

# 统计机器翻译中实例短语对研究

李强<sup>1,†</sup> 李沐<sup>2</sup> 张冬冬<sup>2</sup> 朱靖波<sup>1</sup>

1. 东北大学自然语言处理实验室, 沈阳 110819; 2. 微软亚洲研究院, 北京 100080; †E-mail: liqiangneu@gmail.com

**摘要** 针对由于数据的稀疏性和双语数据规模的局限性造成的大量高质量短语对没有生成的问题, 在基于短语的统计机器翻译系统中, 通过对传统短语抽取算法抽取的短语对进行分解、替换、生成等操作, 生成传统方法无法抽取的实例短语对。在汉英新闻和汉英口语翻译任务上, 与基线系统相比, 该方法在多个测试集上明显提高了翻译系统的翻译质量, 在部分测试集上 BLEU 值可提高 1% 左右。

**关键词** 统计机器翻译; 基于短语; 基于实例; 短语对

**中图分类号** TP391

## Research on Example-Based Phrase Pairs in Statistical Machine Translation

LI Qiang<sup>1,†</sup>, LI Mu<sup>2</sup>, ZHANG Dongdong<sup>2</sup>, ZHU Jingbo<sup>1</sup>

1. NLP Lab, Northeastern University, Shenyang 110819; 2. Microsoft Research Asia, Beijing 100080; † E-mail: liqiangneu@gmail.com

**Abstract** Due to the sparsity of data and the limitation of bilingual data size, many high-quality phrase pairs can't be generated. The example-based phrase pairs proposed by the authors are generated through decomposing, substituting and generating the typical phrase pairs, and the typical phrase pairs are generated by the typical phrase extraction method in phrase-based statistical machine translation. On the Chinese-to-English Newswire and Oral translation tasks, the experimental results demonstrate significant improvements achieved by the proposed methods. Moreover, a gain of about 1% BLEU score increase is yielded on some test sets.

**Key words** statistical machine translation; phrase-based; example-based; phrase pair

基于短语的统计机器翻译模型具有原理简单、性能优异、鲁棒性高等诸多特点, 受到研究与应用人员的广泛青睐, 在当今真实的翻译任务中得到广泛应用<sup>[1-2]</sup>。短语翻译系统使用短语对作为翻译的基本单元, 传统的启发式短语对抽取算法抽取所有与词对齐保持一致的短语对<sup>[3]</sup>。基于统计的机器翻译方法建立在大规模双语及单语语料的基础上。然而, 在当今真实的翻译任务中, 许多翻译任务(如口语翻译任务)并没有充足的双语语料用于抽取高质量的短语对。由于数据的稀疏性、双语数据规模的局限性以及词对齐的错误和与词对齐保持一致规则的约束, 大量有用的、高质量的短语对没有生成, 导致待翻译源语言句子中很多短语片段没有直接对应的翻译结果, 只能在翻译解码的过程中, 通过组

装颗粒度较小的短语对进行译文的生成。因此, 通过有效的方法生成传统短语对抽取算法无法生成的、高质量的短语对, 扩充现有短语对规模, 在翻译过程中, 通过使用新生成的短语对来提高统计机器翻译系统的翻译质量, 是本文研究的主要内容。本文生成短语对的方法借鉴基于实例翻译模型的原理, 因此定义本文方法生成的短语对为实例短语对<sup>[4]</sup>。

在图1中, 存在“经过 几天 紧张 的 布置”这一源语言待翻译句子。在短语表中, 二元短语“几天 紧张”没有对应的翻译结果, 在翻译的过程中, 只能通过“几天”和“紧张”两个短语对应的翻译结果进行拼装, 完成这一短语片段的翻译。由于解码的复杂性, 这一过程大大增加了翻译错误发生的可能

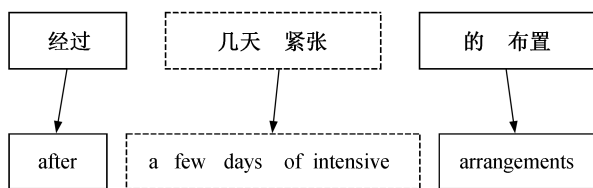


图 1 使用短语对完成解码过程(虚线框中是未生成的短语对)

Fig. 1 Decoding with using phrase pairs (phrase pair in dashed rectangle is not generated)

性。如果在翻译之前,能够通过有效方法生成〈几天 紧张, a few days of intensive〉这一短语对,那么翻译图 1 中句子将变得非常简单。

在基于短语的统计机器翻译系统中,在双语语料规模受限的情况下,本文通过对传统方法生成的短语对进行分解、替换、生成操作,最终生成高质量的实例短语对,以扩充现有短语对的规模。在汉英新闻和口语翻译任务上,与基线系统相比,本文方法在多个测试集上明显提高了翻译系统的翻译质量,在部分测试集上 BLEU 值可提高 1% 左右。

## 1 相关工作

日本的长尾真(Makoto Nagao)于 1981 年提出基于实例的翻译系统,并于 1984 年发表相关学术论文<sup>[4]</sup>。基于实例的基本思想是在已经收集的双语实例库中找出与待翻译部分最相似的翻译实例(通常是句子),再对实例的译文通过串替换、串删除以及串增加等一系列变形操作,实现翻译。基于实例的方法仅仅计算待翻译句子与实例库中句子的相似度,并没有考虑替换词汇之间的相似程度。本文方法可以有效地解决这一问题。Chiang<sup>[5]</sup>提出的基于层次短语的翻译模型有效地缓解了基于短语的翻译模型中无法处理的长距离依赖和调序问题,提升了统计机器翻译系统的翻译性能。但是 Chiang 的方法中使用的层次短语泛化能力过强,对非终结符翻译替换操作没有约束。本文使用词向量的方法,通过计算短语之间的余弦值判断相似度,在相似度大于一定阈值的条件下,执行替换操作,从而提高生成短语对的质量。

## 2 $N$ 元实例短语对

首先,通过对传统算法生成的短语对在测试集上的生成率进行分析,说明本文工作的必要性。在表 1 中, $N$ 元短语对在测试集中的生成率定义为:测

表 1 汉英新闻和汉英口语翻译系统中二元和三元短语对在测试集上的生成率

Table 1 Generation rate of bigram and trigram phrase pairson test sets on Chinese-to-English newswire and oral translation

翻译类别	短语对生成率/%	
	二元	三元
新闻	67.28	27.30
口语	44.51	19.24

试集中所有句子包含的  $N$  元短语的个数为  $n$ , 在  $n$  个短语中通过传统短语抽取算法可生成的  $N$  元短语对个数为  $m$ , 那么生成率为  $m/n$ 。在这里,将源语言长度为  $N$  的短语对称之为  $N$  元短语对。在本文实验的数据中,汉英新闻和汉英口语任务上,二元短语对的生成率分别为 67.28% 和 44.51%,三元短语对的生成率分别为 27.30% 和 19.24%。可以看出,通过提高短语对的生成率,继而增加高质量短语对的数量,是提高翻译系统翻译质量的一个有效途径。

### 2.1 实例短语对

如前所述,在翻译图 1 中句子时,如果存在〈几天 紧张, a few days of intensive〉这一短语对,翻译过程将变得简单,翻译结果将更加可靠。在启发式短语对抽取的过程中,〈几天, a few days〉和〈年 紧张, years of intensive〉已经生成。通过对现有的两个短语对进行分解、替换、生成等操作,可生成“几天 紧张”对应的翻译结果。由于该方法从基于实例的方法得到启发,所以本文将通过分解、替换、生成三步构造的短语对称为实例短语对<sup>[4]</sup>。图 2 为实例短语对生成过程的图解,具体步骤说明如下。

1) 分解:根据与词对齐保持一致的定义,对已有的短语对进行分解,短语对〈年 紧张, years of intensive〉可分解为两个与词对齐保持一致的短语对〈年, years〉和〈紧张, of intensive〉。

2) 替换:短语对〈几天, a few days〉的源语言部分“几天”和分解的短语对的源语言部分“年”相似度较高,为 0.4880;短语对中目标语言“a few days”和“years”的相似度同样较高,为 0.6764。在源语言与目标语言相似度同时较高的情况下,执行替换操作,生成新短语对〈几天 紧张, a few days of intensive〉及词对齐。在这里,被替换与执行替换操作的短语对都为一元短语对。

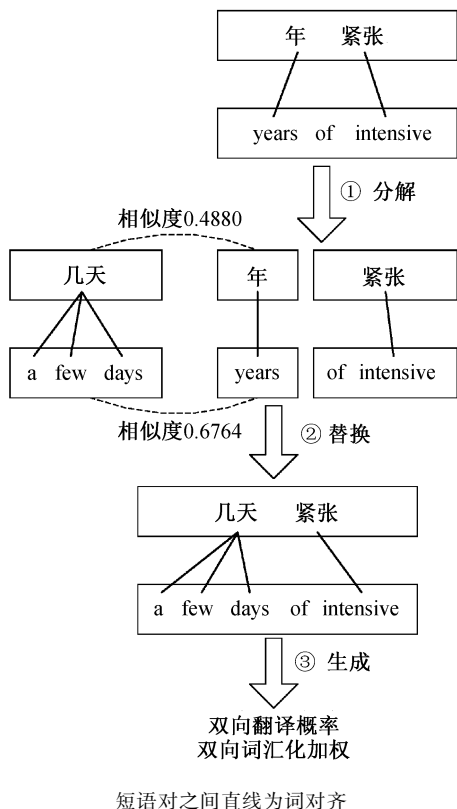


图2 实例短语生成过程

Fig. 2 Example-based phrase pairs generation procedure

3) 生成: 新生成的短语需要加入到现有的短语翻译表中, 所以需要进行概率估计。概率估计主要包括双向的翻译概率和双向的词汇化加权。

执行以上 3 个步骤后, 实例短语对生成完毕。在这里, 存在的挑战如下: 1) 短语对的分解需要根据什么规范; 2) 替换时, 相似度如何计算, 阈值如何设定; 3) 生成时, 如何进行概率估计。下文将对这 3 个问题进行逐条分析和讨论, 从而提出有效的解决方案。

## 2.2 短语分解

短语分解过程中, 分解后的短语对需要满足与词对齐保持一致的定义<sup>[3]</sup>。短语对与词对齐保持一致的定义如下: 短语对  $(\bar{s}, \bar{t})$  与词对齐保持一致, 当且仅当  $\bar{s}$  中的所有单词在词对齐  $A$  中所对应的单词在  $\bar{t}$  范围之内时, 则  $\bar{t}$  中所有单词在词对齐  $A$  中所对应的单词在  $\bar{s}$  范围之内; 同时, 在  $\bar{s}$  与  $\bar{t}$  中, 至少有一个单词对在词对齐  $A$  中。

在上述定义中,  $\bar{s}$  表示源语言短语,  $\bar{t}$  表示目标语言短语。图 2 中第一步分解结果如图 3 所示。在这里, 根据一致性定义,  $\langle$ 年 紧张, years of inten-

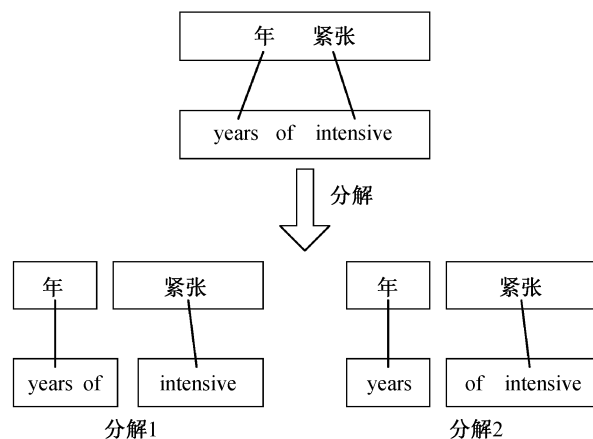


图3 短语对分解结果

Fig. 3 Results of decomposing phrase pairs

sive) 可以分解为  $\langle$ 年, years of) 与  $\langle$ 紧张, intensive) 、 $\langle$ 年, years) 与  $\langle$ 紧张, of intensive) 这样两组分解结果。

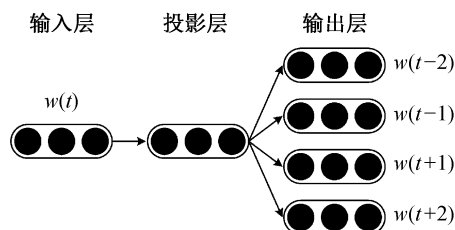
## 2.3 相似度计算

本文使用 Mikolov 等<sup>[6-8]</sup>提出的连续 Skip-gram 模型(Continuous Skip-gram Model)训练词向量(word vector, distributed representations), 通过计算词向量之间的余弦值, 判断相似程度。Skip-gram 模型根据当前词汇预测周围词汇的词向量(图4)。

假设要计算相似度的两个词向量分别为  $(x_1, y_1, z_1)$  与  $(x_2, y_2, z_2)$ , 则余弦值计算公式为

$$\cos(\theta) = \frac{x_1 x_2 + y_1 y_2 + z_1 z_2}{\sqrt{(x_1^2 + y_1^2 + z_1^2)(x_2^2 + y_2^2 + z_2^2)}}$$

在目标语言端, 计算相似度的有可能是短语。在本文中, 短语的词向量为短语中所有词汇词向量的算术平均值。经验性地, 在替换时相似度阈值设置的范围是  $[0.3, 0.5]$ 。



假设每个词的词向量用 3 个实值表示

图4 连续的 Skip-gram 模型

Fig. 4 Continuous Skip-gram Model

## 2.4 概率估计

在短语翻译表训练的过程中,需要对短语对计算双向的短语翻译概率  $\Pr(\bar{t}|\bar{s})$  和  $\Pr(\bar{s}|\bar{t})$  以及双向的词汇化加权  $\Pr_{\text{lex}}(\bar{t}|\bar{s})$  和  $\Pr_{\text{lex}}(\bar{s}|\bar{t})$  [1]。在实例短语对生成的过程中,由于生成了词对齐,所以双向的词汇化加权可以直接按照标准方法进行计算。但是在实例短语对生成的过程中,由于丢失了原始数据的统计信息,所以双向的短语翻译概率无法直接进行估计。本文参考在解码过程中使用的对数-线性模型对模型分数计算的方式,对生成的短语对的翻译概率重新进行估计[9]。双向的短语翻译概率  $\Pr(\bar{t}|\bar{s})$  和  $\Pr(\bar{s}|\bar{t})$  计算公式如下:

$$\Pr(\bar{t}|\bar{s}) = \exp^{\log(\Pr(\bar{t}'|\bar{s}')) - \log(\Pr(\bar{s}'_0|\bar{s}'_0)) + \log(\Pr(\bar{t}_n|\bar{s}_n))},$$

$$\Pr(\bar{s}|\bar{t}) = \exp^{\log(\Pr(\bar{s}'|\bar{t}')) - \log(\Pr(\bar{s}'_0|\bar{t}'_0)) + \log(\Pr(\bar{s}_n|\bar{t}_n))}。$$

通过图 2 中的实例,对上述计算短语翻译概率中的符号进行解释。在图 2 中,  $\bar{s}$  和  $\bar{t}$  分别对应于待生成的短语对〈几天 紧张, a few days of intensive〉的源语言和目标语言;  $\bar{s}'$  和  $\bar{t}'$  对应于原始的短语对〈年 紧张, years of intensive〉;  $\bar{s}'_0$  和  $\bar{t}'_0$  对应于被替换的短语对〈年, years〉;  $\bar{s}_n$  和  $\bar{t}_n$  对应于替换短语对〈几天, a few days〉。

## 2.5 解码

本文提出的方法可通过在解码时使用的对数-线性模型中加入新的特征的方式,嵌入现有的解码器中。新加入的特征包括实例短语对的指示性特征  $h_{\text{sig}}$ 、源语言相似性特征  $h_{\text{src}}$  以及目标语言相似性特征  $h_{\text{tgt}}$ 。  $h_{\text{src}}$  为源语言被替换短语与替换短语之间的相似度,  $h_{\text{tgt}}$  为目标语言被替换短语与替换短语之间的相似度。指示性特征  $h_{\text{sig}}$  如下式所示:

$$h_{\text{sig}} = \begin{cases} 0, & \text{传统短语对,} \\ 1, & \text{实例短语对.} \end{cases}$$

在解码时,将这 3 个特征嵌入到对数-线性模型中,如下式所示:

$$\begin{aligned} \hat{e}'_1 &= \arg \max_{e'_1} \{ \Pr(e'_1 | f'_1) \} \\ &= \arg \max_{e'_1} \left\{ \sum_{m=1}^M \lambda_m h_m(e'_1, f'_1) \right\} \\ &= \arg \max_{e'_1} \left\{ \sum_{m=1}^N \lambda_m h_m(e'_1, f'_1) + \lambda_{\text{sig}} h_{\text{sig}} + \lambda_{\text{src}} h_{\text{src}} + \lambda_{\text{tgt}} h_{\text{tgt}} \right\}。 \end{aligned}$$

在这里,前  $N$  个特征来自于基线系统。其中,  $h_{\text{sig}}$ ,  $h_{\text{src}}$  和  $h_{\text{tgt}}$  为本文提出的新的特征,  $\lambda_{\text{sig}}$ ,  $\lambda_{\text{src}}$  和  $\lambda_{\text{tgt}}$  为对应的特征权重。

## 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,在汉英新闻和汉英口语翻译任务上验证方法的有效性[10]。

### 3.1 实验设置

在汉英新闻和汉英口语翻译任务的实验中,实验数据信息见表 2 和 3。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[11]对双向词对齐结果进行对称化处理。使用最小错误率训练(MERT)方法[12]在开发集上对统计机器翻译系统的模型参数进行优化。通过大小写不敏感的 IBM 版本的 BLEU[13]对翻译性能进行评价。在新闻系统上,替换的源语言和目标语言的相似度阈值为 0.4,在口语系统上,阈值为 0.3。本文分别使用 GIGAWORD 中 Xinhua 语料的中文部分和英文部分训练中文和英文的词向量。

表 2 汉英新闻任务使用数据  
Table 2 Data for Chinese-to-English Newswire translation

分类	数据	句数	词汇数
双语数据	NIST MT 2008 双语句对	2.43 M	56.57 M / 67.12 M
单语数据	GIGAWORD Xinhua	16.28 M	414.03 M
开发集	NIST 06 News	616	17105
	NIST 08 News	691	17409
测试集	NIST 08 Pro News	688	17348
	NIST 12 News	400	10939

表 3 汉英口语翻译任务使用数据  
Table 3 Data for Chinese-to-English Oral translation

分类	数据	句数	词汇数
双语数据	IWSLT 双语句对	30000	0.26 M / 0.34 M
单语数据	双语句对英文	30000	0.34 M
开发集	CSTAR03 Dev	506	3251
测试集	IWSLT04 Test	500	3276
	IWSLT05 Test	506	3497

### 3.2 实验结果

在基于短语的统计机器翻译系统翻译过程中, 解码器更倾向于使用三元及三元以下短语对, 这是因为在解码的过程中, 三元以上短语对在规则装载的过程中命中率较低, 不容易被使用<sup>[1]</sup>。表 4 为不同翻译任务中, 不同元数短语对的使用分布情况。从表 4 可以看出, 一元短语对的使用比例最高, 在两个翻译任务中的使用比例在 60%~70% 之间; 二元短语对的使用比例在 20%~30% 之间; 三元短语对的使用比例最低, 为 10% 左右。

表 5 和 6 分别为汉英新闻和汉英口语翻译任务中, 基线系统与本文方法系统在开发集与测试集上

的 BLEU 值。基于表 4 的分析结果, 所有实验都只加入二元实例短语对。

表 5 中, 加入 3 种不同特征的实验结果与基线系统相比, BLEU 值都有大幅度提高。“+ 指示特征”在 NIST 08 和 NIST 12 上 BLEU 值分别提高 0.88% 和 0.83%。表 6 为小规模口语数据上的实验结果, 从该组实验结果来看, 在语料规模受限、词汇覆盖度不高的情况下, 本文方法可以大幅度提高翻译系统的翻译性能, 如“+ 指示特征”在 IWSLT 05 测试集上 BLEU 值提高 1.07%。

表 7 为在汉英两种翻译任务“+ 指示特征”实验中, 在测试集上使用不同长度短语对的统计情况。可以看出, 在解码器使用的所有短语对中, 本文方法生成的二元实例短语对占 5% 和 7%, 达到一个理想的比例, 说明本文提出的方法可生成高质量的短语翻译对。

使用率定义如下: 待翻译句子数为  $i$ , 使用本文生成的二元实例短语对翻译的句子数为  $j$ , 则使用率为  $j/i$ 。在汉英新闻和口语实验中, 使用率分别为 50% 和 25%。因为在口语数据中, 待翻译句子普遍较短, 所以口语使用率较低。IWSLT 04 测试集的源语言每句平均含有的词汇数为 6.55, 同时, 由

表 4 汉英翻译任务中, 解码时不同元数短语对使用分布情况

Table 4 On Chinese-to-English translation, statistics of  $n$ -gram phrase pairs selected by decoder

方法	短语对使用分布情况/%		
	一元	二元	三元
汉英新闻	70	23	7
汉英口语	60	27	12

表 5 汉英新闻翻译任务中, 开发集和测试集上的翻译性能

Table 5 BLEU score for development and test sets on Chinese-to-English Newswire translation

方法	BLEU/%			
	NIST 06	NIST 08	NIST 08 Pro	NIST 12
基线系统	35.80	30.51	27.79	31.89
+ 指示特征	36.19 (+0.39)	31.39* (+0.88)	28.26 (+0.47)	32.72* (+0.83)
+ 相似特征	36.27 (+0.47)	31.08* (+0.57)	28.20 (+0.41)	32.37 (+0.48)
+ 指示&相似特征	36.18 (+0.38)	31.25* (+0.74)	28.02 (+0.23)	32.76* (+0.87)

说明: “+ 指示特征”为使用  $h_{\text{sig}}$  的实验结果, “+ 相似特征”为使用  $h_{\text{src}}$  和  $h_{\text{tgt}}$  的实验结果, “+ 指示&相似特征”为使用  $h_{\text{sig}}$ ,  $h_{\text{src}}$  和  $h_{\text{tgt}}$  的实验结果。括号内数字表示 BLEU 值提高的幅度。

注: \*表示在  $p < 0.05$  的情况下, 本文方法显著优于基线系统的实验结果。

表 6 汉英口语翻译任务中, 开发集和测试集上的翻译性能  
Table 6 BLEU score for development and test sets on Chinese-to-English Oral translation

方法	BLEU/%		
	CSTAR 03	IWSLT 04	IWSLT 05
基线系统	46.52	47.63	54.17
+ 指示特征	47.51* (+0.99)	48.56* (+0.93)	55.24* (+1.07)
+ 相似特征	47.13* (+0.61)	48.44* (+0.81)	54.83* (+0.66)
+ 指示&相似特征	47.53* (+1.01)	48.48* (+0.85)	54.88* (+0.71)

说明: “+ 指示特征”为使用  $h_{\text{sig}}$  的实验结果, “+ 相似特征”为使用  $h_{\text{src}}$  和  $h_{\text{tgt}}$  的实验结果, “+ 指示&相似特征”为使用  $h_{\text{sig}}$ ,  $h_{\text{src}}$  和  $h_{\text{tgt}}$  的实验结果。括号内数字表示 BLEU 值提高的幅度。

注: \*表示在  $p < 0.05$  的情况下, 本文方法显著优于基线系统的实验结果。

表 7 汉英翻译任务中“+指示特征”实验解码时不同元数短语对使用分布情况

Table 7 Statistics of  $n$ -gram phrase pairs selected by decoder in “+ indicator feature” experiment on Chinese-to-English translation

数据	短语对使用分布情况/%				
	一元	二元	二元实例	三元	使用率
NIST 08 News	64	23	5	8	50
IWSLT 04	60	24	7	9	25

于该数据中绝大多数为常用词汇, 也导致使用率值较低。

通过比较基线系统与本文方法的 1-best 翻译结果来说明本文方法的有效性, 如图 5 所示。待翻译的句子片段为“经过 几天 紧张 的 布置”, 基线系统的翻译结果为“the arrangements after a few

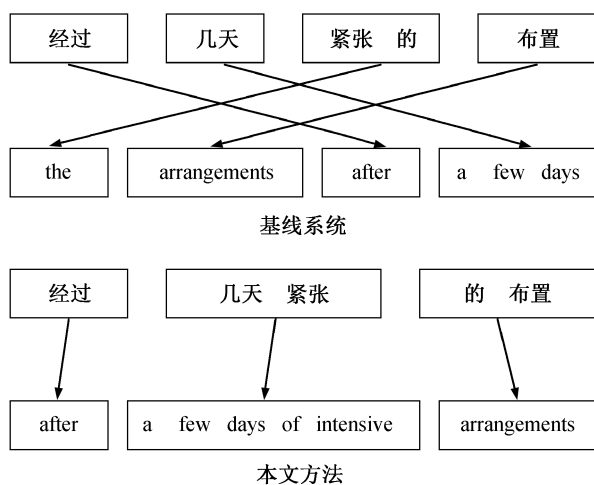


图 5 基线系统与本文方法翻译实例  
Fig. 5 Translation results of baseline system and the proposed methods

days”。在这个翻译结果中, 由于选择短语的不准确, 源语言中“紧张”没有被翻译出来。在本文提出的方法中, 由于生成实例短语〈几天 紧张, a few days of intensive〉, 因此在翻译的过程中, 仅仅需要顺序组合, 便可生成优质的翻译结果“after a few days of intensive arrangements”。

## 4 总结与展望

本文在双语语料规模受限的条件下, 通过对传统方法生成的短语对进行分解、替换、生成操作, 最终生成高质量的实例短语对, 达到扩充现有短语对的规模、提高基于短语的统计机器翻译系统翻译质量的目的。在汉英新闻和口语翻译任务上, 与基线系统相比, 本文方法在多个测试集上明显提高了翻译系统的翻译质量, 在部分测试集上 BLEU 值可提高 1% 左右。

未来, 本文提出的方法将被应用到 Chiang<sup>[5]</sup>提出的基于层次短语的统计机器翻译模型中, 对层次短语规则过强的泛化能力进行约束, 以达到准确匹配的目的, 继而提高翻译系统的翻译质量。

## 参考文献

- [1] Koehn P, Och F J, Marcu D. Statistical phrase-based translation // Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1. Edmonton: Association for Computational Linguistics, 2003: 48-54
- [2] Och F J, Ney H. The alignment template approach to statistical machine translation. Computational Linguistics, 2004, 30(4): 417-449
- [3] Och F J, Tillmann C, Ney H. Improved alignment

- models for statistical machine translation // Proceedings of the 1999 Conference on Empirical Methods in Natural Language Processing. Maryland: Association for Computational Linguistics, 1999: 20–28
- [4] Nagao M. A framework of a mechanical translation between Japanese and English by analogy principle // Elithorn A, Banerji R. Artificial and human intelligence. Amsterdam: NATO Publications, 1984: 173–180
- [5] Chiang D. Hierarchical phrase-based translation. Computational Linguistics, 2007, 33(2): 201–228
- [6] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space // International Conference on Learning Representations Workshop. Scottsdale: arXiv preprint arXiv, 2013: 1–12
- [7] Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality // Proceedings of NIPS. Lake Tahoe: arXiv preprint arXiv, 2013: 3111–3119
- [8] Mikolov T, Yih W, Zweig G. Linguistic regularities in continuous space word representations // Proceedings of NAACL HLT. Atlanta: Association for Computational Linguistics, 2013: 746–751
- [9] Och F J, Ney H. Discriminative training and maximum entropy models for statistical machine translation // Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Philadelphia: Association for Computational Linguistics, 2002: 295–302
- [10] Xiao T, Zhu J, Zhang H, et al. NiuTrans: an open source toolkit for phrase-based and syntax-based machine translation // Proceedings of the ACL 2012 System Demonstrations. Jeju: Association for Computational Linguistics, 2012: 19–24
- [11] Och F J, Ney H. Improved statistical alignment models // Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. Hong Kong: Association for Computational Linguistics, 2000: 440–447
- [12] Och F J. Minimum error rate training in statistical machine translation // Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1. Sapporo: Association for Computational Linguistics, 2003: 160–167
- [13] Papineni K, Roukos S, Ward T, et al. BLEU: a method for automatic evaluation of machine translation // Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Philadelphia: Association for Computational Linguistics, 2002: 311–318