

统计机器翻译删词问题研究*

李强¹, 何燕龙², 栾爽³, 肖桐^{1,4}, 朱靖波^{1,4}

(1. 东北大学自然语言处理实验室, 辽宁 沈阳 110819; 2. 中国民族语文翻译中心, 北京 100080;
3. 辽宁大学, 辽宁 沈阳 110036; 4. 杭州雅拓网络技术有限公司, 浙江 杭州 310012)

摘要: 本文对基于短语的统计机器翻译模型的删词问题进行研究与分析, 使用人工评价的方式将删词错误分为三类。本文通过两种方法, 即基于频次的方法和基于词性标注的方法, 对源语言句子中关键词汇进行识别。通过对传统的短语对抽取算法中引入源语言对空关键词汇的约束来缓解删词错误问题。自动评价方法以及人工评价方法证明, 本文提出的方法在汉英翻译任务以及英汉翻译任务中显著的缓解了删词错误问题, 同时得到一个精简的短语翻译表。

关键词: 统计机器翻译; 删词问题; 人工评价

中图分类号: TP391

文献标识码: A

Research on Word Deletion Issue in Statistical Machine Translation

LI Qiang¹, HE Yan-Long², LUAN Shuang³, XIAO Tong^{1,4}, ZHU Jing-Bo^{1,4}

(1. Northeastern University NLP Lab, Shenyang, Liaoning 110819, China;

2. China Ethnic Languages Translation Bureau, Beijing 100080, China;

3. Liaoning University, Shenyang, Liaoning 110036, China;

4. Hangzhou Yatuo Network Technology Co., Ltd., Hangzhou, Zhejiang 310012, China)

Abstract: This paper addresses the word deletion issue in phrase-based machine translation. From the standpoint for human reading, we categorize the causes of word deletion errors into three classes. We address this issue by introducing constraints on unaligned words of source language in phrase extraction. Two methods are presented for the design of the constraints, including a frequency-based method and a part-of-speech-based method. Automatic and human evaluations demonstrate promising improvements in translation quality on both the Chinese-to-English and the English-to-Chinese translation tasks, with a gain of compact phrase tables.

Key words: Statistical Machine Translation; Word Deletion Issue; Human Evaluation

1 简介

基于短语的统计机器翻译模型(Phrase-based SMT)广泛的应用于当今真实的翻译任务中^{[1][2]}。基于短语的统计机器翻译模型具有翻译模型原理简单, 翻译性能优异等诸多特点, 受到研究与应用人员的广泛青睐。短语翻译系统使用短语对作为翻译的基本单元。在短语对生成的过程中, 传统短语对抽取算法抽取所有与词对齐保持一致的短语对^[3]。在真实的自动词对齐结果中, 由于双语语种间句子结构的不同以及词对齐的错误, 很多源语言句子中的词汇在目标语言句子中并没有与之相对应词对齐结果, 本文将这种源语言词汇定义为对空词汇。在双语训练数据中, 由于对空词汇的存在, 源语言的一个短语可能会有多个目标语言短语与之保持一致, 反之亦然。由于词对齐的错误, 一些源语言词汇被错误的对应到目标语言某个词汇上。在翻译过程中一旦选择这一错误翻译, 从人工评价角度来看源语言词汇未被翻译, 本文将这种源语言词汇定义为错译词汇。由对空词汇或错译词汇造成的源语言中某个词

* 收稿日期: 2014年6月15日

定稿日期: 2014年7月26日

基金项目: 国家自然科学基金(61272376; 61300097); 中国博士后基金(2013M530131)

汇未被翻译的现象，本文定义为删词错误问题。

在人工对机器翻译结果评价的过程中，删词错误现象会严重影响评价人员对机器翻译结果的理解，与此同时，删词错误会严重影响句子的流畅度，对机器翻译结果的可用性造成非常大的影响。例如，当翻译“等待他搜集情报。”这一句子时，如果机器翻译系统给出“wait for it.”这样的翻译结果，而“搜集”与“情报”这两个源语言关键词汇没有进行翻译，那么用户将无法理解原文的意思。在现有的统计机器翻译架构上，为了快速完成翻译系统参数的优化，研究人员使用自动评价的方式对系统翻译性能进行评价，而这些自动评价指标没有特殊考虑删词错误问题。删词问题的研究需要使用人工评价方式对源语言句子中未被翻译的关键词汇进行判断，需要大量的人工操作，实验的可重复性较低。因此，在当前广泛应用、性能优异的统计机器翻译系统中，删词问题没有得到充分的分析与研究。Vilar 等人证实了在汉英翻译任务上，删词错误在所有翻译错误类型中占有近 30%的比例，但目前鲜有工作对这一问题提出较优的解决方法^[4]。

本文针对删词错误问题进行了深入研究，从人工评价的角度上将删词错误类型分成三类，即词汇翻译内容选取错误，边缘对空关键词汇翻空以及中间对空关键词汇翻空。在此基础上，设计了一种简单有效的方法对较为严重的删词错误类型进行求解。首先，给定一个源语言句子，本文将句子中出现的词汇分成两类，即：关键词汇与非关键词汇。在这里，本文定义表达句子主要信息、不能译空的词汇为关键词汇，其它词汇为非关键词汇。本文提出了两种方法进行关键词汇的自动识别，即基于频次的方法与基于词性标注的方法。继而本文提出了一种受限的短语对抽取算法，即在短语对抽取的过程中，对边缘对空的关键词汇加以限制，以达到缓解删词错误的目的。最终，本文方法生成含有较少删词错误信息的短语翻译表。本文将提出的方法应用到当前性能优异的基于短语的统计机器翻译系统中，人工评价和自动评价证实，在汉英翻译任务和英汉翻译任务中，与基线系统相比本文提出的方法在翻译性能上有显著的提高。与此同时，翻译系统得到一个精简的短语翻译表。

2 本文方法

标记:	关键	关键	关键
翻译:	<i>future</i>	<i>belong to</i>	<i>children</i>
中文:	未来	就	属于 孩子
翻译结果:	<i>the future</i>	<i>belongs to</i>	<i>the children</i>
标记:	<i>key</i>	<i>key</i>	<i>key</i>

示例短语对	传统	本文
属于 <i>belongs to</i>	YES	YES
就属于 <i>belongs to</i>	YES	YES
属于孩子 <i>belongs to</i>	YES	NO

图 1. 传统短语对抽取方法与本文短语对抽取方法抽取的示例短语对

首先，本文通过对传统的短语对抽取方法的分析来阐述删词错误问题。对于给定的一个句对，图 1 中“传统”一列显示了传统的短语对抽取方法可能抽取噪声短语对的一个示例。在图 1 中，源语言词汇“孩子”在词对齐中没有与之对应的目标语词汇，即该词为对空词汇。对于传统的短语对抽取算法来说，源语言短语“属于 孩子”有多个与之保持一致的目标语短语。在这些短语对中，包含噪声短语对（属于 孩子，*belongs to*），即“孩子”在对应的目标语短语中并没有翻译结果。在机器翻译系统解码过程中，一旦选择了这一错误短语对，

删词错误便隐含其中。

在短语对抽取的过程中，如果考虑关于源语言句子中词汇更多相关的统计信息或语言学信息，上述问题可以得到缓解。表 1 中显示图 1 所示句对中示例词汇在整个训练语料中的对空占比及对应词性情况。从表 1 中可以看出，“孩子”的对空占比为 14.45%，也就是说，在绝大多数情况下，“孩子”被确定的翻译为目标语言中的某个词汇。与此同时，“孩子”的词性为名词，在一个好的翻译结果中，名词基本不可以被翻译为空。通过对表 1 的观察，“孩子”可以被定义为表达源语言句子信息的关键词汇。通过对源语言关键词汇进行识别后，可以在短语对抽取过程中对关键词汇加以约束来缓解删词错误问题。

表 1. 示例词汇对空占比与词性统计

词汇	对空比例（百分比）	词性
未来	10.41 %	名词
就	37.23 %	副词
属于	18.19 %	动词
孩子	14.45 %	名词
the	58.68 %	冠词
future	40.13 %	名词
belongs	27.61 %	动词
to	45.36 %	介词
children	33.29 %	名词

2.1 删词错误分类

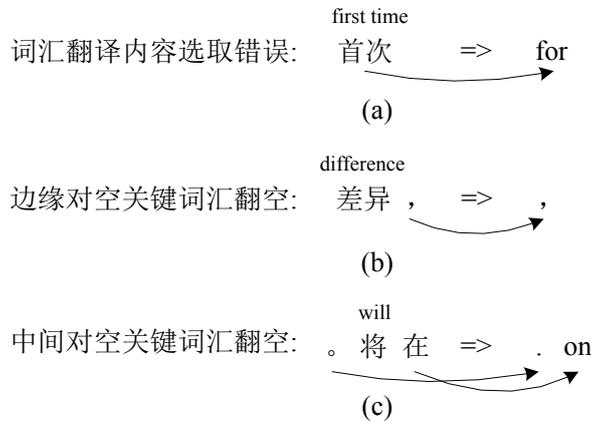


图 2. 三种删词错误类型实际情况

本文的主要内容是对删词问题进行分析与研究。通过使用人工评价的方式对真实机器翻译结果删词错误现象进行分析，本文将删词错误类型分为以下三类：

- 1) 词汇翻译内容选取错误：在这种错误类型中，源语言句子中词汇在目标语中有对应的翻译结果，但该翻译结果并不是当前词汇的正确翻译。该错误由自动词对齐错误造成，传统短语对抽取的过程回避了这一错误，导致错误蔓延。从人工对翻译结果进行评价的角度来看，源语言词汇没有被翻译。
- 2) 边缘对空关键词汇翻空：在短语对抽取的过程中，一个与词对齐保持一致的短语对的源语言短语的边缘存在对空词汇，而该词汇为表达源语言句子句义信息的关键词汇，该词汇没有进行翻译，造成错误的删词现象。

- 3) 中间对空关键词汇翻空：在短语对抽取的过程中，一个与词对齐保持一致的短语对的源语言短语的中间某个词汇对空，从该词汇向两边扩展存在含有词对齐信息的词汇，同时该词汇为关键词汇，该词汇没有进行翻译。

图 2 为三种删词错误类型的真实情况。在图 2 (a) 中，源语言短语中词汇“首次”被错误的翻译为“for”，该问题由于自动词对齐的错误造成；在图 2 (b) 中，源语言短语中词汇“差异”为短语“差异，”的边缘对空关键词汇，在目标语中没有与之对应的翻译结果；在图 2 (c) 中，源语言短语中关键词汇“将”对空，同时该词汇两边的词汇在目标语中有对应翻译结果，词汇“将”没有对应翻译结果。

表 2 中显示上文所列三种删词错误类型的错误分布情况。从表 2 中可以看出，在三种错误类型中，词汇翻译内容选取错误与边缘对空关键词汇翻空错误总数占有所有错误类型的 98% 以上。然而，词汇翻译内容选取错误涉及到整个翻译框架中的“择词”问题，该问题较为复杂，非传统删词问题，本文不对这种错误类型进行讨论。本文主要对图 2 (b) 情况进行分析与讨论，继而提出合理的缓解该类删词错误的解决方案。图 3 中内容为机器翻译系统输出的真实的删词错误示例。在图 3 中，对于源语言短语“搜集 情报 .”，翻译系统为其提供错误的翻译结果“.”，而在使用的短语对 (搜集 情报 ., .) 中，“搜集”和“情报”两个词汇并没有词对齐，即边缘对空词汇“搜集”与“情报”翻空造成了该类删词错误。

表 2. 三种删词错误类型错误分布统计，错误统计数据来源于随机选取的 200 行训练数据

错误分类	汉英翻译		英汉翻译	
	频次	百分比	频次	百分比
词汇翻译内容选取错误	173	60.70 %	161	75.23 %
边缘对空关键词汇翻空	107	37.54 %	52	24.30 %
中间对空关键词汇翻空	5	1.75 %	1	0.47 %

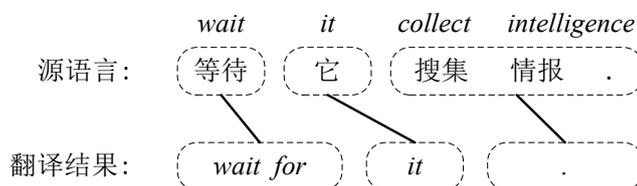


图 3. 删词错误示例

2.2 关键词汇与非关键词汇

对于表达待翻译文本关键信息、不能译空的词汇，本文称之为“关键词汇”，其他词汇则为“非关键词汇”。本文通过两种方法对关键词汇和非关键词汇进行识别。

表 3. 不同词性删词错误统计，统计数据来源于训练数据中随机抽取的 100 个句子

词性	汉英翻译		英汉翻译	
	频次	比例	频次	比例
名词	24	48.00 %	42	58.33 %
动词	20	40.00 %	17	23.61 %
其他	6	12.00 %	13	18.06 %

第一种方法为基于频次的关键词汇识别方法。受到 Li 等人相关工作的启发，本文使用源语言词汇的对空占比作为指标来定义关键词汇^[5]。给定一个源语言词汇 w ，该词汇的对空占比定义为：

$$\Pr(w_{unalign}) = \frac{\text{count}(w_{unalign})}{\text{count}(w)}$$

在该公式中， $\text{count}(w_{unalign})$ 表示在双语训练数据中源语言词汇 w 被对空的次数， $\text{count}(w)$ 表示源语言词汇在源语言数据中出现的次数。如果源语言词汇 w 为关键词汇，那么对空占比 $\Pr(w_{unalign})$ 应当小于等于阈值 τ 。相对频率大于阈值 τ 的源语言词汇被定义为非关键词汇。表 1 中对一些示例词汇的对空占比进行统计。从表 1 中可以看出，对于关键词汇与非关键词汇的识别， $\Pr(w_{unalign})$ 是一个非常有效的指标。例如，中文词汇“就”与英文词汇“to”的对空占比很高 (> 35%)，这两个词汇应该被定义为非关键词汇。在本文工作中， τ 值默认设置为 0.2。

第二种方法为基于词性标注的方法。表 3 为对不同词性删词错误进行人工评价的统计情况。从表 3 中可以看出，在训练数据中，名词和动词删除错误所占比例超过 80%。因此，本文定义源语言词性为名词或动词的词汇为关键词汇，其他词汇为非关键词汇：

$$\begin{aligned} KEY &= \{noun, verb\} \\ USUAL &= \neg KEY \end{aligned}$$

在这里， KEY 表示关键词汇， $USUAL$ 表示非关键词汇。

2.3 受限的短语对抽取方法

输入： 句对 (f_1^J, e_1^I) 的词对齐 A , 源语言非关键词汇列表 $SUEL$ 。

输出： 短语对集合 P

```

1  for  $len = 1$  to  $I$  do
2      for  $i_1 = 1$  to  $I$  do
3           $i_2 = i_1 + len$ 
4          if  $(f_{i_1}^{j_2}, e_{i_1}^{i_2})$  与词对齐 A 保持一致 then
5              if  $f_{i_1}^{j_2}$  的所有边缘对空词汇在  $SUEL$  中 then
6                  加入  $(f_{i_1}^{j_2}, e_{i_1}^{i_2})$  至短语对集合  $P$  中
7              end
8          end
9      end
10 end
```

图 4. 受限的短语对抽取算法

上文方法可以对源语言句子中关键词汇与非关键词汇进行识别，在本节中，通过使用源语言非关键词汇列表对短语对的抽取过程加以限制，本文提出一种受限的短语对抽取算法，算法详细内容见图 4。算法内容解释如下：给定双语平行句对的词对齐 A 以及源语言的非关键词汇列表 $SUEL$ ，该方法抽取所有与词对齐保持一致的短语对（图 4 中行 4），同时确保短语对中源语言短语边缘对空词汇在非关键词列表 $SUEL$ 中（图 4 中行 5）。图 1 中“本文”

一列显示本文提出的方法抽取的示例短语对集合。如前文所述, (属于 孩子, belongs to) 是与词对齐保持一致的短语对, 但该短语对隐含删词错误。在使用受限的短语对抽取算法时, 由于边缘对空词汇“孩子”没有在本文定义的非关键词汇列表 *SUEL* 中, 该短语对不会被抽取。该算法的主要目的是在抽取与词对齐保持一致的短语对的过程中, 过滤掉可能隐含删词错误的短语对, 从而缓解翻译过程中的删词错误问题。

3 评价

本文使用 NiuTrans 开源统计机器翻译系统中基于短语的统计机器翻译模型构建实验平台, 方法的有效性在汉英翻译任务与英汉翻译任务上同时进行验证^[6]。

3.1 实验设置

在汉英翻译任务与英汉翻译任务的实验中, 使用 NIST MT 2008 机器翻译评测提供的 4.46M 双语平行句对进行翻译模型的训练。使用 GIZA++ 工具对双语训练数据进行双向词对齐, 使用 *grow-diag-final-and* 启发式算法对双向词对齐结果进行对称化处理^[7]。此外, 两个翻译任务的实验中分别使用使用 GIGAWORD 中英语、汉语的 Xinhua 部分和双语数据的英语、汉语部分训练 5 元语言模型。使用最小错误率训练 (MERT) 方法对统计机器翻译模型的系统参数进行优化^[8]。关于开发集和测试集, 汉英翻译任务使用 NIST 2006 Newswire 的测试集作为翻译系统权重调优的开发集 (1181 句), 将 NIST 2004, 2005 和 2008 Newswire 的测试集进行合并, 作为评价实验系统翻译性能的测试集 (3561 句)。对于英汉翻译任务, 使用 SSMT2007 测试集作为翻译系统参数优化的开发集 (995 句), 使用 NIST 2008 测试集作为评价机器翻译系统翻译性能的测试集 (1859 句)。关于使用数据更多的信息详见表 4 与表 5。翻译性能通过使用大小写不敏感 (case-insensitive) 的 IBM 版本的 BLEU 以及翻译错误率 (TER) 评价指标进行评价^{[9][10]}。

表 4. 汉英翻译任务使用数据

数据	行数	词汇数
双语训练数据	4.46 M	115.0 M/133.7 M
目标语单语训练数据	66.15 M	1221.3 M
NIST 06 Newswire	1181	28378
NIST 04	1788	47743
NIST 05	1082	29278
NIST 08 Newswire	691	17409

表 5. 英汉翻译任务使用数据

数据	行数	词汇数
双语训练数据	4.46 M	133.7 M/115.0 M
目标语单语训练数据	66.69 M	1463.32 M
SSMT2007	995	26774
NIST 08	1859	46296

3.2 实验结果

表 6 和表 7 分别为汉英翻译任务与英汉翻译任务中, 基线翻译系统与使用本文方法的翻译系统在开发集与测试集上的 BLEU 值和 TER 值。由于 TER 指标较好的体现了删词错误现象的严重程度, 所以, 本实验的主要目的是保证 BLEU 值在不发生较大波动的情况下,

翻译结果在 TER 指标上的性能取得明显的改善。在表 6 和表 7 中, *表示在 $p < 0.05$ 的情况下, 显著优于基线系统的实验结果。

表 6. 汉英翻译任务中, 短语翻译表大小、BLEU4 值以及 TER 值

方法	百分比	开发集		测试集	
		BLEU	TER	BLEU	TER
基线系统	100 %	36.31	57.63	37.30	58.35
基于频次的方法	73.69 %	36.22	56.97*	37.51	57.14*
基于词性标注的方法	69.99 %	35.66	57.30*	36.94	57.76*
边缘无对空词汇	28.96 %	32.75	57.62	34.11	58.27

表 7. 英汉翻译任务中, 短语翻译表大小、基于字的 BLEU5 值以及 TER 值

方法	百分比	开发集		测试集	
		BLEU	TER	BLEU	TER
基线系统	100 %	33.36	73.62	31.10	75.03
基于频次的方法	83.17 %	33.44	72.92*	31.04	74.18*
基于词性标注的方法	72.39 %	33.87	73.05*	31.44	73.64*
边缘无对空词汇	30.00 %	31.06	74.64	29.87	75.74

在表 6 中, 本文提出的基于频次的方法在测试集上得到与基线系统可比的 BLEU 值; 与此同时, 在 TER 评价指标上, 本文方法在测试集上翻译结果质量获得显著的改善, 即取得 1.21 个 TER 点的下降 (TER 值越低, 错误率越低, 翻译性能越好)。更有趣的是, 与基线翻译系统的短语翻译表相比, 基于频次的方法可以有效的减小短语翻译表的大小, 即短语翻译表的大小下降了 26.31%。表 6 中“边缘无对空词汇”行实验结果表明, 有效的处理好边缘对空词汇是统计机器翻译系统成功的关键之一。如果在短语对抽取的过程中, 与词对齐保持一致的短语对的源语言和目标语言的边缘词汇不允许对空, 虽然短语翻译表的大小可大幅度减小 71.04%, 但是在测试集上的翻译性能下降了 3.19 个 BLEU 点。该实验结果表明, 允许源语言句子中某些词汇翻空对于构造好的机器翻译结果来说非常必要。除此之外, 表 6 中同样给出基于词性标注方法的实验性能。基于词性标注的方法与基线系统相比, 短语翻译表的大小下降了 30.01%, 同时翻译结果在 TER 指标上获得 0.59 个点的降低。

在英汉翻译任务上, 实验结果与汉英翻译任务类似, 具体实验结果见表 7。在表 7 中, 基于频次的方法使用的短语翻译表的大小相比基线翻译系统减小 16.83%, 与此同时, 在 TER 指标上获得 0.85 个点的降低。基于词性标注的方法使用的短语翻译表的大小减小 27.61%, 同时获得 1.39 个 TER 点的降低。边缘无对空词汇的方法短语翻译表的大小大幅度减小了 70.00%, 但是在测试集上的翻译性能下降了 1.23 个 BLEU 点。

在汉英翻译任务以及英汉翻译任务上, 除使用 BLEU 和 TER 自动评价指标外, 本文同时在 100 个随机选取的句子中进行人工评价来证实本文方法的有效性。本文通过专业翻译人员对机器翻译结果进行打分来评价翻译结果翻译质量的好坏。在人工评价的过程中, 同时提供基线系统、基于频次的方法、基于词性标注的方法的翻译结果, 翻译结果隐藏来源标签。翻译人员通过对翻译结果的选词准确度以及流畅度进行打分, 分数范围在 0 至 10 之间。如果翻译结果与待翻译句子完全不相关, 则为 0 分; 如果翻译结果能够如实的表达待翻译文本的原意, 同时在翻译结果中覆盖所有待翻译句子中关键词汇的内容, 则为 10 分。由于篇幅有限, 这里对人工评价标准的具体细节不做介绍。在人工打分的过程中, 同时对待翻译句子中关键词汇的删词现象进行标注, 对于句子中的非关键词汇, 本文不做分析。表 8 中的人工评价结果表明, 在人工评价分数以及删词错误个数的评价中, 本文提出的方法优于基线翻译

系统。表 9 中实验结果显示不同词性删词错误的个数以及比例的统计。从表 9 中可以看出，与表 3 中结果相比，本文方法明显降低了名词和动词删词错误的个数。

表 8. 汉英翻译任务以及英汉翻译任务中，人工打分及删词错误个数统计结果

方法	汉英翻译		英汉翻译	
	人工打分	删词错误	人工打分	删词错误
基线系统	5.64	50	4.30	72
基于频次的方法	5.86	38	4.34	57
基于词性标注的方法	5.66	42	4.45	59

表 9. 翻译过程中，不同词性的错误个数和百分比的统计结果。↓表示与表 3 中结果相比下降的个数

词性	汉英翻译		英汉翻译	
	个数	百分比	个数	百分比
名词	19 (↓5)	50.00 %	32 (↓10)	56.14 %
动词	13 (↓7)	34.21 %	15 (↓2)	26.32 %
其他	6 (↓0)	15.79 %	10 (↓3)	17.54 %

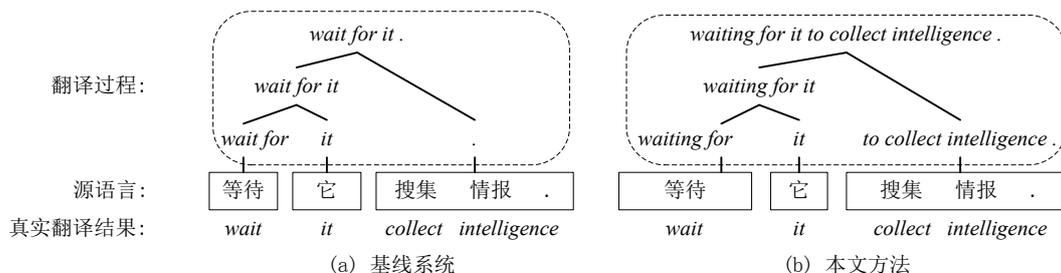


图 5. 基线系统 (a) 与本文方法 (b) 的 1-best 解码路径对比

本文通过比较基线系统与本文方法的 1-best 翻译结果的解码路径来说明本文方法之所以有效的原因。图 5 给出基线系统与本文方法的一个真实的翻译结果对比。在图 5 (a) 中，对源语言短语“搜集 情报 .”来说，“搜集”和“情报”是源语言句子的边缘对空词汇，传统的短语对抽取方法抽取（搜集 情报 ., .）这一与词对齐保持一致的短语对。在解码过程中，针对源语言待翻译短语“搜集 情报 .”，解码器从众多的短语对中选择（搜集 情报 ., .）这一短语对来进行翻译，从而导致表达句义的关键词汇“搜集”和“情报”在目标语中对应的翻译结果不存在，造成了严重的删词错误问题。在图 5 (b) 中，由于边缘对空关键词汇“搜集”和“情报”不在本文方法定义的非关键词列表 *SUEL* 中，所以短语对（搜集 情报 ., .）并没有被本文方法所抽取，在翻译的过程中不存在这样一条短语对供解码器选择，从而在源头上避免了在翻译过程中使用隐含删词错误的短语对。所以在针对源语言关键词汇的删词错误问题上，本文方法明显优于基线翻译系统。

表 10. 汉英翻译任务中，软约束方法实验结果

方法	百分比	开发集		测试集	
		BLEU	TER	BLEU	TER
基线系统	100 %	36.31	57.63	37.30	58.35
基于频次的方法	100 %	36.41	57.25*	37.39	58.00*
基于词性标注的方法	100 %	36.41	56.51*	37.62	57.09*

上文提出的方法是对传统短语翻译表中隐含删词错误信息的短语对进行过滤，可称上文方法为“硬约束”方法。使用硬约束方法有一个不可回避的问题，即在过滤噪声短语对的同时，正确的短语对有可能被同时过滤。在这里，本文将提出的方法使用“软约束”的方式进行实验。为了达到软约束的目的，在实验的过程中通过在短语翻译表中加入一个新的指示性特征，用以指示当前短语对是否可以被本文提出方法进行抽取。表 10 为本文方法在软约束的实验设置下得到的实验结果，该组实验在汉英翻译任务上进行。与基线系统相比，基于词性标注的软约束方法在测试集上获得 1.26 个 TER 点的降低，基于频次的方法获得 0.35 点的降低。基于频次的软约束方法从翻译性能角度来看，效果不如硬约束方法；但是在基于词性标注的方法中，软约束方法的实验性能明显优于硬约束的方法，在 BLEU 和 TER 指标上，分别获得 0.68 点提高和 0.67 点的降低。从这组实验结果来看，在基于词性标注的方法中，硬约束方法过滤掉很多质量较好的短语对。

4 相关工作

在机器翻译任务中，删词问题的分析研究已经取得了一些相关成果。如，Zhang 等在原始训练语料中移除对空频次较高的词汇，使用处理后的语料进行模型的训练^[11]。虽然该方法简单、易于实现，但是该方法也同时向短语翻译表中引入新的噪声。Zhang 等人的工作同时证明，在处理删词错误这类问题上，通过在短语对抽取的过程中过滤噪声短语对的方法在效果上明显优于提高词对齐质量的方法^{[12][13][14][15]}。Menezes 和 Quirk 提出基于 treelet 翻译方法的扩展，该方法在源语言中通过引入结构化的词汇插入和删除的调序模板对删词问题进行讨论^[16]。Parton 等人提出了一种混合的方法 APES 来解决翻译结果充分性错误^[17]。Li 等人提出了三种方法来处理源语言中冗余的词汇，使用多种方法对短语翻译表中短语对的翻译概率重新进行估计^[5]。

虽然此前的工作提出了几种方法来缓解机器翻译任务中的删词错误问题，但是之前的工作没有对删词错误造成的原因进行分类。本文将删词错误分为三类，即词汇翻译内容选取错误，边缘对空关键词汇翻空和中间对空关键词汇翻空。针对这三类删词错误类型，本文选择较为严重的边缘对空关键词汇翻空进行解决，提出了针对这个问题一系列的解法，同时获得较好的实验结果。本文提出的方法与语言无关，可以将本文方法应用到任意语言对上缓解删词错误问题。同时，在图 4 提出的受限的短语对抽取算法中，可以同时加入目标语言非关键词汇列表的约束，从而缓解翻译过程中目标语言冗余词汇的插入问题。本文提出的方法简单且易于实现，可有效缓解源语言句子中关键词汇的删除错误问题。

5 总结与展望

本文针对删词问题作了详细的研究与分析。本文通过人工分析的方法将删词错误类型划分为三类。针对三类删词错误问题，本文选择较为严重的边缘对空关键词汇翻空进行解决。首先，本文提出两种方法，即基于频次的方法与基于词性标注的方法，来对源语言句子中的关键词汇与非关键词汇进行识别；继而，本文方法使用源语言非关键词汇列表对短语对抽取的过程加以约束，提出受限的短语对抽取算法。本文方法同时在硬约束与软约束条件下进行实验，人工评价与自动评价均证实，本文提出的方法在汉英翻译任务和英汉翻译任务上，显著地提高了机器翻译系统的翻译质量，与此同时，硬约束的方法可以获得一个精简的短语翻译表。本文提出的方法有效的缓解了删词错误问题。

在未来的工作中，我们还将研究更为有效的模型和方法来对源语言句子中关键词汇与非关键词汇进行识别，同时也将对获取高质量的短语翻译表的短语对抽取算法进行研究，从而达到进一步缓解删词错误的目的。

参考文献

- [1] Philipp Koehn, Fran J. Och and Daniel Marcu. Statistical phrase-based translation[C]//Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1. Association for Computational Linguistics, 2003:48-54.
- [2] Franz J. Och and Heymann Ney. The alignment template approach to statistical machine translation[J]. Computational Linguistics, 2004, 30(4):417-449.
- [3] Franz J. Och, Christoph Tillmann and Heymann Ney. Improved alignment models for statistical machine translation[C]//Proceedings of the 1999 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 1999:20-28.
- [4] David Vilar, Jia Xu, Luis Fernando D'Haro and Hermann Ney. Error analysis of statistical machine translation output[C]//International Conference on Language Resources and Evaluation. 2006:697-702.
- [5] Chi-Ho Li, Dongdong Zhang, Mu Li, Ming Zhou and Hailei Zhang. 2008. An empirical study in source word deletion for phrase-based statistical machine translation[C]//Proceedings of the Third Workshop on Statistical Machine Translation. Association for Computational Linguistics, 2008:1-8.
- [6] Tong Xiao, Jingbo Zhu, Hao Zhang and Qiang Li. NiuTrans: an open source toolkit for phrase-based and syntax-based machine translation[C]//Proceedings of the ACL 2012 System Demonstrations. Association for Computational Linguistics, 2012:19-24.
- [7] Franz J. Och and Hermann Ney. Improved statistical alignment models[C]//Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2000:440-447.
- [8] Franz J. Och. Minimum error rate training in statistical machine translation[C]//Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1. Association for Computational Linguistics, 2003:160-167.
- [9] Kishore Papineni, Salim Roukos, Todd Ward and Weijing Zhu. BLEU: a method for automatic evaluation of machine translation[C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2002:311-318.
- [10] Matthew Snover, Bonnie Dorr, Richard Schwartz, Linnea Micciulla and John Marhouf. A study of translation edit rate with targeted human annotation[C]//Proceedings of the 7th Conference of the Association for Machine Translation in the Americas. 2006:223-231.
- [11] Yuqi Zhang, Evgeny Matusov and Hermann Ney. Are unaligned words important for machine translation?[C]//Proceedings of the 13th Annual Conference of the EAMT. 2009:226-233.
- [12] Ulf Hermjakob. Improved word alignment with statistics and linguistic heuristics[C]//Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2009:229-237.
- [13] Jingbo Zhu, Qiang Li and Tong Xiao. Improving syntactic rule extraction through deleting spurious links with translation span alignment[J]. Natural Language Engineering, 2013, 1-23.
- [14] Yang Liu, Qun Liu and Shouxun Lin. Discriminative word alignment by linear modeling[J]. Computational Linguistics, 2010, 36(3):303-339.
- [15] Yonggang Deng, and Bowen Zhou. Optimizing word alignment combination for phrase table training[C]//Proceedings of the ACL-IJCNLP 2009 Conference Short Papers. Association for Computational Linguistics, 2009:229-232.
- [16] Arul Menezes and Chiris Quirk. Syntactic Models for Structural Word Insertion and Deletion[C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2008:735-744.
- [17] Kristen Parton, Nizar Habash, Kathleen McKeown, Gonzalo Iglesias and Adrià de Gispert. Can Automatic Post-Editing Make MT More Meaningful?[C]//Proceedings of the Conference on EAMT. 2012:111-118.

作者简介: 李强 (1988——), 男, 博士研究生, 主要研究领域为机器翻译。Email:liqiangneu@gmail.com; 何燕龙 (1968——), 男, 高级工程师, 主要研究领域为少数民族语言翻译。Email:heyanglong@vip.sina.com; 栾爽 (1988——), 女, 硕士研究生, 主要研究领域为英语语言文学。Email:luanshuang1988@163.com; 肖桐 (1982——), 男, 博士, 主要研究领域为机器翻译。Email:xiaotong@mail.neu.edu.cn; 朱靖波 (1973——), 男, 教授, 博士生导师, 主要研究领域为机器翻译。Email:zhujingbo@mail.neu.edu.cn。