Yet Another Symmetrical and Real-time Word Alignment Method: Hierarchical Sub-sentential Alignment using F-measure

Hao Wang, Yves Lepage

Graduate School of Information, Production and Systems
Waseda University
2-7 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0135, Japan
`{oko_ips@ruri.,yves.lepage@}waseda.jp`

Abstract

Symmetrization of word alignments is the fundamental issue in statistical machine translation (SMT). In this paper, we describe an novel reformulation of Hierarchical Sub-sentential Alignment (HSSA) method using F-measure. Starting with a soft alignment matrix, we use the F-measure to recursively split the matrix into two soft alignment sub-matrices. A direction is chosen as the same time on the basis of Inversion Transduction Grammar (ITG). In other words, our method simplifies the processing of word alignment as recursive segmentation in a bipartite graph, which is simple and easy to implement. It can be considered as an alternative of *grow-diag-final-and* heuristic. We show its application on phrase-based SMT systems combined with the state-of-the-art approaches. In addition, by feeding with word-to-word associations, it also can be a real-time word aligner. Our experiments show that, given a reliable lexicon translation table, this simple method can yield comparable results with state-of-the-art approaches.

1 Introduction

Since most of state-of-the-art Statistical Machine Translation (SMT) approaches require word-to-word aligned data on a parallel corpus, word alignment is a fundamental issue to perform this task rapidly. In order to extract translation fragments for various purposes, e.g., word pairs (Brown et al., 1988), phrase pairs (Koehn et al., 2003), hierarchical rules (Chiang, 2005), tree-to-tree correspon-

dences (Zhang et al., 2007), reliable and accurate word aligners are essential.

There exist several problems in state-of-the-art methods for word alignment. Present word alignment approaches are usually based on IBM models (Brown et al., 1993), which parameters are estimated using the Expectation-Maximization (EM) algorithm. Sometimes, they are augmented with an HMM-based model (Vogel et al., 1996). Since IBM Models is the restriction to one-to-many alignments, some multi-word units cannot be correctly aligned. It is necessary to train models in both directions, and merge the outcome of mono-directional alignments using some symmetrization methods can overcome this deficiency to some degree.

It results, even using the standard open source tool aligner, called `GIZA++`[1] (Och, 2003), which consist of the widely used IBM models and their extensions, still will spend lots of time to obtain word alignments. A recent development of word alignment approach `fast_align`[2] (Dyer et al., 2013), based on the variation of the IBM model 2, has been reported faster than baseline `GIZA++` but with comparable results. However, both mentioned approaches generate asymmetric alignments. In order to obtain the symmetrical word alignments, these approaches symmetrize the alignments in both forward and reverse directions using a symmetrization heuristic called *grow-diag-final-and* (Och, 2003). Starting with the intersection alignment points that occur in both of the two directional alignments, *grow-diag-final-and* expands the alignment in the union of

---

[1] http://www.statmt.org/moses/giza/`GIZA++`.html
[2] https://github.com/clab/fast_align

the alignment in either of the two directional alignments. Although it has been shown to be most effective for phrase extraction for phrased-based SMT (Wu and Wang, 2007), there lacks a principled explanation.

Recently, development in mining large parallel patent or document collections increase the needs in fast methods for word alignment. Besides, in the real scenario of Computer Assisted Translation (CAT) (Kay, 1997), in conjunction with SMT system (Farajian et al., 2014) for translation or post-editing (reference) (Guerberof, 2009), real-time word alignment methods become necessary.

In this paper, we propose a novel method based on the use of F-measure for symmetrization of word alignment, at the same time which can be regarded as an real-time word alignment approach. We justify this approach with mathematical principles. The paper is organized as follows: in Section 2, we discuss the motivation. In Section 3, we summarize the related work like Viterbi alignment and inversion transduction grammar. In Section 4, we formulate our method and give a mathematical justification. The Section 5 reports experiments and results. Finally, we give some conclusion and points for the future research.

2 Motivation

There exist several purposes that drive us to introduce such a new method which differs the previous approach. Absolutely, time cost is the first our consideration. Consider the case when huge parallel documents are handed to the computer. It will be a very interesting question that how to align these parallel sentences in a large number of documents while have spent the minimal time. Nowadays, since most of the public available word aligners are based on EM algorithm in order to get the global optimal alignments, the real-time cost of the processing of word alignment can not be estimated.

Another realistic problem is, in the most real situation of machine translation task, a bilingual lexicon dictionary even longer phrase translation fragments table is given or available, while reusing the pre-built knowledge base, rather than aligning data using some machine learning technique to guess the probable Viterbi alignment again, is a more advisable solution to employ a real-time aligner to align words automatically.

There are also some drawbacks of the previous approach likes IBM models and their variations. All these models are based on restricted alignments in the sense that a source word can be aligned to at most one target word. This constraint is necessary to reduce the computational complexity of the models, but it makes it impossible to align phrase in the target language (English) such as *"a car"* to a single word in the source language (Japanese/Chinese) "车/车". Beside, a variation of IBM model 2 was used in `fast_align`. It introduces a "tension" to model the overall accordance of word orders, but it has proved by (Ding et al., 2015) that it performs not well when applied to the very distinct language pairs, e.g., English and Japanese.

3 Related Work

3.1 Viterbi alignment and symmetrization

The basic idea of the previous approaches is to develop a model treating the word alignment as a hidden variables (Och, 2003), by applying some statistical estimation theory to obtain the most possible/Viterbi alignments. The problem of translation can be defined as:

$$Pr(f_1^J|e_1^I) = \sum_{a_1^J} Pr(f_1^J, a_1^J|e_1^I) \qquad (1)$$

Here we use the symbol $Pr(\cdot)$ to denote general probability distributions. a_1^J is a "hidden" alignment which is mapping from a source position j to a target position a_j. It is always possible to find a best alignment by maximizing the likelihood on the given parallel training corpus.

$$\hat{a}_1^J = \text{argmax}_{a_1^J} Pr(f_1^J, a_1^J|e_1^I) \qquad (2)$$

Since Viterbi alignment model is based on conditional probabilities, it only returns one directional alignment in each direction ($F \rightarrow E$ and vice-versa). In other words, this process is asymmetric. The complementary part of Viterbi alignment model before phrase extraction is *grow-diag-final-and*, in which the symmetrical word alignments are generated using simple growing heuristics. Given two sets of alignments $\hat{a}_1^J$ and $\hat{b}_1^J$, in order to increase the
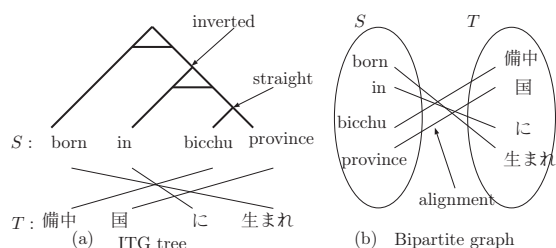
Figure 1: Alignments representations using ITG and bipartite graph. None of the structure contains cycles. The Japanese phrase 備中 国 に 生まれ means *born in bicchu province* in English.

quality of the alignments, they combine $\hat{a}_1^J$ and $\hat{b}_1^J$ into one alignment matrix A using *grow-diag-final-and* algorithm.

A widely used approach to get word alignments is estimating the alignment using IBM models because word alignments are the by-production of estimating lexicon translation probabilities. However, this generative story looks like a "*chicken or the egg*" problem. On the one hand, given alignments with probabilities it is possible to compute translation probabilities. On the other hand, if knowing which words are a probable translation of another one makes it possible to guess which alignment is probable and which one is improbable.

3.2 ITG-based word alignment

Since the search space of word alignment will grow exponentially with the length of source and target sentences (Brown et al., 1993), Wu (1997) proposed an approach to constraining the search space for word alignment, namely inversion transduction grammars (ITG). Generally, ITG is a family of grammars in which the right part of the rule is either two non-terminals or a terminal sequence. ITG is a special case of synchronous context-free grammar, also called Bracketing Transduction Grammar (BTG). There are three simple generation rules, S (straight), I (inverted) and terminal (T).

$$S: \quad \gamma \to [XY] \tag{3}$$
$$I: \quad \gamma \to < XY > \tag{4}$$
$$T: \quad \gamma \to w = (s,t) \tag{5}$$

The algorithm used by (Wu, 1997) synchronously parses the source and the target sentence to build a synchronous parse tree. This ITG tree indicates the same underlying structure but the ordering of constituents may be different. Due to its simplicity and effectiveness of modelling bilingual correspondence, ITG can be used to model the bilingual sentences in very distinct ordering. In fact, an ITG-style Tree is a bitree consists of one tree in the source side and another tree in the target side (see Figure 1.a), here, two trees are compressed as a single tree. Besides, an ITG-style Tree is also able to be displayed in a soft alignment matrix (see Figure 2) with the representation of bipartite graph (see Figure 1.b) .

3.3 Hierarchical sub-sentential alignment

Hierarchical sub-sentential alignment (HSSA) is yet another alignment approach, introduced by (Lardilleux et al., 2012). This method does not rely on the EM algorithm as other alignment models. With a recursive binary segmentation process of searching the segment point in a soft alignment matrix (as Figure 2) between a source sentence and its corresponding target sentence, this approach aims to minimize *Ncut* score (Zha et al., 2001), which can yield acceptable and accurate 1-to-many or many-to-1 word alignments.

In order to build soft alignment matrices before the step of aligning words, Lardilleux et al. (2012) employed `Anymalign`[3] to obtain the prepared translation table of lexicon translation probabilities. Since the training times and the quality of translation table changed considerably depending on the timeouts for `Anymalign`, an easy and fair comparison to state-of-the-art approaches is difficult.

Given the grey-scale graph of soft alignment, Hierarchical Sub-sentential Alignment (hereafter referred to as HSSA) approach takes all cells in the soft alignment matrix into consideration and seeks the precise criterion for a good partition same as image segmentation. It makes use of a popular modern clustering algorithm called *normalized cuts* (Zha et al., 2001; Shi and Malik, 2000), i.e., spectral clustering, or $Ncut$ for short, to binary segment the matrix recursively.

In the following section, we will refine the proposal of hierarchical sub-sentential alignment. We will not use the notion of *Ncut*, so as to give a sim-
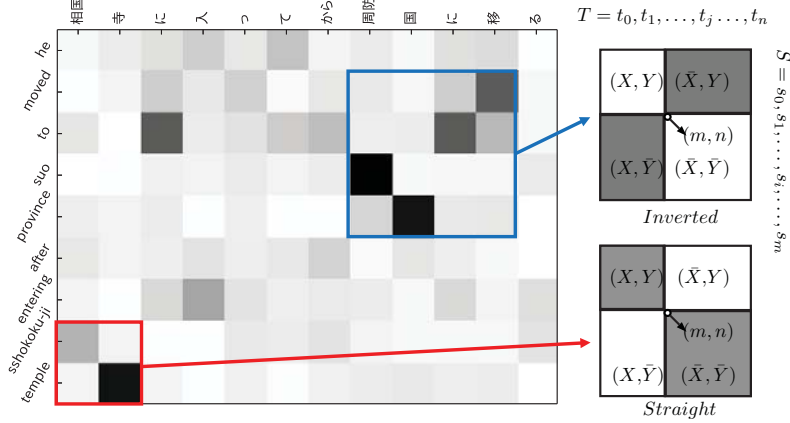
---

[3]https://anymalign.limsi.fr/

Figure 2: Translation strengths on a logarithmic scale in a English-Japanese sentence pair matrix as a grey graph.

ple and convincing justification using F-measure for this symmetrical word alignment approach.

4 Proposed Method

4.1 Soft alignment matrices

We propose to regard the alignment associations between a source sentence S and a target word T as a contingency matrix (Matusov et al., 2004; Moore, 2005) as in Figure 2, noted as $\mathcal{M}(I, J)$, in which I is the length of source sentence in words and J for target side. We define a function w which measuring the strength of the translation link between any source and target pair of words (s_i, t_j). The symmetric alignment between word s_i and t_j presents a greyed cell (i, j) in this matrix. In this paper, the score $w(s_i, t_j)$ is defined as the geometric mean of the bidirectional lexical translation probabilities $p(s_i|t_j)$ and $p(t_j|s_i)$. For a given sub-sentential alignment $A(X, Y) \subseteq I \times J$, we define the weight of this alignment W(X,Y) as the summation of association scores between each source and target words of a block (X, Y) in such a matrix.

$$W(X,Y) = \sum_{s \in X} \sum_{t \in Y} w(s,t) \qquad (6)$$

Since we have to calculate all cells in the block (X, Y), the time complexity here is in $O(I \times J)$.

4.2 Reformulation: from Ncut to F-measure

Ncut can be computed as the following formula same as in (Zha et al., 2001), :

$$Ncut(X,Y) = \frac{cut(X,Y)}{cut(X,Y)+2\times W(X,Y)} + \frac{cut(\overline{X},\overline{Y})}{cut(\overline{X},\overline{Y})+2\times W(\overline{X},\overline{Y})}$$

$$cut(X,Y) = W(X,\overline{Y}) + W(\overline{X},Y) \qquad (7)$$

Actually, minimizing $Ncut(X, Y)$ is equivalent to maximizing the arithmetic mean of the F-measure (also called F-score) of X relatively to Y and $\bar{X}$ relatively to $\bar{Y}$. It can be derived as following. In general, F_1-measure (Kim et al., 1999) of block (X, Y) is defined as the harmonic mean of precision P and recall R:

$$\frac{1}{F_1(X,Y)} = \frac{1}{2} \times \left(\frac{1}{P(X,Y)} + \frac{1}{R(X,Y)}\right) \qquad (8)$$

To interpret sentence pair matrices as contingency matrices, it suffices to read trans- lation strengths as reflecting the contribution of a source word to a target word and reciprocally. With this interpretation, the precision (P) and the recall (R) for two sub-parts of the source and the target sentences can easily be expressed using the sum of all the translation strengths inside a block. These two measures can thus be defined as following Equations.

$$P(X,Y) = \frac{W(X,Y)}{W(X,Y) + W(X,\overline{Y})} \qquad (9)$$

$$R(X,Y) = \frac{W(X,Y)}{W(X,Y) + W(\overline{X},Y)} \qquad (10)$$

146

Now, it suffices to replace precision and recall by their values in terms of cut to derive the following formula.

$$\frac{1}{F_1(X,Y)}= \frac{1}{2} \times (\frac{W(X,Y)+W(X,\overline{Y})}{W(X,Y)} + \frac{W(X,Y)+W(\overline{X},Y)}{W(X,Y)}) \tag{11}$$

$$= \frac{2 \times W(X,Y)+W(\overline{X},Y)+W(X,\overline{Y})}{2 \times W(X,Y)} \tag{12}$$

$$= \frac{2 \times W(X,Y)+cut(\overline{X},Y)}{2 \times W(X,Y)} \tag{13}$$

By using Equation 13 and Equation 7, for (X,Y), we obtain:

$$F_1(X,Y)= 1 - \frac{W(\overline{X},Y)+W(X,\overline{Y})}{2 \times W(X,Y)+W(\overline{X},Y)+W(X,\overline{Y})} \tag{14}$$

$$= 1 - Ncut_{left}(X,Y) \tag{15}$$

In a contingency matrix, where balanced F_1-score can be used regularly for binary classification, especially on the scenario of binary segmentation of bilingual sentence pair under the ITG framework. With this interpretation, for the *straight* case of ITG, we can get the F_1-score for the remaining block $(\overline{X},\overline{Y})$ as,

$$F_1(\overline{X},\overline{Y}) = 1 - Ncut_{right}(X,Y) \tag{16}$$

Absolutely, an equivalent way of writing is:

$$Ncut(X,Y) = 2 \times [1 - \frac{F_1(X,Y)+F_1(\overline{X},\overline{Y})}{2}] \tag{17}$$

To summarize, minimizing *Ncut* equals finding the best point with the maximum value in the matrix of arithmetic means of F_1-score. This in fact makes sense intuitively if we look for the best possible way for parts of the source and target sentences to correspond. These parts should cover one another in both directions as much as possible, that is to say, they should exhibit the best recall and precision at the same time.

4.3 Reducing time complexity

In order to reduce the time complexity in calculate the value of $W(\overline{X},\overline{Y})$, we make use of a specialized data structure for fast computation. For each given sentence pair, a summed area table (SAT) was created for fast calculating the summation of cells in the

corresponding soft alignment matrix $\mathcal{M}(I,J)$. The preprocessing step is to build a new $(I+1, J+1)$ matrix $\mathcal{M}'$, where each entry is the sum of the sub-matrix to the upper-left of that entry. Any arbitrary sub-matrix sum can be calculated by looking up and mixing only 4 entries in the SAT.

Assume X,Y starts from point (i_0, j_0), where $X,\overline{X}$ and $Y,\overline{Y}$ are splitting at i_1 and j_1 separately. We have,

$$W(X,Y) = \sum_{\substack{i_0 < i < i_1 \\ j_0 < j < j_1}} w(i,j)$$
$$= \mathcal{M}'(i_1,j_1) - \mathcal{M}'(i_0,j_1)$$
$$- \mathcal{M}'(i_1,j_0) + \mathcal{M}'(i_0,j_0)$$

Time complexity here is reduced from $O(I \times J)$ to $O(1)$ when calculating the summation of cells in the block of X,Y, and similar to the remaining. Due to data sparsity, a simple Laplace smoothing was used here to handle the unseen alignments with a very small smoothing parameter $\alpha = 10^{-7}$.

5 Experiments

5.1 Alignment Experiments

We evaluate the performance of our proposed methods. We conduct the experiments on KFTT corpus[4], in which applied Japanese-to-English word alignment. We report the performance of various alignment approach in terms of precision, recall and alignment error rate (AER) as (Och, 2003) defined. The quality of an alignment $A = \{(j,a_j)|a_j > 0\}$ is then computed by appropriately redefined precision and recall measures:

$$Recall = \frac{|A \cap S|}{|S|}, Precision = \frac{|A \cap P|}{|P|}, S \subseteq P \tag{18}$$

and the following alignment error rate:

$$AER(S,P;A) = 1 - \frac{|A \cap S| + |A \cap P|}{|A| + |S|} \tag{19}$$

The details are shown in Table 1. Figure 3 plots the average run-time of the currently available alignment approaches as a function of the number of input English-French sentence pairs. The HSSA approach is far more efficient. In total, aligning the

[4]http://www.phontron.com/kftt/index.html

	#	MatchRef	Prec	Rec	AER
Ref	33,377				
GIZA++ + GDFA	31,342	18,641	59.48	55.85	42.39
fast_align + GDFA	25,368	14,076	55.49	42.17	52.08
GIZA++ + HSSA	43,257	15,209	35.16	45.57	60.31
fast_align + HSSA	43,070	14,950	34.71	44.79	60.89

Table 1: Word alignments comparison on Japanese-English data in terms of matches number, precision, recall and alignment error rate (AER). GDFA: an abbreviation of *grow-diag-final-and*. HSSA: an abbreviation of hierarchical sub-sentential alignment.

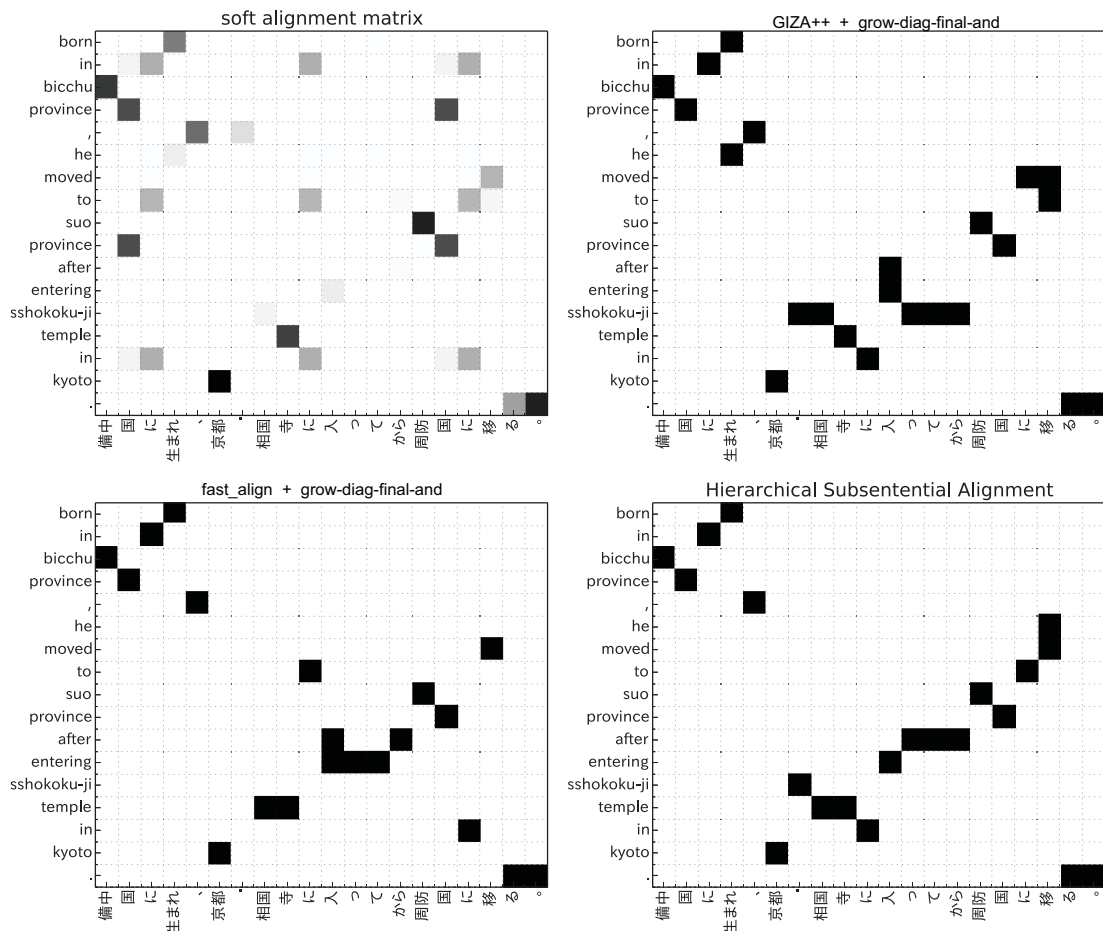

Figure 3: Comparison of alignments output by various tools. The test sentence pair is sampled from KFTT corpus. We fed HSSA with the lexical translation table relying on the output of GIZA++. In this example, our proposed approach (GIZA++ + HSSA) generates a better alignment than GIZA++ + GDFA or fast_align + GDFA.
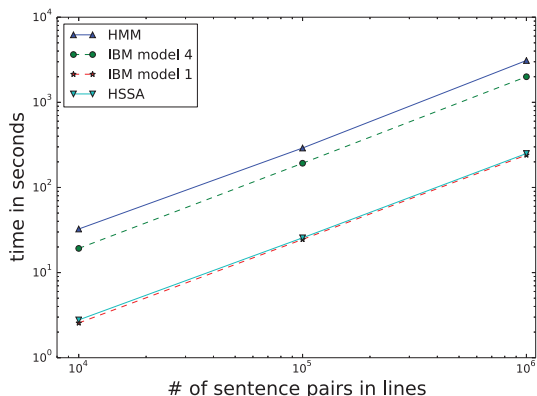
Figure 4: Average word alignment run-time (in seconds) as a function of the size of a corpus (in sentence pairs). Remember that, given the lexical translation probabilities, HSSA runs only in one iteration.

10K sentence pairs in the corpus completed in nearly 20 second with the HSSA approach but required more time with the other EM-based approaches.

In Table 1, Precision of our proposed approach are lower than baseline system, but Recall are better than `fast_align` + GDFA. However, it has been proved (Fraser and Marcu, 2007; Ganchev et al., 2008) that AER does not imply a better translation accuracy (see Table 3).

5.2   Translation Experiments

In this section, we first describe the data used in our experiments. We then perform to extract the lexical translation probabilities. Finally, we conduct translation experiments using both the baseline system (`GIZA++`) and the system using HSSA approach combined with to show, given a reliable lexical translation table for soft alignment matrix, the effectiveness of our proposed integrated system. We also investigate the time cost and the influence on the SMT frameworks.

In order to evaluate the proposed method, we conducted translation experiments on two corpora: Europarl Corpus and KFTT corpus. For English-Japanese (en-ja) and Japanese-English (ja-en), we evaluated on the KFTT corpus. For English-Finnish (en-fi), Spanish-Portuguese (es-pt) and English-French (en-fr), we measure the translation metrics

	Train	Tune	Test
Europarl v7	183K	1K	2K
KFTT	330K	1.2K	1.2K

Table 2: Statistics on the parallel corpus used in the experiments (K=1,000 lines).

on Europarl Corpus v7[5]. The baseline systems are using `GIZA++` to train as generally.

In our experiments, standard phrase-based statistical machine translation systems were built by using the `Moses` toolkit (Koehn et al., 2007), Minimum Error Rate Training (Och, 2003), and the KenLM language model (Heafield, 2011). Default training pipeline for phrase-based SMT in is adopt with default *distortion-limit* 6. For the evaluation of machine translation quality, some standard automatic evaluation metrics have been used, like BLEU (Papineni et al., 2002), NIST (Doddington, 2002) and RIBES (Isozaki et al., 2010) in all experiments. When compared with the baseline system (`GIZA++` + GDFA), there is no significant difference on the final results of machine translation between using the alignments output by the proposed approach and `GIZA++`.

6   Conclusion

In this paper, we studied an ITG-based bilingual word alignment method which recursively segments the sentence pair on the basis of a soft alignment matrix. There are several advantages in our proposed method. Firstly, when combining the proposed method with word association probabilities (lexical translation table), it is more reasonable to obtain symmetrical alignments using the proposed method rather than *grow-diag-final-and*. In other words, this method provides an alternative to *grow-diag-final-and* for symmetrization of word alignments. It achieves a similar speed compared to the simplest IBM model 1. Second, HSSA points a new way to real-time word alignment. For the tasks of processing same domain document, HSSA makes it possible to reuse the pre-built crossing-language information, likes bilingual lexical translation table. In our experiment, it has demonstrated that our proposed method achieves comparable accuracies compared

---
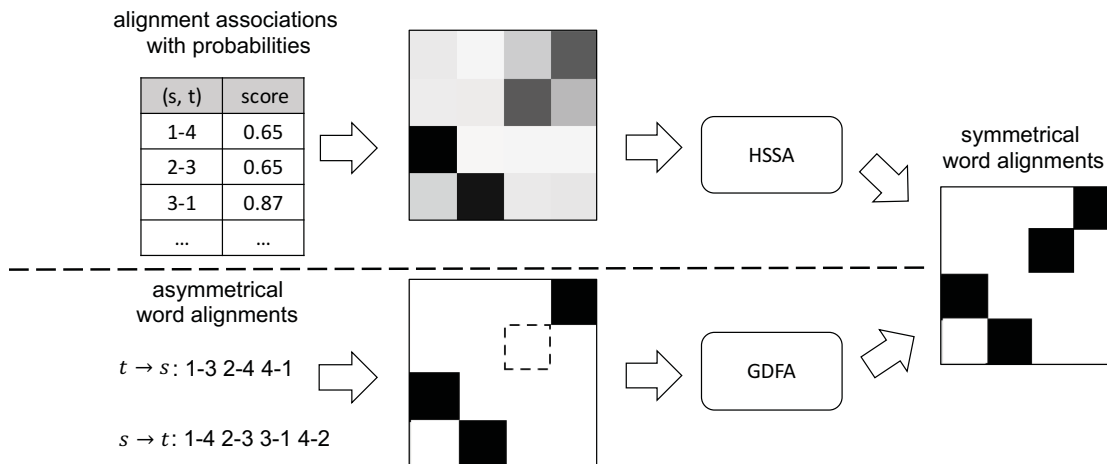
[5]http://www.statmt.org/europarl/

Figure 5: Our proposed approach starts from alignment associations with some probabilities, which is different from the standard phrase-based SMT pipeline.

				BLEU	NIST	TER	WER	RIBES
en-fr	GIZA++	+	GDFA	54.40	9.483	34.37	30.19	91.22
	GIZA++	+	HSSA	**54.42**	9.542	34.07	30.08	91.25
	fast_align	+	GDFA	54.10	9.438	34.63	30.45	91.14
	fast_align	+	HSSA	54.05†	9.417	34.72	30.58	91.11
es-pt	GIZA++	+	GDFA	49.34	9.182	35.97	31.74	90.62
	GIZA++	+	HSSA	49.32	8.980	36.99	32.44	90.30
	fast_align	+	GDFA	**49.70**	92.06	35.46	31.30	90.79
	fast_align	+	HSSA	49.51	9.203	35.59	31.38	90.79
en-fi	GIZA++	+	GDFA	**36.61**	6.608	47.08	41.36	87.03
	GIZA++	+	HSSA	35.15‡	6.448	47.71	42.18	86.60
	fast_align	+	GDFA	36.11	6.669	46.69	41.29	87.01
	fast_align	+	HSSA	35.88†	6.492	47.32	41.69	86.75
en-ja	GIZA++	+	GDFA	**21.59**	5.632	74.12	74.99	68.10
	GIZA++	+	HSSA	21.22	5.585	74.26	73.30	67.84
	fast_align	+	GDFA	20.80‡	5.592	74.50	74.33	68.13
	fast_align	+	HSSA	21.23	5.590	74.35	75.48	68.01
ja-en	GIZA++	+	GDFA	**18.78**	5.730	71.25	68.30	65.87
	GIZA++	+	HSSA	18.38	5.659	70.61	68.40	65.53
	fast_align	+	GDFA	18.23	5.628	71.26	68.01	65.25
	fast_align	+	HSSA	18.24	5.659	70.61	68.27	65.46

Table 3: Comparison of translation results using various configurations, GIZA++ or fast_align with *grow-diag-final-and* (GDFA) or hierarchical subsentential alignment (HSSA). Bold surfaces indicate the best BlEU score in each group. No significant difference between directly GIZA++ + GDFA with our proposed method except en-fi. Statistical significantly difference in BLEU score at ‡: $p < 0.01$ and †: $p < 0.05$ compared with GIZA++ + GDFA.

with a state-of-the-art baseline. Finally, compared with original HSSA, the advantages of our implementation includes well-formulated, shorter computation times spent, armed with smoothing technique.

For future work, we think of designing a beam-search variation to make it possible to generate several parsing derivations during recursive segmentation. This will allow us to investigate recombinations of different derivations in order to obtain more possible alignments.

Acknowledgments

References

Peter F. Brown and John Cocke and Stephen A. Della-Pietra and Vincent J. Della-Pietra and Frederick Jelinek and Robert L. Mercer and Paul Rossin. 1988. A Statistical Approach to Language Translation. *Proceedings of the International Conference on Computational Linguistics (COLING)*.

Peter F. Brown, Vincent J. Della Pietra, Stephen A Della Pietra and Robert L. Mercer 1988. The mathematics of statistical machine translation: Parameter estimation. *Computational linguistic*. volume 19 (2): pp.263–311. MIT Press

M. Amin Farajian, Nicola Bertoldi and Marcello Federico. 2014. Online word alignment for online adaptive machine translation. *EACL*. 84.

Philipp Koehn. 2003. Statistical machine translation. *Cambridge University Press* , 2009.

Percy Liang, Ben Taskar and Dan Klein. 2006. Alignment by agreement. *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*. Association for Computational Linguistics.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistic*. 29.1: pp.19-51

Och, Franz Josef, and Hermann Ney. "A systematic comparison of various statistical alignment models." .

Philipp Koehn, Franz Josef Och and Daniel Marcu. 2003. *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. Association for Computational Linguistics.

Kuzman Ganchev, Joao V Grac a and Ben Taskar. 2008. Better alignments= better translations? *ACL-08: HLT*. Association for Computational Linguistics. page 986.

Martin Kay 1997. The proper place of men and machines in language translation. *machine translation*, 12, 1-2. Springer. pp.3–23

Min Zhang, Hongfei Jiang and AiTi Aw and Jun Sun and Sheng Li and Chew Limu Tan. 2003. A tree-to-tree alignment-based model for statistical machine translation. *MT-Summit-07*. pp.535–542

Chenchen Ding, Masao Utiyama and Eiichiro Sumita. 2015. Improving fast align by Reordering. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*.

David Chiang. 2005. A hierarchical phrase-based model for statistical machine translation. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics.

Kishore Papineni, et al. 2002. BLEU: a method for automatic evaluation of machine translation. *In Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics*. p. 311-318.

Philipp Koehn, et al. 2007. Moses: Open source toolkit for statistical machine translation. *Proceedings of the 45th annual meeting of the ACL on interactive poster and demonstration sessions*. Association for Computational Linguistics, 2007.

Adrien Lardilleux, François Yvon and Yves Lepage. 2012. Hierarchical sub-sentential alignment with Anymalign. *16th annual conference of the European Association for Machine Translation (EAMT)*.

Stephan Vogel, Hermann Ney and Christoph Tillmann. 1996. HMM-based word alignment in statistical translation. *Proceedings of the 16th conference on Computational linguistics*-Volume 2. Association for Computational Linguistics.

George Doddington. 2002. *Automatic evaluation of machine translation quality using n-gram co-occurrence statistics*. Proceedings of the second international conference on Human Language Technology Research. Morgan Kaufmann Publishers Inc.

C Dyer, V Chahuneau and NA Smith 2013. A Simple, Fast, and Effective Reparameterization of IBM Model. *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* pp.644-649.

Ashok K. Chandra, Dexter C. Kozen, and Larry J. Stockmeyer. 1981. Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.

Hongyuan Zha, Xiaofeng He, Chris Ding, Horst Simon, and Ming Gu. 2001. Bipartite graph partitioning and data clustering. *In Proc. of the 10th international conference on Information and knowledge management*, pages 25–32, Atlanta.

Franz Josef Och. 2003. *Minimum error rate training in statistical machine translation.*. Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1. Association for Computational Linguistics.

Hua Wu and Haifeng Wang. 2007. Comparative study of word alignment heuristics and phrase-based SMT. *Proceedings of the MT Summit XI.*

Alexander Fraser and Daniel Marcu. 2007. Measuring word alignment quality for statistical machine translation. *Computational Linguistics*. 33(3)

Dekai Wu. 1997. Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. *Computational linguistics*.23.3: pp.377–403.

Jianbo Shi and Jitendra Malik 2000. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on.*22.8. IEEE. pp.888–905

Hongyuan Zha, Xiaofeng He, Chris Ding and Horst Simon and Ming Gu. 2001. Bipartite graph partitioning and data clustering *Proceedings of the tenth international conference on Information and knowledge management*. Association for Computational Linguistics. pp.25–32.

Evgeny Matusov, Richard Zens, and Hermann Ney. 2004. Symmetric word alignments for statistical machine translation. *Proceedings of the 20th international conference on Computational Linguistics*. Association for Computational Linguistics. pp.219.

J. Makhoul, F. Kubala, R. Schwartz and R. Weischedel,. 1999. Performance measures for information extraction. *In Proceedings of DARPA broadcast news workshop*. pp. 249-252.

J. Makhoul, F. Kubala, R. Schwartz and R. Weischedel,. 1999. Performance measures for information extraction. *In Proceedings of DARPA broadcast news workshop*. pp. 249-252.

Jae Dong Kim, Ralf D. Brown, Peter J. Jansen and Jaime G. Carbonell. 2005. Symmetric probabilistic alignment for example-based translation. *In Proceedings of the Tenth Workshop of the European Assocation for Machine Translation (EAMT-05)*, May.

Robert C Moore. 2005. *Association-based bilingual word alignment. Proceedings of the ACL Workshop on Building and Using Parallel Texts*. Association for Computational Linguistics,

Ana Guerberof 2009. Productivity and quality in MT post-editing *MT Summit XII-Workshop: Beyond Translation Memories: New Tools for Translators MT.*

Hideki Isozaki, Tsutomu Hirao,Kevin Duh, Katsuhito Sudoh and Hajime Tsukada. 2010. Automatic evaluation of translation quality for distant language pairs. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing.* Association for Computational Linguistics. pp. 944–952.

Kenneth Heafield. 2011. KenLM: Faster and smaller language model queries *Proceedings of the Sixth Workshop on Statistical Machine Translation.* Association for Computational Linguistics.