Using RBMT Systems to Produce Bilingual Corpus for SMT

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

Abstract

This paper proposes a method using the existing Rule-based Machine Translation (RBMT) system as a black box to produce synthetic bilingual corpus, which will be used as training data for the Statistical Machine Translation (SMT) system. We use the existing RBMT system to translate the monolingual corpus into synthetic bilingual corpus. With the synthetic bilingual corpus, we can build an SMT system even if there is no real bilingual corpus. In our experiments using BLEU as a metric, the system achieves a relative improvement of 11.7% over the best RBMT system that is used to produce the synthetic bilingual corpora. We also interpolate the model trained on a real bilingual corpus and the models trained on the synthetic bilingual corpora. The interpolated model achieves an absolute improvement of 0.0245 BLEU score (13.1% relative) as compared with the individual model trained on the real bilingual corpus.

1 Introduction

Within the Machine Translation (MT) field, by far the most dominant paradigm is SMT, but many existing commercial systems are rule-based. In this research, we are interested in answering the question of whether the existing RBMT systems could be helpful to the development of an SMT system. To find the answer, let us first consider the following facts:

- Existing RBMT systems are usually provided as a black box. To make use of such systems, the most convenient way might be working on the translation results directly.
- SMT methods rely on bilingual corpus. As a data driven method, SMT usually needs large bilingual corpus as the training data.

Based on the above facts, in this paper we propose a method using the existing RBMT system as a black box to produce a synthetic bilingual corpus[1], which will be used as the training data for the SMT system.

For a given language pair, the monolingual corpus is usually much larger than the real bilingual corpus. We use the existing RBMT system to translate the monolingual corpus into synthetic bilingual corpus. Then, even if there is no real bilingual corpus, we can train an SMT system with the monolingual corpus and the synthetic bilingual corpus. If there exist n available RBMT systems for the desired language pair, we use the n systems to produce n synthetic bilingual corpora, and n translation models are trained with the n corpora respectively. We name such a model the *synthetic model*. An interpolated translation model is built by linear interpolating the n synthetic models. In our experiments using BLEU (Papineni et al., 2002) as the metric, the interpolated synthetic model achieves a relative improvement of 11.7% over the best RBMT system that is used to produce the synthetic bilingual corpora.

[1] In this paper, to be distinguished from the real bilingual corpus, the bilingual corpus generated by the RBMT system is called a synthetic bilingual corpus.

Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 287–295, Prague, June 2007.

Moreover, if a real bilingual corpus is available for the desired language pair, we build another translation model, which is named the *standard model*. Then we can build an *interpolated model* by interpolating the standard model and the synthetic models. Experimental results show that the interpolated model achieves an absolute improvement of 0.0245 BLEU score (13.1% relative) as compared with the standard model.

The remainder of this paper is organized as follows. In section 2 we summarize the related work. We then describe our method Using RBMT systems to produce bilingual corpus for SMT in section 3. Section 4 describes the resources used in the experiments. Section 5 presents the experiment result, followed by the discussion in section 6. Finally, we conclude and present the future work in section 7.

2 Related Work

In the MT field, by far the most dominant paradigm is SMT. SMT has evolved from the original word-based approach (Brown et al., 1993) into phrase-based approaches (Koehn et al., 2003; Och and Ney, 2004) and syntax-based approaches (Wu, 1997; Alshawi et al., 2000; Yamada and Knignt, 2001; Chiang, 2005). On the other hand, much important work continues to be carried out in Example-Based Machine Translation (EBMT) (Carl et al., 2005; Way and Gough, 2005), and many existing commercial systems are rule-based.

Although we are not aware of any previous attempt to use an existing RBMT system as a black box to produce synthetic bilingual training corpus for general purpose SMT systems, there exists a great deal of work on MT hybrids and Multi-Engine Machine Translation (MEMT).

Research into MT hybrids has increased over the last few years. Some research focused on the hybrid of various corpus-based MT methods, such as SMT and EBMT (Vogel and Ney, 2000; Marcu, 2001; Groves and Way, 2006; Menezes and Quirk, 2005). Others tried to exploit the advantages of both rule-based and corpus-based methods. Habash et al. (2006) built an Arabic-English generation-heavy MT system and boosted it with SMT components. METIS-II is a hybrid machine translation system, in which insights from SMT, EBMT, and RBMT are used (Vandeghinste et al., 2006). Seneff et al. (2006) combined an interlingual translation framework with phrase-based SMT for spoken language translation in a limited domain. They automatically generated a corpus of English-Chinese pairs from the same interlingual representation by parsing the English corpus and then paraphrasing each utterance into both English and Chinese.

Frederking and Nirenburg (1994) produced the first MEMT system by combining outputs from three different MT engines based on their knowledge of the inner workings of the engines. Nomoto (2004) used voted language models to select the best output string at sentence level. Some recent approaches to MEMT used word alignment techniques for comparison between the MT systems (Jayaraman and Lavie, 2005; Zaanen and Somers, 2005; Matusov et al. 2006). All the above MEMT systems operate on MT outputs for complete input sentences. Mellebeek et al. (2006) presented a different approach, using a recursive decomposition algorithm that produces simple chunks as input to the MT engines. A consensus translation is produced by combining the best chunk translation.

This paper uses RBMT outputs to improve the performance of SMT systems. Instead of RBMT outputs, other researchers have used SMT outputs to boost translation quality. Callision-Burch and Osborne (2003) used co-training to extend existing parallel corpora, wherein machine translations are selectively added to training corpora with multiple source texts. They also created training data for a language pair without a parallel corpus by using multiple source texts. Ueffing (2006) explored monolingual source-language data to improve an existing machine translation system via self-training. The source data is translated by a SMT system, and the reliable translations are automatically identified. Both of the methods improved translation quality.

3 Method

In this paper, we use the synthetic and real bilingual corpus to train the phrase-based translation models.

3.1 Phrase-Based Models

According to the translation model presented in (Koehn et al., 2003), given a source sentence f, the best target translation e_{best} can be obtained using the following model

$$
\begin{aligned}
\mathrm{e}_{\mathrm{best}} &= \arg\max_{\mathrm{e}} p(\mathrm{e} \mid \mathrm{f}) \\
&= \arg\max_{\mathrm{e}} p(\mathrm{f} \mid \mathrm{e}) p_{\mathrm{LM}}(\mathrm{e}) \omega^{\mathrm{length(e)}}
\end{aligned} \tag{1}
$$

Where the translation model $p(\mathrm{f} \mid \mathrm{e})$ can be decomposed into

$$
\begin{aligned}
& p(\overline{f}_1^I \mid \overline{e}_1^I) \\
&= \prod_{i=1}^{I} \phi(\overline{f}_i \mid \overline{e}_i) d(a_i - b_{i-1}) p_{\mathrm{w}}(\overline{f}_i \mid \overline{e}_i, a)^{\lambda}
\end{aligned} \tag{2}
$$

Where $\phi(\overline{f}_i \mid \overline{e}_i)$ is the phrase translation probability. a_i denotes the start position of the source phrase that was translated into the ith target phrase, and b_{i-1} denotes the end position of the source phrase translated into the (i-1)th target phrase. $d(a_i - b_{i-1})$ is the distortion probability. $p_{\mathrm{w}}(\overline{f}_i \mid \overline{e}_i, a)$ is the lexical weight, and λ is the strength of the lexical weight.

3.2 Interpolated Models

We train synthetic models with the synthetic bilingual corpus produced by the RBMT systems. We can also train a translation model, namely standard model, if a real bilingual corpus is available. In order to make full use of these two kinds of corpora, we conduct linear interpolation between them.

In this paper, the distortion probability in equation (2) is estimated during decoding, using the same method as described in Pharaoh (Koehn, 2004). For the phrase translation probability and lexical weight, we interpolate them as shown in (3) and (4).

$$
\phi(\overline{f} \mid \overline{e}) = \sum_{i=0}^{n} \alpha_i \phi_i(\overline{f} \mid \overline{e}) \tag{3}
$$

$$
p_{\mathrm{w}}(\overline{f} \mid \overline{e}, a) = \sum_{i=0}^{n} \beta_i p_{\mathrm{w},i}(\overline{f} \mid \overline{e}, a) \tag{4}
$$

Where $\phi_0(\overline{f} \mid \overline{e})$ and $p_{\mathrm{w},0}(\overline{f} \mid \overline{e}, a)$ denote the phrase translation probability and lexical weight trained with the real bilingual corpus, respectively. $\phi_i(\overline{f} \mid \overline{e})$ and $p_{\mathrm{w},i}(\overline{f} \mid \overline{e}, a)$ ($i = 1,...,n$) are the phrase translation probability and lexical weight estimated by n synthetic corpora produced by the RBMT systems. α_i and β_i are interpolation coefficients, ensuring $\sum_{i=0}^{n} \alpha_i = 1$ and $\sum_{i=0}^{n} \beta_i = 1$.

4 Resources Used in Experiments

4.1 Data

In the experiments, we take English-Chinese translation as a case study. The real bilingual corpus includes 494,149 English-Chinese bilingual sentence pairs. The monolingual English corpus is selected from the English Gigaword Second Edition, which is provided by Linguistic Data Consortium (LDC) (catalog number LDC2005T12). The selected monolingual corpus includes 1,087,651 sentences.

For language model training, we use part of the Chinese Gigaword Second Edition provided by LDC (catalog number LDC2005T14). We use 41,418 documents selected from the ZaoBao Newspaper and 992,261 documents from the XinHua News Agency to train the Chinese language model, amounting to 5,398,616 sentences.

The test set and the development set are from the corpora distributed for the 2005 HTRDP [2] evaluation of machine translation. It can be obtained from Chinese Linguistic Data Consortium (catalog number 2005-863-001). We use the same 494 sentences in the test set and 278 sentences in the development set. Each source sentence in the test set and the development set has 4 different references.

4.2 Tools

In this paper, we use two off-the-shelf commercial English to Chinese RBMT systems to produce the synthetic bilingual corpus.

We also need a trainer and a decoder to perform phrase-based SMT. We use Koehn's training scripts [3] to train the translation model, and the SRILM toolkit (Stolcke, 2002) to train language model. 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 de-

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

[3] It is located at http://www.statmt.org/wmt06/shared-task/baseline.html.

velopment set. The translation quality is evaluated using a well-established automatic measure: BLEU score (Papineni et al., 2002). We use the same method described in (Koehn and Monz, 2006) to perform the significance test.

5 Experimental Results

5.1 Results on Synthetic Corpus Only

With the monolingual English corpus and the English side of the real bilingual corpus, we translate them into Chinese using the two commercial RBMT systems and produce two synthetic bilingual corpora. With the corpora, we train two synthetic models as described in section 3.1. Based on the synthetic models, we also perform linear interpolation as shown in section 3.2, without the standard models. We tune the interpolation weights using the development set, and achieve the best performance when $\alpha_1 = 0.58$, $\alpha_2 = 0.42$, $\beta_1 = 0.58$, and $\beta_2 = 0.42$. The translation results on the test set are shown in Table 1. Synthetic model 1 and 2 are trained using the synthetic bilingual corpora produced by RBMT system 1 and RBMT system 2, respectively.

Method	BLEU
RBMT system 1	0.1681
RBMT system 2	0.1453
Synthetic Model 1	0.1644
Synthetic Model 2	0.1668
Interpolated Synthetic Model	0.1878

Table 1. Translation Results Using Synthetic Bilingual Corpus

From the results, it can be seen that the interpolated synthetic model obtains the best result, with an absolute improvement of the 0.0197 BLEU (11.7% relative) as compared with RBMT system 1, and 0.0425 BLEU (29.2% relative) as compared with RBMT system 2. It is very promising that our method can build an SMT system that significantly outperforms both of the two RBMT systems, using the synthetic bilingual corpus produced by two RBMT systems.

5.2 Results on Real and Synthetic Corpus

With the real bilingual corpus, we build a standard model. We interpolate the standard model with the two synthetic models built in section 5.1 to obtain interpolated models. The translation results are shown in Table 2. The interpolation coefficients are both for phrase table probabilities and lexical weights. They are also tuned using the development set.

From the results, it can be seen that all the three interpolated models perform not only better than the RBMT systems but also better than the SMT system trained on the real bilingual corpus. The interpolated model combining the standard model and the two synthetic models performs the best, achieving a statistically significant improvement of about 0.0245 BLEU (13.1% relative) as compared with the standard model with no synthetic corpus. It also achieves 26.1% and 45.8% relative improvement as compared with the two RBMT systems respectively. The results indicate that using the corpus produced by RBMT systems, the performance of the SMT system can be greatly improved. The results also indicate that the more the RBMT systems are used, the better the translation quality is.

Interpolation Coefficients			BLEU
Standard model	Synthetic Model 1	Synthetic Model 2	
1	—	—	0.1874
0.90	0.10	—	0.2056
0.86	—	0.14	0.2040
0.70	0.12	0.18	0.2119

Table 2. Translation Results Using Standard and Synthetic Bilingual Corpus

5.3 Effect of Synthetic Corpus Size

To explore the relationship between the translation quality and the scale of the synthetic bilingual corpus, we interpolate the standard model with the synthetic models trained with synthetic bilingual corpus of different sizes. In order to simplify the procedure, we only use RBMT system 1 to translate the 1,087,651 monolingual English sentences to produce the synthetic bilingual corpus.

We randomly select 20%, 40%, 60%, 80%, and 100% of the synthetic bilingual corpus to train different synthetic models. The translation results of the interpolated models are shown in Figure 1. The results indicate that the larger the synthetic bilingual corpus is, the better translation performance would be.

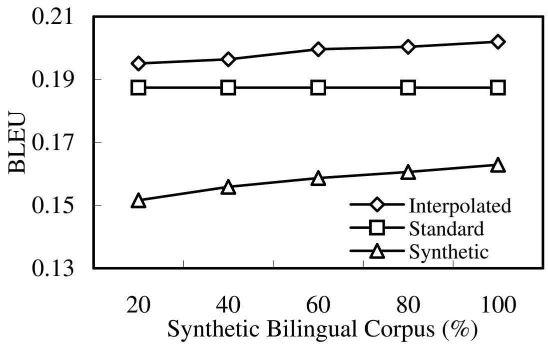


Figure 1. Comparison of Translation Results Using Synthetic Bilingual Corpus of Different Sizes

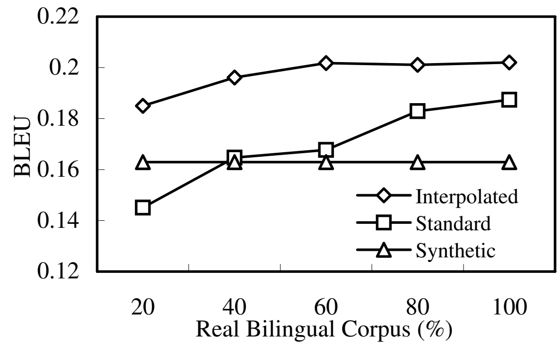


Figure 2. Comparison of Translation Results Using Real Bilingual Corpus of Different Sizes

5.4 Effect of Real Corpus Size

Another issue is the relationship between the SMT performance and the size of the real bilingual corpus. To train different standard models, we randomly build five corpora of different sizes, which contain 20%, 40%, 60%, 80%, and 100% sentence pairs of the real bilingual corpus, respectively. As to the synthetic model, we use the same synthetic model 1 that is described in section 5.1. Then we build five interpolated models by performing linear interpolation between the synthetic model and the five standard models respectively. The translation results are shown in Figure 2.

From the results, we can see that the larger the real bilingual corpus is, the better the performance of both standard models and interpolated models would be. The relative improvement of BLEU scores is up to 27.5% as compared with the corresponding standard models.

5.5 Results without Additional Monolingual Corpus

In all the above experiments, we use an additional English monolingual corpus to get more synthetic bilingual corpus. We are also interested in the results without the additional monolingual corpus. In such case, the only English monolingual corpus is the English side of the real bilingual corpus. We use this smaller size of monolingual corpus and the real bilingual corpus to conduct similar experiments as in section 5.2. The translation results are shown in Table 3.

From the results, it can be seen that our method works well even if no additional monolingual corpus is available. We achieve a statistically significant improvement of about 0.01 BLEU (5.2% relative) as compared with the standard model without using the synthetic corpus.

Interpolation Coefficients			BLEU
Standard model	Synthetic Model 1	Synthetic Model 2	
1	—	—	0.1874
—	1	—	0.1560
—	—	1	0.1522
0.80	0.10	0.10	0.1972

Table 3. Translation Results without Additional Monolingual Corpus

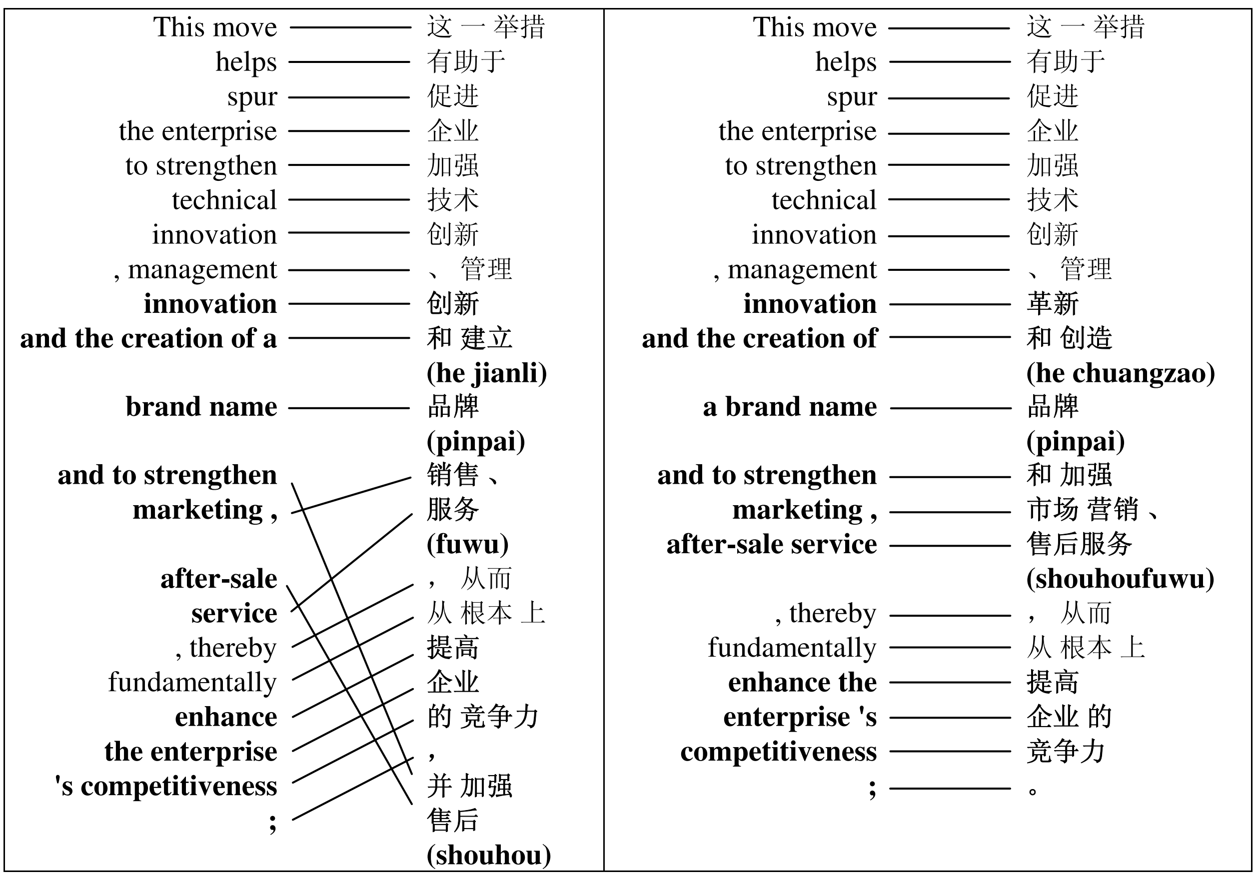
	Standard Model	Synthetic Model 1	Synthetic Model 2
Standard Model	6,105,260	—	—
Synthetic Model 1	356,795	12,062,068	—
Synthetic Model 2	357,489	881,921	9,216,760

Table 4. Numbers of Phrase Pairs

In order to further analyze the translation results, we examine the overlap and the difference among the phrase tables. The analytic results are shown in Table 4. More phrase pairs are extracted by the synthetic models, about twice by the synthetic model 1 in particular, than those extracted by the standard model. The overlap between each model is very low. For example, about 6% phrase pairs extracted by the standard model make appearance in both the standard model and the synthetic model 1. This also explains why the interpolated model outperforms that of the standard model in Table 3.

Methods	English Sentence / Chinese Translations	BLEU
	This move helps spur the enterprise to strengthen technical innovation, management innovation and the creation of a brand name and to strengthen marketing, after-sale service, thereby fundamentally enhance the enterprise's competitiveness;	
Standard model	这 一 举措 有助于 促进 企业 加强 技术 创新 、 管理 创新 和 建立 品牌 销售 、 服务 ， 从而 从 根本 上 提高 企业 的 竞争力 ， 并 加强 售后	0.5022
RBMT System 1	这种 行动 帮助 刺激 企业 加强 技术 地 革新 ， 管理 革新 和 创造 一个 名牌 并且 加强 销售 ， 在 销售 服务 ， 基本上 进而 提高 企业 的 竞争 。	0.1535
RBMT System 2	这项 行动 帮助 刺激 这家 企业 加强 技术 发明 、 管理 创新 和 一 个 商标 的 创造 并 加强 市场 销售 ， 因此 售后服务 根本 增强 这家 企业 的 竞争 。	0.1485
Interpolated Model	这 一 举措 有助于 促进 企业 加强 技术 创新 、 管理 革新 和 创造 品牌 和 加强 市场 营销 、 售后服务 ， 从而 从 根本 上 提高 企业 的 竞争力 。	0.7198

Table 5. Translation Example



(a) Results Produced by the Standard Model (b) Results Produced by the Interpolated Model

Figure 3. Phrase Pairs Used for Translation

6 Discussion

6.1 Model Interpolation vs. Corpus Merge

In section 5, we make use of the real bilingual corpus and the synthetic bilingual corpora by performing model interpolation. Another available way is directly combining these two kinds of corpora to train a translation model, namely corpus merge. In order to compare these two methods, we use RBMT system 1 to translate the 1,087,651 monolingual English sentences to produce synthetic bilingual corpus. Then we train an SMT system with the combination of this synthetic bilingual corpus and the real bilingual corpus. The BLEU score of such system is 0.1887, while that of the model interpolation system is 0.2020. It indicates that the model interpolation method is significantly better than the corpus merge method.

6.2 Result Analysis

As discussed in Section 5.5, the number of the overlapped phrase pairs among the standard model and the synthetic models is very small. The newly added phrase pairs from the synthetic models can assist to improve the translation results of the interpolated model. In this section, we will use an example to further discuss the reason behind the improvement of the SMT system by using synthetic bilingual corpus. Table 5 shows an English sentence and its Chinese translations produced by different methods. And Figure 3 shows the phrase pairs used for translation. The results show that imperfect translations of RBMT systems can be also used to boost the performance of an SMT system.

	Phrase Pairs	Phrase Pairs Used	New Pairs Used
Standard Model	6,105,260	5,509	—
Interpolated Model	73,221,525	5,306	1993

Table 6. Statistics of Phrase Pairs

Further analysis is shown in Table 6. After adding the synthetic corpus produced by the RBMT systems, the interpolated model outperforms the standard models mainly for the following two reasons: (1) some new phrase pairs are added into the interpolated model. 37.6% phrase pairs (1993 out of 5306) are newly learned and used for translation. For example, the phrase pair "after-sale service <-> 售后服务 (shouhoufuwu)" is added; (2) The probability distribution of the phrase pairs is changed. For example, the probabilities of the two pairs "a brand name <-> 品牌 (pinpai)" and "and the creation of <-> 和 创造 (he chuangzao)" increase. The probabilities of the other two pairs "brand name <-> 品牌 (pinpai)" and "and the creation of a <-> 和建立 (he jianli)" decrease. We found that 930 phrase pairs, which are also in the phrase table of the standard model, are used by the interpolated model for translation but not used by the standard model.

6.3 Human Evaluation

According to (Koehn and Monz, 2006; Callison-Burch et al., 2006), the RBMT systems are usually not adequately appreciated by BLEU. We also manually evaluated the RBMT systems and SMT systems in terms of both adequacy and fluency as defined in (Koehn and Monz, 2006). The evaluation results show that the SMT system with the interpolated model, which achieves the highest BLEU scores in Table 2, achieves slightly better adequacy and fluency scores than the two RBMT systems.

7 Conclusion and Future Work

We presented a method using the existing RBMT system as a black box to produce synthetic bilingual corpus, which was used as training data for the SMT system. We used the existing RBMT system to translate the monolingual corpus into a synthetic bilingual corpus. With the synthetic bilingual corpus, we could build an SMT system even if there is no real bilingual corpus. In our experiments using BLEU as the metric, such a system achieves a relative improvement of 11.7% over the best RBMT system that is used to produce the synthetic bilingual corpora. It indicates that using the existing RBMT systems to produce a synthetic bilingual corpus, we can build an SMT system that outperforms the existing RBMT systems.

We also interpolated the model trained on a real bilingual corpus and the models trained on the synthetic bilingual corpora, the interpolated model achieves an absolute improvement of 0.0245 BLEU score (13.1% relative) as compared with the individual model trained on the real bilingual corpus. It indicates that we can build a better SMT system by leveraging the real and the synthetic bilingual corpus.

Further result analysis shows that after adding the synthetic corpus produced by the RBMT systems, the interpolated model outperforms the standard models mainly because of two reasons: (1) some new phrase pairs are added to the interpolated model; (2) the probability distribution of the phrase pairs is changed.

In the future work, we will investigate the possibility of training a reverse SMT system with the RBMT systems. For example, we will investigate to train Chinese-to-English SMT system based on natural English and RBMT-generated synthetic Chinese.

References

Hiyan Alshawi, Srinivas Bangalore, and Shona Douglas. 2000. Learning Dependency Translation Models as Collections of Finite-State Head Transducers. *Computational Linguistics*, 26(1): 45-60.

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.

Chris Callison-Burch and Miles Osborne. 2003. Bootstrapping Parallel Corpora. In *Proceedings of the Human Language Technology conference / North American chapter of the Association for Computational Linguistics (HLT/NAACL-2003) Workshop on Building and Using Parallel Texts: Data Driven Machine Translation and Beyond*, pages 44-49.

Chris Callison-Burch, Miles Osborne and Philipp Koehn, 2006. Re-evaluating the Role of Bleu in Machine Translation Research. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, pages 249-256.

Michel Carl, Paul Schmidt, and Jorg Schutz. 2005. Reversible Template-based Shake & Bake Generation. In *Proceedings of the 10th Machine Translation Summit Workshop on Example-Based Machine Translation*, pages 17-25.

David Chiang. 2005. A Hierarchical Phrase-Based Model for Statistical Machine Translation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL-2005)*, pages 263-270.

Robert Frederking and Sergei Nirenburg. 1994. Three Heads Are Better Than One. In *Proceedings of the 4th Applied Natural Language Processing Conference (ANLP-1994)*, pages 95-100.

Declan Groves and Andy Way. 2006. Hybridity in MT: Experiments on the Europarl Corpus. In *Proceedings of the 11th Annual Conference of the European Association for Machine Translation (EAMT-2006)*, pages 115-124.

Nizar Habash, Bonnie Dorr, and Christof Monz. 2006 Challenges in Building an Arabic-English GHMT System with SMT Components. In *Proceedings of the 11th Annual Conference of the European Association for Machine Translation (EAMT-2006)*, pages 56-65.

Shyamsundar Jayaraman and Alon Lavie. 2005. Multi-engine Machine Translation Guided by Explicit Word Matching. In *Proceedings of the 10th Annual Conference of the European Association for Machine Translation (EAMT-2005)*, pages 143-152.

Philipp Koehn. 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 (AMTA-2004)*, pages 115-124.

Philipp Koehn and Christof Monz. 2006. Manual and Automatic Evaluation of Machine Translation between European Languages. In *Proceedings of the Human Language Technology conference / North American Chapter of the Association for Computational Linguistics (HLT/NAACL-2006) Workshop on Statistical Machine Translation*, pages 102-121.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical Phrase-Based Translation. In *Proceedings of the Human Language Technology conference / North American Chapter of the Association for Computational Linguistics (HLT/NAACL-2003)*, pages 127-133.

Daniel Marcu. 2001. Towards a Unified Approach to Memory- and Statistical-based Machine Translation. In *Proceedings of the Association for Computational Linguistics / European Chapter of the Association for Computational Linguistics (ACL/EACL-2001)*, pages 378-385.

Evgeny Matusov, Nicola Ueffing, and Hermann Ney. 2006. Computing Consensus Translation from Multiple Machine Translation Systems Using Enhanced Hypotheses Alignment. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, pages 33-40.

Bart Mellebeek, Karolina Owczarzak, Josef Van Genabith, and Andy Way. 2006. Multi-engine Machine Translation by Recursive Sentence Decomposition. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas (AMTA-2006)*, pages 110-118.

Arul Menezes and Chris Quirk. 2005. Dependency treelet translation: the convergence of statistical and example-based machine-translation? In *Proceedings of the 10th Machine Translation Summit Workshop on Example-Based Machine Translation*, pages 99-108.

Tadashi Nomoto. 2004. Multi-Engine machine translation with voted language model. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-2004)*, pages 494-501.

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 (ACL-2003)*, pages 160-167.

Franz Josef Och and Hermann Ney. 2004. The Alignment Template Approach To Statistical Machine Translation. *Computational Linguistics*, 30(4):417-449.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL-2002)*, pages 311-318.

Stephanie Seneff, Chao Wang, and John Lee. 2006. Combining Linguistic and Statistical Methods for Bi-Directional English Chinese Translation in the Flight Domain. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas (AMTA-2006)*, pages 213-222.

Andreas Stolcke. 2002. SRILM - An Extensible Language Modeling Toolkit. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP-2002)*, pages 901-904.

Nicola Ueffing. 2006. Using Monolingual Source-Language Data to Improve MT Performance. *In Proceedings of the International Workshop on Spoken Language Translation (IWSLT-2006)*, pages 174-181.

Vincent Vandeghinste, Ineka Schuurman, Michael Carl, Stella Markantonatou, and Toni Badia. 2006. Metis-II: Machine Translation for Low-Resource Languages. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (L-REC-2006)*, pages 1284-1289.

Stephan Vogel and Hermann Ney. 2000. Construction of a Hierarchical Translation Memory. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING-2000)*, pages 1131-1135.

Andy Way and Nano Gough. 2005. Comparing Example-Based and Statistical Machine Translation. *Natural Language Engineering*, 11(3): 295-309.

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

Kenji Yamada and Kevin Knight. 2001. A Syntax Based Statistical Translation Model. In *Proceedings of the Association for Computational Linguistics / European Chapter of the Association for Computational Linguistics (ACL/EACL-2001)*, pages 523-530.

Menno van Zaanen and Harold Somers. 2005. DEMOCRAT: Deciding between Multiple Outputs Created by Automatic Translation. In *Proceedings of the 10th Machine Translation Summit*, pages 173-180.