Improving Statistical Word Alignment with Various Clues

Dengjun Ren, Hua Wu, Haifeng Wang

Toshiba (China) Research and Development Center
5/F., Tower W2, Oriental Plaza
No.1, East Chang An Ave., Dong Cheng District
Beijing, 100738, China
{rendengjun, wuhua, wanghaifeng}@rdc.toshiba.com.cn

Abstract

This paper proposes a method to improve word alignment by combining various clues. Our method first trains a baseline statistical IBM word alignment model. Then we improve it with various clues, which are mainly based on features such as lemmatization, translation dictionary, named entities, and chunks. We incorporate these features into an unified framework. Experimental results show that our method improves word alignment quality by achieving a relative error rate reduction of 39.8%. We also conduct phrase-based machine translation based on the word alignment results. Using BLEU as an evaluation metric, our method achieves an absolute improvement of about 0.02 (about 18% relative) over a baseline method.

Introduction

Word alignment was first proposed as an intermediate result of statistical machine translation (Brown et al., 1993). In recent years, many researchers have employed statistical models (Wu, 1997; Och and Ney, 2003; Cherry and Lin, 2003; Zhang and Gildea, 2005) or association measures (Smadja et al., 1996; Ker and Chang, 1997; Ahrenberg et al., 1998; Tufis and Barbu, 2002) to build alignment links.

One of the main problems in the existing methods is the null alignment. In two different languages, some words in one language have no counterparts in the other. And such information is not available in bilingual dictionaries. In order to solve this problem, the basic IBM model (Brown et al., 1993) trained a probability for all null alignments without consideration of each individual word, which made the accuracy of the null alignment relatively low. Wu and Wang (2004) showed that the accuracy of word alignment decreased if the null alignment was considered in evaluation. Moore (2004) tried to improve null alignment by re-weighting it in the Expectation Maximization (EM) training procedure, but failed to improve word alignment results.

The second problem is the alignment of multi-word units such as phrasal compounds, idiomatic expressions, and complex terms. It is difficult to align them because it depends on the context. Wu and Wang (2004) showed that the accuracy of multi-word alignments is much lower than that of single-word alignments. Previous methods improved multi-word alignment by using either iterative procedures (Smadja et al., 1996; Melamed, 1997) or preprocessing steps for the identification of token N-grams (Ahrenberg et al., 1998; Tiedemann, 1999). Wu and Wang (2004) used a rule-based translation system to identify and disambiguate the multi-word units and improved the multi-word alignment results. Tiedemann (2003) used chunks and n-grams.

The third issue is how to make use of more linguistic information. The basic statistical word alignment method works on the word level of the plain text. In recent years, some discriminative methods are proposed to integrate various syntactic and lexical clues into the alignment models to improve alignment quality (Liu et al., 2005; Moore et al., 2006; Blunsom and Cohn, 2006; Taskar et al., 2005). In these methods, part-of-speech (POS), association measure between bilingual words, and translation dictionaries are usually used. More linguistic information, such as named entity and chunk information, may be useful for word alignment.

In this paper, we propose an unified method to address all of the three problems mentioned above. The method first trains a weight for each null alignment to improve word alignment. Secondly, we use dictionaries, including human crafted translation dictionaries and automatically trained dictionaries, to improve both precision and recall of word alignment. Finally, we use linguistic tools, including named entity recognizers and chunkers, to improve multi-word alignment. Using all of these clues, we propose a method to combine them to improve word alignment.

Experimental results show that our null alignment model can achieve an error rate reduction of 12.35% as compared with the baseline. And the dictionaries and linguistic features such as named entities and chunks can further improve the word alignment by achieving an error rate reduction of 39.8%. We also apply the aligned corpus for phrase-based statistical machine translation. Using BLEU as an evaluation metric, our method improves translation quality by achieving an absolute improvement of about 0.02 (18% relative) over the baseline method.

The remainder of this paper is organized as follows. The next section describes our method combining multiple clues to improve word alignment. And then we will show the experimental results on both word alignment and statistical machine translation. After that, we will compare our methods to some related work. In the last section, we will conclude this paper and present the future work.

Methodology

Och and Ney (2003) proved that the statistical word alignment models proposed by (Brown et al., 1993) outperform the heuristic methods based on the association measures. However, the statistical models still have some deficiencies. For example, the models use simple methods to handle null alignments, and cannot handle multi-word alignment and do not take context into account.

In this section, we use IBM model 4 (Brown et al., 1993) as a baseline, and use various clues to improve word alignment quality.

Definition

For convenience, we use the following definitions in this paper.

- **f** represents a source language sentence $f_1, f_2, ..., f_m$.
- **e** represents a target language sentence $e_1, e_2, ..., e_n$.
- The link (e_i, f_j) represents that e_i is aligned to f_j.
- $A_{f->e}$ is defined as the alignment set in the source to target direction produced by the IBM model 4.
- $A_{e->f}$ is defined as the alignment set in the target to source direction produced by the IBM model 4.
- $A_{\cap}$ is the intersection set of $A_{e->f}$ and $A_{f->e}$.
- $A_{\cup}$ is the union set of $A_{e->f}$ and $A_{f->e}$.

Morphological Analysis

The IBM models need bilingual corpus for training. Since a large bilingual corpus is not always available, it is subject to the problem of data sparseness. One possible way to solve this problem is to perform morphological analysis on the bilingual corpus. In this paper, we use the lemmatized form of English words in the bilingual corpus to perform statistical word alignment. For example, the lemmatized form of "verified" is "verify".

Based on the alignment results, we assign a weight to each alignment link as follows.

$$W_0(e_i, f_j) = \frac{p(e_i \mid f_j) + p(f_j \mid e_i)}{2} \quad (1)$$

Where $p(x \mid y)$ describes the translation probability obtained from the alignment results $A_{e->f}$ and $A_{f->e}$ trained with the lemmatized corpus.

Null Alignment Model

Some words in one language have no counterparts in the other. And such kind of information is not available in translation dictionaries. Although Brown et al. (1993) train a probability for all the null alignments, it does not condition the null translation probability on individual words.

In order to solve this problem, we estimate the confidence score for each null alignment link. The score is based on an association measure (Taskar et al., 2005), namely Dice coefficient, which is shown in Equations (2) and (3).

$$W_1(e, NULL) = \frac{2 * count(e, NULL)}{\sum_{f'} count(e, f') + \sum_{e'} count(e', NULL)} \quad (2)$$

$$W_1(NULL, f) = \frac{2 * count(NULL, f)}{\sum_{f'} count(NULL, f') + \sum_{e'} count(e', f)} \quad (3)$$

Where $count(e, f)$ is the occurring frequency of the word alignment link (e, f) in the union alignment set $A_{\cup}$.

Translation Dictionary

Handcraft Dictionary

For some language pairs, there exist handcraft translation dictionaries of high quality. In order to improve alignment accuracy, we use these dictionaries as a clue in this paper. For each entry in the dictionary, we also assign a weight for it, which is shown in (4).

$$W_2(e, f) = \frac{2}{|\{f' \mid (e, f') \in HD\}| + |\{e' \mid (e', f) \in HD\}|} \quad (4)$$

Where $|\{f' \mid (e, f') \in HD\}|$ and $|\{e' \mid (e', f) \in HD\}|$ describe the number of alternative translations in the handcraft dictionary HD for the word e and f, respectively.

Automatically Trained Dictionary

Although the handcraft translation dictionary has high quality translation, it cannot cover word or phrase translations in all kinds of specific domains. Thus, we also automatically train a translation dictionary from the alignment results obtained with IBM model 4. To build the translation dictionary, we first get the intersection set $A_{\cap}$. Then the alignment links in $A_{\cap}$ are extended by iteratively adding word alignment links from $A_{\cup}$ into it as described in (Och and Ney, 2003). Finally, to filter some noise caused by the error alignment links, we only retain those translation pairs whose translation probabilities are above a threshold or co-occurring frequencies are above a threshold.

We estimate a weight for each entry in this dictionary using the same method as described in Equation (2) or (3), which is rewritten as shown in (5).

$$W_3(e, f) = \frac{2 * count(e, f)}{\sum_{f'} count(e, f') + \sum_{e'} count(e', f)} \quad (5)$$

Named Entity

It is difficult to align named entities because of the following reasons. First, they often consist of several words, forming multi-word units. Second, it is difficult for existing dictionaries to contain all of them because they are dynamic words, which results in out-of-vocabulary (OOV) problem. Third, most of the named entities do not frequently occur in the corpus, which results in data sparseness problem.

Fortunately, for some languages, named entities reorganization tools are available. Meulder and Daelemans (2003) showed that their method can achieve a precision of 88.99% and a recall of 88.54% on English named entity recognition. Sun et al. (2002) showed that recognizers can achieve a precision of 82.28% and a recall of 85.53% on Chinese named entity recognition. In this paper, we use available tools to recognize the named entities in the source language and the target language. The types of the named entities include time, data, number, person names, organization, and locations.

Two named entities $N_{\mathsf{f}} = f_j, f_{j+1}, ..., f_{j+t}$ and $N_{\mathsf{e}} = e_i, e_{i+1}, ..., e_{i+s}$ are consistent if N_{e} and N_{f} belong to the same named entity type (such as persons) and one of the following conditions is satisfied:

- There is no other named entity in the sentence pair having the same type as that of N_{e} and N_{f}.
- For any non-null alignment link (e_x, f_y) in the alignment set A [1], $i \le x \le i + s$ if and only if $j \le y \le j + t$.

Given $e \in N_{\mathsf{e}}$, $f \in N_{\mathsf{f}}$, we assign the weight for the link (e, f) as follows:

$$W_4(e, f) = \begin{cases} 1 & \text{if } N_{\mathsf{e}} \text{ and } N_{\mathsf{f}} \text{ are consistent} \\ 0 & \text{Otherwise} \end{cases} \quad (6)$$

Chunk

In this paper, chunks are defined as phrases where syntactically related words become members of the same phrase. For some languages, shallow parsers are available to identify chunks with high precision. For example, Koeling (2000) showed that their method can achieve a precision of 93.45% and a recall of 93.51% on English chunking.

In this paper, we only use base chunks of the source language. If we use shallow parsers to identify the chunks in both source language and target language, it is difficult to align them because of structure divergence (Al-Adhaileh, 2002). Thus, we only use a shallow parser in one language (for examples, source language) to identify the chunk, and then obtain the corresponding chunk in another language based on the word alignments.

Two chunks $C_{\mathsf{f}} = f_j, f_{j+1}, ..., f_{j+t}$ and $C_{\mathsf{e}} = e_i, e_{i+1}, ..., e_{i+s}$ are consistent if the following condition is satisfied:

- For any non-null alignment link (e_x, f_y) in the alignment set A [2], $i \le x \le i + s$ if and only if $j \le y \le j + t$

Given $e \in C_{\mathsf{e}}$, $f \in C_{\mathsf{f}}$, we assign the weight for the link (e, f) as follows:

$$W_5(e, f) = \begin{cases} 1 & \text{if } C_{\mathsf{e}} \text{ and } C_{\mathsf{f}} \text{ are consistent} \\ 0 & \text{Otherwise} \end{cases} \quad (7)$$

[1] For the alignment set, we can use the intersection set $A_{\cap}$ or other alignment set with high precision. In this paper, we use the alignment results after applying the clue of translation dictionary. In our experimental result, it achieves higher alignment precision than other methods except the intersection method.

[2] We employ the same alignment set that is used to judge the consistency of two named entities.

Combination of Clues

With all of the above clues, we combine them to improve word alignment quality. First, we create a word alignment matrix for the bilingual sentence pair $(\mathbf{e}, \mathbf{f})$ as shown in Table 1.

	f_0	f_1	f_2	$\cdots$	f_m
e_0	--	c_{01}	c_{02}	$\cdots$	c_{0m}
e_1	c_{10}	c_{11}	c_{12}	$\cdots$	c_{1m}
e_2	c_{20}	c_{21}	c_{22}	$\cdots$	c_{2m}
…	…	…	…	…	…
e_n	c_{n0}	c_{n1}	c_{n2}	$\cdots$	c_{nm}

Table 1. The Alignment Matrix

In Table 1, c_{ij} describes the alignment association strength of e_i and f_j. It can be estimated as described in (8).

$$c_{ij} = \sum_{k=0}^{5} W_k(e_i, f_j) \quad (8)$$

Where $W_k(e_i, f_j)$ is the corresponding weight described in the above subsections.

With the alignment matrix, we use a best-first strategy to add the word alignment links. We first select the link with the highest score c_{ij} to the final word alignment set A, and then select the link with the second high score, and so on. The procedure is repeated until no alignment link can be added whose alignment score is above a fixed threshold[3].

Experimental Results on Word Alignment

Data

In this section, we take English-Chinese word alignment as a case study. The English-Chinese bilingual training data is provided by Chinese Linguistic Data Consortium (CLDC)[4]. The catalog number is CLDC-LAC-2003-004. It contains about 150,000 sentence pairs, with about 3 million English words and about 5 million Chinese characters.

The development set and the test set for word alignment are from the corpora distributed for the 2005 HTRDP evaluation of machine translation[5]. It can also be obtained from CLDC (catalog number 2005-863-001). The test set contains 505 sentence pairs, with 6,866 sure links and 4,106 possible links in the reference word alignment set.

Tools and Resources

For English, we choose the system OAK developed by

[3] This threshold will be tuned using a development set.

[4] http://www.chineseldc.org/EN/index.htm

[5] The full name of HTRDP is National High Technology Research and Development Program of China, also named as 863 program.

New York University[6]. It is used to perform tokenization, lemmatization, named entity recognition, and chunking for English sentences.

For Chinese, we choose LTP (Language Technology Platform), which is developed by the Information Retrieval Laboratory, Harbin Institute of Technology[7]. It is used to perform word segmentation, and named entity recognition.

In our experiments, we use OAK to recognize the English chunks, and then recognize Chinese chunks based on both the word alignment results and English chunks.

We also use a handcraft Chinese-English dictionary included in HowNet (Dong and Dong, 2006), which is a Chinese conceptual database[8]. This dictionary includes 55,462 entries. The handcraft English-Chinese dictionary is collected from various resources, comprising 64,234 entries.

Evaluation Metrics

The reference of the test data provided by CLDC includes possible links and sure links. So we use the same evaluation metrics as described in (Och and Ney, 2000). If we use A to indicate the alignments identified by the proposed methods, and S and P to denote the sure and possible links in the reference alignments, the precision, recall, and alignment error rate (AER) are calculated as described in Equations (9), (10) and (11).

$$precision = \frac{|\mathrm{A} \cap \mathrm{S}|}{|\mathrm{S}|} \tag{9}$$

$$recall = \frac{|\mathrm{A} \cap \mathrm{P}|}{|\mathrm{P}|} \tag{10}$$

$$AER = 1 - \frac{|\mathrm{A} \cap \mathrm{S}| + |\mathrm{A} \cap \mathrm{P}|}{|\mathrm{A}| + |\mathrm{S}|} \tag{11}$$

Results

With the data described above, we perform bi-directional (source to target and target to source) word alignment based on IBM model 4, and obtain two alignment results on the test set. Based on these two results, we get the "refined" combination as described in (Och and Ney, 2000). The result of the "refined" method is used as the baseline in this paper. The tool used to train the baseline model is the GIZA++ toolkit[9].

In our experiments, we first lemmatize the English sentences. With the lemmatized sentence pair, we use the GIZA++ toolkit to train the alignment models and get the "refined" results. This method is denoted as "Lemma".

Based on the statistical results on lemma, we add the clues one by one. The word alignment results are shown in Table 2. In the table, "Null", "Dic", "NER", and "Chunk" represent null alignment, translation dictionary, named entity recognition, and chunking, respectively.

From the results, it can be seen that word alignment is improved by using morphological analysis to get the lemmatized sentence pairs. It achieves a relative error rate reduction of 6.03% as compared with the baseline. The null alignment model further improves the word alignment by achieving a relative error rate reduction of 12.35%. Using translation dictionaries, the recall and precision are greatly improved because the dictionary built automatically can cover most of the alignment links and the handcraft dictionary with high quality can filter the links of $\mathrm{A}_{\cup}$. The method including the clues of the named entities and chunks further improves the alignment results, achieving a relative error reduction of 10.72% as compared with the method "Lemma+Null+Dic". The method combining all of the clues achieves a relative error rate reduction of 39.8% as compared with the baseline.

	Precision	Recall	AER
Baseline	0.6535	0.6881	0.3319
Lemma	0.6769	0.7033	0.3119
Lemma +Null	0.7067	0.7120	0.2909
Lemma +Null+Dic	0.7806	0.7718	0.2238
Lemma+Null +Dic+NER	0.7035	0.8861	0.2159
Lemma+Null+Dic +NER+Chunk	0.7246	0.8933	0.1998

Table 2. Word Alignment Results

In order to further analyze the experimental results, we classify the word alignment links into single-word links and multi-word links. The former includes the alignment links that have no multi-word units. The latter includes at least one multi-word unit in the alignment link. The results of the single-word links and multi-word links are shown in Tables 3 and 4.

From the results, it can be seen that it is more difficult to align multi-word units. After using the dictionaries, our method greatly improves both single-word alignment and multi-word alignment, achieving relative error rate reductions of 37.61% and 27.17% as compared with the baseline. With named entities and chunks, the alignment for multi-word units is further improved by 7.76% as compared with "Lemma+Null+Dic". By combining all the clues, multi-word alignment is greatly improved by achieving a relative error rate reduction of 32.82% as compared with the baseline. This indicates that our method is effective to align multi-word units.

	Precision	Recall	AER
Baseline	0.7308	0.6870	0.2917
Lemma+Null+Dic	0.7895	0.8486	0.1820
Lemma+Null+Dic +NER	0.7820	0.8623	0.1798
Lemma+Null+Dic +NER+Chunk	0.7842	0.8767	0.1720

Table 3. Single-Word Alignment Results

	Precision	Recall	AER
Baseline	0.5135	0.6491	0.4266
Lemma+Null+Dic	0.6480	0.7361	0.3107
Lemma+Null+Dic +NER	0.6572	0.7585	0.2957
Lemma+Null+Dic +NER+Chunk	0.6508	0.7893	0.2866

Table 4. Multi-Word Alignment Results

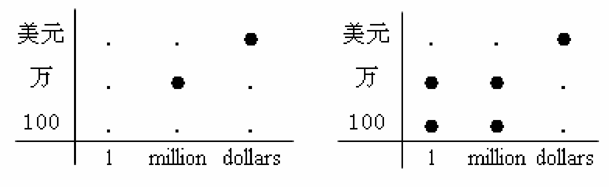
[6] http://nlp.cs.nyu.edu/oak
[7] http://ltp.ir-lab.org/Sharing_Plan.htm
[8] http://www.keenage.com/html/e_index.html
[9] http://www.fjoch.com/ GIZA++.html

Result Analysis by Using Examples

This section uses specific examples to illustrate the alignment improvement achieved by our method. The example for named entities is shown as follows:

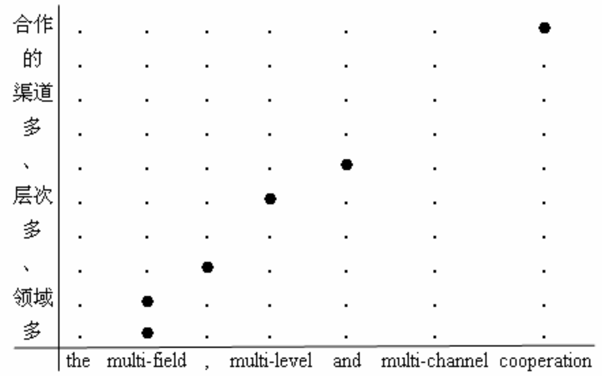
(1) 1 million dollars
    100 万 美元
    100 ten-thousand dollar

The alignment results of using the named entity clue are shown in Figure 1. Using the named entity clue, we first identify the two named entities "1 million dollars" and "100 万 美元" , which belong to the same named entity type "currency". And then we modify the alignment in Figure 1(a) to 1(b), which correctly aligns the number "1 million" to the Chinese words "100 万".

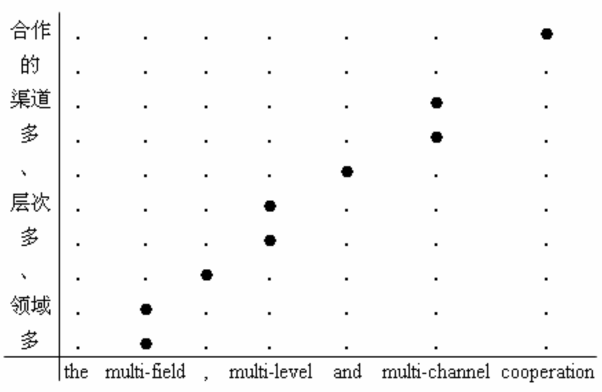


(a) Without NER (b) With NER

Figure 1. Alignment Examples w/o NER



(a) Without Chunks



(b) With Chunks

Figure 2. Alignment Examples w/o Chunks

Example (2) shows a sentence pair using the chunk clue.

(2) the multi-field, multi-level, and multi-channel cooperation
    多 领域 、多 层次 、多 渠道 的 合作
    multi field , multi level , multi channel of cooperation

The alignment results of using the chunk clue are shown in Figure 2. With the chunk clue, we can identify "the multi-field, multi-level and multi-channel cooperation" as one chunk. According to the alignment in Figure 2(a), we also identify the Chinese chunk "多 领域 、 多 层次 、 多 渠道 的 合作". With the two chunks and the alignment information, we can modify the alignment in Figure 2(a) to the alignment in Figure 2(b).

Translation Experiments

In this section, we perform English to Chinese translation to investigate whether the improved word alignment leads to better translation quality.

The training data is the same as that for word alignment. We use two kinds of testing data in this experiment. One is the test set used for word alignment evaluation. For each English sentence in this test set, there is only one reference translation. Here, we name it "WA Test Set". The other test set is from the corpora distributed for the 2005 HTRDP evaluation of machine translation, which is also distributed by CLDC, with catalog number of 2005-863-001. The test set contains 494 English sentences, with each sentence having four reference translations. It is named "MT Test Set".

The development set is also from the corpora distributed for the 2005 HTRDP evaluation of machine translation, which includes 278 sentences, with four reference translations for each source sentence.

We use the SRILM toolkit (Stolcke, 2002) to train a language model on the Chinese Gigaword Second Edition provided by LDC (catalog number LDC2005T14).

Translation Results

We conduct phrase-based statistical machine translation (SMT) from English to Chinese. To perform phrase-based SMT, we need a trainer and a decoder. For training, we use Koehn's training scripts[10]. For the decoder, we use Pharaoh (Koehn, 2004). We run the decoder with its default settings (maximum phrase length 7) and then use Koehn's implementation of minimum error rate training (Och, 2003) to tune the feature weights on the development set. The translation quality was evaluated using a well-established automatic measure: BLEU score (Papineni et al., 2002).

The translation results are shown in Table 5. In the translation experiments, our method combines all of the clues to get the alignment results. Based on the alignment results, we extract the phrase pairs used by the Pharaoh decoder.

	WA Test Set	MT Test Set
Baseline	0.1137	0.1426
Our Method	0.1346	0.1690

Table 5. English to Chinese Translation Results

From the results, it can be seen that our method outperforms the baseline on both of the test sets. Using BLEU as a metric, our method achieves an absolute improvement of 0.0209 (18.28% relative) and 0.0264 (18.51% relative) as compared with the baseline on the WA Test Set and MT Test Set, respectively.

[10] http://www.statmt.org/wmt06/shared-task/baseline.html

Comparison with Related Work

Many researchers have used morpho-syntactic information to improve performance of a phrase-based statistical machine translation. Popovic and Ney (2004 & 2005) used stem-suffix and lemma-POS to improve translation quality. Gupta and Federico (2006) compared lemma-based methods and stem-based methods on statistical machine translation, and found out that these two methods only outperform the word-based method when the training data is limited to less than 1 million words.

In this section, we will investigate the lemma-based and stem-based methods on our corpus. The stemming tool for English is Porter[11]. In order to examine the effect of the sizes of training corpus on both word alignment and translation, we randomly select 30k, 90k, and 150k sentence pairs (the entire training set) from our training corpus. The results on word alignment are shown in Figure 3. The translation results on the WA test set and MT test set are shown in Figure 4 and 5, respectively.

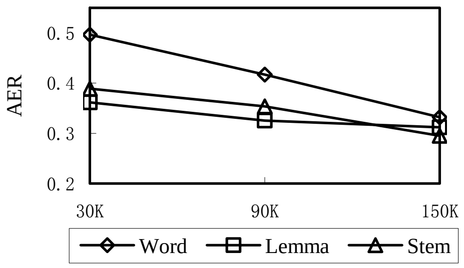


Figure 3. Word Alignment Results by Using Different Sizes of Training Corpus

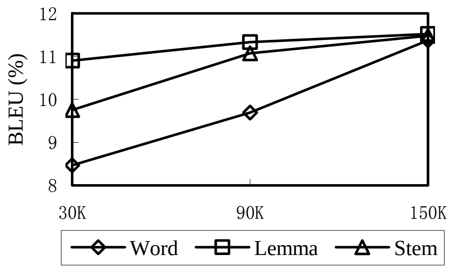


Figure 4. Translation Results on the WA Test Set by Using Different Sizes of Training Corpus

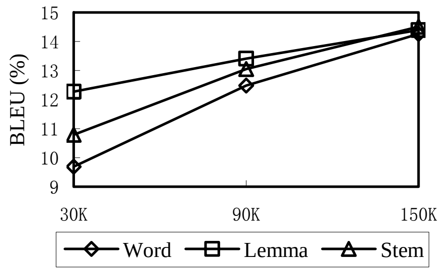


Figure 5. Translation Results on the MT Test Set by Using Different Sizes of Training Corpus

From the word alignment results, it can be seen that the lemma-based method and the stem-based method consistently reduce AER as compared with the word-based method. On smaller training corpus, lemma-based method outperforms the stem-based method.

From the translation results, it can be seen that the lemma-based method and the stem-based method outperform the word-based method when only smaller training corpus are available. Although the stem-based method and the lemma-based achieve lower AER, they do not achieve much improvement on translation quality when a larger training corpus is available. This result again confirms that large gains in alignment performance can achieve relatively small gains in translation performance (Lopez and Resnik, 2006).

In conclusion, lemma-based and stem-based methods are effective to alleviate the problem of data sparseness. This result is similar to that in (Gupta and Federico 2006). However, the lemma-based method outperforms the stem-based method on our corpus, which is different from that in (Gupta and Federico 2006). This may be caused by the different kind of language pairs used.

Conclusion and Future Work

This paper proposed a method to improve statistical word alignment by combining various clues. Our method first trained a baseline statistical IBM word alignment model and then improved it with different clues. The clues are mainly based on features such as lemmatization, translation dictionary, named entities, and chunks. We incorporated these features into the statistical alignment models. Experimental results showed that our method improved word alignment quality by achieving a relative error rate reduction of 39.8%. The results also indicated that our method combining translation dictionaries, named entity recognition, and chunks greatly improved the alignment of multi-word units.

We also conducted phrase-based statistical machine translation based on the word alignment results. Using BLEU as an evaluation metric, our method achieved an absolute improvement of about 0.02 (about 18% relative) over a baseline method for English to Chinese translation.

Further analysis indicates that our lemma-based generally outperform stem-based method on both word alignment and translation quality. And both lemma-based and stem-based methods are effective to alleviate the problem of data sparseness.

In future work, we will incorporate the various features into a discriminative framework to automatically train the weights for the clues.

References

Ahrenberg, L., Merkel, M., and Andersson, M. (1998). A Simple Hybrid Aligner for Generating Lexical Correspondences in Parallel Tests. In Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (pp. 29--35).

Al-Adhaileh, H.M., Kong, T.E., and Yusoff, Z. (2002). A Synchronization Structure of SSTC and its Applications in Machine Translation. In Proceedings of the COLING-2002 Post-Conference Workshop on Machine Translation in Asia (pp. 1--8).

Blunsom, P. and Cohn, T. (2006). Discriminative Word Alignment with Conditional Random Fields. In Proceedings of the 21st International Conference on

[11] http://www.tartarus.org/~martin/PorterStemmer/

Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics (pp. 65--72).

Brown, P.F., Della Pietra, A.F., Della Pietra, V.J., and Mercer, L.R. (1993). The Mathematics of Statistical Machine Translation: Parameter Estimation. Computational Linguistics, 19(2): 263--311.

Cherry, C. and Lin, D. (2003). A Probability Model to Improve Word Alignment. In Proceedings of the 41st Annual Meeting of the Association for Compu-tational Linguistics (pp. 88--95).

Dong, Z. and Dong, Q. (2006) HowNet and the Computation of Meaning. World Scientific Publishing, Singapore.

Gupta, D. and Federico, M. (2006). Exploiting Word Transformation in Statistical Machine Translation from Spanish to English. In Proceedings of 11th Annual Conference of the European Association for Machine Translation (pp. 75--80).

Ker, J.S. and Chang, S.J. (1997). A Class-based Approach to Word Alignment. Computational Linguistics, 23(2): 313--343.

Koehn, P. (2004). Pharaoh: A Beam Search Decoder For Phrase-Based Statistical Machine Translation Models. In Proceedings of the 6th Conference of the Association for Machine Translation in the Americas (pp. 115--124).

Koeling, R. (2000). Chunking with Maximum Entropy Models. In Proceedings of Computational Natural Language Learning (pp. 139--141).

Liu, Y., Liu, Q., and Lin, S. (2005). Log-Linear Models for Word Alignment. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistic (pp. 459--466).

Lopez A. and Resnik, P. (2006). Word-Based Alignment, Phrase-Based Translation: What's the Link? In Proceedings of the 12th Meeting of the Association for Machine Translation in the Americas (pp. 90--99).

Melamed, I.D. (1997). Automatic Discovery of Non-Compositional Compounds in Parallel Data. In Proceedings of the 2nd Conference on Empirical Methods in Natural Language Processing (pp. 97--108).

Meulder, D.F. and Daelemans, W. (2003). Memory-Based Named Entity Recognition using Unannotated Data. In Proceedings of Conference on Computational Natural Language Learning (pp. 208--211).

Moore C.R., Yih, W., and Bode, A. (2006). Improved Discriminative Bilingual Word Alignment. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics (pp. 513--520).

Moore, C.R. (2004). Improving IBM Word Alignment Model 1. In Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (pp. 519--526).

Och F.J. (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).

Och, F.J. and Ney, H. (2003). A Systematic Comparison of Various Statistical Alignment Models. Computational Linguistics, 29(1): 19--51.

Och, F.J. and Ney, H. (2000). Improved Statistical Alignment Models. In Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (pp. 440--447).

Papineni, K., Roukos, S., Ward, T., and Zhu, W. (2002). BLUE: a Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (pp. 311--318).

Popovic, M. and Ney, H. (2004). Towards the Use of Word Stems and Suffixes for Statistical Machine Translation. In Proceedings of the 4th International Conference on Language Resources and Evaluation (pp. 1585--1588).

Popovic, M. and Ney, H. (2005). Exploiting Phrasal Lexica and Additional Morpho-syntactic Language Resources for Statistical Machine Translation with Scarce Training Data. In Proceedings of 10th Annual Conference of the European Association for Machine Translation (pp. 212--218).

Smadja, F., McKeown, K.R., and Hatzivassiloglou, V. (1996). Translating Collocations for Bilingual Lexicons: a Statistical Approach. Computational Linguistics, 22(1): 1--38.

Stolcke, A. (2002). SRILM - An Extensible Language Modeling Toolkit. In Proceedings of the International Conference on Spoken Language Processing (pp. 901--904).

Sun, J., Gao, J., Zhang, L., Zhou, M., and Huang, C. (2002). Chinese Named Entity Identification Using Class-based Language Model. In Proceedings of the 19th International Conference on Computational Linguistics (pp. 967--973).

Taskar, B., Lacoste-Julien, S., and Klein, D. (2005), A Discriminative Matching Approach to Word Alignment. In Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (pp. 73--80).

Tiedemann, J. (1999). Word Alignment Step by Step. In Proceedings of the 12th Nordic Conference on Computational Linguistics (pp. 216--227).

Tiedemann, J. (2003). Combining Clues for Word Alignment. In Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (pp. 339--346).

Tufis, D. and Barbu, A.M. (2002). Lexical Token Alignment: Experiments, Results and Application. In Proceedings of the 3rd International Conference on Language Resources and Evaluation (pp. 458--465).

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

Wu, H. and Wang, H. (2004). Improving Statistical Word Alignment with a Rule-Based Machine Translation System. In Proceedings of 20th International Conference on Computational Linguistics (pp. 29--35).

Zhang, H. and Gildea, D. (2005). Stochastic Lexicalized Inversion Transduction Grammar for Alignment. In Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (pp. 475--482).