Synonym-Based Reordering Model for Statistical Machine Translation

Zhenxin Yang[1,2(✉)], Miao Li[1], Lei Chen[1], and Kai Sun[2]

[1] Institute of Intelligent Machines,
Chinese Academy of Sciences, Hefei 230031, China
mli@iim.ac.cn, alan.cl@163.com
[2] University of Science and Technology of China, Hefei 230026, China
{xinzyang, sasunkai}@mail.ustc.edu.cn

Abstract. Reordering model is the crucial component in statistical machine translation (SMT), since it plays an important role in the generation of fluent translation results. However, the data sparseness is the key factor that greatly affects the performance of reordering model in SMT. In this paper, we exploit synonymous information to alleviate the data sparseness and take Chinese-Mongolian SMT as example. First, a synonym-based reordering model with Chinese synonym is proposed for Chinese-Mongolian SMT. Then, we flexibly integrate synonym-based reordering model into baseline SMT as additional feature functions. Besides, we present source-side reordering as the preprocessing module to verify the extensibility of our synonym-based reordering model. Experiments on the Chinese-Mongolian dataset show that our synonym-based reordering model achieves significant improvement over baseline SMT system.

Keywords: Synonym · Reordering · Statistical machine translation · Feature function

1 Introduction

Statistical machine translation (SMT), which translates one language into another language automatically by statistical training corpus, breaks the barrier of different languages [1]. Since the fluency of translation results is one of the key evaluation metric in SMT, modeling word order between source and target sentences has been a research focus since the emerging of statistical machine translation [2].

The past decade has witnessed the rapid development of phrase reordering models. Among them, lexicalized reordering models [3–6] and syntax-based reordering models [7–10] have been widely used in practical phrase-based SMT systems.

Lexicalized reordering models have been widely researched in the phrase-based machine translation systems. Generally speaking, they take advantage of lexical information to predict the orientation of current phrase pair by using word alignment sentences. They often distinguish three orientation types of current phrase pair with respect to context: *monotone*, *swap*, and *discontinuous*. However, lexicalized reordering models suffer the data sparseness problem when translating with low-resource languages, such as Mongolian.

D.-S. Huang et al. (Eds.): ICIC 2016, Part III, LNAI 9773, pp. 369–378, 2016.
DOI: 10.1007/978-3-319-42297-8_35

Source-side reordering models [7–9] as the pre-processing modules are the effective syntax-based reordering models. These approaches first parse the source language sentences to create parse trees. Then, syntactic reordering rules, either hand-written or extracted automatically, are applied to these parse trees to adjust the source word order to match the target word order. This method is usually done in a pre-processing step, and then followed by a standard phrase-based SMT system to finish the translation. However, syntax-based reordering models will be affected by the accuracy of the parser easily.

The data sparseness and order difference bring great challenge to phrase reordering in Chinese-Mongolian SMT. Generally speaking, Mongolian is the low-resource minority language in China, so the training data used in SMT system is scarce. Besides, Mongolian is the Subject-Object-Verb (SOV) structure while Chinese is the Subject-Verb-Object (SVO) structure. Figure 1 illustrates word order difference between Chinese and Mongolian.

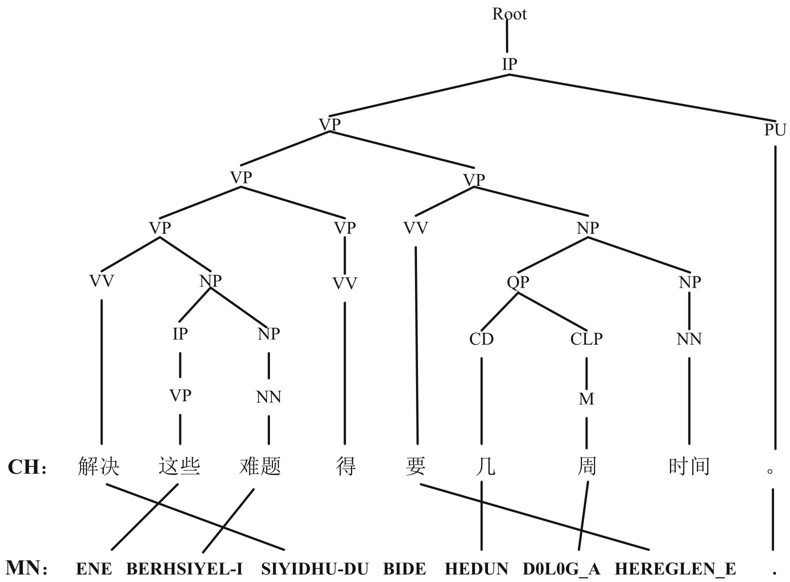


Fig. 1. Illustration of word order difference.

To address the challenge, a novel and effective synonym-based reordering model is proposed to alleviate the data sparseness by making full use the Chinese synonymous information. We integrate synonym-based reordering model into baseline SMT flexibly by additional feature functions. Finally, to further verify the effectiveness and extensibility of our method, some comparable experiments are conducted to demonstrate the effectiveness of our method.

The remainder of this article is organized as follows. Section 2 describes the background of baseline SMT and Sect. 3 provides synonym-based reordering model and its integration method. Extensive experiments are presented in Sect. 4. Section 5 describes the related work. Finally, we conclude in Sect. 6.

2 Background

In this section, we provide some background on baseline statistical machine translation system [11], which has emerged as the dominant paradigm in Chinese-Mongolian SMT research.

Given an input sentence $f = f_1\ f_2\ \ldots\ f_m$ which is to be translated into a target sentence $e = \mathrm{e}_1\ e_2\ \ldots\ e_{n,}$ the baseline SMT searches the most probable translation e^* according to the following decision rule:

$$e^* = \arg\max_{e} p(e|f) = \arg\max_{e}\{\sum_{j=1}^{J} \lambda_j h_j(f,e)\} \tag{1}$$

where $h_j(f,e)$ are J arbitrary feature functions over sentence pairs and λ_j are corresponding feature weights. The feature functions are soft constraints to encourage the decoder to choose correct translations during the decoding step. The feature weights can be learned by Minimum Error Rate Training [12] (MERT) with a separate development data.

Given a set of source sentences F_1^s with corresponding reference translations R_1^s, the goal of MERT is to find a parameter set which minimizes an automated evaluation criterion under a log-linear model:

$$\widehat{\lambda}_1^M = \arg\min_{\lambda_1^M}\left\{\sum_{s=1}^{S} Err\left(R_s, \widehat{E}(F_s;\lambda_1^M)\right)\right\} \tag{2}$$

$$\widehat{E}(F_s;\lambda_1^M) = \arg\max_{E}\left\{\sum_{s=1}^{S} \lambda_m h_m(E, F_s)\right\} \tag{3}$$

The translation results are evaluated automatically by BLEU metric [13], which has been demonstrated by showing that it correlates with human judgment. BLEU compares n-grams of the candidate with the n-grams of the reference translation and count the number of matches. Given the precision p_n of n-grams of size up to N (usually N = 4), the length of the test set in words c and the length of the reference translation in words r, the BLEU is computed as follows:

$$BLEU = \min(1, e^{1-r/c}) * \exp\sum_{n=1}^{N} \log p_n \tag{4}$$

3 Reordering Model with Synonym

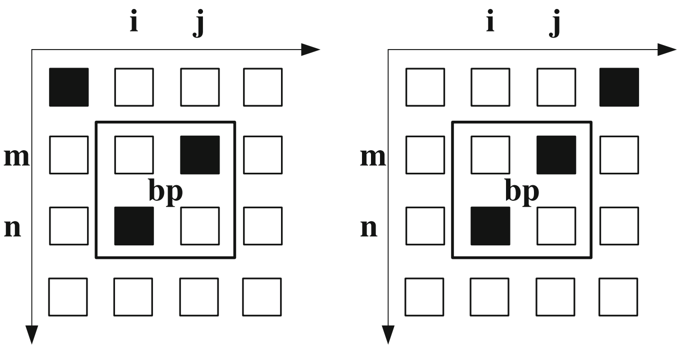
In this section, we will introduce lexicalized reordering model used in baseline SMT and its drawback at first and then describe our synonym-based reordering model and its integration method.

3.1 Lexicalized Reordering Model

Lexicalized reordering model [4] calculates the reordering orientation of current phrase pairs based on previous word and next word during the phrase extraction. The model takes advantage of lexical information to predict the orientation of current phrase pair by using word alignment sentences. More formally, given a source sentence S, a target sentence T and its word alignment set A, we use a_s^t to denote the alignment from source position s to target position t, S_i^j to denote the phrase between position i and j in S and T_m^n to denote the phrase between position m and n in T. A source phrase S_i^j and a target phrase T_m^n form a phrase pair, lexicalized reordering model aims to predict orientations of a given phrase pair. The most widely used orientations in practice are *monotone* (M), *swap* (S), and *discontinuous* (D), the phrase pair can be classified to one of three orientations with respect to the previous word:

- *Monotone*: if $a_{i-1}^{m-1} \in A \cap a_{j+1}^{m-1} \notin A$
- *Swap*: if $a_{j+1}^{m-1} \in A \cap a_{i-1}^{m-1} \notin A$
- *Discontinuous*: if $(a_{i-1}^{m-1} \in A \cap a_{j+1}^{m-1} \in A) \cup (a_{i-1}^{m-1} \notin A \cap a_{j+1}^{m-1} \notin A)$

The orientations with respect to the next word are analogous. Hence, six lexicalized reordering feature functions are exploited in the baseline SMT. Figure 2 provides an example of the orientations with previous word.



(a) *monotone* **(b)** *swap*

Fig. 2. Reordering orientation with previous word.

The probability of a phrase pair bp in the orientation o is calculated as follows:

$$p(o|bp) = \frac{Count(o, bp) + \alpha}{\sum_{o'} (Count(o', bp) + \alpha)} \tag{5}$$

where α is the smoothing value, and we set α to 0.5 in practice.

We can conclude that maximum likelihood estimation can hardly train the lexicalized reordering model accurately because of the data sparseness in Chinese-Mongolian SMT.

3.2 Synonym-Based Reordering Model

In this subsection, a novel and effective synonym-based method will be presented to alleviate the data sparseness in phrase reordering. Inspired by the work [14], which used source language information to handle unknown words in SMT, we exploit monolingual information rather than bilingual information to enrich the reordering instances. Much reordering instances will lead the better probabilities estimation.

Intuitively, words with same meaning have similar reordering rules. Therefore, we use the Chinese synonym to calculate reordering probabilities, which assume that a word and its synonyms can share their reordering instances extracted from training data. We utilize the synonym dictionary named HIT IR-Lab Tongyici Cilin[1] (Extended) to train a synonym-based reordering model.

Synonym-based approach is modeled only condition on source language and synonym dictionary. If A and B are synonymous phrases, the count of B is also used as additional information when calculating the reordering probability of A. The training procedure is similar with the model conditioned on both source and target languages. The probability of a source phrase *sp* in the orientation *o* is calculated as follows:

$$p(o|sp) = \frac{Count(o, sp) + \alpha}{\sum_{o'} (Count(o', sp) + \alpha)} \tag{6}$$

where α is the smoothing value, and we set α to 0.5 like previous equation.

Note that we only have synonym rather than synonymous phrase, so our dictionary is helpless for a phrase with more than one word. However, large percentage of translation units used during decoding are word-by-word or phrases including two words even in a phrase-based system [15]. Therefore, we only consider synonymous information for phrases including one or two words, and ignore the situation of long, uncommonly used phrases. We use a simple method to expand our synonym to synonymous phrase which has only two words.

The key idea of our synonym-based reordering model is that synonymous phrases share the reordering instances, so the sufficient data is used for probabilities calculation. For a phrase including two words, if one of the words in phrase has a synonym, we just use its synonym to replace this word to construct a synonymous phrase. For example, if we have a phrase "飞快 地(feikuai de)", and the word "飞快(feikuai)" has a synonym "迅速(xunsu)" in the dictionary, then the previous phrase has a synonymous phrase "迅速 地(xunsu de)". Both two phrases are found in our training data. The data sparseness can be alleviated since both two phrases can share the reordering instances.

3.3 Integration into Baseline SMT

The model assigns three distinct parameters (λ_m, λ_s, λ_d) for the three feature functions with respect to previous word:

[1] http://ir.hit.edu.cn/.

$$f_m = \sum\nolimits_{i=1}^{n} \log p(o_i = M|\cdots) \tag{7}$$

$$f_s = \sum\nolimits_{i=1}^{n} \log p(o_i = S|\cdots) \tag{8}$$

$$f_d = \sum\nolimits_{i=1}^{n} \log p(o_i = D|\cdots) \tag{9}$$

Three feature functions with respect to the next word are analogous. Hence, the above six additional feature functions are added into the log-linear model of the baseline SMT system. Essentially, they are soft constraints to encourage the decoder to choose translations with the correct word order.

4 Experiment

4.1 Data and Setup

The training set which contains 67288 Chinese-Mongolian parallel sentences is obtained from the 5th China Workshop on Machine Translation (CWMT). The test set contains 400 sentences, where each source sentence has four reference sentences translated by native linguistics experts, since the correct translation results are not unique. Besides, the development set is the same with the test set.

We employed GIZA++[2] and the grow-diag-final-and [11] balance strategy to generate the final symmetric word alignments. A 3-gram language model with modified Kneser-Ney smoothing [16] is built by SRILM toolkit. The final translation quality is evaluated in terms of BLEU metric. The feature weights are learned by using Minimum Error Rate Training [12] (MERT) algorithm. We use toolkit ICTCLAS[3] for Chinese word segmentation. The open-source toolkit Moses is used for each translation task. Besides, maximum phrase length is set to 7 when extracting bilingual phrase pairs. We run each experiment 3 times and get the average BLEU score [13] as the experimental result.

4.2 Evaluation for Synonym-Based Approach

We compare the synonym-based reordering model with the baseline. Besides, in order to compare our proposed model with previous work, we re-implement two approaches, which are denoted as system C and system D respectively. Table 1 shows the results. System A is baseline SMT, which is widely used in Chinese-Mongolian SMT research. System B denotes that we integrate our proposed synonym-based reordering model into baseline SMT by six additional feature functions. System C [6] is a improvement of lexicalized reordering model by weighted alignment matrices. System D [9] is a source-side reordering model which is also a research focus in SMT.

[2] http://code.google.com/p/giza-pp/.

[3] http://ictclas.nlpir.org/.

Table 1. Experimental results of different translation methods.

System	BLEU(%)
A	24.05
B	**25.27**
C	24.73
D	25.01

From Table 1, it can be noted that our synonym-based reordering model can substantially improve the translation quality compared with the baseline SMT. Besides, our proposed method is better than some typical related work, demonstrating the synonym-based reordering model can alleviate the data sparseness in phrase reordering.

4.3 Source-Side Reordering Model as Pre-processing Module

To further evaluate the effectiveness and extensibility of the proposed synonym-based reordering model, we incorporate a source-side reordering model [9] as a pre-processing module into system B. A phrase structure tree with syntactic information is acquired by stanford parser [17], then we use the manual rules to reorder the source language. The reordering rules described in Table 2, where VP denotes verb phrase, PP denotes prepositional phrase, NP denotes noun phrase, VV denotes verb.

Table 2. Source-side reordering rule.

No.	Original rule	Reordering rule
(1)	VP—> VV PP	VP—> PP VV
(2)	VP—> VV NP	VP—> NP VV

The subtrees are constructed to match with the original linguistic rules. The corresponding reordering rules on these subtrees are exploited to swap their left branches and the right branches. Then, the reordering of the source sentence is achieved.

Figure 3 is a source-side reordering of Fig. 1. The final BLEU scores are given in Table 3.

Table 3. Translation results of combination of two models.

System	BLEU(%)
A	24.05
E	**25.53**

From Table 3, it can be noted that the combination of synonym-based reordering model and source-side reordering model can further improve the translation quality with a maximum improvement of 1.48 BLEU score points increment over baseline. Hence, our synonym-based approach is easily incorporate with other module.

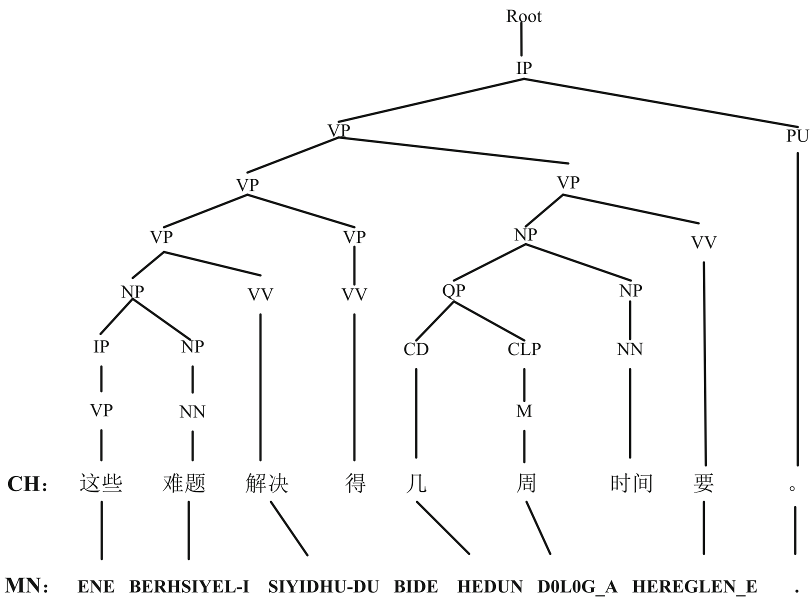


Fig. 3. The reordered phrase structure tree.

5 Related Work

Phrase reordering for statistical machine translation has attracted much attention in recent years. Within the phrase-based SMT framework, such research efforts can be roughly divided into two groups: (1) lexicalized reordering models, which model reordering relationships between adjacent phrases; (2) syntax-based reordering models, which apply some syntactic reordering rules on the phrase structure subtree to select right word order.

In lexicalized reordering work, the classifier will make decision on next phrase's relative position with the context. The classifier can be trained with maximum likelihood like Moses lexicalized reordering [4] and hierarchical lexicalized reordering model [5] or be trained under maximum entropy framework [18, 19]. Some improvement methods on lexicalized reordering, such as graph-based lexicalized reordering model [20], weighted alignment matrices reordering [6], are proposed to further improve the performance of reordering model. Similar work [21] based on maximum entropy exploited Mongolian morphological information as features to train a robust maximum entropy based reordering model. However, all these methods mentioned above are suffered a serious data sparseness in Chinese-Mongolian SMT.

Another direction is to employ syntactic information for word order selection. These methods need a statistical parser to produce the grammatical structure of sentences. Among them, source-side reordering is one of the research focus. In source-side approaches, the source sentence is reordered by heuristics, so that the word order of source and target sentences is similar. Liang et al. [9] proposed a rule-based reordering, which applied some syntactic reordering rules on the phrase structure subtree to reorder source language. Similar work [10] used dependence information to extracted reordering rules automatically. Yang et al. [22] utilized a ranking model based on word order precedence in the target language to reposition nodes in the syntactic parse tree of a source sentence. Visweswariah et al. [23] presented an automatic method to learn rules using only a source side parse tree and automatically generated alignments. However, syntax-based models are easily affected by the accuracy of the statistical parser.

Inspired by the work [14], which used source language information to handle unknown words in statistical machine translation, we make full use of monolingual information to alleviate the data sparseness in reordering model. To the best of our knowledge, our work is the first attempt to exploit the source synonym information to alleviate the data sparseness in reordering model, and we have successfully incorporated the proposed model into an SMT system with significant improvement of BLEU metrics.

6 Conclusion

In this paper, we presented a novel synonym-based reordering model to improve translation quality. We first utilize Chinese synonymous information to alleviate the data sparseness by enrich reordering instances. Then a novel and effective synonym-based reordering model is integrated into current baseline SMT by additional feature functions. The new additional features embedded in the log-linear model can encourage the decoder to produce more fluent translation results. Besides, we employ source-side reordering as a pre-processing module to verify the extensibility of our approach. Our experimental results demonstrate that our approach can significantly improve the translation quality.

Our contributions can be summarized as follows:

(1) To our knowledge, our work is the first attempt to utilize synonym dictionary to alleviate the data sparseness in reordering model while the dictionary is always used for translation model on SMT research.
(2) We investigate the flexible and efficient integration method, which integrates synonym-based reordering model as additional features into current SMT system.
(3) Our model is a general method, besides Chinese-Mongolian SMT, the method can also be adopted for other language pairs.

Acknowledgement. This work is supported by the National Natural Science Foundation of China under No. 61572462, No. 61502445, the Informationization Special Projects of Chinese Academy of Science under No. XXH12504-1-10.

References

1. Tu, M., Zhou, Y., Zong, C.: Exploring diverse features for statistical machine translation model pruning. IEEE/ACM Trans. Audio Speech Lang Process. **23**(11), 1847–1857 (2015)
2. Farzi, S., Faili, H., Khadivi, S.: A syntactically informed reordering model for statistical machine translation. J. Exp. Theor. Artif. Intell. **27**(4), 449–469 (2015)

3. Tillmann, C.: A unigram orientation model for statistical machine translation. In: HLT-NAACL 2004: Short Papers, pp. 101–104. Association for Computational Linguistics (2004)
4. Koehn, P., Hoang, H., Birch, A., et al: Moses: open source toolkit for statistical machine translation. In: ACL, pp. 177–180. Association for Computational Linguistics (2007)
5. Galley, M., Manning, C.D.: A simple and effective hierarchical phrase reordering model. In: EMNLP, pp. 848–856. Association for Computational Linguistics (2008)
6. Ling, W., Luis, T., Graa, J., Coheur, L., Trancoso, I.: Reordering modeling using weighted alignment matrices. In: ACL-HLT, pp. 450–454. Association for Computational Linguistics (2011)
7. Yeon-Soo, L.E.E.: Utilizing global syntactic tree features for phrase reordering. IEICE Trans. Inf. Syst. **97**(6), 1694–1698 (2014)
8. Chen, L., Li, M., He, M., Liu, H.: Dependency parsing on source language with reordering information in SMT. In: IALP, pp. 133–136 (2012)
9. Liang, F., Chen, L., Li, M.: Nasun-urtu: a rule-based source-side reordering on phrase structure subtrees. In: IALP, pp. 173–176 (2011)
10. Cai, J., Utiyama, M., Sumita, E., et al.: Dependency-based pre-ordering for Chinese-English machine translation. In: ACL, pp. 155–160. Association for Computational Linguistics (2014)
11. Koehn, P., Och, F.J., Marcu, D.: Statistical phrase-based translation. In: NAACL-HLT, pp. 48–54. Association for Computational Linguistics (2003)
12. Och, F.J.: Minimum error rate training in statistical machine translation. In: ACL, pp. 160–167. Association for Computational Linguistics (2003)
13. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a method for automatic evaluation of machine translation. In: ACL, pp. 311–318. Association for Computational Linguistics (2002)
14. Zhang, J., Zhai, F., Zong, C.: A substitution-translation-restoration framework for handling unknown words in statistical machine translation. J. Comput. Sci. Technol. **28**(5), 907–918 (2013)
15. Hoang, H., Koehn, P.: Improving mid-range reordering using templates of factors. In: EMNLP, pp. 372–379. Association for Computational Linguistics (2009)
16. Chen, S.F., Goodman, J.: An empirical study of smoothing techniques for language modeling. In: ACL, pp. 310–318. Association for Computational Linguistics (1996)
17. Levy, R., Manning, C.: Is it harder to parse Chinese, or the Chinese treebank? In: ACL, pp. 439–446. Association for Computational Linguistics (2003)
18. Xiong, D., Liu, Q., Lin, S.: Maximum entropy based phrase reordering model for statistical machine translation. In: ACL, pp. 521–528. Association for Computational Linguistics (2006)
19. He, Z., Meng, Y., Yu, H.: Maximum entropy based phrase reordering for hierarchical phrase-based translation. In: EMNLP, pp. 555–563. Association for Computational Linguistics (2010)
20. Ling, W., Graça, J., de Matos, D.M., Trancoso, I., Black, A.W.: Discriminative phrase-based lexicalized reordering models using weighted reordering graphs. In: IJCNLP, pp. 47–55. Association for Computational Linguistics (2011)
21. Yang, Z., Li, M., Zhu, Z., et al.: A maximum entropy based reordering model for Mongolian-Chinese SMT with morphological information. In: IALP, pp.175–178 (2014)
22. Yang, N., Li, M., Zhang, D., Yu, N.: A ranking-based approach to word reordering for statistical machine translation. In: ACL, pp. 912–920. Association for Computational Linguistics (2012)
23. Visweswariah, K., Navratil, J., Sorensen, J., et al.: Syntax based reordering with automatically derived rules for improved statistical machine translation. In: ICCL, pp. 1119–1127 (2010)