

机器翻译技术及应用

王海峰

东芝（中国）研究开发中心 北京 100738

wanghaifeng@rdc.toshiba.com.cn

1 引言

语言是人类进行交流的工具，而语言不通则阻碍着人们之间的交流。于是，人类在克服语言交流障碍的过程中发展了语言之间的翻译，并产生了翻译学。而用机器来帮助甚至替代人工翻译也是人类长久以来的一个梦想。最原始的机器翻译思想可以追溯到 17 世纪关于机械词典的构想。Weaver 1949 年的著名备忘录则标志基于现代计算机的机器翻译正式登上历史舞台。

机器翻译既涉及到人类对自身语言和思维方式的认知，又涉及到人工智能、信息论、知识工程、软件工程等很多技术领域，是一个交叉性很强的学科。机器翻译的发展，既需要基于很多学科的综合发展，同时又有自身的发展规律，形成了独立的学科体系。

本文将在简要回顾机器翻译发展史的基础上，全面论述机器翻译方法，然后介绍牵引机器翻译技术发展的两驾马车：科学评测和实际应用，最后进行总结¹。

2 机器翻译发展回顾

发展至今，机器翻译经历了 1954 年之前的草创期、1954-1966 年的高期望期、1967-1976 年的沉寂期、1976-1989 年的复苏期、以及 1989 年之后的发展期等五个不同的历史时期。

1954 年美国 IBM 公司和乔治敦大学合作开发的机器翻译原型系统第一次向人们展示了一个现实的机器翻译系统，从而使各国从事机器翻译研究的学者们都大受鼓舞，并且吸引了大量政府机构的资金投入。这一代机译系统都是采用直接法的，主要以词典为驱动，辅以较低层次上的语法分析和语义特征的使用。

1966 年的 ALPAC 报告指出，机器翻译速度慢，准确率低，比人工翻译花费多，且不能马上实用，从此，各类机器翻译项目纷纷下马，机器翻译研究走入低潮。

从七十年代中期开始，基于规则的第二代机器翻译的方法逐渐成熟，机器翻译研究重焕生机，第二代方法的主导思想是面向句法、基于规则的转换策略。

八十年代末开始，在国际交流日益增多的环境下，在计算机、互联网等相关技术日新月异的基础上，机器翻译的技术与应用也取得了蓬勃发展。技术方面，理性主义的基于规则的机器翻译方法仍在不断发展完善，经验主义的基于语料库的机器翻译方法也开始出现并迅速发展。应用方面，随着需求分析的不断深入、产品定位的逐渐明确，机器翻译产品也在不断为更多用户所接受，在很多领域真正走向了实用。

我国机器翻译研究始于 1957 年，并于 1959 年完成俄汉机译实验。以后二十多年时间里，机器翻译研发在我国进展缓慢，直至七十年代末开始复苏并逐渐走向繁荣。我国学者冯志伟于七十年代末完成了汉译多种外语的实验。之后，我国的中科院计算所、哈尔滨工业大学、国防科技大学等单位陆续开始了机器翻译的研究，并同样经过了基于规则方法、基于语料库

¹ 无论采用何种方法，机器翻译系统都需要相应的语言资源作支撑，由于本专辑中另有文章专门论述资源问题，本文不再赘述这部分内容。

方法以及多种方法相融合的发展历程。

3 机器翻译方法

广义地讲，机器翻译涉及自然语言处理技术的方方面面，几乎所有自然语言处理方面的研究成果都可以直接或间接的应用于机器翻译。狭义地讲，机器翻译方法通常可以分为三类：基于规则的机器翻译（Rule-Based Machine Translation，简称 RBMT）、基于实例的机器翻译（Example-Based Machine Translation，简称 EBMT）和统计机器翻译（Statistical Machine Translation，简称 SMT），其中后两种方法又可以统称为基于语料库的方法（Corpus-based Machine Translation，简称 CBMT）。

3.1 基于规则的机器翻译

基于规则的机器翻译方法认为翻译的过程是需要对源语言的分析 and 源语言意义的表示，然后再生成等价的目标语言的过程。根据翻译过程的不同，规则方法可分为两种主要方法：基于转换的方法的翻译过程包括三个阶段：分析得到一种源语言的抽象表示；把源语言的抽象表示转换为目标语言的抽象表示；由目标语言的抽象表示生成目标语言。基于中间语言的方法在对源语言分析后产生的是中间语言，而目标语言的生成是直接由这种中间语言开始的。

基于规则的方法从七十年代中期开始到八十年代末在机器翻译界一直占有主导地位。现在，规则方法虽然已经较少在主流学术会议上出现了，但一个有趣的现象是，目前真正为用户所使用的机器翻译产品大多都是基于规则的。同时，规则方法也并不是停滞不前了，而是仍有不断提高的潜力[1]。

使用基于规则的机器翻译方法，很容易开发出一个小型实验系统，但在开发面向真实应用的大型系统时，问题则要复杂得多。在基于规则的机器翻译系统中，随着系统规模不断扩大，规则之间相互影响甚至冲突的可能性也会随之增加，而这种相互影响和冲突将使系统性能的提升变得越来越困难，甚至是步履维艰、停滞不前。因此，一个大型的基于规则的商品化机器翻译系统，不仅要解决机器翻译方法论问题，还要从知识工程和软件工程的角度将系统组织好，其中的规则往往是多层次、细粒度的。规则层次和知识粒度的细化，可以有效控制规则之间的相互影响与冲突，并使规则系统具有良好的可扩展性[1]。

3.2 基于实例的机器翻译

基于实例的机器翻译的本质是“以翻译实例为基础，基于相似原理的机器翻译”，其利用的主要知识源是预处理过的双语语料和翻译词典[2]。基于实例的翻译过程通常包括三步：在翻译实例库中搜索匹配片段；确定相应的译文片段；重新组合译文片段以得到最终翻译。

在基于实例的翻译系统中，翻译实例的存储形式直接影响着实例匹配和译文生成的方法及系统的时间和空间开销，并影响着最终译文质量。翻译实例可以不经过任何加工以本来面目存储，也可以表示成完全结构化的形式，而近来的研究表明，半结构化的翻译实例表示方法在翻译实例的预处理难度、翻译的时空效率以及翻译质量之间找到了一个良好的平衡点[3]。基于实例的方法通常采用编辑距离来计算输入句子和翻译实例之间的相似度，而为了提高匹配效率，可以使用贪心算法等一些快速算法[4]。基于实例的翻译通常有两类译文生成算法：基于语义和基于统计的方法。前者译文片段之间的衔接考虑得不够，因此译文流利度会受到影响。后者一定程度上解决了译文片段的衔接问题，但没有考虑译文片段和输入句子的语义关系，因而影响了译文生成的正确性。而综合利用多种特征的模型（如线性对数模型）则可以得到流利度和正确性都较好的结果[5]。

目前来看,实例方法不如规则方法在产品中应用广泛,也不如统计方法在学术会议上广受重视,但实例方法也依然活跃并富有潜力。在最近两届机器翻译峰会(MT Summit)上,实例方法仍然是与会学者大量讨论的话题,而且MT Summit X还专门组织了一个基于实例机器翻译的Workshop,著名的机器翻译专业国际期刊“Machine Translation”也在2005到2006年连续出了三期专刊研究基于实例的机器翻译。从译文质量的角度,实例方法也毫不逊色甚至还能做得更好[4]。

了解基于实例的翻译技术的同时,也有必要关注一下另一个在原理上与基于实例的机器翻译非常相似的技术——翻译记忆(Translation Memory,简称TM)。二者都以翻译实例为基础,不同之处在于,基于实例的机器翻译以全自动翻译为目标,而翻译记忆思想看重的是计算机对人的辅助,本质上是一种辅助翻译,它从实例库中检索出一个或多个相似的翻译实例,以一种友好的形式提交给用户,从而达到帮助用户翻译的目的。近年来,翻译记忆技术也在越来越多的融合各种自动翻译技术[6]。

3.3 统计机器翻译

统计机器翻译也是基于双语语料库的,但与基于实例的方法在翻译过程中直接使用翻译实例不同,统计方法通过事先的训练过程将双语语料库中隐含的翻译知识抽象成统计模型,而翻译过程通常就是基于这些统计模型的解码过程[7]。

统计机器翻译中使用的统计模型通常包括翻译模型和语言模型。翻译模型是联系两种语言的纽带,描述的是一种语言的串翻译成另一种语言的对应串的可能性,通常表示为一个条件概率[8]。语言模型则只与目标语言相关,反映的是一个串在目标语言中的合理程度,通常表示为目标语言的N-gram模型及其变型。解码算法的性能直接影响翻译过程的效率和翻译质量,通常使用的有栈解码、柱解码以及A*算法等。

相比于语言模型和解码,翻译模型是目前统计机器翻译研究中涉及最多的内容。一般的,翻译模型可以分为基于词的模型[8]、基于短语的模型[9]和基于语法的模型[10,11]三种。目前,基于短语和基于语法的模型性能已经明显优于基于词的模型。无论采用哪种模型,词对齐的研究一直广受重视,研究内容包括统计模型本身[12]、各种资源的综合利用[13]、受限领域的词对齐[14]、语料不足时的词对齐[15]、以及词对齐的评价[16]等等。

目前,在三种主流机器翻译方法中,统计机器翻译方法正受到学术界更多的关注。一方面,在具有广泛影响力的NIST机器翻译评测中,统计机器翻译系统连连胜出;另一方面,在ACL等学术会议上,也是多数机器翻译方面的论文都集中在统计方法上。

3.4 各种方法的融合

虽然统计方法由于其良好的数学模型、无指导的学习能力、良好的鲁棒性等优点而备受重视,但规则方法对语言规律良好的概括和描述能力及实例方法对相似句子相当精确的翻译也无疑是颇有价值的。事实上,多种方法相融合的机器翻译策略正在成为机器翻译发展的重要方向。例如,基于语法的模型越来越多的出现在统计机器翻译的研究中;Groves和Way提出了Statistical EBMT和Example-based SMT[17]的概念,等等。而更多的机器翻译系统则是在一种翻译策略为主体的前提下,在系统的不同模块分别使用了不同的技术,例如[4]就在基于实例的机器翻译框架下使用了基于规则的分析 and 基于统计的生成。

3.5 讨论

基于规则的、基于实例的、以及统计机器翻译方法是目前公认的对机器翻译方法的分类方式。通常认为,统计方法和实例方法都是数据驱动的方法。统计机器翻译方法在训练阶段

使用双语语料获得统计模型，翻译时使用事先训练好的统计模型。基于实例的机器翻译方法本质上是一种基于实例推理的方法，它在翻译阶段直接使用翻译实例。与数据驱动的方法相对，传统的基于规则的机器翻译方法中，通常用人工撰写的规则来描述语言规律。

那么，三种机器翻译方法之间到底是什么关系呢？本质上，这三种方法并不是互斥的，而是着眼点不同，从不同角度强调问题的不同方面。统计是从数学角度强调统计建模能力，实例是从机器学习的角度强调对翻译实例的抽象程度，而规则是从语言现象入手来描述语言的成分构成规律。严谨的统计模型、大单元的翻译实例、多层次的成分构成完全是可以融合共生的。

4 机器翻译评测

用计算机翻译人类语言固然很难，对翻译结果进行恰当的评价也不简单。同样一句话，不同的人翻译可能会产生不同的结果；同样一个翻译结果，不同的人也可能会给出不同的评价。这种客观存在的不确定性为准确公平地评价机器翻译结果带来了困难，而合理的评测体系又是进行科学研究和实验所需要的。因此，机器翻译评测方法也一直广受重视。

4.1 机器翻译评测方法

对机器翻译系统的评测比较复杂，可能包括系统性能的方方面面，如译文质量、翻译速度、空间开销、稳定性、易用性等等。所有这些中，译文质量无疑是机器翻译引擎最核心的指标。这里只讨论译文质量的评价。

根据评测的自动化程度，机器翻译评测可以分为人工评测和自动评测。顾名思义，人工评测由人来对翻译结果进行评价；而自动评测则由计算机自动完成对翻译结果的评价，自动评测往往需要预先给出评价模型及设置参考答案。

根据评测内容，机器翻译评测可以分为语言点评测和全文评测。前者根据事先设定的语言点是否被准确翻译来评价翻译质量，这种评价有些类似于作文考试中，参考语言点来为作文评分的方法。后者则不预先设定评价点，而是对译文进行整体评价。

现在的人工评测通常采用忠实度和流利度两个评价指标。忠实度用来衡量译文忠实于原文的程度，反映的是译文的含义相对于原文是否正确。忠实度又可分为若干等级，例如：“译文准确完整地表达了原文含义”为5分；“完全错误或没译出来”为0分。流利度则用来衡量译文本身是否流利，反映译文是否符合目标语言的表达习惯。流利度也可分为若干等级，例如：“流畅地道的译文”为5分；“完全不通的译文”为0分。

上世纪九十年代初，我国学者俞士汶提出了一种基于测试点的机器翻译自动评价方法[18]，并曾应用于两次863机器翻译评测。现在的机器翻译自动评测，通常衡量的是被评价译文与参考译文之间的总体相似度，而且这种相似度的计算只涉及字符本身，而不需要进行深层的语言理解。这类方法中，影响最大、使用最广的BLEU[19]根据N-gram的比较来进行评价。BLEU的思想虽然很简单，但其评价结果却与人工评价有着良好的相关性。NIST打分的基本原理与BLEU类似，但与BLEU使用几何平均不同，NIST使用了算术平均，同时NIST还对低频N-gram赋予了更高的权值。其它常用的评价指标包括METEOR、Word Error Rate (WER)、Position independent word Error Rate (PER)、Translation Edit Rate (TER)等。最新研究表明，即使没有人工参考译文，使用机器自动生成的伪参考译文也能得到很好的评价结果[20]。

4.2 机器翻译评测举例

NIST机器翻译评测²是国际上影响最大的机器翻译评测。与中文相关的，NIST评测过去主要开展了汉语到英语的机器翻译的评测，2008年的评测也将开展英语到汉语的评测。NIST评测主要针对文本语料的翻译，国际上另有一些口语翻译方面的评测，如IWSLT³和TC-STAR⁴的评测。

国内方面，由国家863计划组织的中文信息处理与智能人机接口评测曾有6次进行了机器翻译评测⁵，973项目曾开展过一次汉英机器翻译的评测，2007年8月在哈尔滨工业大学召开的第三届统计机器翻译研讨会（SSMT2007）也进行了机器翻译评测⁶。

5 机器翻译应用

机器翻译技术有着广阔的应用需求，本章将分类介绍机器翻译的各种应用，而不拟提及具体的公司及产品名称。

按自动化程度，机器翻译可分为自动翻译和辅助翻译。前者由于没有人工参与，难以直接得到很高质量的译文，但可以用于信息的快速浏览和获取。后者通过人机互助，可以得到高质量的译文用于出版、发布等信息传播目的。辅助翻译根据人机分工不同又可分为机助人译和人助机译两种。

根据处理对象不同，机器翻译可以分为文本翻译和语音翻译。前者处理文本输入，而后者处理语音输入。

从使用目的角度，机器翻译可以分为四种用途：传播信息、获取信息、交流信息、和检索信息。一般来说，辅助翻译系统或严格领域受限的高质量自动翻译系统可以生成高质量译文用于信息传播；通用的全自动翻译系统可以生成虽不完美但基本易懂的译文用于信息获取；受限领域的口语翻译系统可以帮助人通过语音或即时通讯系统进行信息交流；跨语言信息检索系统可以帮助人完成不同语言之间的信息检索。

机器翻译系统可以有多种产品形态，如独立的翻译软件包、翻译引擎授权、翻译服务提供、与硬件捆绑等等。

6 结语

现代的机器翻译研究已有半个多世纪的历史，其间产生过令人振奋的成果，也有过令人沮丧的时候，但无论多么艰深，人类对于机器翻译的探索和渴求始终也没有停止过。身兼文理、统合技术与艺术的学科魅力吸引了无数的研究者献身其中；促进跨语言交流的显著作用和巨大的应用价值吸引了大量的机构投资其中。作为世界上使用人数最多的语言，中文相关的机器翻译一直广受重视，而随着其研究的不断深入、开发的不断完善，其性能也必将不断提高，进而带来巨大的社会效益和经济效益。

参考文献

- 1 Jiang Zhu, Haifeng Wang. 2005. The Effect of Adding Rules into the Rule-based MT System. In Proceedings of MT SUMMIT X, pages 298-304. Phuket Island, Thailand.
- 2 M. Nagao. A Framework of a Mechanical Translation between Japanese and English by

² <http://www.nist.gov/speech/tests/mt/>

³ <http://iwslt07.itc.it/>

⁴ <http://www.tc-star.org/>

⁵ <http://www.863data.org.cn>

⁶ http://mitlab.hit.edu.cn/EvaluationGuidelines_Cn.pdf

-
- Analogy Principle, In: A. Elithorn, R. Banerji (eds), *Artificial and Human Intelligence*, pages 173-180. NATO publications.
- 3 王海峰, 刘占一, 吴华. 2007. 半结构化的基于实例的机器翻译. 见: 孙茂松, 陈群秀主编. 内容计算的研究与应用前沿. 清华大学出版社. 1-9.
 - 4 Zhanyi Liu, Haifeng Wang, Hua Wu. 2006. Example-Based Machine Translation Based on Tree-string Correspondence and Statistical Generation. *Machine Translation*, 20(1): 25-41.
 - 5 Zhanyi Liu, Haifeng Wang, Hua Wu. 2007. Log-linear Generation Models for Example-based Machine Translation. In *Proceedings of MT SUMMIT XI*, pages 305-312. Copenhagen, Denmark.
 - 6 Hua Wu, Haifeng Wang, Zhanyi Liu, Kai Tang. 2005. Improving Translation Memory with Word Alignment Information. In *Proceedings of MT SUMMIT X*, pages 313-320. Phuket Island, Thailand.
 - 7 Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, Paul S. Roossin, 1990. A Statistical Approach to Machine Translation. *Computational Linguistics*, 16(2):79-85.
 - 8 Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Robert L. Mercer 1993. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, 19(2):263-311.
 - 9 Philipp Koehn, Franz Josef Och, Daniel Marcu. 2003. Statistical Phrase-Based Translation. In *Proceedings of HLT-NAAC- 2003*, pages 127-133. Edmonton, Canada.
 - 10 Dekai Wu. 1997. Stochastic Inversion Transduction Grammars and Bilingual Parsing of Parallel Corpora. *Computational Linguistics*, 23(3):377-403.
 - 11 David Chiang. 2005. A hierarchical Phrase-based Model for Statistical Machine Translation. In *Proceedings of ACL 2005*, pages 263-270. Ann Arbor, Michigan.
 - 12 Franz Josef Och, Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1):19-51
 - 13 Dengjun Ren, Hua Wu, Haifeng Wang. 2007. Improving Statistical Word Alignment with Various Clues. In *Proceedings of MT SUMMIT XI*, pages 391-397. Copenhagen, Denmark.
 - 14 Hua Wu, Haifeng Wang, Zhanyi Liu. 2005. Alignment Model Adaptation for Domain-Specific Word Alignment. In *Proceedings ACL 2005*, pages 467-474. Ann Arbor, Michigan.
 - 15 Haifeng Wang, Hua Wu, Zhanyi Liu. 2006. Word Alignment for Languages with Scarce Resources Using Bilingual Corpora of Other Language Pairs. In *Proceedings COLING-ACL2006 Main Conference Poster Sessions*, pages 874-881. Sydney, Australia.
 - 16 Necip Fazil Ayan, Bonnie J. Dorr. 2006. Going Beyond AER: An Extensive Analysis of Word Alignments and Their Impact on MT. In *Proceedings of COLING-ACL 2006*, pages 9-16. Sydney, Australia.
 - 17 Declan Groves, Andy Way. 2005. Hybrid Example-Based SMT: the Best of Both Worlds?. In *Proceedings of ACL 2005 Workshop on Building and Using Parallel Texts: Data-Driven Machine Translation and Beyond*, pages 183-190. Ann Arbor, Michigan.
 - 18 Shiwen Yu. 1993. Automatic Evaluation of Output Quality for Machine Translation Systems. *Machine Translation*, 8(1-2): 117-126.
 - 19 Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of ACL 2002*, pages 311-318. Philadelphia, Pennsylvania.
 - 20 Joshua S. Albrecht, Rebecca Hwa. 2007. A Re-examination of Machine Learning Approaches for Sentence-Level MT Evaluation. In *Proceedings of ACL-2007*, pages 880-887. Prague, Czech Republic.