A Retrieve-and-Rewrite Initialization Method for Unsupervised Machine Translation

Shuo Ren[†‡]*, **Yu Wu** [§], **Shujie Liu**[§], **Ming Zhou**[§], **Shuai Ma**[†‡]
[†]SKLSDE Lab, Beihang University, Beijing, China
[‡]Beijing Advanced Innovation Center for Big Data and Brain Computing, China
[§]Microsoft Research Asia, Beijing, China
[†]{shuoren,mashuai}@buaa.edu.cn [§]{Wu.Yu,shujliu,mingzhou}@microsoft.com

Abstract

The commonly used framework for unsupervised machine translation builds initial translation models of both translation directions, and then performs iterative back-translation to jointly boost their translation performance. The initialization stage is very important since bad initialization may wrongly squeeze the search space, and too much noise introduced in this stage may hurt the final performance. In this paper, we propose a novel retrieval and rewriting based method to better initialize unsupervised translation models. We first retrieve semantically comparable sentences from monolingual corpora of two languages and then rewrite the target side to minimize the semantic gap between the source and retrieved targets with a designed rewriting model. The rewritten sentence pairs are used to initialize SMT models which are used to generate pseudo data for two NMT models, followed by the iterative back-translation. Experiments show that our method can build better initial unsupervised translation models and improve the final translation performance by over 4 BLEU scores.

1 Introduction

Recent work has shown successful practices of unsupervised machine translation (UMT) (Artetxe et al., 2017; Lample et al., 2017, 2018; Artetxe et al., 2018b; Marie and Fujita, 2018; Ren et al., 2019; Lample and Conneau, 2019). The common framework is to build two initial translation models (i.e., source to target and target to source) and then do iterative back-translation (Sennrich et al., 2016a; Zhang et al., 2018) with pseudo data generated by each other. The initialization stage is important because bad initialization may wrongly squeeze the search space, and too much noise introduced in this stage may hurt the final performance.

Previous methods for UMT (Lample et al., 2018; Artetxe et al., 2018b; Marie and Fujita, 2018; Ren et al., 2019) usually use the following n-gram embeddings based initialization. They first build phrase translation tables with the help of unsupervised cross-lingual n-gram embeddings (Conneau et al., 2017; Artetxe et al., 2018a), and then use them to build two initial Phrase-based Statistical Machine Translation (PBSMT) (Koehn et al., 2003) models with two language models. However, there are two problems with their initialization methods. (1) Some complex sentence structures of original training sentences are hard to be recovered with the n-gram translation tables. (2) The initial translation tables inevitably contain much noise, which will be amplified in the subsequent process.

In this paper, we propose a novel retrieve-and-rewrite initialization method for UMT. Specifically, we first retrieve semantically similar sentence pairs from monolingual corpora of two languages with the help of unsupervised cross-lingual sentence embeddings. Next, with those retrieved similar sentence pairs, we run GIZA++ (Och and Ney, 2003) to get word alignments which are used to delete unaligned words in the target side of the retrieved sentences. The modified target sentences are then rewritten with a designed sequence-to-sequence rewriting model to minimize the semantic gap between the source and target sides. Taking the pairs of the source sentences and corresponding rewritten targets as pseudo parallel data, we then build two initial PBSMT models (source-to-target and target-to-source), which are used to generate pseudo parallel data to warm up NMT models, followed by an iterative back-translation training process. Our code is released at https://github.com/Imagist-Shuo/RRforUNMT.git.

Our contributions are threefold. (1) We propose a novel method to initialize unsupervised MT models with a retrieve-and-rewrite schema, which can

*Contribution during internship at MSRA.

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 3498–3504
July 5 - 10, 2020.

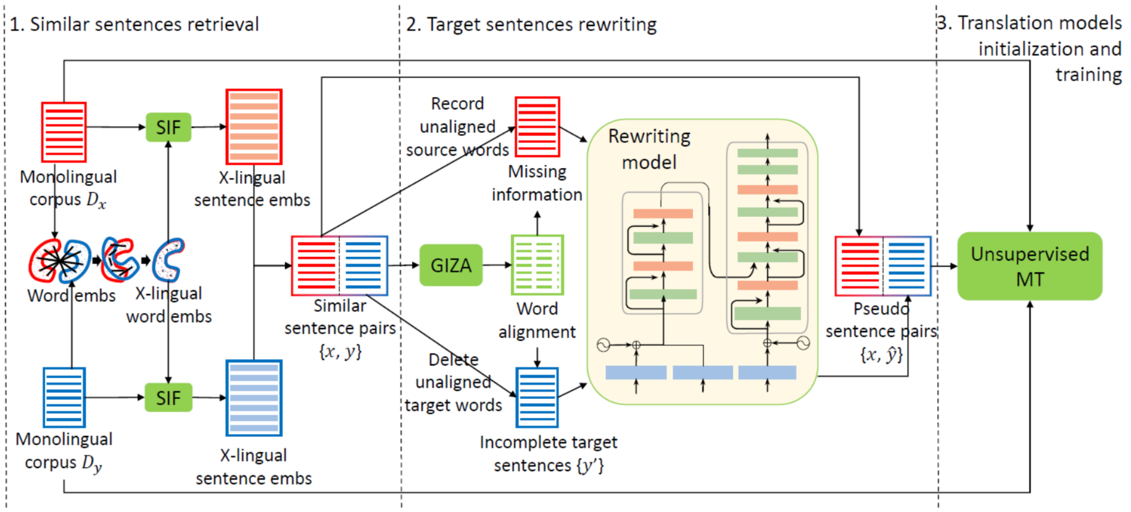

Figure 1: Method overview. (In the figure, "embs" means "embeddings" and "x-lingual" means "cross-lingual".)

preserve the rich sentence structure and provide high-quality phrases. (2) We design an effective seq-to-seq architecture based on the Transformer to rewrite sentences with semantic constraints. (3) Our method significantly outperforms the previous non-pre-training based UMT results on en-fr and en-de translation tasks, and give the first unsupervised en-zh translation results on WMT17.

2 Method

Our method can be divided into **three steps** as shown in Figure 1. First, we do *similar sentences retrieval* (§2.1) from two monolingual corpora with the help of unsupervised cross-lingual sentence embeddings. Next, to minimize the semantic gap between the source and retrieved targets, we do *target sentences rewriting* (§2.2) by deleting unaligned words in the target side, and generate complete and better-aligned targets via our rewriting model with the help of missing information provided by the source. After that, we treat the rewritten pairs as the pseudo parallel data for *translation models initialization and training* (§2.3).

2.1 Similar Sentences Retrieval

Given two monolingual corpora D_x and D_y of two languages X and Y respectively, we first build unsupervised cross-lingual word embeddings of X and Y using fastText (Bojanowski et al., 2017) and vecmap (Artetxe et al., 2018a), and then we obtain cross-lingual sentence embeddings based on the cross-lingual word embeddings via SIF (Arora et al., 2017). After that, we use the marginal-based scoring (Artetxe and Schwenk, 2018) to retrieve

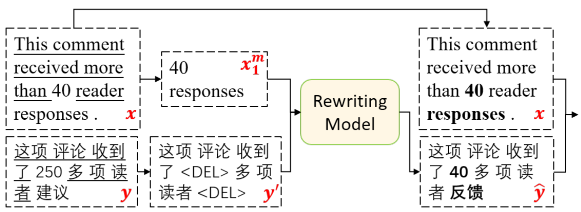

Figure 2: Example of rewriting. The unaligned words, i.e., 250 and 建议(suggestion), proposed by GIZA++ have been removed in y', which is then rewritten by the model to the right target $\hat{y}$ (40 and 反馈(responses)). More examples of the sentences before and after rewriting are shown in Appendix B.

similar sentences from two corpora[1]. Examples retrieved from monolingual English and Chinese corpora are shown in Figure 1 in the Appendix A.

2.2 Target Sentences Rewriting

As shown in Figure 2, having retrieved similar sentence pairs $\{x, y\}$, we first run GIZA++ (Och and Ney, 2003) on these pairs and obtain the word alignment information. Then, for each target sentence y, we remove the unaligned words from it according to lexical translation probabilities of GIZA++ output. We replace each deleted word with $\langle$DEL$\rangle$ in y to get the incomplete target sentence y'. Meanwhile, we record the unaligned words in the source as x_1^m where m is the number of the unaligned source words. Next, we feed y' and x_1^m into a sequence-to-sequence model to generate the refined target sentence $\hat{y}$. The rewritten pairs $\{x, \hat{y}\}$ are

[1]For each source sentence, we choose 30 nearest neighbors in the target language, which have approximately similar lengths to the source (within the difference of ±5 words), and keep the neighbors with the scores more than 0.6.

used as training data to train initial UMT systems.

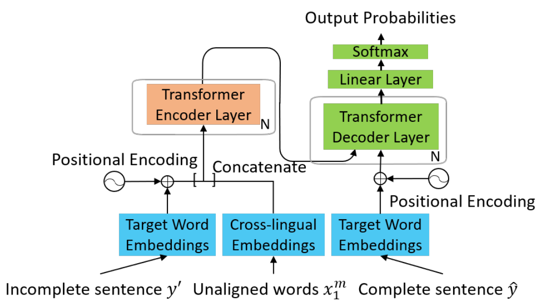

Figure 3: The architecture of the rewriting model. We modify the input of the Transformer encoder into two parts. **The first part** is the incomplete target sentence y', which is the same as the original Transformer input, and **the second part** is a sequence of unaligned source words x_1^m, for which we remove positional encoding because the order of these words is not a concern.

Our rewriting model is a modification of Transformer (Vaswani et al., 2017) shown as Figure 3. We initialize the embedding layer of the second input part with pre-trained cross-lingual word embeddings because its content should be independent of languages. We keep it fixed during training. Thus the second part is like a memory recording semantic information of words. We concatenate the readout embeddings of both parts with a separator, and feed them to the Transformer encoder, so that the attention mechanism will take effect on both parts together. **For model training**, due to the lack of references, we need to build training data for the rewriting model from monolingual corpus D_y. Firstly, we remove 20 to 30 percent of words from a given sentence $y \in D_y$, and replace them with $\langle$DEL$\rangle$ to get y'. Next, we randomly swap contiguous words in y' with the probability of 0.2 to introduce some noises. Then we record the removed words as set s_1^m and randomly drop/add some words from/to this set. We then treat y' and s_1^m as the inputs, and y as the output to train the model. **For model inference**, we feed the incomplete sentence y' and unaligned source words x_1^m into the trained model and generate the refined sentence $\hat{y}$. Note there seems to be a bias between the training and inference that s_1^m during training are in the same language as y, while during inference, they are from the source language X. But the bias has been eliminated since the second input part of the encoder is the readout cross-lingual embeddings, which is independent of languages.

2.3 Translation Models Initialization and Training

Once we get $\{x, \hat{y}\}$ generated above, we use them to train initial PBSMT models, and use the SMT models to produce pseudo data to setup two NMT models, followed by the iterative back-translation.

3 Experiments

3.1 Setup

Dataset

In our experiments, we consider three language pairs, English-French (en-fr), English-German (en-de) and English-Chinese (en-zh). For en, fr and de, we use 50 million monolingual sentences in NewsCrawl from 2007 to 2017. As for zh, we use the Chinese side from WMT17 en-zh parallel data.[2] For the convenience of comparison, we use *newstest* 2014 as the test set for en-fr, *newstest* 2016 for en-de, and *newstest* 2017 for en-zh. The data preprocessing is described in Appendix D.

Baselines

Our method is compared with eight baselines of unsupervised MT systems listed in the upper area of Table 1. The first three baselines are unsupervised NMT models, and the fourth baseline is an unsupervised PBSMT model. The fifth baseline is an extract-and-edit schema for unsupervised neural machine translation. The sixth and seventh baselines are hybrid models of NMT and PBSMT. And the last baseline is a pre-training based method.

3.2 Results

Overall Results

The comparison results are reported in Table 1. From the table, we find that our method significantly outperforms the best non-pre-training based baseline with an average of 4.63 BLEU scores on all pairs. Note that Lample and Conneau (2019) is based on pre-training, which uses much more monolingual data than our method. Even so, we reach comparable results on the en-fr pair.

Comparison of Initial SMT Models

We compare the performance of SMT models initialized with different methods in Table 2. All

[2]Note that we only retrieve similar sentences from sampled 20 million sentences in each monolingual corpus and use Hierarchical Navigable Small World (HNSW) (Malkov and Yashunin, 2018) to build embedding index for space and time efficiency. During the iterative back-translation process in §2.3, we use the whole monolingual corpora.

Method	fr2en	en2fr	de2en	en2de	zh2en	en2zh
(Artetxe et al., 2017)	15.6	15.1	-	-	-	-
(Lample et al., 2017)	14.3	15.1	13.3	9.6	-	-
(Yang et al., 2018)	15.6	17.0	14.6	10.9	-	-
(Artetxe et al., 2018b)	25.9	26.2	23.1	18.2	-	-
(Wu et al., 2019)	26.9	27.6	23.3	19.6	-	-
(Lample et al., 2018)	27.7	28.1	25.2	20.2	-	-
(Ren et al., 2019)	28.9	29.5	26.3	21.7	11.2	18.7
(Lample et al.,2019)*	33.3	33.4	**34.3**	**26.4**	-	-
Ours	**33.3**	**34.0**	31.6	26.0	**15.3**	**23.9**

Table 1: Comparison of the final test BLEU. en2zh: character-level BLEU. *: pre-training based method.

three baselines initialize their SMT models with phrase tables inferred from n-gram embeddings and language models. From the table, we find that our proposed method gives better initialization to SMT models. Even the SMT models trained with only the retrieved sentences reach higher performance than previous methods, which verifies that the noise within the retrieved sentences is random to a greater extent and can be easily eliminated by SMT models, which is consistent with Khayrallah and Koehn (2018). With the target sentences rewritten by our rewriting model, the quality of extracted phrases can be further improved. We also try to directly train NMT models with the rewritten pseudo data, but only get the BLEU scores under 10, which means there is still much noise for SMT to eliminate in the pseudo pairs.

Initialization Method	fr2en	en2fr	de2en	en2de
(Ren et al., 2019)	15.34	11.74	11.03	8.14
(Lample et al., 2018)	17.50	-	15.63	-
(Artetxe et al., 2018b)	21.16	20.13	13.86	10.59
Only retrieval	21.36	20.23	15.96	12.03
+ target rewriting	**25.21**	**23.58**	**20.41**	**15.98**

Table 2: BLEU of different initial SMT models.

Discussion of Rewriting Model

We build two test sets to quantify the performance of our rewriting models. The first test set denoted as **"in-domain"**, is from our synthetic training data. As described before, we build training samples using monolingual data according to the rules in §2.2. We select 8M sentences from the monolingual corpus of a certain language for model training and randomly sample 8k sentences as development and test sets respectively. In addition, we also test our rewriting model on *newstest2014* (en-fr), which is denoted as **"out-domain"**. We first run GIZA++ on the parallel sentences in the original test set to find the golden alignments between source and target words. Next, we randomly delete up to 30% words in the target side and record their aligned source words. Then we feed the incomplete target sentence and the recorded source words into our model to recover the original target. The BLEU scores on both test sets are listed in Table 3, which shows our rewriting model has good performance.

Test sets	en as target	fr as target
In-domain	59.87	58.71
Out-domain	48.52	47.63

Table 3: Test BLEU scores of the rewriting models.

4 Related Work

Unsupervised machine translation becomes a hot research topic in recent years. The pioneering methods are based on NMT models (Transformer) (Artetxe et al., 2017; Lample et al., 2017; Yang et al., 2018) trained with denoising auto-encoder (Vincent et al., 2010) and iterative back-translation. The following work shows that SMT methods and the hybrid of NMT and SMT can be more effective (Artetxe et al., 2018b; Lample et al., 2018; Marie and Fujita, 2018; Ren et al., 2019; Artetxe et al., 2019). They build the initial PBSMT models with language models and phrase tables inferred from unsupervised cross-lingual n-gram embeddings. Recently, Lample and Conneau (2019) propose a pre-training method and achieve state-of-the-art performance on unsupervised en-fr and en-de translation tasks. But they use much more monolingual data from Wikipedia than previous work and this paper. We must also mention the work of Wu et al. (2019). They similarly use retrieval and rewriting framework for unsupervised MT. However, ours is different from theirs in two aspects. First, we efficiently calculate the cross-lingual sentence embeddings via a training-free method SIF rather than a pre-trained language model. Second, our rewriting method is based on the word alignment information which is more explicit than their max pooling, and our rewriting model is more simple but effective so that the rewriting results can be directly used without extra training techniques.

5 Conclusion

In this paper, we propose a novel method for unsupervised machine translation with a retrieve-and-rewrite schema. We first retrieve similar sentences from monolingual corpora and then rewrite the targets with a rewriting model. With the pseudo parallel data, we better initialize PBSMT models and significantly improve the final iteration performance as the experiments show.

Acknowledgments

This work is supported in part by National Key R&D Program of China AAA0102301, and NSFC 61925203 & U1636210 & 61421003.

References

Sanjeev Arora, Yingyu Liang, and Tengyu Ma. 2017. A simple but tough-to-beat baseline for sentence embeddings.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018a. A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 789–798.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018b. Unsupervised statistical machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium. Association for Computational Linguistics.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2019. An effective approach to unsupervised machine translation. In *Proceedings of the 57th Annual Meeting of ACL*.

Mikel Artetxe, Gorka Labaka, Eneko Agirre, and Kyunghyun Cho. 2017. Unsupervised neural machine translation. *arXiv preprint arXiv:1710.11041*.

Mikel Artetxe and Holger Schwenk. 2018. Margin-based parallel corpus mining with multilingual sentence embeddings. *arXiv preprint arXiv:1811.01136*.

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146.

Alexis Conneau, Guillaume Lample, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2017. Word translation without parallel data. *arXiv preprint arXiv:1710.04087*.

Howard Johnson, Joel Martin, George Foster, and Roland Kuhn. 2007. Improving translation quality by discarding most of the phrasetable. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*.

Huda Khayrallah and Philipp Koehn. 2018. On the impact of various types of noise on neural machine translation. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pages 74–83.

Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 48–54. Association for Computational Linguistics.

Guillaume Lample and Alexis Conneau. 2019. Cross-lingual language model pretraining. arXiv:1901.07291.

Guillaume Lample, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2017. Unsupervised machine translation using monolingual corpora only. *arXiv preprint arXiv:1711.00043*.

Guillaume Lample, Myle Ott, Alexis Conneau, Ludovic Denoyer, et al. 2018. Phrase-based & neural unsupervised machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5039–5049.

Yury A Malkov and Dmitry A Yashunin. 2018. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE transactions on pattern analysis and machine intelligence*.

Benjamin Marie and Atsushi Fujita. 2018. Unsupervised neural machine translation initialized by unsupervised statistical machine translation. *arXiv preprint arXiv:1810.12703*.

Franz Josef Och and Hermann Ney. 2003. A systematic comparison of various statistical alignment models. *Computational linguistics*, 29(1):19–51.

Shuo Ren, Zhirui Zhang, Shujie Liu, Ming Zhou, and Shuai Ma. 2019. Unsupervised neural machine translation with smt as posterior regularization. *arXiv preprint arXiv:1901.04112*.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016a. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 86–96.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016b. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1715–1725.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 6000–6010.

Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408.

Jiawei Wu, Xin Wang, and William Yang Wang. 2019. Extract and edit: An alternative to back-translation for unsupervised neural machine translation. *arXiv preprint arXiv:1904.02331*.

Zhen Yang, Wei Chen, Feng Wang, and Bo Xu. 2018. Unsupervised neural machine translation with weight sharing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 46–55.

Zhirui Zhang, Shujie Liu, Mu Li, Ming Zhou, and Enhong Chen. 2018. Joint training for neural machine translation models with monolingual data. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

A Examples of Retrieval

Examples retrieved from monolingual English and Chinese corpora are shown in Figure 5. With this method, we can retrieve not only highly similar sentences like the first case, but also sentence pairs with rich sentence structures like the second one. The rest retrieved pairs, though containing some noise, also provide high-quality alignments after rewriting according to our observation.

1	[src]: This comment received more than 40 reader responses . [trg]: 这 项 评论 收到 了 250 多 项 读者 建议 。
2	[src]: Cholera is contracted by consuming food or water contaminated with the fecal bacteria Vibrio cholerae . [trg]: 霍乱 是 由 霍乱 弧菌 引起 的 急性 肠道 传染病 , 主要 经由 不 洁净 的 水 和 食物 传播 。
3	[src]: The audience of this meeting is from officials and non-profit organizations . [trg]: 这次 会议 吸引 了 来自 政府 和 非 政府 部门 的 众多 听众 。 [Note]: A from B and C -> 来自 B 和 C 的 A
4	[src]: The Sydney Opera House is one of the most iconic landmarks in the world . [trg]: 山顶 是 世界 上 最 具 特色 的 景点 之一。 [Note]: A is one of B in the world -> A 是 世界 上 的 B 之一

Figure 4: Examples of similar sentences retrieved by our method. The underlined words are already aligned. The note is a hierarchical translation rule, which belongs to a rich sentence structure.

B Examples of Rewriting

We list some rewriting cases from *en* to *zh* in this section. Figure 6 shows some retrieved sentence pairs before and after being rewritten, to demonstrate the effectiveness of our retrieval method and rewriting model. From the first case, we see that the unaligned word "CPSC" is replaced with the right one "她" (she); unrelated words "锂 离子" (lithium-ion) and "消 费 者" (consumer) are removed; "设备" (device) and "爆炸" (explosion) are added into the rewritten sentence. From the second case, we see that the unaligned word "小组" (group) is replaced with the right one "科学家们" (scientists); unrelated words "迎来" (welcome) and "天文学" (astronomy) are removed; "最大" (biggest) and "突破" (breakthrough) are added in the rewritten sentence. The two cases show that our rewriting model can produce the target sentences that are better aligned with the given sources.

C Examples of Translation

Figure 5 shows some translation results generated by our unsupervised MT models to exemplify the final performance. The cases verify that our method empowers the models to learn rich sentence structure such as the hierarchical translation rules of "be A that B" → "是 B 的 A" in the first case and "act as if A" → "表现 的 好像 A 一样" in the second one. This means that our initialization method can preserve the rich sentence structures of the original monolingual sentences, thus giving better initialization for initial UMT models.

D Data Preprocessing

We use Moses scripts[3] for tokenization and truecasing. For Chinese tokenization, we use our in-house tool. For SMT, we use the Moses implementation of hierarchical PBSMT systems with Salm (Johnson et al., 2007). For the rewriting and NMT models, we use the modified version of the public implementation[4] of the Transformer (Vaswani et al., 2017) base model. The rewriting model is based on word level with the vocabulary size of 200,000, while the unsupervised NMT model is based on BPE (Sennrich et al., 2016b) level with the vocabulary size of 60,000. The BPE vocabulary space is shared for each language pair.

[3]https://github.com/moses-smt/mosesdecoder

[4]https://github.com/tensorflow/tensor2tensor

Source	Batteries in **some** of the **devices** are overheating , causing a fire or an **explosion** , **she** said .
Retrieved target Rewritten target	~~CPSC~~ 说 : “ ~~锂~~ ~~离子~~ 电池 ~~包~~ 会 过热 , 给 ~~消费者~~ 造成 燃烧 和 火灾 危害 。 ” 她 说 ： “ 设备 电池 会 过热 ， 爆炸 造成 燃烧 和 火灾 危害 。 ”
Human reference	她 说 ： “ 一些 设备 中 的 电池 会 过热 ， 从而 造成 火灾 或 爆炸 。 ”

Source	**Scientists** have spotted gravitational waves in a **historic discovery hailed** as “ the **biggest scientific breakthrough** of the century ” .
Retrieved target Rewritten target	~~国际~~ 研究 ~~小组~~ ~~说~~ ， 对 这些 引力 波 的 首次 探测 ~~将~~ ~~迎来~~ ~~天文学~~ 的 新纪元 。 研究 科学家 们 ， 对 这些 引力 波 的 首次 探测 是 最大 历史性 突破 的 新纪元 。
Human reference	科学家 们 探测 到 了 引力 波 ， 这 一 历史性 发现 被 誉 为 “ 本 世纪 最大 的 科学 突破 ” 。

Figure 5: Cases of the WMT17 English-Chinese translation results. The underlined words are in hierarchical rules.

Source	He was the brother that went with the flow .
Output	他 是 一个 跟随 流动 的 兄弟 。
Reference	他 是 一个 随从 大家 意见 的 人 。

Source	The next day one newspaper here described Mr Erdogan as acting as if nothing bad had ever happened .
Output	在 未来 的 日子 里 , 这里 有 个 报 称 埃尔多安 先生 表现 的 好像 没有 不好 的 事情 发生 过 一样 。
Reference	次日 , 此间 一家 报纸 写 到 , 埃尔多安 先生 则 表现 的 好像 什么 都 没 发生 一样 。

Figure 6: Cases of the retrieved and rewritten sentences. The bold words are unaligned source words while the strikethrough words are unaligned target words. Human references are given by a translation expert.