利用依存限制抽取长距离调序规则

涂兆鹏　刘群　林守勋
中国科学院计算技术研究所 北京 100190
Email: {tuzhaopeng, liuqun, sxlin}@ict.ac.cn

摘　要： 长距离调序仍然是大多数统计机器翻译系统的一个重要问题。层次短语模型提供了一个很好的解决方案，它使用层次短语规则可以很好地表示局部调序和长距离调序。但是，使用传统的算法抽取长距离层次规则将会导致规则表数量急剧增加，从而加大解码内存和时间消耗。为了解决这个问题，我们提出了一种利用依存限制抽取长距离调序规则的新方法。我们的实验表明，我们的方法可以比基准系统高出 0.74 个 BLEU 点。

关键字： 统计机器翻译，层次短语模型，长距离调序，快速匹配

Extract Long Distance Reordering Rules with Dependency Restriction

Zhaopeng Tu, Qun Liu and Shouxun Lin
Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190
Email: {tuzhaopeng, liuqun, sxlin}@ict.ac.cn

Abstract: Long distance reordering is still a key problem for most statistical machine translation (SMT) systems. Hierarchical phrase-based model offers an alternative to address this problem by using hierarchical rules that could characterize both local and long distance reordering. However, extracting long distance reordering rules with traditional algorithm will make decoder time-and-memory consuming. We propose a new algorithm to extract long distance reordering rules with an extra dependency restriction. Our experiments show that our method achieves 0.74 point improvement in BLEU score.

Keywords: statistical machine translation, hierarchical phrase-based model, long distance reordering, quick match

1 前言

过去十年，我们见证了机器翻译领域的快速发展。基于短语的系统[Och and Ney, 2004; Koehn et al., 2003]通过使用短语翻译替代字翻译来提高翻译质量。基于句法的系统(比如[Liu et al., 2006; Galley et al., 2006; Huang et al., 2006; Chiang, 2007; Shen et al., 2008; Mi et al., 2008])通过加入句法信息进一步提高翻译质量。在这些系统中，层次短语模型[Chiang, 2007]非常吸引人，因为它使用上下文无关语法规则来综合基于短语模型和基于句法模型的优势。Chiang[2007]表明使用层次短语模型可以比当前最好的短语模型高出 1 到 3 个 BLEU 点的提高。

层次短语规则能表示局部调序和长距离调序。由于层次规则是从初始规则中泛化而来的，如果要抽取隐含长距离调序信息的规则，则必须先抽取长跨度的初始短语。这将会生成巨大的规则表，从而导致极大的解码系统内存和时间消耗。为了避免这个问题，Chiang[2007]限制了初始短语的最大跨度的阈值。但是，这样会削弱模型的长距离调序能力，因为规则无法表示跨度大于阈

值的长距离调序。此外，层次短语模型使用 Cocke-Younger-Kasami (CKY) 算法以从小跨度到大距离来生成推导。当解码时跨度超过阈值时，不包含任何调序信息的粘贴规则将是唯一选择。

依存树能在一定程序上反映调序信息。Quirk et al. [2005]在源端使用依存树以训练一个调序模型。Shen et al. [2008]引入了一个依存语言模型以刻画目标端依存结构中的长距离词之间的关系。Ding and Palmer[2005]使用依存树上定义的概率同步依存插入语法。

受上述工作的启发，我们提出了一个基本但有效的方法以在层次短语模型上抽取长距离调序规则。首先，我们对训练语料的源端进行依存分析。然后，我们抽取源端为一棵完整依存子树或几棵完整依存子树集合的长距离调序规则。实验表明，我们的方法可以得到 0.74 个 BLEU 点的提高，并且规则表数量增加不大。

2 层次短语模型

层次短语模型 [Chiang, 2007]是基于上下文无关语法的。正式地，层次短语模型的规则可以定义如下：

$$X \rightarrow \langle \gamma, \alpha, \sim \rangle$$

其中，X 是非终结符，γ 和 α 是源端和目标端的字符串（由终结符和非终结符组成），~表示 γ 和 α 之间非终结符间的对齐。

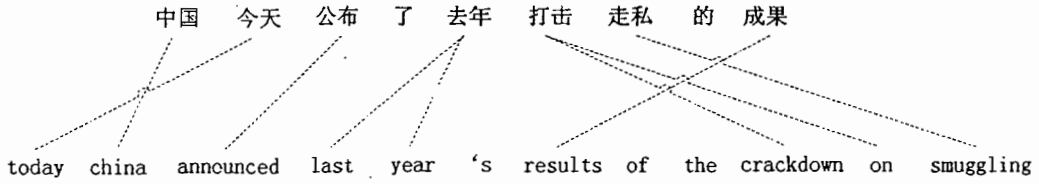

图 1. 一个中文句子，它的英文翻译，和它们之间的对齐。

层次短语模型的规则抽取可以分为两步。首先，抽取满足对齐一致性 [Och and Ney, 2004] 的初始短语；然后，将初始短语中的子短语替换为非终结符得到层次短语。比如对于图 1 中所示的对齐句对，我们可以首先抽取一个满足对齐一致性的初始短语：

$$\text{中国 今天 公布} \rightarrow \text{today china announced}$$

然后我们可以通过将子初始短语

$$\text{公布} \rightarrow \text{announced}$$

替换为非终结符得到一条包含一个非终结符的规则：

$$\text{中国 今天 } X_1 \rightarrow \text{today china } X_1$$

这里 X 表示非终结符，下标表示源端和目标端中非终结符的联系。

另外，层次短语包含了两条粘合规则：

$$S \rightarrow \langle S_1 X_2, S_1 X_2 \rangle$$
$$S \rightarrow \langle X_1, X_1 \rangle$$

粘合规则是用来将一系列部分翻译顺序拼接起来。

3 长距离调序规则

3.1 定义

层次短语模型可以很好地表达局部调序和长距离调序。但使用传统的规则抽取方法抽取长距离调序规则将会生成极大的规则表，从而影响翻译速度及所占内存。我们认为一个可能的原因是对于长距离调序规则来说，对齐一致性的约束较弱。对于覆盖较多词的跨度，里面会包含很多满足对齐一致性的子短语，从而生成指数级的长距离调序规则。

一个解决方法是在抽取长距离调序规则时，对于子短语加入更强的限制，以减少满足条件的子短语，从而减少抽取的长距离调序规则。为了解决这一问题，我们在抽取长距离调序规则时加入依存限制，以抽取数量可以接受的高质量长距离调序规则。

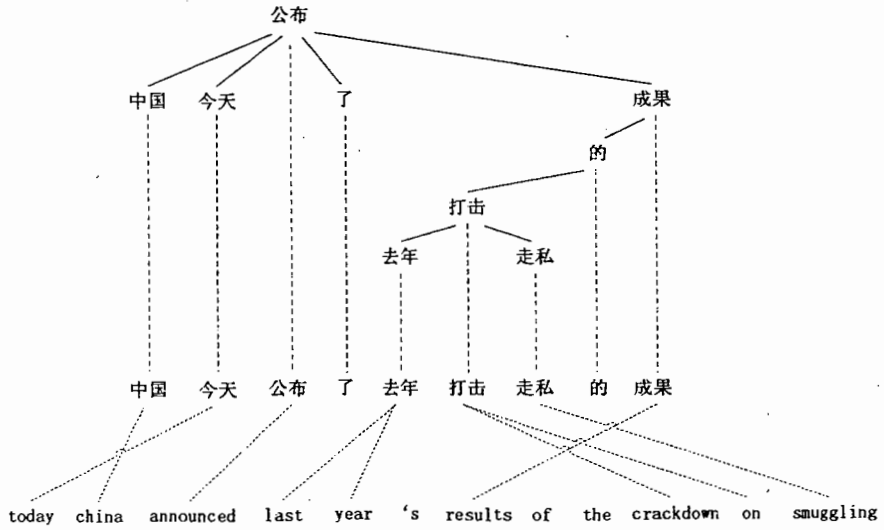

图 2. 一个中文依存树，它的英文翻译，和它们之间的对齐。为了更清楚地表示中英文之间的联系，我们同样给出了中文句子。

图 2 显示了一个中文句子 “中国 今天 公布 了 去年 打击 走私 的 成果” 的依存树。箭头由孩子节点指向它的父节点，或称为头节点。比如在图 2 中，“公布” 是 “中国” 的父节点或头节点。依存树可以反映词语间，尤其是较长距离的词语间的关系[Quirk, 2005; Shen, 2008; Ding and Palmer, 2005]。比如图 2 中，“成果” 直接依存于 “公布” 。此外，我们观察到同时满足对齐一致性和依存结构完整性的初始短语是一个非常好的整体。比如从图 2 抽取的初始短语 (去年 打击 走私 的 成果， last year 's of the crackdown on smuggling)。

为此，我们限定长距离调序规则的源端必须是完整的依存结构。完整的依存结构是指一棵或多棵完整依存子树的集合。严格定义如下[Shen et al., 2008]：

定义 1：对于一个句子 $S = w_1 w_2 ... w_n$ ， $d_1 d_2 ... d_n$ 表示每个词的头节点（父节点），对于根节点

w_i，我们定义 $d_i = 0$。一个依存结构 $d_i ... d_j$ 是带头节点集合 H 的完整依存结构，当且仅当

$$\bullet \exists h \notin [i, j], s.t. \forall k \in H, d_k = h$$
$$\bullet \forall k \in [i, j] \text{ and } k \notin H, d_k \in [i, j]$$
$$\bullet \forall k \notin [i, j], d_k \notin [i, j]$$

图 3 给出了两个完整依存结构的例子，(a)和(b)的头节点集合分别是（中国，今天)和(成果)。我们可以发现(a)和(b)同样满足对齐一致性。

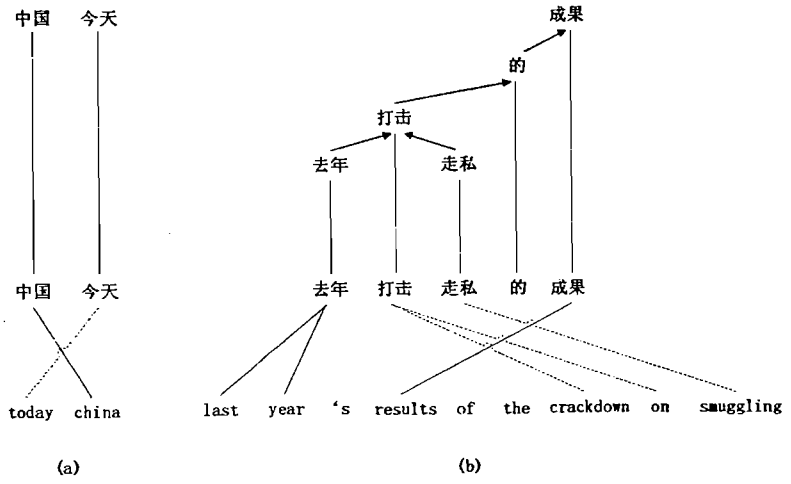

图 3. 完整依存结构的示例。(a)和(b)的头节点集合分别是（中国，今天)和(成果)。

假设层次短语模型传统算法中初始短语的最大跨度为 7（论文中为 10，这里为叙述方便作此假设)，则对于跨度为 9 的源端“中国 去年 公布 了 去年 打击 走私 的 成果”，传统抽取算法无法处理。而我们可以通过将同时满足对齐一致性和完整依存结构限制的图 3 中(a)和(b)结构泛化成非终结符得到长距离调序规则（X_1 公布 了 X_2，X_1 announced X_2)。

由于长距离调序规则覆盖的词语较多，我们可以抽取包含多个终结符的规则。我们使用 LDDR_n 表示包含 n 个非终结符的长距离调序规则。

3.2 规则快速匹配

层次短语模型使用自底向上的 CKY 算法来生成推导。对于一个长度为 l 的跨度，传统的规则匹配算法是枚举出所有可能的候选规则，然后在规则表中查找。假设每条规则最多含有 m 个非终结符，则将会有 $O(l^{2m})$ 个候选规则。对于 l>10 的跨度，枚举所有候选规则是非常耗时的。

受 Lopez[2007]工作的启发，我们使用前缀树结构存储规则，并构建词图表示候选规则 ()。如图 4 所示，对于输入 $a\ b\ c\ d$，所有的候选规则只能以 a 或变量 X 起始。我们首先查找所有以 a 起始的候选规则，在规则表中我们找到了以 a 开始的规则；起始为 a 的候选规则后面只能接 b 或变量 X，然后我们在规则表中发现以 a 起始的规则后面只有接 b 的规则，所以所有 $a\ X$ 起始的候选规则均不存在于规则表中。

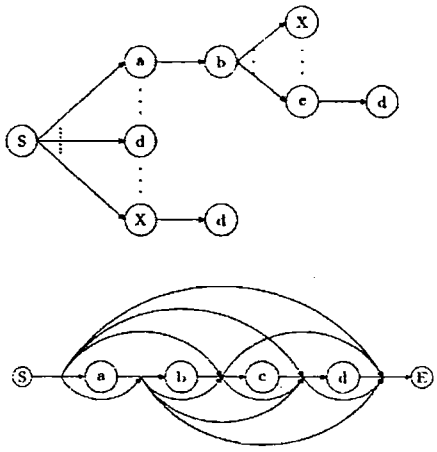

图 4. 前缀树规则表和词组候选规则。每条曲线箭头表示一个变量。

4 实验

4.1 数据准备

我们使用 FBIS 语料 (240K 句对)作为训练语料，并使用移进-归约的依存分析器 [Huang et al., 2009]对源端进行依存分析。为了得到更好的依存分析结果，我们过滤源句子超过 40 的句对，则剩下的句对数为 190K。我们在训练数据上运行 GIZA++ [Och and Ney, 2000]以生成对齐句对。我们使用 SRI 工具[Stolcke, 2002]在新华语料的 GIGAWORD 部分训练一个 4 元的语言模型，我们采用改进的 Kneser-Ney 平滑方法[Kneser and Ney, 1995]的。

所有的实验均是在汉一英测试集上执行的。我们用最小错误率训练 (MERT, [Och and Ney, 2002])方法在 NIST 2002 数据集上调参，并在 NIST 2005 数据集上测试。我们使用大小写不敏感的 BLEU 标准 [Papineni et al., 2002]测试翻译质量。

我们使用修改的层次短语模型来完成翻译。我们在层次短语模型上加入了一个新的特征——长距离调序规则计数，以将之和普通规则区分开。当跨度小于 10 时，我们使用传统抽取算法抽取规则；当大于 10 时，我们使用章节 3.1 所定义的方法抽取长距离调序规则。

4.2 结果

	规则表	比率	BLEU 值
baseline	1.7M		30.11
LDDR_2	1.7M+190K	9.3%	30.68*
LDDR_3	1.7M+230K	10.2%	30.85**

表 1. 规则表大小和 BLEU 值。比率表示 1-best 结果中长距离调序规则所占的比率。LDDR_*n* 表示使用最多含有 *n* 个非终结符的长距离调序规则。*和**分别表示 $p<0.05$ 和 $p<0.01$ 情况下的显著性提高。

表 1 列出了规则表大小和 BLEU 值。我们可以发现新增的长距离调序规则的数量是可以接受的（<10%）。当长距离调序规则所含的最大非终结符数目增加时，规则数量增加并不明显。一个可能的原因是仅有较少的初始短语同时满足对齐一致性和完整依存结构两个限制。我们发现使用长距离调序规则可以得到 0.74 个 BLEU 点的提高。

方法	枚举规则 / 构建词图	规则匹配	总共
传统匹配方法	1.76	0.40	2.16
快速匹配方法	0.05	0.15	0.20

表 2. 不同规则匹配方法的平均时间（秒/句）。

NIST05 测试集包含 1082 个句子，平均长度为 28 个单词。规则表包含 1.7M 的普通规则和 190K 的长距离调序规则。表 2 显示了不同规则匹配方法消耗的时间。我们发现传统规则匹配方法的大部分时间花在枚举规则上。由于使用了长距离调序规则，传统方法需要枚举整个句子所有的候选规则，所以候选规则数量极其多。这也导致规则匹配所需时间稍长。而当我们使用快速匹配方法时，基本上不用花费时间构造词图，而规则匹配的时间也仅需要 0.15 秒/句，较之传统方法极大的减少了时间。这是由于我们在快速匹配时采用动态规则的方法，匹配过程舍弃了大部分不可能存在于规则表的候选规则。

5 结语

我们提出了一个基本但有效的方法以抽取长距离调序规则。相应地，我们设计了新的规则匹配算法以快速匹配长距离调序规则。实验表明使用长距离调序可以在生成较少数量长距离调序规则的情况下，得到 0.74 个 BLEU 点的提高。

尽管如此，我们的方法仍然依赖于词语对齐和依存分析。将来我们会设计新的算法以减轻对词语对齐和依存分析的依赖，比如，使用对齐矩阵[Liu et al., 2009]和依存森林。

参 考 文 献

[1] Adam Lopez. 2007. Hierarchical phrase-based translation with suffix arrays. In Proceedings of EMNLP.

[2] Andreas Stolcke. 2002. Srilm - an extensible language modeling toolkit. In Proceedings of ICSLP.

[3] David Chiang. 2007. Hierarchical phrase-based translation. Computational Linguistics, pages 201–228.

[4] Franz J. Och and Hermann Ney. 2000. Improved statistical alignment models. In Proceedings of ACL.

[5] Franz J. Och and Hermann Ney. 2002. Discriminative training and maximum entropy models for statistical machine translation. In Proceedings of ACL.

[6] Franz J. Och and Hermann Ney. 2004. The alignment template approach to statistical machine translation. Computational Linguistics, 30(4):417–449.

[7] Haitao Mi, Liang Huang, and Qun Liu. 2008. Forest-based translation. In Proceedings of ACL.

[8] Kishore Papineni, Salim Roukos, Todd Ward, and WeiJing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In Proceedings of ACL.

[9] Liang Huang, Kevin Knight, and Aravind Joshi. 2006. Statistical syntax-directed translation with extended domain of

locality. In Proceedings of AMTA.

[10] Liang Huang, Wenbin Jiang, and Qun Liu. 2009. Bilingually-constrained (monolingual) shift-reduce parsing. In Proceedings of EMNLP.

[11] Libin Shen, Jinxi Xuand, and Ralph Weischedel. 2008. A new string-to-dependency machine translation algorithm with a target dependency language model. In Proceedings of ACL.

[12] Michel Galley, Jonathan Graehl, Kevin Knight, Daniel Marcu, Steve DeNeefe, Wei Wang, and Ignacio Thayer. 2006. Scalable inference and training of context-rich syntactic translation models. In Proceedings of ACL.

[13] Philipp Koehn, Franz Joseph Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In Proceedings of NAACL.

[14] R. Kneser and H. Ney. 1995. Improved backing-off for m-gram language modeling. In Proceedings of Acoustics, Speech, and Signal.

[15] Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-string alignment template for statistical machine translation. In Proceedings of ACL.

[16] Yang Liu, Tian Xia, Xinyan Xiao, and Qun Liu. 2009. Weighted alignment matrices for statistical machine translation. In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pages 1017–1026, Singapore, August. Association for Computational Linguistics.