Improving Statistical Machine Translation with Monolingual Collocation

Zhanyi Liu[1], Haifeng Wang[2], Hua Wu[2], Sheng Li[1]
[1]Harbin Institute of Technology, Harbin, China
[2]Baidu.com Inc., Beijing, China
zhanyiliu@gmail.com
{wanghaifeng, wu_hua}@baidu.com
lisheng@hit.edu.cn

Abstract

This paper proposes to use monolingual collocations to improve Statistical Machine Translation (SMT). We make use of the collocation probabilities, which are estimated from monolingual corpora, in two aspects, namely improving word alignment for various kinds of SMT systems and improving phrase table for phrase-based SMT. The experimental results show that our method improves the performance of both word alignment and translation quality significantly. As compared to baseline systems, we achieve absolute improvements of 2.40 BLEU score on a phrase-based SMT system and 1.76 BLEU score on a parsing-based SMT system.

1 Introduction

Statistical bilingual word alignment (Brown et al. 1993) is the base of most SMT systems. As compared to single-word alignment, multi-word alignment is more difficult to be identified. Although many methods were proposed to improve the quality of word alignments (Wu, 1997; Och and Ney, 2000; Marcu and Wong, 2002; Cherry and Lin, 2003; Liu et al., 2005; Huang, 2009), the correlation of the words in multi-word alignments is not fully considered.

In phrase-based SMT (Koehn et al., 2003), the phrase boundary is usually determined based on the bi-directional word alignments. But as far as we know, few previous studies exploit the collocation relations of the words in a phrase. Some researches used soft syntactic constraints to predict whether source phrase can be translated together (Marton and Resnik, 2008; Xiong et al., 2009). However, the constraints were learned from the parsed corpus, which is not available for many languages.

In this paper, we propose to use monolingual collocations to improve SMT. We first identify potentially collocated words and estimate collocation probabilities from monolingual corpora using a Monolingual Word Alignment (MWA) method (Liu et al., 2009), which does not need any additional resource or linguistic preprocessing, and which outperforms previous methods on the same experimental data. Then the collocation information is employed to improve Bilingual Word Alignment (BWA) for various kinds of SMT systems and to improve phrase table for phrase-based SMT.

To improve BWA, we re-estimate the alignment probabilities by using the collocation probabilities of words in the same cept. A cept is the set of source words that are connected to the same target word (Brown et al., 1993). An alignment between a source multi-word cept and a target word is a many-to-one multi-word alignment.

To improve phrase table, we calculate phrase collocation probabilities based on word collocation probabilities. Then the phrase collocation probabilities are used as additional features in phrase-based SMT systems.

The evaluation results show that the proposed method in this paper significantly improves multi-word alignment, achieving an absolute error rate reduction of 29%. The alignment improvement results in an improvement of 2.16 BLEU score on phrase-based SMT system and an improvement of 1.76 BLEU score on parsing-based SMT system. If we use phrase collocation probabilities as additional features, the phrase-based

This work was partially done at Toshiba (China) Research and Development Center.

Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 825–833,
Uppsala, Sweden, 11-16 July 2010.

SMT performance is further improved by 0.24 BLEU score.

The paper is organized as follows: In section 2, we introduce the collocation model based on the MWA method. In section 3 and 4, we show how to improve the BWA method and the phrase table using collocation models respectively. We describe the experimental results in section 5, 6 and 7. Lastly, we conclude in section 8.

2 Collocation Model

Collocation is generally defined as a group of words that occur together more often than by chance (McKeown and Radev, 2000). A collocation is composed of two words occurring as either a consecutive word sequence or an interrupted word sequence in sentences, such as "by accident" or "take ... advice". In this paper, we use the MWA method (Liu et al., 2009) for collocation extraction. This method adapts the bilingual word alignment algorithm to monolingual scenario to extract collocations only from monolingual corpora. And the experimental results in (Liu et al., 2009) showed that this method achieved higher precision and recall than previous methods on the same experimental data.

2.1 Monolingual word alignment

The monolingual corpus is first replicated to generate a parallel corpus, where each sentence pair consists of two identical sentences in the same language. Then the monolingual word alignment algorithm is employed to align the potentially collocated words in the monolingual sentences.

According to Liu et al. (2009), we employ the MWA Model 3 (corresponding to IBM Model 3) to calculate the probability of the monolingual word alignment sequence, as shown in Eq. (1).

$$p_{\mathrm{MWAModel3}}(S, A \mid S) \propto \prod_{i=1}^{l} n(\phi_i \mid w_i) \cdot \prod_{j=1}^{l} t(w_j \mid w_{a_j}) \cdot d(j \mid a_j, l) \quad (1)$$

Where $S = w_1^l$ is a monolingual sentence, ϕ_i denotes the number of words that are aligned with w_i. Since a word never collocates with itself, the alignment set is denoted as $A = \{(i, a_i) \mid i \in [1, l] \,\&\, a_i \neq i\}$. Three kinds of probabilities are involved in this model: word collocation probability $t(w_j \mid w_{a_j})$, position collocation probability $d(j \mid a_j, l)$ and fertility probability $n(\phi_i \mid w_i)$.

In the MWA method, the similar algorithm to bilingual word alignment is used to estimate the parameters of the models, except that a word cannot be aligned to itself.

Figure 1 shows an example of the potentially collocated word pairs aligned by the MWA method.

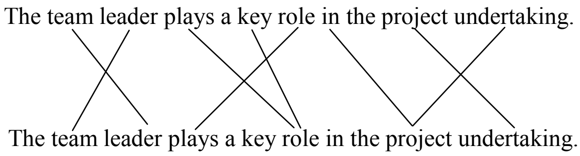

Figure 1. MWA Example

2.2 Collocation probability

Given the monolingual word aligned corpus, we calculate the frequency of two words aligned in the corpus, denoted as $freq(w_i, w_j)$. We filtered the aligned words occurring only once. Then the probability for each aligned word pair is estimated as follows:

$$p(w_i \mid w_j) = \frac{freq(w_i, w_j)}{\sum_{w'} freq(w', w_j)} \quad (2)$$

$$p(w_j \mid w_i) = \frac{freq(w_i, w_j)}{\sum_{w'} freq(w_i, w')} \quad (3)$$

In this paper, the words of collocation are symmetric and we do not determine which word is the head and which word is the modifier. Thus, the collocation probability of two words is defined as the average of both probabilities, as in Eq. (4).

$$r(w_i, w_j) = \frac{p(w_i \mid w_j) + p(w_j \mid w_i)}{2} \quad (4)$$

If we have multiple monolingual corpora to estimate the collocation probabilities, we interpolate the probabilities as shown in Eq. (5).

$$r(w_i, w_j) = \sum_k \alpha_k r_k(w_i, w_j) \quad (5)$$

α_k denotes the interpolation coefficient for the probabilities estimated on the k^{th} corpus.

3 Improving Statistical Bilingual Word Alignment

We use the collocation information to improve both one-directional and bi-directional bilingual word alignments. The alignment probabilities are re-estimated by using the collocation probabilities of words in the same cept.

3.1 Improving one-directional bilingual word alignment

According to the BWA method, given a bilingual sentence pair $E = e_1^l$ and $F = f_1^m$, the optimal alignment sequence A between E and F can be obtained as in Eq. (6).

$$A^* = \arg\max_A p(F, A \mid E) \quad (6)$$

The method is implemented in a series of five models (IBM Models). IBM Model 1 only employs the word translation model to calculate the probabilities of alignments. In IBM Model 2, both the word translation model and position distribution model are used. IBM Model 3, 4 and 5 consider the fertility model in addition to the word translation model and position distribution model. And these three models are similar, except for the word distortion models.

One-to-one and many-to-one alignments could be produced by using IBM models. Although the fertility model is used to restrict the number of source words in a cept and the position distortion model is used to describe the correlation of the positions of the source words, the quality of many-to-one alignments is lower than that of one-to-one alignments.

Intuitively, the probability of the source words aligned to a target word is not only related to the fertility ability and their relative positions, but also related to lexical tokens of words, such as common phrase or idiom. In this paper, we use the collocation probability of the source words in a cept to measure their correlation strength. Given source words $\{f_j \mid a_j = i\}$ aligned to e_i, their collocation probability is calculated as in Eq. (7).

$$r(\{f_j \mid a_j = i\}) = \frac{2\sum_{k=1}^{\phi_i - 1}\sum_{g=k+1}^{\phi_i} r(f_{[i]k}, f_{[i]g})}{\phi_i * (\phi_i - 1)} \quad (7)$$

Here, $f_{[i]k}$ and $f_{[i]g}$ denote the k^{th} word and g^{th} word in $\{f_j \mid a_j = i\}$; $r(f_{[i]k}, f_{[i]g})$ denotes the collocation probability of $f_{[i]k}$ and $f_{[i]g}$, as shown in Eq. (4).

Thus, the collocation probability of the alignment sequence of a sentence pair can be calculated according to Eq. (8).

$$r(F, A \mid E) = \prod_{i=1}^{l} r(\{f_j \mid a_j = i\}) \quad (8)$$

Based on maximum entropy framework, we combine the collocation model and the BWA model to calculate the word alignment probability of a sentence pair, as shown in Eq. (9).

$$p_r(F, A \mid E) = \frac{\exp(\sum_i \lambda_i h_i(F, E, A))}{\sum_{A'} \exp(\sum_i \lambda_i h_i(F, E, A'))} \quad (9)$$

Here, $h_i(F, E, A)$ and λ_i denote features and feature weights, respectively. We use two features in this paper, namely alignment probabilities and collocation probabilities.

Thus, we obtain the decision rule:

$$A^* = \arg\max_A \{\sum_i \lambda_i h_i(F, E, A)\} \quad (10)$$

Based on the GIZA++ package[1], we implemented a tool for the improved BWA method. We first train IBM Model 4 and collocation model on bilingual corpus and monolingual corpus respectively. Then we employ the hill-climbing algorithm (Al-Onaizan et al., 1999) to search for the optimal alignment sequence of a given sentence pair, where the score of an alignment sequence is calculated as in Eq. (10).

We note that Eq. (8) only deals with many-to-one alignments, but the alignment sequence of a sentence pair also includes one-to-one alignments. To calculate the collocation probability of the alignment sequence, we should also consider the collocation probabilities of such one-to-one alignments. To solve this problem, we use the collocation probability of the whole source sentence, $r(F)$, as the collocation probability of one-word cept.

3.2 Improving bi-directional bilingual word alignments

In word alignment models implemented in GIZA++, only one-to-one and many-to-one word alignment links can be found. Thus, some multi-word units cannot be correctly aligned. The symmetrization method is used to effectively overcome this deficiency (Och and Ney, 2003). Bi-directional alignments are generally obtained from source-to-target alignments A_{s2t} and target-to-source alignments A_{t2s}, using some heuristic rules (Koehn et al., 2005). This method ignores the correlation of the words in the same alignment unit, so an alignment may include many unrelated words[2], which influences the performances of SMT systems.

[1] http://www.fjoch.com/GIZA++.html

[2] In our experiments, a multi-word unit may include up to 40 words.

In order to solve the above problem, we incorporate the collocation probabilities into the bi-directional word alignment process.

Given alignment sets A_{s2t} and A_{t2s}. We can obtain the union $A_{s\leftrightarrow t} = A_{s2t} \cup A_{t2s}$. The source sentence f_1^m can be segmented into m' cepts $\bar{f}_1^{m'}$. The target sentence e_1^l can also be segmented into l' cepts $\bar{e}_1^{l'}$. The words in the same cept can be a consecutive word sequence or an interrupted word sequence.

Finally, the optimal alignments $\bar{A}$ between $\bar{f}_1^{m'}$ and $\bar{e}_1^{l'}$ can be obtained from $A_{s\leftrightarrow t}$ using the following decision rule.

$$(\bar{e}_1^{l'}, \bar{f}_1^{m'}, \bar{A})^* = \underset{\bar{A} \subseteq A_{s\leftrightarrow t}}{\arg\max} \{ \prod_{(\bar{e}_i, \bar{f}_j) \in \bar{A}} p(\bar{e}_i, \bar{f}_j)^{\tau_1} \cdot r(\bar{e}_i)^{\tau_2} \cdot r(\bar{f}_j)^{\tau_3} \} \quad (11)$$

Here, $r(\bar{f}_j)$ and $r(\bar{e}_i)$ denote the collocation probabilities of the words in the source language and target language respectively, which are calculated by using Eq. (7). $p(\bar{e}_i, \bar{f}_j)$ denotes the word translation probability that is calculated according to Eq. (12). τ_i denotes the weights of these probabilities.

$$p(\bar{e}_i, \bar{f}_j) = \frac{\sum_{e \in \bar{e}_i} \sum_{f \in \bar{f}_j} (p(e|f) + p(f|e))/2}{|\bar{e}_i| * |\bar{f}_j|} \quad (12)$$

$p(e|f)$ and $p(f|e)$ are the source-to-target and target-to-source translation probabilities trained from the word aligned bilingual corpus.

4 Improving Phrase Table

Phrase-based SMT system automatically extracts bilingual phrase pairs from the word aligned bilingual corpus. In such a system, an idiomatic expression may be split into several fragments, and the phrases may include irrelevant words. In this paper, we use the collocation probability to measure the possibility of words composing a phrase.

For each bilingual phrase pair automatically extracted from word aligned corpus, we calculate the collocation probabilities of source phrase and target phrase respectively, according to Eq. (13).

$$r(w_1^n) = \frac{2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} r(w_i, w_j)}{n*(n-1)} \quad (13)$$

Here, w_1^n denotes a phrase with n words; $r(w_i, w_j)$ denotes the collocation probability of a word pair calculated according to Eq. (4). For the phrase only including one word, we set a fixed collocation probability that is the average of the collocation probabilities of the sentences on a development set. These collocation probabilities are incorporated into the phrase-based SMT system as features.

Corpora	Chinese words	English words
Bilingual corpus	6.3M	8.5M
Additional monolingual corpora	312M	203M

Table 1. Statistics of training data

5 Experiments on Word Alignment

5.1 Experimental settings

We use a bilingual corpus, FBIS (LDC2003E14), to train the IBM models. To train the collocation models, besides the monolingual parts of FBIS, we also employ some other larger Chinese and English monolingual corpora, namely, Chinese Gigaword (LDC2007T38), English Gigaword (LDC2007T07), UN corpus (LDC2004E12), Sinorama corpus (LDC2005T10), as shown in Table 1.

Using these corpora, we got three kinds of collocation models:

CM-1: the training data is the additional monolingual corpora;
CM-2: the training data is either side of the bilingual corpus;
CM-3: the interpolation of CM-1 and CM-2.

To investigate the quality of the generated word alignments, we randomly selected a subset from the bilingual corpus as test set, including 500 sentence pairs. Then word alignments in the subset were manually labeled, referring to the guideline of the Chinese-to-English alignment (LDC2006E93), but we made some modifications for the guideline. For example, if a preposition appears after a verb as a phrase aligned to one single word in the corresponding sentence, then they are glued together.

There are several different evaluation metrics for word alignment (Ahrenberg et al., 2000). We use precision (P), recall (R) and alignment error ratio (AER), which are similar to those in Och and Ney (2000), except that we consider each alignment as a sure link.

Experiments		Single word alignments			Multi-word alignments		
		P	R	AER	P	R	AER
Baseline		0.77	0.45	0.43	0.23	0.71	0.65
Improved BWA methods	CM-1	0.70	0.50	0.42	0.35	0.86	0.50
	CM-2	0.73	0.48	0.42	0.36	0.89	0.49
	CM-3	0.73	0.48	0.41	0.39	0.78	0.47

Table 2. English-to-Chinese word alignment results

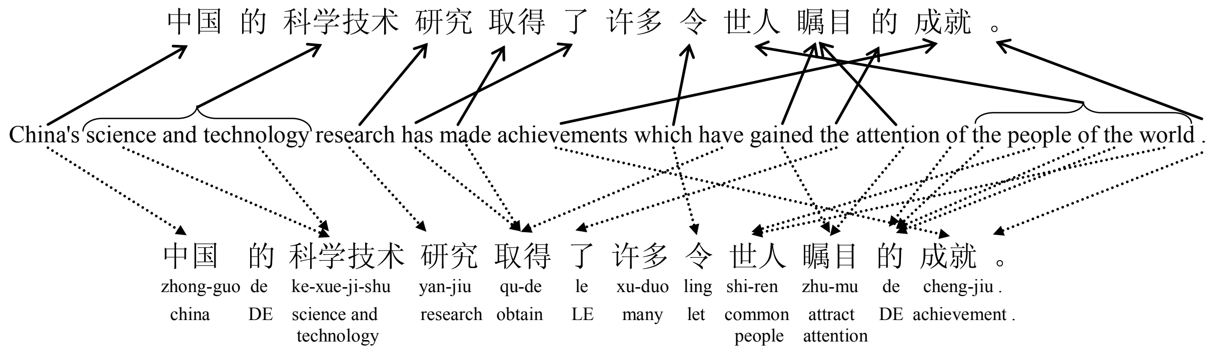

Figure 2. Example of the English-to-Chinese word alignments generated by the BWA method and the improved BWA method using CM-3. "⟶" denotes the alignments of our method; "⋯▸" denotes the alignments of the baseline method.

$$P = \frac{|S_g \cap S_r|}{|S_g|} \tag{14}$$

$$R = \frac{|S_g \cap S_r|}{|S_r|} \tag{15}$$

$$AER = 1 - \frac{2*|S_g \cap S_r|}{|S_g| + |S_r|} \tag{16}$$

Where, S_g and S_r denote the automatically generated alignments and the reference alignments.

In order to tune the interpolation coefficients in Eq. (5) and the weights of the probabilities in Eq. (11), we also manually labeled a development set including 100 sentence pairs, in the same manner as the test set. By minimizing the AER on the development set, the interpolation coefficients of the collocation probabilities on CM-1 and CM-2 were set to 0.1 and 0.9. And the weights of probabilities were set as $\tau_1 = 0.6$, $\tau_2 = 0.2$ and $\tau_3 = 0.2$.

5.2 Evaluation results

One-directional alignment results

To train a Chinese-to-English SMT system, we need to perform both Chinese-to-English and English-to-Chinese word alignment. We only evaluate the English-to-Chinese word alignment here. GIZA++ with the default settings is used as the baseline method. The evaluation results in Table 2 indicate that the performances of our methods on single word alignments are close to that of the baseline method. For multi-word alignments, our methods significantly outperform the baseline method in terms of both precision and recall, achieving up to 18% absolute error rate reduction.

Although the size of the bilingual corpus is much smaller than that of additional monolingual corpora, our methods using CM-1 and CM-2 achieve comparable performances. It is because CM-2 and the BWA model are derived from the same resource. By interpolating CM1 and CM2, i.e. CM-3, the error rate of multi-word alignment results is further reduced.

Figure 2 shows an example of word alignment results generated by the baseline method and the improved method using CM-3. In this example, our method successfully identifies many-to-one alignments such as "the people of the world ⟶ 世人". In our collocation model, the collocation probability of "the people of the world" is much higher than that of "people world". And our method is also effective to prevent the unrelated

Experiments		Single word alignments			Multi-word alignments			All alignments		
		P	R	AER	P	R	AER	P	R	AER
Baseline		0.84	0.43	0.42	0.18	0.74	0.70	0.52	0.45	0.51
Our methods	WA-1	0.80	0.51	0.37	0.30	0.89	0.55	0.58	0.51	0.45
	WA-2	0.81	0.50	0.37	0.33	0.81	0.52	0.62	0.50	0.44
	WA-3	0.78	0.56	0.34	0.44	0.88	0.41	0.63	0.54	0.40

Table 3. Bi-directional word alignment results

words from being aligned. For example, in the baseline alignment "has made ... have ⇢取得", "have" and "has" are unrelated to the target word, while our method only generated "made → 取得", this is because that the collocation probabilities of "has/have" and "made" are much lower than that of the whole source sentence.

Bi-directional alignment results

We build a bi-directional alignment baseline in two steps: (1) GIZA++ is used to obtain the source-to-target and target-to-source alignments; (2) the bi-directional alignments are generated by using "grow-diag-final". We use the methods proposed in section 3 to replace the corresponding steps in the baseline method. We evaluate three methods:

WA-1: one-directional alignment method proposed in section 3.1 and grow-diag-final;
WA-2: GIZA++ and the bi-directional bilingual word alignments method proposed in section 3.2;
WA-3: both methods proposed in section 3.

Here, CM-3 is used in our methods. The results are shown in Table 3.

We can see that WA-1 achieves lower alignment error rate as compared to the baseline method, since the performance of the improved one-directional alignment method is better than that of GIZA++. This result indicates that improving one-directional word alignment results in bi-directional word alignment improvement.

The results also show that the AER of WA-2 is lower than that of the baseline. This is because the proposed bi-directional alignment method can effectively recognize the correct alignments from the alignment union, by leveraging collocation probabilities of the words in the same cept.

Our method using both methods proposed in section 3 produces the best alignment performance, achieving 11% absolute error rate reduction.

Experiments			BLEU (%)
Baseline			29.62
Our methods	WA-1	CM-1	30.85
		CM-2	31.28
		CM-3	31.48
	WA-2	CM-1	31.00
		CM-2	31.33
		CM-3	31.51
	WA-3	CM-1	31.43
		CM-2	31.62
		CM-3	31.78

Table 4. Performances of Moses using the different bi-directional word alignments (Significantly better than baseline with $p < 0.01$)

6 Experiments on Phrase-Based SMT

6.1 Experimental settings

We use FBIS corpus to train the Chinese-to-English SMT systems. Moses (Koehn et al., 2007) is used as the baseline phrase-based SMT system. We use SRI language modeling toolkit (Stolcke, 2002) to train a 5-gram language model on the English sentences of FBIS corpus. We used the NIST MT-2002 set as the development set and the NIST MT-2004 test set as the test set. And Koehn's implementation of minimum error rate training (Och, 2003) is used to tune the feature weights on the development set.

We use BLEU (Papineni et al., 2002) as evaluation metrics. We also calculate the statistical significance differences between our methods and the baseline method by using paired bootstrap re-sample method (Koehn, 2004).

6.2 Effect of improved word alignment on phrase-based SMT

We investigate the effectiveness of the improved word alignments on the phrase-based SMT system. The bi-directional alignments are obtained

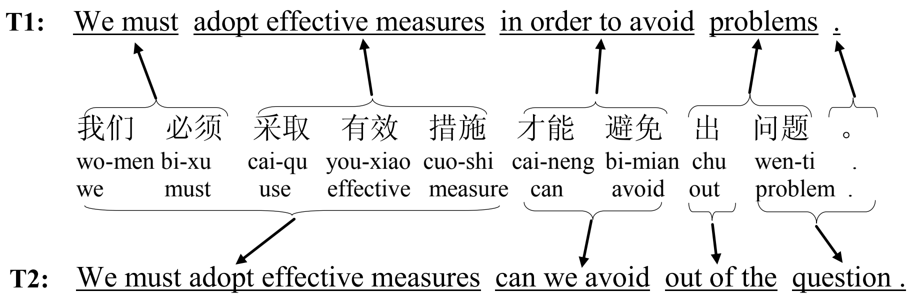

Figure 3. Example of the translations generated by the baseline system and the system where the phrase collocation probabilities are added

Experiments	BLEU (%)
Moses	29.62
+ Phrase collocation probability	30.47
+ Improved word alignments + Phrase collocation probability	32.02

Table 5. Performances of Moses employing our proposed methods (Significantly better than baseline with $p < 0.01$)

using the same methods as those shown in Table 3. Here, we investigate three different collocation models for translation quality improvement. The results are shown in Table 4.

From the results of Table 4, it can be seen that the systems using the improved bi-directional alignments achieve higher quality of translation than the baseline system. If the same alignment method is used, the systems using CM-3 got the highest BLEU scores. And if the same collocation model is used, the systems using WA-3 achieved the higher scores. These results are consistent with the evaluations of word alignments as shown in Tables 2 and 3.

6.3 Effect of phrase collocation probabilities

To investigate the effectiveness of the method proposed in section 4, we only use the collocation model CM-3 as described in section 5.1. The results are shown in Table 5. When the phrase collocation probabilities are incorporated into the SMT system, the translation quality is improved, achieving an absolute improvement of 0.85 BLEU score. This result indicates that the collocation probabilities of phrases are useful in determining the boundary of phrase and predicting whether phrases should be translated together, which helps to improve the phrase-based SMT performance.

Figure 3 shows an example: T1 is generated by the system where the phrase collocation probabilities are used and T2 is generated by the baseline system. In this example, since the collocation probability of "出 问题" is much higher than that of "问题 。", our method tends to split "出 问题 。" into "(出 问题) (。)", rather than "(出) (问题 。)". For the phrase "才能 避免" in the source sentence, the collocation probability of the translation "in order to avoid" is higher than that of the translation "can we avoid". Thus, our method selects the former as the translation. Although the phrase "我们 必须 采取 有效 措施" in the source sentence has the same translation "We must adopt effective measures", our method splits this phrase into two parts "我们 必须" and "采取 有效 措施", because two parts have higher collocation probabilities than the whole phrase.

We also investigate the performance of the system employing both the word alignment improvement and phrase table improvement methods. From the results in Table 5, it can be seen that the quality of translation is future improved. As compared with the baseline system, an absolute improvement of 2.40 BLEU score is achieved. And this result is also better than the results shown in Table 4.

7 Experiments on Parsing-Based SMT

We also investigate the effectiveness of the improved word alignments on the parsing-based SMT system, Joshua (Li et al., 2009). In this system, the Hiero-style SCFG model is used (Chiang, 2007), without syntactic information. The rules are extracted only based on the FBIS corpus, where words are aligned by "MW-3 & CM-3". And the language model is the same as that in Moses. The feature weights are tuned on the development set using the minimum error

Experiments	BLEU (%)
Joshua	30.05
+ Improved word alignments	31.81

Table 6. Performances of Joshua using the different word alignments (Significantly better than baseline with $p < 0.01$)

rate training method. We use the same evaluation measure as described in section 6.1.

The translation results on Joshua are shown in Table 6. The system using the improved word alignments achieves an absolute improvement of 1.76 BLEU score, which indicates that the improvements of word alignments are also effective to improve the performance of the parsing-based SMT systems.

8 Conclusion

We presented a novel method to use monolingual collocations to improve SMT. We first used the MWA method to identify potentially collocated words and estimate collocation probabilities only from monolingual corpora, no additional resource or linguistic preprocessing is needed. Then the collocation information was employed to improve BWA for various kinds of SMT systems and to improve phrase table for phrase-based SMT.

To improve BWA, we re-estimate the alignment probabilities by using the collocation probabilities of words in the same cept. To improve phrase table, we calculate phrase collocation probabilities based on word collocation probabilities. Then the phrase collocation probabilities are used as additional features in phrase-based SMT systems.

The evaluation results showed that the proposed method significantly improved word alignment, achieving an absolute error rate reduction of 29% on multi-word alignment. The improved word alignment results in an improvement of 2.16 BLEU score on a phrase-based SMT system and an improvement of 1.76 BLEU score on a parsing-based SMT system. When we also used phrase collocation probabilities as additional features, the phrase-based SMT performance is finally improved by 2.40 BLEU score as compared with the baseline system.

Reference

Lars Ahrenberg, Magnus Merkel, Anna Sagvall Hein, and Jorg Tiedemann. 2000. Evaluation of Word Alignment Systems. In *Proceedings of the Second International Conference on Language Resources and Evaluation*, pp. 1255-1261.

Yaser Al-Onaizan, Jan Curin, Michael Jahr, Kevin Knight, John Lafferty, Dan Melamed, Franz-Josef Och, David Purdy, Noah A. Smith, and David Yarowsky. 1999. Statistical Machine Translation. Final Report. In *Johns Hopkins University Workshop*.

Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, and Robert. L. Mercer. 1993. The Mathematics of Statistical Machine Translation: Parameter estimation. *Computational Linguistics*, 19(2): 263-311.

Colin Cherry and Dekang Lin. 2003. A Probability Model to Improve Word Alignment. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pp. 88-95.

David Chiang. 2007. Hierarchical Phrase-Based Translation. *Computational Linguistics*, 33(2): 201-228.

Fei Huang. 2009. Confidence Measure for Word Alignment. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP*, pp. 932-940.

Philipp Koehn. 2004. Statistical Significance Tests for Machine Translation Evaluation. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 388-395.

Philipp Koehn, Amittai Axelrod, Alexandra Birch Mayne, Chris Callison-Burch, Miles Osborne, and David Talbot. 2005. Edinburgh System Description for the 2005 IWSLT Speech Translation Evaluation. In *Processings of the International Workshop on Spoken Language Translation 2005*.

Philipp Koehn, Franz J. Och, and Daniel Marcu. 2003. Statistical Phrase-based Translation. In *Proceedings of the Human Language Technology Conference and the North American Association for Computational Linguistics*, pp. 127-133.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the ACL, Poster and Demonstration Sessions*, pp. 177-180.

Zhifei Li, Chris Callison-Burch, Chris Dyer, Juri Ganitkevitch, Sanjeev Khudanpur, Lane Schwartz, Wren Thornton, Jonathan Weese, and Omar Zaidan. 2009. Demonstration of Joshua: An Open Source Toolkit for Parsing-based Machine Translation. In *Proceedings of the 47th Annual Meeting of the As-*

sociation for Computational Linguistics, Software Demonstrations, pp. 25-28.

Yang Liu, Qun Liu, and Shouxun Lin. Log-linear Models for Word Alignment. 2005. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 459-466.

Zhanyi Liu, Haifeng Wang, Hua Wu, and Sheng Li. 2009. Collocation Extraction Using Monolingual Word Alignment Method. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 487-495.

Daniel Marcu and William Wong. 2002. A Phrase-Based, Joint Probability Model for Statistical Machine Translation. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, pp. 133-139.

Yuval Marton and Philip Resnik. 2008. Soft Syntactic Constraints for Hierarchical Phrase-Based Translation. In *Proceedings of the 46st Annual Meeting of the Association for Computational Linguistics*, pp. 1003-1011.

Kathleen R. McKeown and Dragomir R. Radev. 2000. Collocations. In Robert Dale, Hermann Moisl, and Harold Somers (Ed.), *A Handbook of Natural Language Processing*, pp. 507-523.

Franz Josef Och and Hermann Ney. 2000. Improved Statistical Alignment Models. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pp. 440-447.

Franz Josef Och. 2003. Minimum Error Rate Training in Statistical Machine Translation. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pp. 160-167.

Franz Josef Och and Hermann Ney. 2003. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics*, 29(1): 19-52.

Kishore Papineni, Salim Roukos, Todd Ward, and Weijing Zhu. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of 40th annual meeting of the Association for Computational Linguistics*, pp. 311-318.

Andreas Stolcke. 2002. SRILM - An Extensible Language Modeling Toolkit. In *Proceedings for the International Conference on Spoken Language Processing*, pp. 901-904.

Dekai Wu. 1997. Stochastic Inversion Transduction Grammars and Bilingual Parsing of Parallel Corpora. *Computational Linguistics*, 23(3): 377-403.

Deyi Xiong, Min Zhang, Aiti Aw, and Haizhou Li. 2009. A Syntax-Driven Bracketing Model for Phrase-Based Translation. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP*, pp. 315-323.