文章编号：1003-0077(2011)06-0063-09

基于句法的统计机器翻译模型与方法

刘 群

（中国科学院 计算技术研究所 中国科学院 智能信息处理重点实验室，北京 100190）

摘 要：该文总结了我们近几年来在基于句法的统计机器翻译方面所做的研究工作，特别是基于源语言句法的一系列统计机器翻译模型与方法，具体包括：基于最大熵括号转录语法的翻译模型，基于源语言短语结构树的树到串翻译模型及其相应的基于树的翻译方法，基于森林的翻译方法和句法分析与解码一体化翻译方法，基于源语言依存树的翻译模型。

关键词：统计机器翻译；基于句法的翻译模型；基于句法的翻译方法

中图分类号：TP391 **文献标识码**：A

Syntax-based Statistical Machine Translation Models and Approaches

LIU Qun

(Key Laboratory of Intelligent Information Processing
Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: This paper gives an overview of our research work on syntax-based statistical machine translation in recent year. Our work especially focused on a series of syntax-based statistical machine translation models and approaches, which includes: the translation model based on maximum entropy bracketing transduction grammar, source phrase structure based tree-to-string translation model and the translation approaches based on it—tree-based approach, forest-based approach, and joint parsing and decoding approach, and the source dependency based translation model.

Key words: statistical machine translation; syntax-based translation model; syntax-based translation approach

1 引言

自从 IBM 在 20 世纪 90 年代初正式提出统计机器翻译方法以来，统计机器翻译的发展经历了一个从沉寂到复苏再到迅猛发展的过程。在 2003 年前后，基于短语的统计机器翻译方法已经成熟，并超越了早前 IBM 提出的基于词的方法。但基于短语的方法缺陷也是很明显的：泛化能力差，短语中的词不能被相近的词替换；调序能力差，短语内调序还可以，但短语间调序和更远距离调序效果都不好。因此，研究者普遍意识到需要针对句子结构进行直接建模，才有可能解决上述问题。为此出现了各种各样的基于句法的统计机器翻译方法。

基于句法的统计机器翻译模型可以分为两大类：第一大类是形式上基于句法的模型。这类模型利用了某种形式的句子结构，但这种结构是通过语料库学习自动获得的，而不是语言学意义上的句法结构；第二大类是语言学上基于句法的模型，这类模型需要利用语言学意义上的句法结构。第二大类又可分为三小类：第一小类是树到串模型，这类模型只在源语言端利用语言学意义上的句法结构；第二小类是串到树模型，这类模型只在目标语言端利用语言学意义上的句法结构；第三小类是树到树模型，要在源语言端和目标语言端同时利用语言学意义上的句法结构。

我们从 2006 年开始，在基于句法的统计机器翻译模型方法方面开展了一系列深入的研究工作，提出了三个不同形式的翻译模型：最大熵括号转录语

收稿日期：2011-09-19 定稿日期：2011-10-10

基金项目：国家自然科学基金课题(60736014)；国家高技术研究发展计划课题(2006AA010108)

作者简介：刘群(1966—)，男，研究员，博士生导师，主要研究方向为自然语言处理与机器翻译。

法模型、树到串模型[①]和依存到串模型。其中，对于树到串模型，根据解码时输入形式的不同，我们又提出了三种不同形式的翻译方法：基于树的翻译方法、基于森林的翻译方法和基于串的翻译方法（句法分析与解码一体化方法），这三种翻译方法由简单到复杂，时空复杂度逐渐增加，性能越来越高。

在基于短语的翻译模型中，翻译知识的基本表示形式是双语短语表，也就是一个源语言短语翻译成一个目标语言短语，如表1所示：

表1 双语短语示例

源语言短语	目标语言短语	翻译概率
举行 了 会谈	hold a meeting	0.7

这种双语短语无法表现句子的内部结构。要引入句子的内部结构，最直观的做法就是引入上下文无关规则，在机器翻译中，就是需要引入同步上下文无关语法，如表2所示：

表2 同步上下文无关语法规则示例

源语言规则	目标语言规则	翻译概率
NP→NR 的 NP （小王的书）	NP → NP of NR (the book of Xiao Wang)	0.7

由于同步上下文无关语法形式较为复杂，而且训练这类模型需要双语句子结构对齐的语料库，在实现上面临很多问题，目前还达不到预期的效果。因此目前研究较多的各种基于句法的统计翻译模型都采用同步上下文无关语法的某种简化形式。

2 基于最大熵括号转录语法的翻译模型（最大熵括号转录语法模型）

Dekai Wu 提出了一种简化的同步上下文无关语法——反向转录语法（Inversion Transduction Grammar，ITG）[33]，这是同步上下文无关语法的二叉化形式，类似于乔姆斯基范式是上下文无关语法的二叉化形式。ITG 只有三种类型的规则，如表3所示：

在 ITG 基础上建立统计翻译模型，还会面临一个问题，就是如何获得各个短语的短语标记。为了回避这个问题，Dekai Wu 对 ITG 进一步简化成了括号转录语法（Bracketing Transduction Grammar，BTG）[25]。BTG 就是在 ITG 中只定义一个标记（通常用 X 表示）。于是，在 BTG 中，上述三类规则简化为表4：

表3 反向转录语法（ITG）的规则类型

ITG 规则		对应源语言规则	对应目标语言规则
短语规则	A → [B C]	A → B C	A → B C
	A → < B C >	A → B C	A → C B
词汇规则	A → x / y	A → x	A → y

表4 括号转录语法（BTG）的规则类型

ITG 规则		对应源语言规则	对应目标语言规则
短语规则	X → [X_1 X_2]	X → X_1 X_2	X → X_1 X_2
	X → < X_1 X_2 >	X → X_1 X_2	X → X_2 X_1
词汇规则	X → x / y	X → x	X → y

实际上，由于只有唯一的短语标记 X，所以这里的两类短语规则就是两条规则，并无其他形式。这两条短语规则分别表示两个小短语在组合成一个大短语的情况下，翻译时是否要交换位置。Dekai Wu 1996 提出了一种用于统计机器翻译的随机括号转录语法模型（SBTG），在这个模型中，BTG 的两条短语规则被赋予不同的概率。可以想象，这个模型的描述能力太弱，很难准确刻画短语的语序调整规律。

针对 SBTG 描述能力弱的确定，我们在 SBTG 的基础上引入最大熵模型[1]，采用最大熵方法来计算两个短语组合翻译时顺序或逆序的概率，我们称该模型为最大熵括号转录语法模型（MEBTG），如图1所示。

在 MEBTG 模型中，任何两个短语组合时，顺序翻译和逆序翻译的概率之和为1，具体这两个概率的计算由一个最大熵模型来决定，这个最大熵模型的特征来自于这两个短语本身和这两个短语出现的上下文。MEBTG 模型的训练也很简单。从词语对齐的双语语料库中即可获得大规模的训练数据，无需任何额外的人工标注。

MEBTG 模型是一种形式上基于句法的模型。该模型一方面在短语模型的基础上引入了二叉化的形式化句法结构，克服了短语模型的固有缺陷，又可以跟短语模型很好地兼容，同时与 SBTG 模型相比又大大增强了模型的区分能力，取得了很好的效果，

① 为了简明：从此以下，“树到串模型”特指“短语结构树到串模型”。

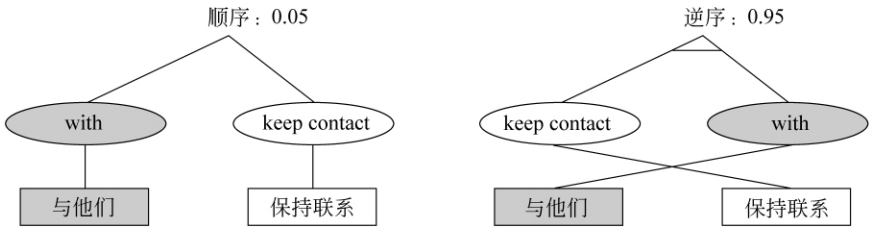


图 1　最大熵括号转录语法模型中规则概率示例

模型性能超过了基于短语的模型。我们基于该模型开发的一个系统，在 NIST2006 评测中作为一个对比系统取得了与第四名主系统完全相同的成绩。

3　基于源语言短语结构树的翻译模型（树到串模型）

树到串模型[2]建立在树到串规则的基础上，树到串规则也可以看作是同步短语结构语法规则的一种变化形式：其源语言端是一棵句法树片段（对应多条上下文无关语法规则），目标语言端是一个词语和变量组成的序列，目标端的每个变量对应于源端句法树片段的某个非终结符叶子节点，如图 2 所示：

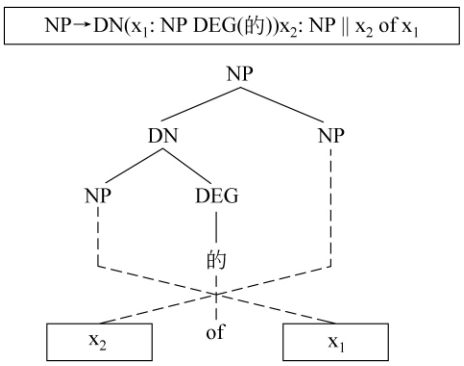


图 2　树到串规则示例

树到串模型建立在源语言端句法分析和双语词语对齐的基础上，要求对双语语料库的源语言进行完整的句法分析，并进行双语句子的词语对齐。在这样处理以后的语料库上，就可以抽取出所有可能的规则，并统计出每条规则出现的次数。模型的定义采用与短语翻译模型类似的方法，通过对数线性模型将多个概率进行对数线性叠加而得到。

树到串模型可以有效地利用源语言端的句法结构信息，在句法分析正确的情况下，对于翻译中的长距离调序问题可以取得较好的效果。

但简单采用树到串模型效果并不太好，一个主要的问题是树到串模型无法兼容非句法短语（不具有完整句法结构的短语）。实验表明，在短语模型中，非句法短语的作用是非常明显的，如果删掉所有非句法短语，短语模型的性能将大大下降。在树到串模型中，所有规则的源语言端都是一棵完整的句法树片段，由此导致无法引入非句法短语，这会导致单纯的树到串模型的实际性能低于短语模型。解决这一问题的办法有多种[2,5]，最简单的办法是[2]把非句法短语用在树到串模型翻译结果的后处理中，替换掉相应源语言短语的译文。更复杂一些的方法[5]是引入树序列到串（tree sequence to string）翻译规则。后面介绍的基于森林的翻译方法也可以一定程度上缓解这一问题。

3.1　基于树的翻译方法

利用树到串模型进行翻译解码，最简单的做法就是首先对源语言句子进行句法分析，然后自底向上对句法树上的每个节点进行规则匹配，记录所有可能的翻译结果，一直到根节点处理结束，就得到整个句子的翻译结果。当然每个节点上都需要进行剪枝。这种翻译方法称为基于树的翻译方法[2]。

基于树的翻译方法简单直观，在不考虑句法分析的情况下，翻译解码速度极快，时间复杂度正比于句子长度。加上源语言句法分析的时间，复杂度是句子长度的三次方。

我们采用基于树的翻译方法，结合非句法短语进行后处理，取得了较好的效果。基于该方法开发的机器翻译系统，在 NIST2006 评测中作为我们的基本系统取得了第五名的成绩。

但这种方法的缺陷也是很明显的，由于句法分析正确率不高，句法分析错误将传递给翻译解码，造成翻译错误。为解决这一问题，我们提出了基于森林的翻译方法。

3.2　基于森林的翻译方法

基于树的翻译方法中，句法分析错误会传递到

翻译解码节点，使得翻译准确率严重下降。为了克服这一问题，直观的做法是在句法分析过程输出多个最优句法分析结果，并利用这多个句法分析结果进行句法分析，这种方法称为 K-best 方法。

但 K-best 方法带来的机器翻译性能提高是很有限的，其原因在于 K-best 结果中存在大量的冗余。简单地理解，一个句子中如果有三处歧义，那么互相组合，就会得到八个不同的句法分析结果。实际上，如果有 n 处歧义，可以组合得到 2n 个。也就是说，即使我们取 1024-best 的句法分析结果，花 1 024 倍的解码时间，也只能多考虑 10 处句法分析歧义。这无疑是非常大的浪费，因为这 1024-best 个句法分析产生的句法树中，绝大部分树片段结构都是重复的。以图 3 为例，这个句子有两个分析结果，两个句法结构树上，阴影部分是不同的，其他部分都完全一样。

为了解决这一问题，我们提出了基于森林的机器翻译方法[3]。其核心思想，就是在统计机器翻译中引入句法压缩森林的结构表示形式。图 3 所示的两个句法分析树就可以表示为图 4 所示的句法压缩森林。

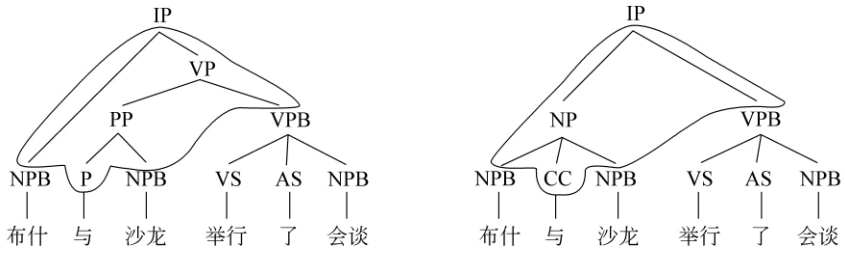


图 3 N-best 句法树的冗余

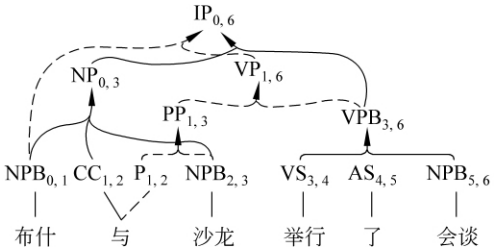


图 4 句法压缩森林示例

句法压缩森林很早就在句法分析中得到应用，Liang Huang[26]提出了基于句法压缩森林的 K-best 句法分析方法。我们的工作[3]最早把句法压缩森林用在了统计机器翻译中。

句法压缩森林的采用，可以在多项式规模的压缩森林中，保留指数级别的 K-best 句法分析结果，使得机器翻译的搜索空间大大增加，很大程度上缓解了句法分析错误带来的影响。

图 5 给出了采用句法压缩森林和采用 K-best 句法分析方法的机器翻译系统性能对比。

可以看到，采用 K-best 方法，当 K 超过 10 以后，机器翻译系统的性能随着平均解码时间的增加上升得非常缓慢，而采用句法压缩森林解码，翻译系统的性能随着平均解码时间的增加迅速提高。两种方法的差别是非常明显的。

我们可以从另外一个角度来看采用句法压缩森林的优势。句法压缩森林是一种非常紧凑的数据结构，句法压缩森林展开后得到的句法树的数量，为句

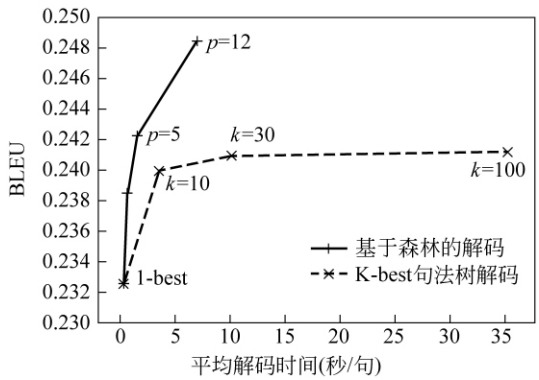


图 5 基于森林的解码和 K-best 句法树解码性能的比较

法压缩森林结点数的指数级别。图 6 说明了在一个基于句法压缩森林的机器翻译系统中，如果把句法压缩森林展开成 K-best 句法树序列，那么实际输出的评分最高的机器翻译译文所对应的句法分析树，在这个 K-best 句法分析树序列中的位置。

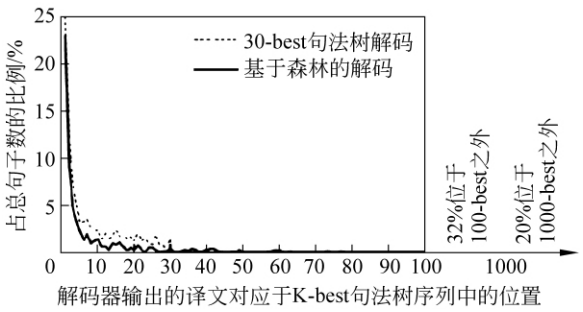


图 6 解码器输出的译文在 K-best 句法树序列中的分布

从图中我们可以看到，32％的机器翻译译文对应的句法分析树位于 100-best 以外，20％的机器翻译译文对应的句法分析树位于 1000-best 以外。

3.3　基于串的翻译方法（句法分析与解码一体化翻译方法）

基于森林的翻译方法使得树到串模型的性能有了大幅度提高。但如果我们仔细分析，还是会发现一个问题。在基于树和基于森林的翻译方法中，句法分析和翻译解码的过程是割裂的。这种割裂不仅是指过程的割裂，而且包括翻译模型的割裂。源语言的句法分析采用的是独立的句法分析器，这个句法分析器所采用句法分析模型（如词汇化概率上下文无关语法模型[27]）通常需要利用一个人工标注的句法树库训练得到。而翻译解码采用的是树到串模型，这个模型是利用双语对齐语料库得到的。人工标注的句法树库通常规模较小，一般只有几万个句子的规模，而双语语料库规模大得多，在汉英机器翻译研究中通常达到数百万句子对的规模。这两个模型的训练语料不仅规模上严重不匹配，实际上所涉及的领域通常也是差别很大的，也就是说，通过句法分析模型得到的高概率句法树，在树到串翻译模型中，概率可能很低，反之亦然。这就会使得我们在句法分析中搜索到的高概率句法树的实际翻译效果并不好，反之，在树到串翻译模型中高概率的翻译结果对应的句法树，在句法分析中因为概率太低又搜索不到。

为了解决这一问题，我们提出了句法分析和解码一体化的翻译方法[4]。其基本思想是，树到串翻译规则的源语言端是一棵句法树片段，可以把这棵句法树片段理解成一条句法分析规则，可以把翻译规则的概率就理解为这条句法分析规则的概率，这样，我们就得到了一部可以用于句法分析的概率语法（严格的说，是概率树替换语法）。在翻译解码过程中，我们并不采用独立的句法分析器，而是直接利用这部概率树替换语法进行句法分析，句法分析的概率就是翻译的概率，这样就避免了句法分析和翻译解码的概率不一致问题，甚至可以在句法分析的同时就完成翻译解码。

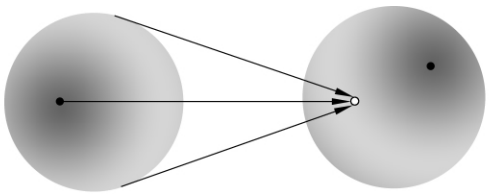
在这种翻译方法中，翻译解码的起点既不是句法分析的 1-best 树，也不是句法压缩森林，而是源语言句子本身，在句法分析的同时完成翻译解码，所以我们又把这种方法称为基于串的翻译方法。

假设我们用两个圆形分别表示句法分析模型和树到串翻译模型的概率分布，其中颜色越深表示搜索空间中该处的概率越大。通常情况下，由于这两个模型来自于不同的训练数据，因此二者概率分布通常并不一致，如图 7 所示：

句法分析模型的概率分布图　　树到串翻译模型的概率分布图

图 7　句法分析模型和树到串翻译模型的概率分布差异

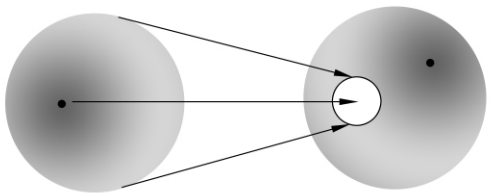
在基于树的翻译方法中，句法分析得到的是句法分析模型中概率最大的点（不考虑搜索误差），而这个点在树到串翻译模型中并不是概率最大的点（图 8）。



句法分析模型的概率分布图　　树到串翻译模型的概率分布图

图 8　基于树的翻译方法中句法分析与翻译解码的搜索空间比较

而在基于森林的方法中，句法分析得到的森林可以覆盖句法分析模型中概率最大的一片区域，因此相对于基于树的翻译方法，可以找到更接近树到串模型中最优位置的点（图 9）。

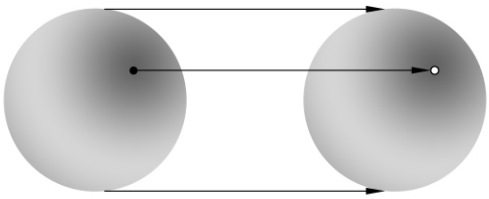


句法分析模型的概率分布图　　树到串翻译模型的概率分布图

图 9　基于森林的翻译方法中句法分析与翻译解码的搜索空间比较

而在句法分析与解码一体化方法中，句法分析模型与树到串翻译模型共享相同的概率空间，句法分析搜索得到的最优点就是翻译模型的最优点（图 10）。

实验表明，采用基于树的方法、基于森林的方法和基于串的方法，翻译系统的性能可以稳步提高，如图 11 所示：



句法分析模型的概率分布图　　树到串翻译模型的概率分布图

图10　基于串的翻译方法中句法分析与翻译解码的搜索空间比较

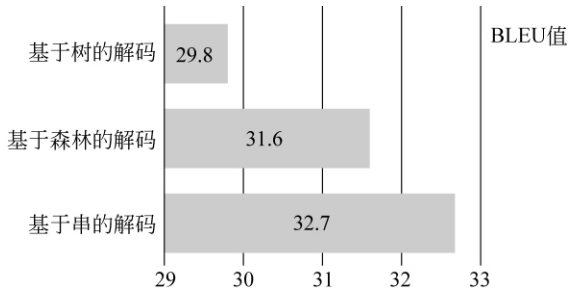


图11　树到串模型的三种解码方法的性能对比

不过，这三种方法的搜索空间也是逐步增大的，因此带来的时空消耗也增长非常明显。特别是句法分析和翻译解码一体化方法，由于得到的规则数量远远大于传统的基于小规模树库训练出来的句法分析器，因此解码搜索空间非常大而且解码时间也会大大延长。

4　基于源语言依存树的翻译模型（依存到串模型）

依存树是一种有效的句法结构表现形式，与短语结构树相比，依存树具有表达简洁、冗余信息少、分析速度快等优点，因此在自然语言处理的很多问题中得到了成功的应用。但在统计机器翻译中，基于依存树的模型一直不是很成功。

在基于短语结构树的翻译模型中，我们所定义的规则的源语言端都要求是完整的句法树片段，也就是所每一个节点或者作为叶节点出现，或者必须带上其所有的子节点出现。这对于基于短语结构树的翻译模型是合适的，但由于依存树的每一个节点都是句子中的一个词，因此这个规定对于依存树来说就太严格了，如图12所示。

在这个例子中，如果根节点“举行”带上其所有子节点作为一条规则，那么这样的规则只能翻译类似“＊世界杯＊在＊成功＊举行＊”这样的句子，可以看到，这样的规则过于具体，泛化能力太差，很难匹配到实际的句子。

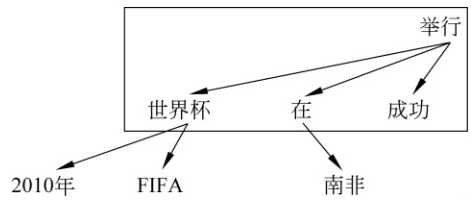


图12　包含所有子节点的依存到串规则抽取示例

我们在2007年提出的模型[28]采用了基于依存树权（treelet）建模的方法，也就是不要求每个非叶子节点都必须带上其子节点，只要是依存树上任何一个联通子图都可以用于建模，如图13所示。

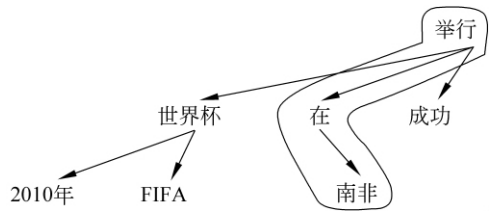


图13　依存树权到串规则抽取示例

这种模型的好处是非常灵活，但带来的问题是规则之间的组合形式会变得非常复杂，不同的规则组合后译文如何排序没有合理的描述方法，另外一条规则的译文内部会形成多个间隔（gap），而这些间隔的填充任意性太强，也无法在模型中进行准确的刻画，这都大大影响了翻译的效果。

我们最近提出的依存到串模型[29]，依据以下两个原则提取规则。

（1）每条规则只有一个层次，也就是说规则没有嵌套结构；

（2）提取规则时根节点的所有子节点都必须出现在规则中，这一点明显区别于基于依存树权的模型。

提取初始规则后，我们还需要对每一条初始规则进行泛化。每条初始规则的节点分为三类：根节点、带结构的叶节点、不带结构的叶节点，对这三类节点可以分别用词性标记进行泛化，一共可以组合出八类泛化规则。

采用前面的例子给出规则提取过程，如图14所示，我们把方框中的树片段结构提取出一条规则。

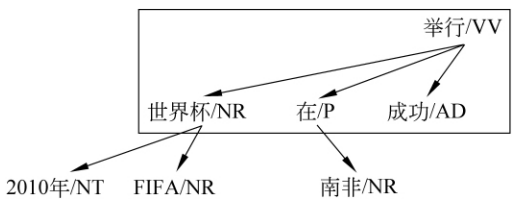


图14　抽取带词性标记的依存到串规则的一个翻译实例

对这条规则中的三类节点进行泛化以后，我们可以得到八条规则，如图 15 所示。

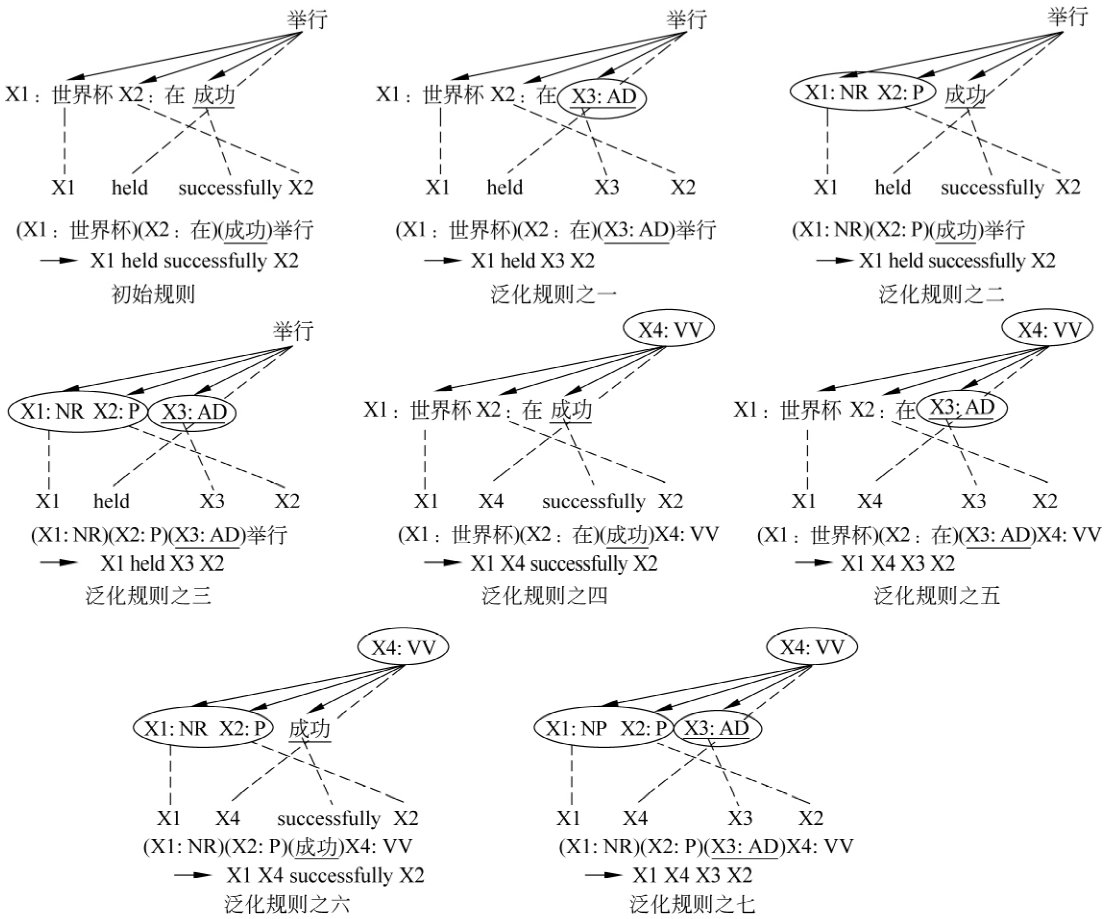


图 15　依存到串规则的抽取和泛化示例

注图中带下划线的节点表示该节点在匹配时不能被扩展

实验表明，我们采用这种依存到串翻译模型实现的机器翻译系统效果非常好，性能超过了层次短语模型[29]。

5　研究工作影响及相关工作

在最大熵括号转录语法模型方面，熊德意博士在本研究组毕业后，在新加坡 I2R 研究所后继续在该模型基础上做了一系列改进工作[19-22]。ACM Computing Surveys 上一篇机器翻译综述也引用了这项工作[23]。

树到串翻译模型我们最早发表在 COLING-ACL2006 上[2]，该论文获得了大会颁发的 Meritorious Asian NLP Paper Award。国际上与我们这项工作相关的研究有：宾州大学 Liang Huang 比我们稍晚提出了与我们非常类似的模型[8]；新加坡 I2R 研究所 Jun Sun 等人将树到串模型扩展为不连续的树序列对齐形式[12]；新加坡 I2R 研究所 Hui Zhang 等人对树到串规则匹配的索引结构进行了改进[13]；微软亚洲研究院 Dongdong Zhang 改进了我们的调序方法[14]；IBM 的 Bing Zhao 在我们的工作基础上提出放宽树到串规则匹配的约束条件来改进翻译的效果[15]；南加州大学 Liang Huang 和本研究组 Haitao Mi 在 EMNLP2010 提出了一种高效的树到串增量式解码算法[7]；东京大学 XianchaoWu 等人在 ACL2010 上发表论文[30]，在一种更复杂的语法形式 HPSG 下实现了更高精度的树到串规则提取，等等。

我们提出的基于森林的翻译方法[3]是首次将句法压缩森林应用在统计机器翻译中，这一做法后来在统计机器翻译中被普遍采用。与之相关的工作有：新加坡 I2R 研究所 Hui Zhang 等人对我们的工作进行了扩展，将句法森林应用到树序列到串的翻译模型[11]；本研究组 Yang Liu 等人将句法森林用于树到树模型[6]；约翰霍普金斯大学（JHU）Zhifei Li 等人为翻译森林结构及其上面定义的运算提供

了更加理论化的形式描述[16]；南加州大学信息科学研究所 USC-ISI 的 John DeNero 等人将翻译森林应用于系统融合[17]；Google 的 Kumar 等人将词格和翻译森林应用于最小错误率训练和最小贝叶斯风险解码[18]。其中，前两项工作是对我们的工作的直接扩展，后三项工作主要是建立在目标语言端的翻译森林基础上，与我们在源语言端句法森林的工作略有不同，但采用森林作为机器翻译中的数据表示形式，也是受到了我们的工作的启发。用句法森林取代单一句法树作为机器翻译的中间数据表示形式已经成为研究界经常采用的做法。

在统计机器翻译中较早引入依存树的是微软研究院 Quirk 等人的工作[31]，由于他们的工作非常复杂，其他研究者很难跟进。我们在 2007 年提出了另一个简单的基于依存树权(treelet)的模型[28]，这个模型的性能也不是很理想。近年来统计机器翻译领域利用依存树信息比较有影响的工作是 Libin Shen 的工作[32]，他是在层次短语模型的基础上加上了目标段的依存树信息。虽然这个工作比较成功，但它不是一个单纯的模型，而是在层次短语模型上的一个扩展，并且是在一个依存语言模型的辅助下才能取得较好的效果。另一方面，这个模型利用的是目标端的依存树信息，而不是源端的依存树信息。这与我们的工作也有较大差别。

6 总结和未来工作

本文介绍了近年来我们提出的多个基于句法的统计机器翻译模型以及相关的多种机器翻译方法。其中一些工作，如最大熵括号转录语法模型、树到串模型、基于森林的机器翻译方法等，都已经产生了较大影响。依存到串模型是我们最近提出的新模型，该模型简洁而且效果很好，我们认为还有较大潜力，有可能成为通向基于语义的翻译模型的一个桥梁。

另外在统计机器翻译中起重要作用的一个模型是语言模型。目前在统计机器翻译中采用的主流语言模型还是 N-gram 模型。虽然 N-gram 模型简单并且效果不错，但毕竟该模型无法考虑句子的结构信息，无法评价一个句子是否是一个语言学上合法的句子，其缺陷是非常明显的。目前在这方面已经有了一些工作，但都还没有普遍采用。我们期待在这方面能有进一步的进展。

致谢

本文是对本研究组多年来部分工作的一个综述[1-6]，所有这些工作都是在这些文章上共同署名的我的同事、学生和我共同完成的，其中做出主要贡献的几个人包括：刘洋、熊德意、米海涛、黄亮、谢军、吕雅娟。虽然他们在这篇综述文章上不作为共同作者署名，但我必须向他们表示衷心的感谢。

参考文献

[1] Deyi Xiong, Qun Liu, Shouxun Lin. Maximum Entropy Based Phrase Reordering Model for Statistical Machine Translation[C]//Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL (COLING-ACL2006), Sydney, Australia, July 17-21. 2006: 521-528.

[2] Yang Liu, Qun Liu, Shouxun Lin. Tree-to-String Alignment Template for Statistical Machine Translation [C]//Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL (COLING-ACL2006), Sydney, Australia, July 17-21. 2006: 609-616.

[3] Haitao Mi, Liang Huang, Qun Liu. Forest-Based Translation[C]//Proceedings of ACL-08: HLT, Columbus, Ohio, USA. 2008: 192-199.

[4] Yang Liu, Qun Liu. Joint Parsing and Translation [C]//Proceedings of COLING 2010, Beijing, August, 2010 .

[5] Yang Liu, Yun Huang, Qun Liu, et al. Forest-to-String Statistical Translation Rules[C]//ACL2007, Prague, Czech, June 2007 .

[6] Yang Liu, Yajuan Lü, Qun Liu. Improving Tree-to-Tree Translation with Packed Forests[C]//Proceedings of ACL-IJCNLP 2009. Singapore, August. 2009: 558-566.

[7] Liang Huang, Haitao Mi. Efficient Incremental Decoding for Tree-to-String Translation[C]//Proceedings of EMNLP 2010, Boston, USA, October.

[8] Liang Huang, Kevin Knight, Aravind Joshi. Statistical Syntax-Directed Translation with Extended Domain of Locality[C]//Proceedings of AMTA, 2006.

[9] Min Zhang, Hongfei Jiang, Aiti Aw, et al. A Tree Sequence Alignment-based Tree-to-Tree Translation Model[C]//ACL08: HLT.

[10] Min Zhang, Hongfei Jiang, Ai Ti Aw, et al. 2007. A Tree-to-Tree Alignment-based Model for Statistical

Machine Translation[C]//MT-Summit-07. 535-542.

[11] Hui Zhang, Min Zhang, Haizhou Li, et al. Forest-based Tree Sequence to String Translation Model [C]//The 47th Annual Meeting of Association for Computational Linguistics and the 4th International Joint Conference of Natural Language Processing (full paper, ACL-IJCNLP 2009), August 2-7 2009, Singapore.

[12] Jun Sun, Min Zhang, Chew Lim Tan. A non-contiguous Tree Sequence Alignment-based Model for Statistical Machine Translation [C]//Proceedings of ACL-IJCNLP 2009.

[13] Hui Zhang, Min Zhang, Haizhou Li, et al. Fast Translation Rule Matching for Syntax-based Statistical Machine Translation[C]//EMNLP 2009, August 6-7 2009, Singapore.

[14] Dongdong Zhang, Mu Li, Chi-Ho Li, et al. Phrase reordering model integrating syntactic knowledge for SMT[C]//EMNLP-CoNLL-2007: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Prague, Czech Republic; 2007: 533-540.

[15] Bing Zhao, Yaser Al-Onaizan, Generalizing local and non-local word-reordering patterns for syntax-based machine translation[C]//EMNLP 2008: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, 25-27 October 2008, Honolulu, Hawaii, USA 2008: 572-581.

[16] Zhifei Li, Jason Eisner, First- and Second-Order Expectation Semirings with Applications to Minimum-Risk Training on Translation Forests [C]//EMNLP2009.

[17] John DeNero, David Chiang, Kevin Knight. Fast Consensus Decoding over Translation Forests[C]//ACL-IJCNLP 2009.

[18] Shankar Kumar, Wolfgang Macherey, Chris Dyer, et al. Efficient Minimum Error Rate Training and Minimum Bayes-Risk Decoding for Translation Hypergraphs and Lattices[C]//ACL-IJCNLP2009.

[19] Deyi Xiong, Min Zhang, Aiti Aw, et al. Refinements in BTG-based Statistical Machine Translation[C]//IJCNLP2008.

[20] Deyi Xiong, Min Zhang, Ai Ti Aw. Linguistically Annotated BTG for Statistical Machine Translation [C]//Proceedings of COLING 2008.

[21] Deyi Xiong, Min Zhang, Ai Ti Aw, et al. A Linguistically Annotated Reordering Model for BTG-based Statistical Machine Translation[C]//Proceedings of ACL 2008.

[22] Deyi Xiong, Min Zhang, Aiti Aw et al. A Syntax-Driven Bracketing Model for Phrase-Based Translation[C]//Preceedings of ACL-IJCNLP 2009.

[23] Adam Lopez. Statistical machine translation [J]. ACM Computing Surveys, 2008, 40(3).

[24] Xianchao Wu, Takuya Matsuzaki, Jun'ichi Tsujii. Fine-grained Tree-to-String Translation Rule Extraction[C]//Proceedings of ACL 2010. Uppsala, Sweden, 2010: 325-334.

[25] Dekai Wu. A polynomial-time algorithm for statistical machine translation [C]//ACL-96: 34th Annual Meeting of the Assoc. for Computational Linguistics. Santa Cruz, CA: Jun. 1996.

[26] Liang Huang, David Chiang. Better k-best Parsing [C]//Proceedings of the 9th International Workshop on Parsing Technologies, Vancouver, B. C, October 2005.

[27] Michael John Collins, Mitchell P. Marcus. Head-driven statistical models for natural language parsing [J]. Journal of Computational Linguistics, 2003, 29 (4).

[28] Deyi Xiong, Qun Liu, Shouxun Lin. A Dependency Treelet String Correspondence Model for Statistical Machine Translation[C]//Second Workshop on Statistical Machine Translation, Prague, Czech, June 2007.

[29] Jun Xie, Haitao Mi, Qun Liu. A novel dependency-to-string model for statistical machine translation [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing, July 27-31, 2011, Edinburgh, Scotland, UK.

[30] Wu, Xianchao and Matsuzaki, Takuya and Tsujii, Jun'ichi. Fine-Grained Tree-to-String Translation Rule Extraction[C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, 2010: 325-334.

[31] Chris Quirk, Arul Menezes, Colin Cherry. Dependency treelet translation: syntactically informed phrasal SMT[C]//Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, Stroudsburg, PA, USA.

[32] Libin Shen, Jinxi Xu, Ralph M. Weischedel: String-to-Dependency Statistical Machine Translation [J]. Computational Linguistics, 36(4): 649-671.

[33] Dekai Wu. Grammarless Extraction of Phrasal Translation Examples from Parallel Texts [C]//TMI-95, Sixth International Conference on Theoretical and Methodological Issues in Machine Translation, v2. Leuven, Belgium, 1995: 354-372.