究一直是基于理性主义的规则方法的舞台。基于规则的机器翻译方法的一般过程可分为分析、转换和生成三个阶段。分析阶段完成对源语言句子的解析，将源语言句子解析成一种深层结构表示。转换阶段将源语言句子的深层结构表示转换为目标语言的深层结构表示。生成阶段完成从目标语言的深层结构表示到目标语言句子的生成过程。

根据深层结构的表示方式，基于规则的机器翻译方法又细分为基于句法的方法、基于语义的方法和基于中间语言的方法。基于句法的方法着重研究源语言句法结构的分析和目标语言句法结构的生成。这种方法一直是基于规则的机器翻译方法的主流，围绕这一方法，人们曾尝试了多种句法理论，如生成转换语法、依存语法等。基于语义的方法尝试对原文进行语义分析，获得原文的语义表示，然后进行语义表示转换并生成目标语言译文。基于中间语言的方法设想有一种中间语言，每一种自然语言都可以映射到该中间语言，执行翻译时，系统首先将源语言句子分析表示为中间语言，然后基于中间语言生成目标语言译文。

基于规则的翻译方法的效果与自然语言分析和生成的水平有着密切关系。由于生成转换语法、依存语法、功能合一语法与词汇功能语法等句法理论相对成熟，因此，基于句法规则的翻译方法在特定条件下能够取得较好的译文质量，在很多应用上都获得了初步的成功。例如，美国的 SYSTRAN 公司和国内的华健公司，他们以基于规则为主要方法的翻译系统一直在为企业和政府提供翻译服务。相比而言，语义分析与中间语言一直停留在初步的研究阶段，到目前为止仍没有一套较为成熟的语义分析系统和被广泛认可的中间语言表示。所以，基于语义和中间语言的规则翻译方法也只是进行了一些很初步的探索。

综上所述，基于规则的机器翻译方法能够充分利用语言学家总结出来的语言规律，具有一定的通用性。但是，如前所述，该方法存在一些难以突破的瓶颈问题。如规则一般仅从规范文本的语言现象中总结获得，规则的质量和效果依赖于语言学家的知识和经验，规则获取的人工成本较高，而且维护大规模规则往往比较困难，新规则与已有规则易发兼容性问题等。这些问题导致基于规则的机器翻译方法很难处理大规模的真实文本，尤其是对于弱规范或不规范的口语语言现象和网络用语，处理能力较差。相对于基于规则的翻译方法，如果系统能够从以前的翻译实例中自动学习到翻译知识，将会为系统自动更新，提高系统的鲁棒性和翻译准确率，带来极大的好处。因而，基于大规模真实文本的翻译方法应运而生。

2.2 基于语料库的机器翻译

20 世纪 80 年代末至 90 年代初, 随着计算机技术的快速发展, 大规模双语语料库的构建以及机器学习方法的兴起, 机器翻译方法逐渐由基于理性主义思维的规则方法转向基于经验主义思维的语料库方法。

2.2.1 基于实例的机器翻译

基于语料库的机器翻译方法可进一步划分为基于实例的翻译方法和基于统计模型的翻译方法。基于实例的机器翻译方法是由日本著名学者长尾真(Makoto Nagao)教授于 1980 年代提出[2]。这种方法试图在平行语料库中找出与源语言文本相似的实例(通常是句子), 并对实例对应的译文进行适当的替换、删除和插入等一系列变形操作, 实现翻译。

基于实例的方法借鉴了类比的原理, 一般由如下部分构成[3]:

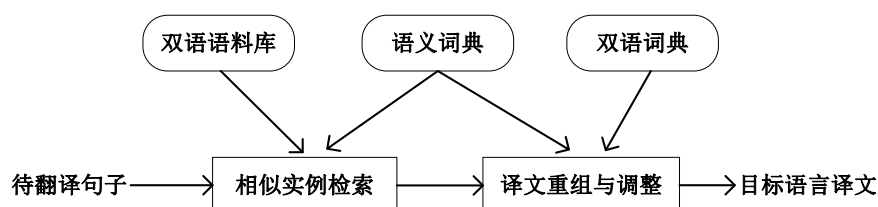


图 1 基于实例的机器翻译方法基本结构图

基于实例的翻译方法直接根据某种相似度(如基于字符串匹配、基于词的匹配或基于向量空间模型等)和语义词典, 从双语语料库中检索出与待翻译句子相似的实例(可能是多个实例), 然后从实例中获得相同片段的译文, 最后根据实例译文的结构重组片段的译文候选, 从而得到最终的目标译文。

可以看出, 基于实例的机器翻译方法无需对句子进行复杂的语言分析, 可直接利用已有的翻译实例库。但是, 由于该方法采用的实例粒度一般是句子, 因此, 只有找到相似度很高的实例时才能体现该方法的优点。因而基于实例的方法无法充分利用翻译实例库(如有很多相似度低的实例中也可以提供一些词或短语的翻译信息)。

2.2.2 统计机器翻译

机器翻译领域因统计方法的出现而充满了活力。统计方法的基本思想就是充分利用机器学习技术从大规模双语平行语料中自动获取翻译规则及其概率参数, 然后利用翻译规则对源语言句子进行解码。20 世纪 90 年代初期, IBM 的研究人员提出了基于词的五翻译模型[4-5], 正式标志着统计机器翻译时代的到来。但是, 基于词的统计机器翻译模型复杂度较高, 翻译质量却难以尽如人意, 因此, 逐渐被一些更加有效的翻译模型所替代。近 10 年来, 一系列充满新意的翻译模型相继被提出。图 2 给出了当前统计机器翻译中一些典型的翻译模型。

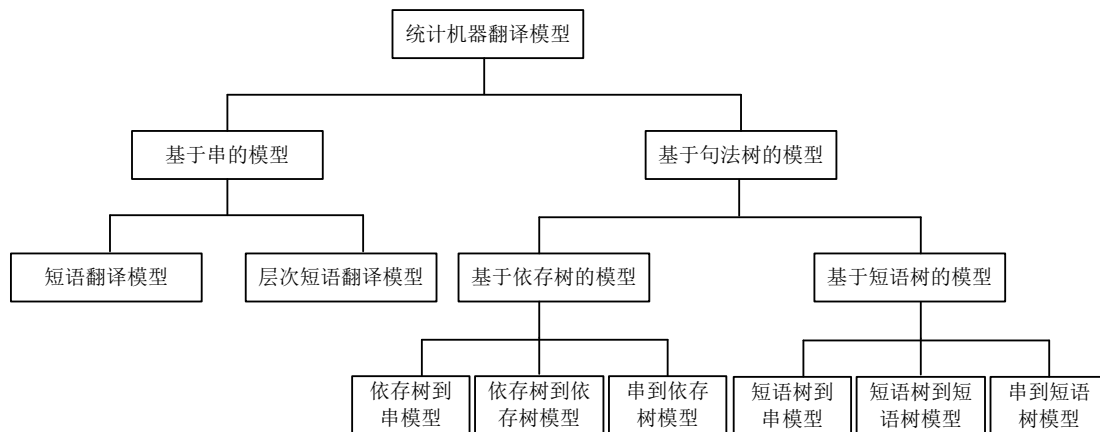


图 2 统计机器翻译研究中的前沿模型

基于短语的翻译模型（phrase-based translation model）是其中最为成熟的模型[6]。这里的“短语”表示连续的词串。该模型的基本思想是：首先从双语句子对齐的平行语料库中抽取短语到短语的翻译规则，在翻译时将源语言句子切分为短语序列，利用翻译规则得到目标语言的短语序列，然后借助调序模型对目标语言短语序列进行排序，最终获得最佳的目标译文。其中，短语调序模型，尤其是长距离的短语调序，一直是短语翻译模型的一个关键问题。

若翻译规则中的短语含有变量，短语翻译模型就发展成为基于层次短语的（hierarchical phrase-based）翻译模型[7]，具有更强的表达能力，能够取得更好的翻译性能。图 3 以汉英翻译为例对比了短语翻译模型与层次短语翻译模型。但是，层次短语翻译模型存在一个问题：变量 X 过于泛化，导致抽取的规则数量庞大，且规则的区别性不够好。

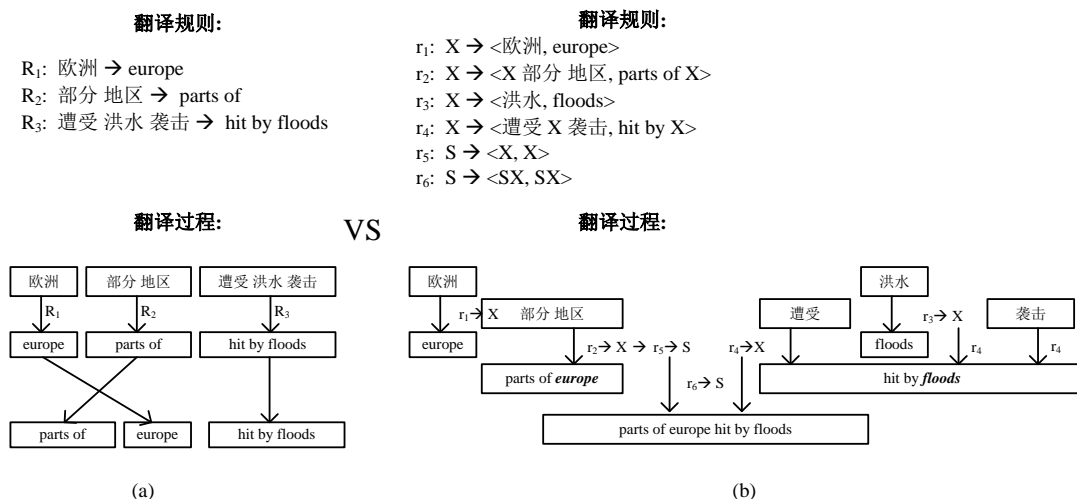


图 3 短语翻译模型（a）与层次短语翻译模型（b）在翻译汉语句子“欧洲 部分 地区 遭受洪水”时的对比。

通过图 3 中的例子我们可以看出，短语模型翻译过程中需要短语调序模型的参与，而在层次短语模型中短语调序隐含于规则当中。

将语言学知识融入到翻译建模过程一致被认为是攻克机器翻译堡垒的正确道路。因此，基于语言学句法（linguistically syntax-based）的统计翻译模型正是目前这一领域研究的热点课题之一，一批基于句法的（syntax-based）翻译模型相继被提出。

根据翻译规则表示形式的不同，这些模型可以被划分为两种：（1）基于依存树的（dependency tree-based）翻译模型；（2）基于短语结构树的（phrasal structural tree-based）翻译模型。不管是基于依存树还是基于短语结构树，如果翻译规则的源端和目标端都是树结构，这种模型称为树到树的（tree-to-tree）翻译模型。在树到树的翻译模型中，由于过度约束的原因，其翻译性能并不卓越。相比而言，源端或目标端放松约束的依存树到串（dependency tree-to-string）的翻译模型[8]、串到依存树（tree-to-dependency tree）的翻译模型[9]、短语树到串（phrasal tree-to-string）的翻译模型[10-11]以及串到短语树（string-to-phrasal tree）的翻译模型[12-13]，都能超越层次短语翻译模型。图 4 对比了基于短语结构树的树到串、树到树和串到树三种翻译模型的翻译规则形式。

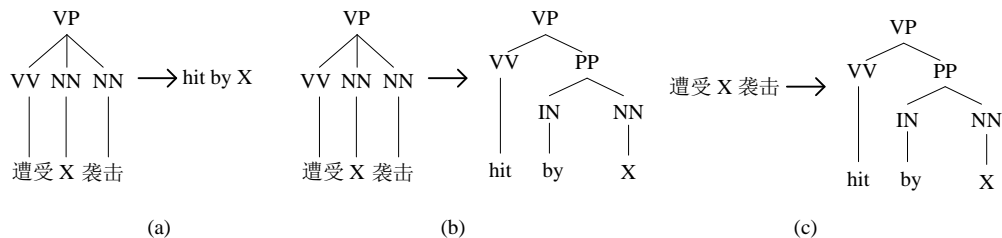


图 4 (a)树到串翻译规则 (b)树到树翻译规则 (c)串到树翻译规则

基于句法的翻译模型存在两个主要问题：（1）使用的树结构由句法分析器产生，独立于平行语料库，与词语对齐不兼容；（2）源语言端和目标语言端的句法知识很难同时被有效地利用。对于问题（1），研究者进行了两种尝试。[14-16]等将句法树或者词对齐视为正确结果，设计算法修改词对齐结果或者句法分析树结构，从而使两者更加兼容；[17-18]则绕过传统句法分析器，直接从平行语料和词对齐出发，利用无监督方法生成与词对齐结果完全兼容的树结构。针对问题（2），目前主要有三种尝试：①在层次短语翻译模型的基础上发展形成的模糊短语树到模糊短语树的翻译模型[19]；②在短语树到串模型的基础上发展形成的短语树到依存树的翻译模型[20]；③在串到短语树模型的基础上发展形成的模糊短语树到短语树的翻译模型[21]。这三种成功的尝试有力地推动了基于语言学句法的翻译方法研究，取得了当前最好的翻译性能。

虽然统计机器翻译方法目前几乎主导着整个机器翻译领域，但是，我们应清楚地看到该方法的不足之处：（1）深度依赖双语平行语料的质量和规模，领域适应性差；（2）对语言的分析不够深入，还未出现真正基于语言理解的统计机器翻译方法；（3）只关注句子层面的双语转换，忽略段落上下文和篇章信息，从而造成译文不连贯，衔接性差等问题。

容易发现，基于理性主义的规则方法和基于经验主义的语料库方法都各有优势和不足。近年来，机器翻译研究正朝着规则方法和语料库方法相结合，兼容并包，博采众长的方向发展。

3. 国内外研究进展及比较

3.1 理论模型

近年来,无论是从各种顶级学术会议上,还是从各类互联网应用中,我们都不难发现机器翻译一直是自然语言处理领域中一个炙手可热的研究方向。从事机器翻译研究的人员越来越多,新的机器翻译方法和模型不断涌现。这些方法和模型主要关注机器翻译中的如下几个方面:(1)领域自适应问题;(2)基于语义的翻译模型;(3)基于篇章的翻译模型;(4)基于无监督学习的翻译模型;(5)基于融合技术的翻译方法。

3.1.1 领域自适应问题研究

对于统计机器翻译来说,由于翻译规则必须从双语平行语料中学习获得,这就导致了用某一领域的平行语料训练得到的翻译模型在翻译另一个领域的文本时很难取得好的译文质量。这就是典型的领域自适应问题,这一问题主要体现在如下两个方面:一是如何学习新领域中未登录词的翻译[22];二是如何在多个候选译文中选择符合新领域意境的译文。后者可以视为领域相关的词义消歧问题,被看作是目前研究的重点。近两三年内,研究者们分别从翻译模型、短语调序模型和语言模型三个关键组成部分对该问题进行了深入探讨。

由于翻译模型是机器翻译的核心,因此,翻译模型的领域自适应问题倍受关注。关于这一问题的研究方法大致可以分为两类:第一类主要是利用领域和主题信息,通过模型融合或者主题模型的方法使得翻译规则的概率符合领域的偏好[23-27];第二类主要是在训练翻译模型时利用领域知识从大规模通用数据中筛选出领域相关的双语平行语料[28]。

语言模型是衡量译文流畅性的重要方法。相对而言,处理语言模型的领域自适应问题所采用的方法比较固定:根据语言模型的困惑度从海量单语语料中选择与所关注领域最相关的那些数据来训练语言模型[29-30]。

短语调序模型是基于短语的统计机器翻译中的一个重要组成部分。短语调序模型的领域自适应问题之前几乎没有受到关注。[31]利用模型融合的方法使得词汇化的短语调序模型具有领域偏好的功能。

实际上,领域自适应问题不仅是自然语言处理中的一个难点,而且是机器学习方法研究及其应用中一个普遍存在的难题。这一问题在未来几年内,仍然是机器翻译研究中的一个热点。2012年夏天部分研究人员在约翰霍普金斯大学举办了一个关于机器翻译领域自适应方法的研讨班,这正说明了机器翻译的研究人员对领域自适应问题的重视程度。

3.1.2 基于语义的翻译模型研究

目前在已有的统计机器翻译方法中,语言学句法知识是用于翻译建模的最深层次的信息。句法知识能够展示一个句子的组织结构,但无法表达一个句子的语义信息(例如施事、受事、与事、动作等),因此,系统在翻译时并未真正理解句子的含义。对于研究者来说,最终的目标是建立基于语言理解的机器翻译系统。因此,从短语、句法走向语义层面,已经成为近

年来机器翻译研究的一个共识。

[32]提出了利用语义角色标记信息对候选译文列表进行重新排序的方法；[33]借助谓词—论元结构调整源语言句子的语序，使之在词序上更接近于目标语言表达。这类方法利用语义信息对翻译过程进行预处理或后处理。

[34]在基于句法的翻译模型中融入语义角色调序特征；[35]用语义角色信息标记句法翻译模型中的非终结符号，使翻译规则更具区分度；[36]关注谓词和论元的译文顺序，并采用语义角色标记作为特征建立短语调序模型，以预测谓词和论元的翻译语序。这类方法将语义信息作为一种特征融入现有的翻译模型，并没有建立一种以语义为框架的翻译模型。

[37-38]尝试建立了一种基于谓词—论元结构转换的语义翻译模型。由于谓词—论元结构描述了句子的骨架和语义结构，而且，与句法结构相比，源语言和目标语言在谓词—论元结构上更趋于一致。因此，基于谓词—论元结构转换的翻译模型能够最大限度地保持原文的语义信息。这种翻译模型的执行过程分为三个步骤：①分析源语言句子的谓词—论元结构；②利用从训练语料中学习得到的概率化转换规则将源语言的谓词—论元结构转化为目标语言结构；③翻译源语言句子的每个论元，填充目标语言结构，得到最终的译文输出。

不同于谓词—论元结构，[39-40]设计了一种图结构进行抽象的语义表示（abstract meaning representation, AMR），并标注了一个 AMR 语义库，建立了基于 AMR 的翻译模型。基于 AMR 的翻译模型[39]首先将源语言句子分析为 AMR 图结构，完成“串到图”（string-to-graph）的转换，然后基于 AMR 图结构生成目标语言译文，实现“图到串”（graph-to-string）的生成过程。

从目前情况来看，基于语义的翻译模型仍处于探索阶段，在未来的几年内必将是机器翻译领域的一个重点研究方向。

3.1.3 基于篇章的翻译模型研究

在相当长的一段时间内，机器翻译方法仅仅关注句子级的翻译（即逐句翻译，句与句之间是独立的），而忽略段落和篇章的上下文信息对句子翻译的影响，往往导致译文之间缺乏衔接性和连贯性。近两三年来，基于篇章的翻译研究受到越来越多的关注。

衔接性和连贯性是篇章表达中的两个重要属性。目前，基于篇章的翻译模型研究大多利用衔接性和连贯性作为约束或特征，以改善翻译模型的性能。[41]依据词汇篇章衔接性质，利用动态、静态和主题缓存技术改善词或短语译文的选择，企图使得相同的词和短语在整个文档中保持译文的一致性；[42-43]直接将同一文档中相同词的译文一致性作为约束。[44]设计了一种模型将词汇衔接作为特征融入解码器中。在利用连贯性方面，[45]设计了一种主题转换模型改善篇章级统计机器翻译的质量。

到目前为止，直接基于篇章理论建立翻译框架的研究还非常少见。最近，[46]在这方面做了初步的探索，他们尝试建立了一种以篇章中修辞结构理论(rhetorical structure theory, RST)为框架的统计机器翻译模型。这一工作还在完善当中，我们相信，越来越多的研究者将会关注并投身到这一方向。2013年国际计算语言学学会年会(ACL)上将举办一个基于篇章的机器

翻译研讨会，这意味着基于篇章的翻译方法可能是未来几年的一个研究热点。

3.1.4 基于无监督学习的翻译模型研究

一般而言，在给定足够多的双语平行语料的情况下，统计机器翻译系统可以获得较好的译文质量。可见，提高统计翻译系统的一个关键前提是准备充足的双语平行语料。然而，在实际情况中，对于很多语言对和特定领域来说，很难获得平行语料，甚至根本没有这样的数据。人工收集成本高代价大，而且很难规模化，领域的平衡性也难以控制。因此，面对双语资源缺乏的领域和语言对，如何建立翻译模型，是人们关注的另一个重要问题。为了解决这一问题，很多研究人员基于现有的翻译模型，通过信息检索等技术从海量网络资源中挖掘平行语料。然而，虽然网络中存在大量潜在的双语资源，但网上的语言资源基本都集中在少数语种和领域上，很难解决大多数领域和语言对的资源获取问题。因此，近两三年来，机器翻译领域产生了一种新的思想：直接从两种语言的大规模单语语料中学习翻译模型。

[47]从机器翻译最初的思想出发，将基于单语语料的翻译问题视为一个密码破译问题。与传统的基于噪声信道模型的翻译方法相比，这种方法的噪声信道输入也是未知的。有效穷举或抽样噪声信道的输入成为基于单语语料的无监督翻译模型的核心思想。以基于词的翻译模型为框架，[47]分别采用了期望最大化(expectation maximum, EM)算法和 Gibbs 采样算法估计模型参数；[48]有效改进了 EM 算法；[49]改进了 Gibbs 采样算法；[50]提出了一种哈希采样算法解决大规模的基于单语语料的翻译模型学习问题。

上述这些方法基于如下的假设：在两种语言相同领域的大规模单语语料中存在大量互为翻译的词语对。于是，基于单语语料的词的翻译模型应运而生。实际上，除了词语翻译对以外，在两种语言相同领域的大规模单语语料中也存在很多短语翻译对。[51]在这个假设的基础上设计了一个面向单语语料的基于短语的统计机器翻译方法。该方法借鉴信息检索的思想从单语语料中抽取双语短语翻译对，利用并拓展了文献[52]中提出的估计短语对的翻译概率和短语调序概率的方法。

这种无监督的翻译方法将在资源缺乏的领域和语言对中发挥重要作用，也必将引起更多研究者的关注。

3.1.5 多引擎融合的翻译方法

任何一种机器翻译模型都有其优势和不足，若能有效地融合不同翻译模型的优势，必将有利于提高翻译质量。基于这种思路，很多研究者从不同的角度探讨了机器翻译中的模型融合问题。已有的融合方法大致分为如下三类：（1）多个统计翻译模型的融合；（2）统计翻译模型与其他基于语料库的翻译方法的融合；（3）统计翻译引擎与规则方法的融合。

多个统计翻译模型的融合方法受到更多的关注。针对这种方法，人们主要研究如何有效地集成基于短语的翻译模型、基于层次短语的翻译模型和基于语言学句法的翻译模型之间的优势。其融合方法一般分为两种：①作为一种后处理，融合各个翻译模型的译文输出，从中搜索最佳译文[53-55]；②在解码过程中融合各个翻译模型的翻译假设，即在更小的粒度上融合各个模型，获得最优翻译结果[56-58]。

在统计机器翻译模型中，通常情况下翻译规则（短语翻译对、同步语法规则等）从双语平行语料中抽取出来之后便与语料没有任何关系了。但是，翻译规则在解码过程中被使用时可能非常严重地依赖于规则所在句对的上下文。因此，近两三年来，不少研究者开始关注统计机器翻译与翻译记忆的结合方法。[59-61]等根据翻译记忆抽取片段翻译信息，用于指导统计翻译的解码过程。

统计方法与规则方法的融合也是当前研究的一个热点。由于基于规则的翻译方法具有与统计翻译方法互补的优势（例如无需双语平行语料库，具有一定的泛化能力等），统计方法与规则方法相结合的混合翻译模型被寄予厚望。目前的混合方法主要采用黑箱法：①利用统计方法对规则翻译系统的翻译结果进行后处理。典型的处理方式是将规则翻译系统的结果视作源语言，正确的译文视作目标语言，采用统计翻译的思想训练翻译模型，对规则翻译系统的译文进行二次翻译[62]；②将基于规则的翻译引擎与统计翻译引擎等价对待，采用多系统融合的方法获得更好的译文[63]。

3.2 技术评测

机器翻译研究者经常会被问到这样一个问题：现在的机器翻译系统译文质量如何？这是一个简单却又很难回答的问题。相对于其他自然语言处理任务，如语音识别，机器翻译任务一般不存在标准答案。只要语义一致，一个源语言句子可以对应多个目标语言译文。那么我们应该如何评测一个机器翻译系统的译文质量呢？

人工评测是最简单的一种方法。对于某个机器翻译系统输出的译文，聘请专家逐个查看每个译文，评判其正确性。一种常用的评判方法是让人工评测者按正确性对译文分级打分。由于“正确性”是一个太过宽泛的概念，通常在人工评测中使用忠实度和流利度作为判断标准。忠实度衡量译文传达了原文意思的程度，保留了原文多少信息，相当于“信、达、雅”中的“信”；流利度则衡量译文是否流畅通顺，是否符合目标语言的表达，相当于“信、达、雅”中的“达”。一般地，都将忠实度和流利度都划分为5个等级（5分制），分值越高，译文越好。

然而，人工评测不仅成本昂贵，而且效率很低。我们经常在设计一种新的算法后希望立刻检验该算法是否对译文质量有所改善。这样人工评测就很难满足这种要求。因此，自动评测是一种理想的方案。

在自动计算译文质量的评价指标时，一般的做法是通过比较机器的译文与一个或多个参考译文的相似程度来定量地给出机器译文的得分。在过去十年左右的时间里，研究者们提出了许多译文质量的自动评价指标，包括 BLEU[64]、NIST[65]、METEOR[66]、TER[67]、GTM[68]、MAXSIM[69]、WoodPecker[70]、TESLA[71]和 MEANT[72]等。以下简要介绍几个常用的评价指标。

BLEU 是目前使用最广泛的自动评测指标，由于其简洁、可靠，被各机器翻译评测组织用作译文质量的官方评价指标。BLEU 指标的原理是：机器译文越接近专业人工的翻译结果，

译文的质量就越好。其计算方法是：统计机器译文与参考译文（一般有 4 个参考译文）中 n 元语法（通常取 $n=4$ ）匹配的数目占有 n 元语法总数的比率。在计算 n 元语法匹配比率的基础上，BLEU 指标还通过引入一个长度惩罚因子防止过短的译文容易获得较高的分值。

NIST 指标在 BLEU 指标的基础上，对 n 元语法依据其出现的频繁程度赋予不同的权重。不同于 BLEU 和 NIST 计算 n 元语法匹配的方法，TER 在词的级别上度量机器译文与参考译文之间的距离，计算将机器译文变为参考译文所需要的编辑次数，并且不同于简单的编辑距离，它允许任意语块的移动。

上述三种指标强调的都是准确率，而很多实验结果表明，在确保译文意思完整性方面，召回率比准确率更重要。METEOR 是一种强调召回率的译文评测指标。相比于前三种评测指标，METEOR 增加了词干还原和同义词的使用。它首先从词形上对译文进行匹配，然后对于未匹配上的词进行词干还原后再进行匹配，并且使用 WordNet 定义的语义类信息。

这些评测指标主要还是在字符串层面度量机器译文与参考译文之间的相似程度，缺乏句法和语义知识的参与。因此，近年来很多研究者开始探讨基于句法结构、语义知识和篇章信息的自动评测指标。依据测试数据集和参考译文数据集之间的词对齐结果和它们的句法分析树，WoodPecker 首先自动从源语言和目标语言中抽取各种检测点类型，包括名词短语，动宾搭配，介词短语和新词等几十种语言学类型。然后，WoodPecker 通过计算测试语料中检测点的参考翻译结果与机器翻译系统翻译结果之间的匹配程度来评估翻译系统在特定语言学现象方面的翻译能力。MEANT 则是一种基于语义框架的译文评测指标，更关注于译文的忠实度。[73-74]探索了基于篇章结构的译文评测方法。[73]在篇章译文的评测中加入了话语连接的特征，而[74]对 BLEU、TER 和 METEOR 等句子级的评测指标进行了扩展，使其满足检测篇章级词汇衔接性的需要。

3.3 系统应用

由于统计机器翻译方法具有很多优势，如开发速度快、周期短、无需人工干预等，在特定领域训练数据充分的情况下译文虽不完美，但也能够达到可理解的水平。因此，近 20 年来统计机器翻译研究发展非常迅速，尤其随着大数据和云计算技术的快速发展，统计机器翻译系统已经走进了人们的日常生活，并在很多特定领域为满足各种社会需求发挥了重要作用。

Google、Microsoft 与国内的百度、有道等互联网公司都为用户提供了免费的在线多语言翻译系统。Google 主要关注以英语为中心的多语言翻译。百度则同时关注以英语和汉语为中心的多语言翻译。Google 在 2013 年的 Google I/O 大会上透露，Google 翻译平台目前能够支持 71 种语言之间的互译，每天为约 2 亿用户提供 10 亿次的翻译服务，处理文本数量相当于 100 万本书籍的内容。另外，即时通讯工具 GoogleTalk、MSN 等以及社交网络 Facebook 也都提供了即时翻译服务。实时语音翻译系统也纷纷出现，如中科院自动化所的紫冬口译⁵和 IBM、Google 推出的产品等。不可否认，机器翻译已经影响着我们的生活。

⁵ <http://mobile.91.com/Soft/Android/interpret.main-1.1.0.html>

在专业应用领域，Asia Online、SDL Language Weaver、SYSTRAN 和国内的华健、格微等翻译公司已经连续多年向企业和政府提供翻译服务。国内的很多科研单位，如中科院自动化研究所开发的以外汉翻译为重点的多语言机器翻译系统（包括：英汉双向、日汉、法汉、德汉、阿拉伯语到汉语、藏汉、（新、老）维汉和蒙汉等翻译），已经在国家多个特定部门获得实际应用。

目前在绝大多数情形下，机器翻译系统提供的译文只是帮助用户理解原文的大致意思，不可能成为直接出版的流畅译文。要得到完全正确流利的译文，还需要专业译员的修饰和编辑。为了提高专业译员的效率，目前很多研究机构和公司都在致力于拉近机器翻译与专业译员的距离，为专业译员提供质量上乘的机器翻译候选结果以便其经过较少的后编辑操作就可获得正确的译文。例如，国际翻译服务提供商 TAUS 向专业译员提供的翻译工具中，统计机器翻译系统 Moses 是其中的核心模块。从目前情况来看，机器翻译技术用于专业领域翻译的时机已趋于成熟。

3.4 性能现状

机器翻译系统的应用已比较广泛，其译文质量的实际水平究竟如何，是大家非常关心的一个问题。以下通过几个在线翻译系统的实例做一些具体分析。

例 1：原文：该研究所设于华府，为非营利研究团体。

Google 在线翻译系统给出的译文：The Institute is based in Washington, a nonprofit research organization.

Baidu 在线翻译系统给出的译文：The Institute in Washington is a non-profit research group.

例 2：原文：他呼吁在未来几周进行密集谈判。

Google 在线翻译系统给出的译文：He called for intensive negotiations in the coming weeks.

Baidu 在线翻译系统给出的译文：He called for intensive negotiations in the next few weeks.

例 3：原文：鲍尔森今天转往北京，与中国主管官员讨论推动去年发起的"美中战略经济对话"进程等议题。

Google 在线翻译系统给出的译文：Paulson transferred to Beijing today, with Chinese officials in charge of promoting discussion last year launched the "US-China Strategic Economic Dialogue" process and other issues.

Baidu 在线翻译系统给出的译文：Paulson today to Beijing, discuss the driving last year launched the "strategic economic dialogue" process issues and Chinese officials.

例 4：原文：表酱紫黑国足了。

Google 在线翻译系统给出的译文: Jiang Zi a black table football.

Baidu 在线翻译系统给出的译文: Table that black country foot.

从上述例子可以看出,当原文表述规范且较为简短时,机器翻译系统给出的译文已基本达到可理解的水平,如例 1 和例 2。当原文较长且结构较复杂时,一些词的翻译就会出现错误,而且译文顺序比较混乱,如例 3。当原文高度口语化或使用非规范的网络用语时,机器译文几乎完全不能传达原文的意思,如例 4。这些例子大致反映了当前机器翻译系统的实际水平。

国际机器翻译评测是促进机器翻译研究不断发展的催化剂,其评测结果在一定程度上体现了机器翻译的实际水平。NIST、WMT 和 IWSLT 是目前国际上最为著名的三大评测。其中, NIST 是美国国家标准技术研究院举办的国际机器翻译评测,它主要关注阿拉伯语、汉语等语言与英语之间的互译,国内很多研究机构 and 高校都曾参加过 NIST 评测,并取得了较好成绩。WMT 是欧洲组织的一个机器翻译评测,主要关注欧洲语言之间的互译。不同于 NIST 和 WMT, IWSLT 关注口语的翻译评测,国内的研究机构和高校(中科院自动化所、中科院计算所、哈工大、厦门大学等)参加过多届 IWSLT 评测,并获得了优异成绩。

除了上述三个评测以外,中国中文信息学会组织的全国机器翻译评测(CWMT)的影响力也越来越大。2013 年 10 月第九届全国机器翻译研讨会将在昆明召开。CWMT 重点关注我国少数民族语言到汉语的翻译,如藏汉、维汉和蒙汉等语言的翻译。

通过对比每年的评测结果可以发现,机器翻译系统评测的分值呈增长的趋势,译文质量越来越好。但与专业译员的翻译结果相比,机器翻译还有很长的路要走,很多问题有待于更深入的研究,如命名实体的识别与翻译问题、领域自适应问题和资源缺乏语言对之间的翻译问题等。

4. 发展趋势与展望

4.1 存在的问题

无论是从产业应用的角度,还是从国际评测的水平来看,在过去十多年的时间里机器翻译取得了很大的进展,但仍面临很多难题需要进一步探索。这些问题可以大致归纳为如下几个方面:

(1) **长句和复杂句式的处理问题。**目前的翻译系统对于长句和复杂句式的处理效果还有很不理想。

(2) **弱规范、非规范化文本的翻译问题。**互联网是产生语言文本最多的地方,而互联网上使用的语言文本具有口语化、社交化等诸多新的特征,弱规范甚至不规范的现象比较严重。而目前的机器翻译系统几乎都是面向规范文本的,如何提高非规范文本的处理能力和翻译效果,是目前面临的难题之一。

(3) **双语资源缺乏问题**。在实际应用中,为了实现高性能的多语言自动翻译,通常会遇到双语资源缺乏这样的瓶颈问题。尤其对于少数民族语言与汉语之间的翻译,大多数少数民族语言的电子文本和知识库规模都比较小,双语数据更少,而且缺乏相应的语言基础处理工具。

(4) **缺乏基于理解的翻译模型**。虽然学术界已经提出一些基于语义的翻译模型,但语义分析的质量还不够理想,很多情况下语义分析所带来的好处远不如由于语义分析而引入的噪声。

(5) **篇章级翻译问题**。利用篇章信息进行机器翻译的研究工作才刚刚开始,很多问题尚缺乏有效的解决办法,如指代消解问题和主语省略等现象的处理。

(6) **增长式学习问题**。对于统计机器翻译系统来说,每增加一定数量的双语平行语料都需要重新训练一次模型,系统尚不能实时地从新增加的双语句对中自动学习翻译知识。这让机器翻译用户觉得系统缺乏智能。

(7) **反馈学习问题**。从某种意义上说,当前的机器翻译系统无法对用户反馈作出任何响应,更缺乏思考能力,不能从用户的反馈中学习并自动改进翻译模型。

(8) **机器翻译评测指标问题**。虽然近年来不少基于句法、语义的评测指标相继被提出,但很多方法都非常复杂,难以操作。因此,如何设计又好又方便的机器翻译评测指标也是一个需要研究的问题。

(9) **应用创新问题**。机器翻译目前的使用方式还比较单一,多作为独立的系统。实际上,机器翻译系统可以在很多应用领域发挥作用,或者与其他应用系统相结合。由此预测,机器翻译系统的产业化创新之路可能成为互联网产业的一个新热点。

(10) **资源共享问题**。很多情况下一个研究单位的双语资源虽然缺乏,但多个单位的双语资源合在一起也许非常可观。那么,如何建立数据资源与系统资源的共享机制,也是我们不能不考虑的一个问题。

问题是理论突破和方法创新的驱动力,我们相信上述的若干问题将有效促进机器翻译的快速发展。

4.2 趋势展望

综合近两三年来顶级学术会议中关于机器翻译的学术文章以及机器翻译当前面临的问题,我们认为未来几年机器翻译研究的重点应该包括如下几个方面:

(1) 翻译模型的领域自适应问题

目前机器翻译系统的一个重要缺陷是:基于某一领域数据建立的翻译模型应用于另一个不同的领域时,系统性能下降得非常显著。虽然近年来研究者们针对翻译模型、语言模型、短语调序模型和领域新词翻译等问题尝试了很多领域自适应方法,但每种方法只是解决了某一方面的问题,而缺少解决领域自适应问题统一的有效框架。因此,领域自适应问题仍将是未来几年内一个非常重要的研究方向。

(2) 基于语义建模的统计机器翻译方法

基于语言理解的翻译方法一直是机器翻译研究者们追求的终极目标,而语言理解的本质是语义理解。[37-38]提出的基于谓词—论元结构转换的翻译模型和[39]提出的基于图结构抽象语义表示的翻译模型极大地增强了我们建立基于语义的翻译方法的信心,但是如何建立适合机器翻译的语义表示和利用方法仍然是一个未解的难题。

(3) 基于篇章分析的翻译模型

近两三年来,不少研究者开始重视篇章分析理论在机器翻译中的作用,人们已做了大量尝试。但是,由于缺乏成熟的篇章分析工具,尤其对汉语而言,因此,仍未建立起基于篇章分析理论的机器翻译模型。我们认为,如何基于篇章分析的研究成果,建立基于篇章结构和语义分析的机器翻译模型将是未来几年的一个研究热点。

(4) 面向弱规范或不规范文本的机器翻译方法

当今时代网络无处不在,而语言文本是网络内容的最主要的表现形式。网络语言的口语化、社交化性质导致了网络文本弱规范、甚至不规范的表达特征。如何处理网络语言的非规范现象,为多语言机器翻译研究提出了巨大挑战。[75]针对微博数据提出了一种面向机器翻译的文本规范化方法。这种方法仅对一些网络用语和标点等进行了恢复处理,很多复杂问题并未涉及。另外,是否必须首先对源语言文本进行规范化处理,还是直接基于原始文本进行翻译,哪种处理办法更好,也是一个有待于深入研究的问题。因此,面向弱规范、不规范以及噪音文本的机器翻译研究必将是未来几年的研究热点。

(5) 基于用户反馈的翻译模型自学习研究

对于机器翻译研究者来说,终极目标是希望翻译系统能够像专业译员一样高质高效地完成翻译过程。对人而言,不管是专业译员,还是普通用户,都有从反馈中不断学习、不断完善的能力,而当前的任何一种翻译模型都缺乏这种从反馈中自我学习、自我更新的能力。

(6) 面向资源缺乏的机器翻译方法研究

统计机器翻译是一种数据驱动的方法,双语平行资源是统计机器翻译的基础,但很多语言对(如藏汉、越汉等)缺少大规模双语资源。为了绕开平行语料这一瓶颈,基于枢轴语的机器翻译模型研究和面向单语语料的无监督翻译模型研究显然是一个不错的选择。

(7) 基于群体智慧的翻译资源获取

我们目前处在一个社交网络构成的社会,群体智慧是其中的一个典型特征。如何充分利用群体智慧,廉价且快速地获得翻译资源,也将成为一个典型的研究方向。

(8) 混合机器翻译方法

虽然研究人员对统计机器翻译模型、基于实例的翻译模型和基于规则的翻译模型,及其它它们之间的融合方法进行了深入研究和实践,但是,目前融合方法中各种模型的耦合方式还比较松散,融合方法无法真正凝聚各个模型的优势。因此,多模型的深度融合可能是未来研究的一个热点问题。

(9) 基于深度学习的翻译方法

最近两三年，深度学习在语音识别和图像处理领域取得了突破性进展。深度学习擅长于无监督的特征学习，能够学到高度抽象的语义表示。而目前的机器翻译模型使用的都是人工设计的特征，那么，是否可以基于深度学习技术建立更加灵活、有效的翻译模型，也许会成为研究者们感兴趣的另一个研究方向。

5. 结束语

近两三年来，机器翻译研究基本处于一个平台期，虽没有突破性进展，但翻译水平却在稳步提升。很多新的尝试已初现成效，如基于谓词—论元结构转换的语义翻译模型、基于篇章理解的翻译方法、机器翻译模型的融合方法等。尽管与我们的最终目标的相比，目前的机器译文质量仍然差强人意，若干理论问题一直没有从根本上得到解决，许多方法和技术仍有待于进一步深入研究和探索，但我们从另一方面看到，机器翻译技术已经在若干领域为人们提供了快捷方便的翻译服务，例如，天气预报翻译、产品说明书翻译、专利翻译、口语即时翻译等，即使在无领域限制的面向网络终端客户的网页在线翻译等方面，也提供了一定的便利，而且计算机辅助的人工翻译和译后编辑（post-editing）功能都为人类的翻译工作提供了一定的帮助。

我们欣慰地看到，人们对机器翻译的未来充满了信心。越来越多的研究者投身到了这一领域，越来越多的政府部门和企业开始拥抱机器翻译，并投资研究和开发机器翻译。我们有理由相信，随着机器学习、语义和篇章理解等相关技术的进展，机器翻译模型一定会发生质的变化，机器翻译系统的水平一定会得到大幅度的提升，机器翻译系统的产业应用必将越来越广。

参考文献

- [1] 宗成庆，统计自然语言处理，清华大学出版社，2008.
- [2] M. Nagao, 1984. A framework of a mechanical translation between Japanese and English by analogy principle. In A. Elithorn and R. Banerji. *Artificial and Human Intelligence*.
- [3] 王厚峰. 基于实例的机器翻译—方法和问题. 《术语标准化与信息技术》2003年第2期.
- [4] P.F. Brown, S.A. Della Pietra, V.J. Della Pietra, J.D. Lafferty, R.L. Mercer. 1992. Analysis, statistical transfer, and synthesis in machine translation, In *Proceedings of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*.
- [5] P.F. Brown, S.D. Pietra, V.J.D. Pietra, R.L. Mercer. 1993. The mathematic of statistical machine translation: Parameter estimation. *Computational linguistics*.
- [6] P. Koehn, H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens. 2007. Moses: Open source toolkit for statistical machine translation, In *Proceedings of ACL2007*.
- [7] D. Chiang. 2007. Hierarchical phrase-based translation. *Computational linguistics*, 33(2): 201-228.
- [8] J. Xie, H. Mi, Q. Liu. 2011. A novel dependency-to-string model for statistical machine translation. In *Proceedings of EMNLP-2011*, pp. 216-226.
- [9] L. Shen, J. Xu, R. Weischedel, 2008. A new string-to-dependency machine translation algorithm

- with a target dependency language model. In Proc. of ACL.
- [10] Y. Liu, Q. Liu, S. Lin. 2006. Tree-to-string alignment template for statistical machine translation, In Proceedings of COLING-ACL 2006, pp. 609-616.
- [11] H. Mi, L. Huang, Q. Liu. 2008. Forest based translation. In Proceedings of ACL-08: HLT: 192-199.
- [12] M. Galley, J. Graehl, K. Knight, D. Marcu, S. Deneefe, W. Wang, I. Thayer. 2006. Scalable inference and training of context-rich syntactic translation models, In Proceedings of COLING-ACL 2006, pp. 961-968.
- [13] D. Chiang, K. Knight, W. Wang. 2009. 11,001 new features for statistical machine translation, In Proceedings of HLT-NAACL2009, pp. 218-226.
- [14] J. DeNero and D. Klein, 2007. Tailoring Word Alignments to Syntactic Machine Translation. In Proc. of ACL 2007.
- [15] D. Burkett and D. Klein, 2012. Transforming Trees to Improve Syntactic Convergence. In Proc. of EMNLP 2012.
- [16] S. Liu, C. Li, M. Li and M. Zhou, 2012. Re-training Monolingual Parser Bilingually for Syntactic SMT. In Proc. of EMNLP 2012.
- [17] F. Zhai, J. Zhang, Y. Zhou and C. Zong, 2012. Tree-based Translation without using Parse Trees. In Proc. of COLING 2012.
- [18] F. Zhai, J. Zhang, Y. Zhou and C. Zong, 2013. Unsupervised Tree Induction for Tree-based Translation. In Transaction on ACL.
- [19] D. Chiang. 2010. Learning to translate with source and target syntax, In Proceedings of ACL2010, pp. 1443-1452.
- [20] H. Mi, Q. Liu. 2010. Constituency to dependency translation with forests, In Proceedings of ACL2010, pp. 1433-1442.
- [21] J. Zhang, F. Zhai, C. Zong. 2011. Augmenting string-to-tree translation models with fuzzy use of source-side syntax. In Proceedings of EMNLP 2011, pp. 204-215.
- [22] H. Daumé III and Jagarlamudi, 2011. Domain Adaptation for Machine Translation by Mining Unseen Words. In Proc. of ACL 2011.
- [23] G. Foster, C. Goutte and R. Kuhn, 2010. Discriminative Instance Weighting for Domain Adaptation in Statistical Machine Translation. In Proc. of EMNLP 2010.
- [24] V. Eidelman, J. Boyd-Graber and P. Resnik. 2012. Topic Models for Dynamic Translation Model Adaptation. In Proc. of ACL 2012.
- [25] J. Su, H. Wu, H. Wang, Y. Chen, X. Shi, H. Dong and Q. Liu, 2012. Translation Model Adaptation for Statistical Machine Translation with Monolingual Topic Information. In Proc. of ACL 2012.
- [26] W. Wang, K. Macherey, W. Macherey, Franz Och and Peng Xu. Improved Domain Adaptation for Statistical Machine Translation. In Proc. of AMTA 2012.
- [27] S. Mansour and H. Ney, 2013. Phrase Training Based Adaptation for Statistical Machine Translation. In Proc. of NAACL 2013.
- [28] A. Axelrod, X. He and J. Gao, 2011. Domain Adaptation via Pseudo In-Domain Data Selection. In Proc. of EMNLP 2011.
- [29] R. C. Moore and W. Lewis. 2010. Intelligent selection of language model training data. In ACL 2010.
- [30] S. Lu, W. Wei, X. Fu and B. Xu, 2012. Translation Model Based Cross-Lingual Language Model Adaptation: from Word Models to Phrase Models. In Proc. of EMNLP 2012.

- [31] B. Chen, G. Foster and R. Kuhn, 2013. Adaptation of Reordering Models for Statistical Machine Translation. In Proc. of NAACL 2013.
- [32] D. Wu and P. Fung, 2009. Semantic roles for smt: A hybrid two-pass model. In Proc. of NAACL 2009.
- [33] X. Wu, K. Sudoh, K. Duh, H. Tsukada and M. Nagata, 2011. Extracting preordering rules from predicate-argument structures. In Proc. of IJCNLP 2011.
- [34] D. Liu and D. Gildea, 2010. Semantic role features for machine translation. In Proc. of COLING 2010.
- [35] W. Aziz, M. Rios and L. Specia, 2011. Shallow semantic trees for smt. In Proc. of WMT 2011.
- [36] D. Xiong, M. Zhang, and H. Li, 2012. Modelling the translation of predicate-argument structure for smt. In Proc. of ACL 2012.
- [37] F. Zhai, J. Zhang, Y. Zhou and C. Zong, 2012. Machine Translation by Modeling Predicate-Argument Structure Transformation. In Proc. of COLING 2012.
- [38] F. Zhai, J. Zhang, Y. Zhou and C. Zong, 2013. Handling Ambiguities of Bilingual Predicate-Argument Structures for SMT. In Proc. of ACL 2013.
- [39] B. Jones, J. Andreas, D. Bauer, K. M. Hermann and K. Knight, 2012. Semantics-Based Machine Translation with Hyperedge Replacement Grammars. In Proc. of COLING 2012.
- [40] L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer and N. Schneider, 2013. Abstract Meaning Representation for Sembanking. In Proc. of Linguistic Annotation Workshop (LAW VII-ID), ACL2013.
- [41] Z. Gong, M. Zhang and G. Zhou, 2011. Cache-based Document-level Statistical Machine Translation. In Proc. of EMNLP 2011.
- [42] T. Xiao, J. Zhu, S. Yao and H. Zhang, 2011. Document-level consistency verification in machine translation. In Proc. of MT Summit XIII 2011.
- [43] F. Ture, D. W. Oard and P. Resnik, 2012. Encouraging Consistent Translation Choices. In Proc. of NAACL 2012.
- [44] D. Xiong, G. Ben, M. Zhang, Y. Lv and Q. Liu, 2013. Modeling Lexical Cohesion for Document-Level Machine Translation. In Proc. of IJCAI 2013.
- [45] D. Xiong and M. Zhang, 2013. A Topic-Based Coherence Model for Statistical Machine Translation. In Proc. of AAAI 2013.
- [46] M. Tu, Y. Zhou and C. Zong, 2013. A Novel Translation Framework Based on Rhetorical Structure Theory. In Proc. of ACL 2013.
- [47] S. Ravi and K. Knight, 2011. Deciphering foreign language. In Proc. of ACL 2011.
- [48] M. Nuhn, A. Mauser and H. Ney, 2012. Deciphering Foreign Language by Combining Language Models and Context Vectors. In Proc. of ACL 2012.
- [49] Q. Dou and K. Knight, 2012. Large Scale Decipherment for Out-of-Domain Machine Translation. In Proc. of EMNLP 2012.
- [50] S. Ravi. Scalable decipherment for machine translation via hash sampling. In Proc. of ACL 2013.
- [51] J. Zhang and C. Zong. Learning a Phrase-based Translation Model from Monolingual Data with Application to Domain Adaptation. In Proc. of ACL 2013.
- [52] A. Klementiev, A. Irvine, C. Callison-Burch and D. Yarowsky, 2012. Toward Statistical Machine Translation without Parallel Corpora. In Proc. of EACL 2012.
- [53] A. I. Rosti, N. F. Ayan, B. Xiang, S. Matsoukas, R. Schwartz, B. J. Dorr, 2007. Combining Outputs from Multiple Machine Translation Systems. In Proc. of NAACL 2007.

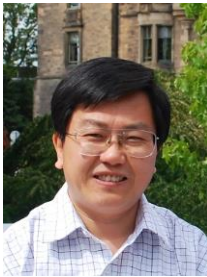
- [54] J. DeNero, S. Kumar, C. Chelba and F. Och, 2010. Model Combination for Machine Translation. In Proc. of NAACL 2010.
- [55] T. Watanabe and E. Sumita, 2011. Machine Translation System Combination by Confusion Forest. In Proc. of ACL 2011.
- [56] Y. Liu, H. Mi, Y. Feng, and Q. Liu, 2009. Joint Decoding with Multiple Translation Models. In Proc. of ACL 2009.
- [57] M. Li, N. Duan, D. Zhang, C. Li and M. Zhou, 2009. Collaborative Decoding: Partial Hypothesis Re-ranking Using Translation Consensus between Decoders. In Proc. of ACL 2009.
- [58] N. Duan, M. Li and M. Zhou, 2011. Hypothesis Mixture Decoding for Statistical Machine Translation. In Proc. of ACL 2011.
- [59] P. Koehn and J. Senellart, 2010. Convergence of Translation Memory and Statistical Machine Translation. In Proc. of AMTA Workshop on MT Research and the Translation Industry, 2010.
- [60] Y. Ma, Y. He, A. Way and J. Genabith, 2011. Consistent Translation using Discriminative Learning - A Translation Memory-inspired Approach. In Proc. of ACL 2011.
- [61] K. Wang, C. Zong and K. Su. Integrating Translation Memory into Phrase-Based Machine Translation. In Proc. of ACL 2013.
- [62] J. Yang, S. Enoue and J. Senellart, 2011. SYSTRAN Chinese-English and English-Chinese Hybrid Machine Translation Systems for CWMT2011. In Proc. of CWMT 2011.
- [63] C. Callison-Burch, P. Koehn, C. Monz, and O. Zaidan, 2011. Findings of the 2011 Workshop on Statistical Machine Translation. In Proc. of WMT 2011.
- [64] K. Papineni, S. Roukos, T. Ward and W. Zhu, 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In Proc. of ACL 2002.
- [65] G. Doddington, 2002. Automatic Evaluation of Machine Translation Quality Using N-gram Co-occurrence Statistics. In Proc. of HLT 2002.
- [66] S. Banerjee and A. Lavie, 2005. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. In Proc. of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization.
- [67] M. Snover, B. Dorr, R. Schwartz, J. Makhoul, L. Micciulla and R. Makhoul, 2006. A Study of Translation Edit Rate with Targeted Human Annotation. In Proc. of AMTA 2006.
- [68] I. D. Melamed, R. Green and J. P. Turian, 2003. Precision and Recall of Machine Translation. In Proc. of NAACL 2003.
- [69] Y. Chan and H. Ng, 2008. MAXSIM: A Maximum Similarity Metric for Machine Translation Evaluation. In Proc. of ACL 2008.
- [70] M. Zhou, B. Wang, S. Liu, M. Li, D. Zhang and T. Zhao, 2008. Diagnostic Evaluation of Machine Translation Systems Using Automatically Constructed Linguistic Check-Points. In Proc. of COLING 2008.
- [71] C. Liu, D. Dahlmeier and H. Ng, 2010. TESLA: Translation Evaluation of Sentences with Linear-Programming-Based Analysis. In Proc. of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR, 2010.
- [72] C. Lo and D. WU, 2011. MEANT: An inexpensive, high-accuracy, semi-automatic metric for evaluating translation utility via semantic frames. In Proc. of ACL 2011.
- [73] T. Meyer, A. Popescu-Belis and N. Hajlaoui, 2012. Machine Translation of Labeled Discourse Connectives. In Proc. of AMTA 2012.
- [74] B.T.M. Wong and C. Kit, 2012. Extending Machine Translation Evaluation Metrics with Lexical

Cohesion to Document Level. In Proc. of EMNLP 2012.

[75] P. Wang and H. Ng, 2013. A Beam-Search Decoder for Normalization of Social Media Text with Application to Machine Translation. In Proc. of NAACL 2013.



张家俊，2011年7月毕业于中国科学院自动化研究所，获博士学位。2011年7月至今在自动化所从事研究和开发工作，现为模式识别国家重点实验室助理研究员。近几年来主要从事机器翻译的研究和开发工作，在本领域权威学术会议上发表论文10余篇，2009年获第23届亚太地区语言、信息与计算(PACLIC)国际学术会议最佳论文奖，2010年中国科学院“朱李月华优秀博士生”奖，2012年获首届自然语言处理与中文计算学术会议(NLP&CC)最佳论文奖。



宗成庆，1998年3月毕业于中国科学院计算技术研究所，获博士学位，现为模式识别国家重点实验室研究员、博士生导师。近年来，他主要从事自然语言处理、机器翻译和文本分类等相关技术的研究和教学工作，作为项目负责人主持国家自然科学基金项目、国家“863”项目、国家支撑计划项目和中科院国际合作项目等10余项，发表论文100余篇，出版专著和译著各一部，获得国家发明专利9项。目前他是国际计算语言学委员会(ICCL)委员、国际计算语言学学会(ACL)汉语特别兴趣组(SIGHAN)候任主席(Chair-Elect)，中国中文信息学会常务理事和机器翻译专业委员会副主任、中国人工智能学会理事、中国计算机学会中文信息技术专委会副主任，担任国际学术期刊 *ACM TALIP* 副主编(Associate Editor)、*IJCPOL* 副主编、*IEEE Intelligent Systems* 编委、*Machine Translation* 编委、*JCST* 编委和《自动化学报》副主编等多种学术职务。享受政府特殊津贴。