Unsupervised Neural Machine Translation with Future Rewarding

Xiangpeng Wei[1,2], **Yue Hu**[1,2]*, **Luxi Xing**[1,2], **Li Gao**[3]
[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2]School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
[3]Platform & Content Group, Tencent, Beijing, China
{weixiangpeng, huyue, xingluxi}@iie.ac.cn
leolgao@tencent.com

Abstract

In this paper, we alleviate the local optimality of back-translation by learning a policy (takes the form of an encoder-decoder and is defined by its parameters) with future rewarding under the reinforcement learning framework, which aims to optimize the global word predictions for unsupervised neural machine translation. To this end, we design a novel reward function to characterize high-quality translations from two aspects: n-gram matching and semantic adequacy. The n-gram matching is defined as an alternative for the discrete BLEU metric, and the semantic adequacy is used to measure the adequacy of conveying the meaning of the source sentence to the target. During training, our model strives for earning higher rewards by learning to produce grammatically more accurate and semantically more adequate translations. Besides, a variational inference network (VIN) is proposed to constrain the corresponding sentences in two languages have the same or similar latent semantic code. On the widely used WMT'14 English-French, WMT'16 English-German and NIST Chinese-to-English benchmarks, our models respectively obtain 27.59/27.15, 19.65/23.42 and 22.40 BLEU points without using any labeled data, demonstrating consistent improvements over previous unsupervised NMT models.

1 Introduction

Neural Machine Translation (Sutskever et al., 2014; Bahdanau et al., 2015) directly models the entire translation process through training an encoder-decoder model that has achieved remarkable performance (Wu et al., 2016; Gehring et al., 2017; Vaswani et al., 2017) when provided with massive amounts of parallel corpora. However, the lack of large-scale parallel data is a serious problem for the vast majority of language pairs.

As a result, several works have recently tried to get rid of the dependence on parallel corpora using unsupervised setting, in which the NMT model only has access to two independent monolingual corpora with one for each language (Lample et al., 2018a; Artetxe et al., 2018b; Yang et al., 2018). Among these works, the encoder and decoder act as a standard auto-encoder (AE) that are trained to reconstruct the inputs from their noised versions. Due to the lack of cross-language signals, unsupervised NMT usually requires pseudo parallel data generated with the back-translation method for achieving the final goal of translating between source and target languages.

Back-translation typically uses beam search (Sennrich et al., 2016a) or just greedy search (Lample et al., 2018a,b) to generate synthetic sentences. Both are approximate algorithms to identify the maximum a posteriori (MAP) output, i.e. the sentence with the highest estimated probability given an input. Although back-translation with MAP prediction has been proved to be successful, it suffers from several apparent issues when trained with maximum likelihood estimation (MLE) only, including exposure bias and loss-evaluation mismatch. Thus, this method often fails to produce the optimal synthetic sentences for the subsequent training.

In this paper, we address the problem mentioned above with future rewarding for unsupervised NMT. The basic idea is to model the future direction of a translation and optimize the global word predictions under the policy gradient reinforcement learning framework. More concretely, we sample N translations via the policy for each input sentence and build a new objective function by combining the cross-entropy loss used in prior works with sequence-level rewards from policy gradient reinforcement learning. We consider the sequence-level reward from two aspects: 1) n-

*Corresponding Author.

Proceedings of the 23rd Conference on Computational Natural Language Learning, pages 281–290
Hong Kong, China, November 3-4, 2019.

gram matching, which is the precision or recall of all sub-sequences of 1, 2, 3 and 4 tokens in generated sequence and is responsible for measuring the accuracy of surface word predictions; 2) semantic adequacy, which is the similarity between the underlying semantic representations of the generated translation and the input sentence. These two aspects of rewards are inspired by the general criteria of what properties a high-quality translation should have and are complementary to each other. Additionally, a variational inference network (VIN) is proposed to model the underlying semantics of monolingual sentences explicitly. It is used to map the source and target languages into a shared semantic space during auto-encoding, as well as constrain the sentences and their translated counterparts have the same or similar semantic code during cross-language training.

The major contributions of this paper can be summarized as follows:

- We propose a novel learning paradigm for unsupervised NMT that models future rewards to optimize the global word predictions via policy gradient reinforcement learning. To enforce the underlying semantic space, we introduce a VIN into our model.
- We introduce an effective reward function that jointly accounts for the n-gram matching and the semantic adequacy of generated translations.
- We conduct extensive experiments on English-French, English-German and NIST Chinese-to-English translation tasks. Experimental results show that the proposed approach achieves significant improvements across different language pairs.

2 Unsupervised Neural Machine Translation

In this section, we first describe the composition of the introduced model and then give details of the newly proposed unsupervised training method.

2.1 Model Composition

The introduced translation model consists of six components: including two encoders with sharing last few layers, two completely independent decoders with one for each language, and two newly introduced VINs with one for each language. For the encoders and decoders, we follow

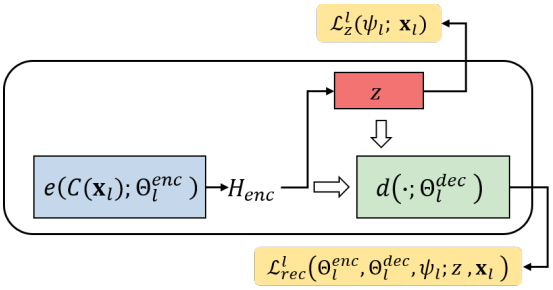

Figure 1: Illustration of Variational Denosing Auto-Encoding. The newly introduced VIN is highlighted in red. Two aspects of losses are respectively abbreviated as $\mathcal{L}_z^l$ and $\mathcal{L}_{rec}^l$.

the recently emerged Transformer (Vaswani et al., 2017). Specifically, each encoder is composed of a stack of four identical layers, and each layer consists of a multi-head self-attention sub-layer and a fully connected feed-forward sub-layer. The encoders of the source and target languages are respectively parameterized as Θ_{src}^{enc} and Θ_{tgt}^{enc}, and the encoding operation is denoted as $e(\mathbf{x}_l; \Theta_l^{enc})$, $\mathbf{x}_l$ is the input sequence of word embeddings, $l \in \{src, tgt\}$. The decoders are also composed of four identical layers. In addition to the two sub-layers in each encoder layer, the decoder inserts a third sub-layer, which performs multi-head attention over the output of the encoder stack, the details we refer the reader to (Vaswani et al., 2017). Similar to encoders, we denote source decoder as Θ_{src}^{dec}, target decoder as Θ_{tgt}^{dec}, and decoding operation as $d(\mathbf{x}_l; \Theta_l^{dec})$, $l \in \{src, tgt\}$. For VINs, each of them is composed of a standard Gaussian distribution $\mathcal{N}(0, \mathbf{1})$ as the *prior*, and a *neural posterior* that is implemented as feed-forward neural network and parameterized by ψ_l, $l \in \{src, tgt\}$.

In this work, the entire model is trained in an unsupervised manner by optimizing two objectives: 1) variational denoising auto-encoding; 2) cross-language training with future rewarding.

2.2 Variational Denoising Auto-Encoding

Firstly, two auto-encoders are respectively trained to learn to reconstruct their inputs. In this form, each encoder should learn to compose the input sentence of its corresponding language, and each decoder is expected to learn to recover the original input sentence from this composition. However, without any constraint, the auto-encoder would make very literal word-by-word copies, without capturing any internal structure of the input sentence involved. To address this issue, prior works

often adapt the same strategy as Denosing Auto-Encoding (DAE) (Vincent et al., 2008), and add some noise to the input sentences (Hill et al., 2016). As shown in Figure 1, we augment the DAE with a variational inference network (VIN) to model underlying semantics of monolingual sentences explicitly, which assumes that there exists a latent variable $\mathbf{z}$ from this semantic space. And this variable, together with the noised input sentence, guides the decoding process. With this assumption, we define the objective function of reconstruction as follow:

$$\mathcal{L}^l_{rec} = \log P_{\Theta_{l\to l}}(\mathbf{x}_l|\mathbf{z}, C(\mathbf{x}_l)) \quad (1)$$

where $\Theta_{l\to l} = \Theta^{enc}_l \circ \Theta^{dec}_l \circ \boldsymbol{\psi}_l$ represents the combination of Θ^{enc}_l, Θ^{dec}_l and $\boldsymbol{\psi}_l$, $l \in \{src, tgt\}$. C denotes a stochastic noise model, in which we apply the same method as in (Lample et al., 2018a).

The continuous latent variable $\mathbf{z}$, acts as the underlying semantics here, is approximated by a *neural posterior* inference network $q_{\boldsymbol{\psi}_l}(\mathbf{z}|\mathbf{x}_l)$. Following (Kingma and Welling, 2014; Kingma et al., 2014), the posterior approximation is regarded as a diagonal Gaussian $\mathcal{N}(\mu, diag(\sigma^2))$, and its mean μ and variance σ^2 are parameterized with deep neural networks. We also reparameterize $\mathbf{z}$ as a function of μ and σ (i.e., $\mathbf{z} = \mu + \sigma \odot \varepsilon$, ε is a standard Gaussian variable that plays a role of introducing noises) rather than using the standard sampling method. We aim to map source and target languages into a shared semantic space and use the following objective function for VINs:

$$\mathcal{L}^l_z = -\mathrm{KL}(q_{\boldsymbol{\psi}_l}(\mathbf{z}|\mathbf{x}_l)||\mathcal{N}(0, \mathbf{1})) \quad (2)$$

where $l \in \{src, tgt\}$. $\mathrm{KL}(Q||P)$ is the Kullback-Leibler divergence between Q and P.

We finally incorporate the auto-encoder and the VIN into an end-to-end neural network, and the overall training objective of auto-encoding is to minimize the following loss function:

$$\mathcal{L}^l_{ae} = -(\mathcal{L}^l_z + \mathcal{L}^l_{rec}) \quad (3)$$

2.3 Cross-language Training with Future Rewarding

In spite of the auto-encoding, the second objective of unsupervised NMT is to constrain the model to be able to map an input sentence from the source (target) language to the target (source) language.

Due to the lack of alignment information between two independent monolingual corpora, the back-translation (Sennrich et al., 2016a) method is used to synthetise a pseudo parallel corpus for cross-language training. More concretely, given an input sentence in one language, which can be firstly translated into the other language (i.e. use the corresponding encoder and the decoder of the other language) by applying the model in inference mode with greedy decoding. And then, the model is trained to reconstruct the original sentence from this translation. The most widely used method in previous works to train the model for sequence generation, called maximum likelihood estimation (MLE for short), it assumes that the ground-truth is provided at each step during training. The objective of MLE is defined as the maximization of the following log-likelihood:

$$\mathcal{L}^{l_1}_{mle} = \log P_{\Theta_{l_2\to l_1}}(\mathbf{x}_{l_1}|\mathbf{z}_p, \tilde{\mathbf{x}}_{l_2}) \quad (4)$$

where $\Theta_{l_2\to l_1} = \Theta^{enc}_{l_2} \circ \Theta^{dec}_{l_1} \circ \boldsymbol{\psi}_{l_2}$ represents the combination of $\Theta^{enc}_{l_2}$, $\Theta^{dec}_{l_1}$ and $\boldsymbol{\psi}_{l_2}$. $\mathbf{z}_p$ is approximated by the introduced VIN (i.e., reparameterized from the Gaussian $q_{\boldsymbol{\psi}_{l_2}}(\mathbf{z}_p|\tilde{\mathbf{x}}_{l_2})$). $\tilde{\mathbf{x}}_{l_2} = d(e(\mathbf{x}_{l_1}; \Theta^{enc}_{l_1}); \Theta^{dec}_{l_2})$ is obtained by greedy decoding in inference mode ($l_1 = src, l_2 = tgt$ or $l_1 = tgt, l_2 = src$).

2.3.1 Future Rewarding

Unfortunately, maximizing $\mathcal{L}^{l_1}_{mle}$ does not always produce the best results on discrete evaluation metrics such as BLEU (Papineni et al., 2002), as the accumulation of errors caused by exposure bias as well as the inconsistency between training and testing measurements lead to the models tend to be short-sighted. We bridge the discrepancy between training and testing modes caused by MLE through learning a policy to model future rewards, which can directly optimize the global word predictions and is made possible with reinforcement learning, as illustrated in Figure 2. To reduce the variance of the model, we use the self-critical policy gradient learning algorithm (Rennie et al., 2017).

For self-critical policy gradient learning, we produce two separate output sequences at each training iteration: $\hat{\mathbf{x}}$, the sampled translation, which is obtained by sampling from the final output probability distribution, and $\hat{\mathbf{x}}^g$, the baseline output, obtained by performing a greedy search. Thus, the objective function of cross-language training can be redefined as the expected advantages of the sampled sequence over the baseline

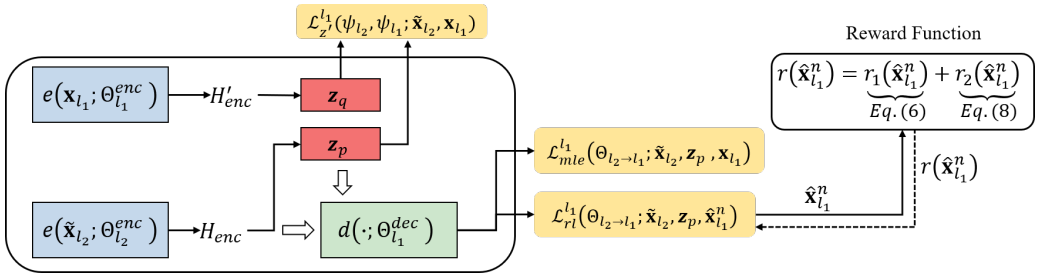

Figure 2: Illustration of the proposed method for cross-language training with future rewarding. Three aspects of losses are respectively abbreviated as $\mathcal{L}^{l_1}_{z'}$, $\mathcal{L}^{l_1}_{mle}$ and $\mathcal{L}^{l_1}_{rl}$. And $\mathcal{L}^{l_1}_{z'}$ is an auxiliary function that constrains the sentences and their translated counterparts in other language have the same or similar semantic codes.

sequence:

$$\begin{aligned}\mathcal{L}^{l_1}_{rl} =& \mathbb{E}_{P_{\Theta_{l_2\to l_1}}(\hat{\mathbf{x}}_{l_1}|\mathbf{z}_p,\tilde{\mathbf{x}}_{l_2})}[r(\hat{\mathbf{x}}_{l_1}) - r(\hat{\mathbf{x}}^g_{l_1})] \\ =& \log P_{\Theta_{l_2\to l_1}}(\hat{\mathbf{x}}_{l_1}|\mathbf{z}_p,\tilde{\mathbf{x}}_{l_2}) \times [r(\hat{\mathbf{x}}_{l_1}) - r(\hat{\mathbf{x}}^g_{l_1})]\end{aligned} \tag{5}$$

where a terminal reward r is observed after the generation reaches the end of each sentence. It is worth noting that considering a baseline reward into training objective can reduce the variance of the model. And we can see that maximizing $\mathcal{L}_{rl}$ is equivalent to maximizing the conditional likelihood of the sampled sequence $\hat{\mathbf{x}}$ if it obtains a higher reward than the baseline $\hat{\mathbf{x}}^g$, thus increasing the expected reward of our model.

2.3.2 Reward

r in Equation 5 denotes the sequence-level reward that evaluates the quality of generated translations. In this subsection, we discuss two major factors that contribute to the success of a translation, that is, n-gram matching and semantic adequacy, and describe how to approximate these factors through computable reward functions.

N-gram matching For a translation generated by a NMT model, we need to measure the accuracy of surface word predictions. For that purpose, the BLEU (Papineni et al., 2002) score is often utilized in previous works. However, the BLEU score has some undesirable properties when used for single sentences, as it was designed to be a corpus measure. Thus, we apply the smoothed version of GLEU (Wu et al., 2016) as the reward for measuring n-gram precision or recall. More concretely, given a generated translation $\hat{\mathbf{x}}_{l_1}$ in one language and the ground-truth reference $\mathbf{x}_{l_1}$, we record all sub-sequences of 1, 2, 3 and 4 tokens in $\hat{\mathbf{x}}_{l_1}$ and $\mathbf{x}_{l_1}$, and start all n-gram counts from 1 instead of 0. Then we compute a recall R_{gleu}, which is the ratio of the number of matching n-grams to the number of total n-grams in $\mathbf{x}_{l_1}$ (ground-truth), and a precision P_{gleu}, which is the ratio of the number of matching n-grams to the number of total n-grams in $\hat{\mathbf{x}}_{l_1}$ (generated output). Finally, the reward of the generated translation $\hat{\mathbf{x}}_{l_1}$ on n-gram matching is defined as:

$$r_1(\hat{\mathbf{x}}_{l_1}) = \min\{R_{gleu}, P_{gleu}\} \tag{6}$$

where r_1 ranges from zero to one and it is symmetrical when switching $\hat{\mathbf{x}}$ and $\mathbf{x}$.

Semantic adequacy We want the model can adequately convey the meaning of the source sentence to the target as much as possible. Thus, we introduce another crucial reward function that is used to measure the semantic adequacy of the generated translations. More concretely, for a generated translation $\hat{\mathbf{x}}_{l_1}$ in one language, we compute the representation of $\hat{\mathbf{x}}_{l_1}$ as:

$$\begin{aligned} e_i &= \text{TFIDF}(w_i), w_i \in \hat{\mathbf{x}}_{l_1} \\ \mathbf{w}_i &= e_i / \text{Sum}(e_1, e_2, ..., e_{T_{\hat{\mathbf{x}}_{l_1}}}) \\ c_{\hat{\mathbf{x}}_{l_1}} &= \sum_{i=1}^{T_{\hat{\mathbf{x}}_{l_1}}} \mathbf{w}_i \hat{\mathbf{x}}^i_{l_1} \end{aligned} \tag{7}$$

Identically, for the corresponding input sentence in another language, its representation $c_{\tilde{\mathbf{x}}_{l_2}}$ can be extracted from the embedding matrix $\tilde{\mathbf{x}}_{l_2}$. As the source and target word embeddings are often mapped to a shared-latent space in unsupervised NMT, we therefore can directly use the following cosine similarity as the reward for semantic adequacy:

$$r_2(\hat{\mathbf{x}}_{l_1}) = \frac{(c_{\hat{\mathbf{x}}_{l_1}}, c_{\tilde{\mathbf{x}}_{l_2}})}{\|c_{\hat{\mathbf{x}}_{l_1}}\| \cdot \|c_{\tilde{\mathbf{x}}_{l_2}}\|} \tag{8}$$

where $(,)$ indicates the dot product operation.

The final reward for a translation $\hat{\mathbf{x}}_{l_1}$ is a linear combination of the rewards discussed above:

$$r(\hat{\mathbf{x}}_{l_1}) = r_1(\hat{\mathbf{x}}_{l_1}) + r_2(\hat{\mathbf{x}}_{l_1}) \quad (9)$$

where $r_1(\hat{\mathbf{x}}_{l_1})$ and $r_2(\hat{\mathbf{x}}_{l_1})$ complement to each other and work jointly to guide the learning of our model. Note that the combination of these two aspects of rewards helps because it can prevent the cases that the generated translation with high n-gram matching but low semantic adequacy to have relatively high rewards, and vice versa.

2.3.3 Overall Objective Function

In addition to the aforementioned MLE objective function (Eq. 4) and the RL objective function (Eq. 5), there is an auxiliary function that constrains the sentences and their translated counterparts have the same or similar semantic code and is defined as:

$$\mathcal{L}_{z'}^{l_1} = -\mathrm{KL}(q_{\psi_{l_1}}(\mathbf{z}_q|\mathbf{x}_{l_1})||q_{\psi_{l_2}}(\mathbf{z}_p|\tilde{\mathbf{x}}_{l_2})) \quad (10)$$

Finally, the overall training objective of cross-language training is to minimize the following loss function with hyperparameters η:

$$\mathcal{L}_{cl}^{l_1} = -((1-\eta)(\mathcal{L}_{mle}^{l_1} + \mathcal{L}_{z'}^{l_1}) + \eta\mathcal{L}_{rl}^{l_1}) \quad (11)$$

where η is a scaling factor. In the beginning of the training $\eta = 0$, while as we move on with the training we can increase the η to slowly reduce the effect of MLE loss. And η is updated as follows:

$$\eta = \min(0.8, \max(0.0, \frac{steps - n_s}{n_e - n_s})) \quad (12)$$

where $steps$ is the global steps that the model has been updated, n_s and n_e are the start and end steps for increasing η respectively.

2.4 Training Procedure

There are two stages in the proposed unsupervised training. In the first stage, we pre-train the proposed model with denoising auto-encoding and cross-language training, until no improvement is achieved on the development set. This ensures that the model starts with a much better policy than random because now the model can focus on the good part of the search space. In the second state, we use an annealing schedule to teach the model to produce stable sequences gradually. That is, after the initial pre-training steps, we continue training the model with future rewarding. During each iteration, we perform one batch of denoising auto-encoding and cross-language training for the source as well as target languages alternately.

For model selection, we randomly extract 3000 source and target sentences to form a development set. Following (Lample et al., 2018a), we translate the source sentences to the target language and then convert the resulting sentences back to the source language. The quality of the model is then evaluated by computing the BLEU score over the original inputs and their reconstructions via this two-step translation process. The performance is finally averaged over two directions, and the selected model is the one with the highest score.

3 Experiments

We mainly evaluate the proposed approach on the widely used English-German, English-French and NIST Chinese-to-English[1] translation tasks.

3.1 Datasets

For English-French and English-German, we use 30M sentences from the WMT monolingual News Crawl datasets from years 2007 through 2017. We use the publicly available implementation of Moses[2] scripts for tokenization. Besides, we use a shared vocabulary for source and target languages with 60K subword tokens based on byte-pair encoding (Sennrich et al., 2016b). We remove sentences longer than 50 subword-tokens. Experimental results are reported on *newstest2014* for English-French translation and *newstest2016* for English-German translation. We adopt the same method as in (Lample et al., 2018b) to obtain cross-lingual embeddings.

For NIST Chinese-to-English translation, our training data consists of 1.6M sentence pairs randomly extracted from LDC corpora[3], which has been widely utilized by previous works. Similar to (Yang et al., 2018), we build the monolingual dataset by randomly shuffling the Chinese and English sentences respectively since the data set is not big enough. We set the vocabulary size to 30K for both Chinese and English. The average BLEU score over *NIST02∼06* is reported

[1]The reason that we do not conduct experiments on English-to-Chinese translation is that we do not get public test sets for English-to-Chinese.

[2]http://www.statmt.org/moses/

[3]LDC2002E18, LDC2003E07, LDC2003E14, the Hansards portion of LDC2004T07, LDC2004T08, and LDC2005T06

	en→fr	fr→en	en→de	de→en	zh→en
Existing Unsupervised NMT					
Artetxe et al. (2018b)	15.13	15.56	-	-	-
Lample et al. (2018a)	15.05	14.31	9.64	13.33	-
Yang et al. (2018)	16.97	15.58	10.86	14.62	14.52
Lample et al. (2018b), NMT	25.14	24.18	17.16	21.00	-
Wu et al. (2019)	27.56	26.90	19.55	23.29	-
Song et al. (2019)	27.41	27.09	18.21	23.37	-
This work					
MLE	25.47	24.51	17.04	21.13	18.26
(+Future Rewarding)	**27.59**	**27.15**	**19.65**	**23.42**	**22.40**

Table 1: Results of the proposed method in comparison to existing unsupervised NMT systems (BLEU).

in this paper. To pre-train cross-lingual embeddings, we utilize the monolingual corpora to train the embeddings for each language independently by using word2vec (Mikolov et al., 2013). Then we apply the public implementation[4] proposed by Artetxe et al. (2017) to map these embeddings into a shared latent space and keep the mapped embeddings fixed during training.

For NIST Chinese-to-English, we apply case-insensitive NIST BLEU computed by the script *mteval-v13a.pl* to evaluate the translation performance. For English-German and English-French, we evaluate the translation performance with the script *multi-belu.pl*.

3.2 Hyper-parameters

We set the following hyper-parameters: word embedding dimension as 512, hidden size of self-attention as 512, hidden size of fully connected layers as 1024 and the head number as 8. We share the last one layer of encoders in both languages. The dropout rate is set as 0.1, 0.3 and 0.2 during the training for En-Fr, En-De and Zh-to-En, respectively. We perform a fixed number of iterations (500K) to train each model, and set $n_s = 300$K, $n_e = 400$K, for gradually increasing the effect of future rewarding. We use the Adam optimizer with a simple learning rate schedule: we start with a learning rate of 10^{-4}, after 300K updates, we begin to halve the learning rate every 100K steps. We set the mini-batch size as 64. At decoding time, we use greedy search.

3.3 Overall Results

Our method is compared with several previous unsupervised NMT systems (Artetxe et al., 2018b; Lample et al., 2018a,b; Yang et al., 2018; Wu et al., 2019; Song et al., 2019). Although, Song et al. (2019) have achieved comparable results with supervised NMT systems with larger monolingual data (Wikipedia data) and bigger model[5], we still list the results that obtained with the same data and model as ours for fair comparison. We also consider a "Baseline" model, with the same architecture as described in Section 2.1 except for the variational inference network and is trained using MLE only. We directly copy the experimental results of previous models reported in their papers and report the BLEU scores on English-French, English-German and NIST Chinese-to-English test sets in Table 1.

As shown in Table 1, our approach achieves BLEU score of 27.59 and 27.15 on En→Fr and Fr→En translations respectively, which outperforms Lample et al. (2018b) by more than 2 BLEU points on both En→Fr and Fr→En. For the En-De, we achieve 19.65 and 23.42 BLEU scores on En→De and De→En respectively, with up to +10.09 BLEU points improvement over previous unsupervised NMT models. For the Chinese-to-English translation, the proposed method leads to a substantial improvement (up to 54%) over the previous system showed in Yang et al. (2018). Compared to baseline, our approach demonstrates significant improvements by more than 2 BLEU points over three benchmarks. These results indicate that the newly proposed training method that models future rewards to optimize global word predictions for unsupervised NMT is promising and enables the model to generate quality translations.

[4]https://github.com/artetxem/vecmap

[5]Our model can also adopt such an advanced pre-training technique, we leave this for feature work.

3.4 Analysis

In this section, we conduct some analysis over the proposed method by taking English-French translation as an example.

3.4.1 Ablation Study

To understand the importance of different components of the proposed system, we perform an ablation study by training multiple versions of our model with some missing components: the variational inference network and the future rewarding method. Results are reported in Table 2. From the table, we can see that removing the future rewarding, and the accuracy drops by 0.98/1.02 BLEU points. Without the variational inference networks, the accuracy decreases with 0.62/0.69 BLEU points. These findings demonstrate that both the future rewarding and the VIN are important, and both contribute to the improvement of translation accuracy. The more critical component is the future rewarding technology, which is vital to optimize the global word predictions.

	en-fr	fr-en
Full Model	27.59	27.15
Without VINs	26.97	26.46
Without Future Rewarding	26.61	26.13

Table 2: Ablation study of our method on English-French translation task.

3.4.2 Qualitative Comparison of Back-translating

We perform qualitative evaluation on the pseudo parallel data generated with the back-translation method. To this end, we conduct a "round-trip" translation (e.g., $src \rightarrow \tilde{tgt} \rightarrow \hat{src}$), where src and $\tilde{tgt}$ form a pseudo parallel corpus, $\hat{src}$ is the reconstruction from $\tilde{tgt}$. We explore three settings for qualitative evaluation: 1) *UNKs*, the ratio of the number of unknown words to the number of total words in $\tilde{tgt}$; 2) the average over all sentences in $\tilde{tgt}$ with respect of their semantic adequacy, denoted as *SA*; 3) the BLEU scores over the original inputs and their reconstructions, denoted as *r-BLEU*. All settings are finally averaged over two directions.

Results are shown in Table 3. The proposed training method introduces significant boosts in all of the three settings, with reducing 1.34% of unknown words, increasing the semantic adequacy by 0.088 and improving r-BLEU points by 5.85. This is in line with our expectations, as the proposed future rewarding method is not optimized to predict the next token, but rather to increase long-term reward.

	UNKs	SA	r-BLEU
Baseline	3.51%	0.794	54.23
+Future Rewarding	2.17%	0.882	60.08

Table 3: Qualitative comparison of the generated pseudo parallel sentences from the models trained with MLE only and with the proposed training method on English-French test set.

Better than Baseline	
S:	He put together a real feast for his fans to mark the occasion.
R:	Pour l'occasion, il a concocté un vrai festin pour ses fans.
B:	Il a mis en scène un vrai festin pour *son public* pour marquer le *souvenir*.
O:	Il a mis un vrai festin pour **ses fans** pour marquer la **circonstance**.
S:	Des scientifiques viennent de mettre en lumière la façon dont les mouvements de la queue d'un chien sont liés à son humeur.
R:	Scientists have shed more light on how the movements of a dog's tail are linked to its mood.
B:	Scientists *come out of light the way the movements of* the tail of a dog are linked to *his spirits*.
O:	Scientists **come to light** the way of the movements of a dog's tail are related to **its mood**.
Worse than Baseline	
S:	The recalled models were built between August 1 and September 10.
R:	Les modèles rappelés ont été construits entre le 1er août et le 10 septembre.
B:	Les modèles *rappelés* ont été construits entre le 1er août et le 10 septembre.
O:	Les modèles **de raconté** ont été construits entre le 1er août et le 10 septembre.
S:	Elles connaissent leur entreprise mieux que personne.
R:	They know their business better than anyone else.
B:	They know their *business* better than *anyone else*.
O:	They know their **company** better than **anyone**.

Table 4: Translation examples from English-French test set (English-to-French is above the dotted line and French-to-English is below the dotted line). **B**: the baseline model; **O**: our proposed model.

3.4.3 Example Translations

Table 4 shows four example translations. The first part shows examples for which the proposed model reached a higher BLEU score than the baseline model. We find that the translation produced

by the baseline model doesn't adequately convey the meaning of the source sentence to the target. By contrast, the proposed future rewarding method enables the model to generate translations that are more diversity while ensuring the meaning of the source sentences, such as "***circonstance***" and "***come to light***". The possible reason is that we apply the semantic adequacy to reward translations that have different syntax structures and expressions but share the same meaning as the ground-truth sentence. The second part contains examples where the baseline achieved better BLEU score than our model, that is, in a few cases, our model chooses inappropriate words that under the same topic as reference words.

4 Related Work

In order to reduce the exposure bias and optimize the metrics used to evaluate sequence modeling tasks (like BLEU, ROUGE or METEOR) directly, reinforcement learning (RL) has been widely used in many of recent works on machine translation (Ranzato et al., 2016; Shen et al., 2016; He et al., 2017; Bahdanau et al., 2017; Li et al., 2017), text summarization (Paulus et al., 2018; Wu and Hu, 2018; Li et al., 2018; Wang et al., 2018), dialogue generation (Li et al., 2016), and question answering (Hu et al., 2018). However, our proposed method is the first use in combination with reinforcement learning for unsupervised NMT to explicitly enhance back-translation.

Recently, motivated by the success of cross-lingual embeddings (Artetxe et al., 2016; Zhang et al., 2017; Conneau et al., 2017), several works have tried to train NMT or SMT models using unsupervised setting, in which the model only has access to unlabeled data. For example, Lample et al. (2018a) propose a model that consists of a single encoder and a single decoder for both languages, respectively responsible for encoding source and target sentences to a shared latent space and to decode from that latent space to the source or target domain. Different from (Lample et al., 2018a), Artetxe et al. (2018b) introduce a shared encoder but two independent decoders with one for each language. Both of these two works mentioned above utilize denoising auto-encoding to reconstruct their noisy inputs and incorporate back-translation into cross-language training procedure. Further, Yang et al. (2018) extend the single encoder by using two independent encoders but sharing some partial weights, which are responsible for alleviating the weakness in keeping language-specific characteristics of the shared encoder. And the entire system is fine-tuned by introducing two global GANs with one for each language. More recently, Artetxe et al. (2018a) and Lample et al. (2018b) propose an alternative approach based on phrase-based statistical machine translation, which profits from the modular architecture of SMT. In addition, Lample et al. (2018b) also introduce a novel cross-lingual embedding training method which is particularly suitable for related languages (e.g., English-French and English-German). Ren et al. (2019) introduce SMT models as posterior regularization, in which SMT and NMT models boost each other through iterative back-translation in a unified EM training algorithm. Wu et al. (2019) propose an alternative for back-translation, , extract-edit, to extract and then edit real sentences from the target monolingual corpora. Lample and Conneau (2019) and Song et al. (2019) propose to pretrain cross-lingual language models for the initialization stage of unsupervised neural machine translation, which is critical to the performance of their proposed model. In contrast to theirs, we propose an effective training method for unsupervised NMT that models future rewards to optimize the global word predictions via neural policy reinforcement learning, which can be applied to arbitrary architectures and language pairs easily.

5 Conclusion

In this paper, we have proposed a novel learning paradigm for unsupervised NMT that models future rewards to optimize the global word predictions via reinforcement learning, in which we design an effective reward function that jointly accounts for the n-gram matching and the semantic adequacy of generated translations. To constrain the corresponding sentences in two languages have the same or similar semantic code, we also introduce a variational inference network into the proposed model.

We test the proposed model on WMT'14 English-French, WMT'16 English-German and NIST Chinese-to-English translation tasks. Experiment results show that our approach leads to significant improvements over various language pairs, especially on distantly-related languages such as Chinese and English.

6 Acknowledgments

We would like to thank the anonymous reviewers for their valuable comments and suggestions. This work was supported by the National Key Research and Development Program of China (No. 2017YFB0803301). Yue Hu is the corresponding author.

References

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2016. Learning principled bilingual mappings of word embeddings while preserving monolingual invariance. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016*, pages 2289–2294.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2017. Learning bilingual word embeddings with (almost) no bilingual data. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017*, pages 451–462.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2018a. Unsupervised statistical machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, pages 3632–3642.

Mikel Artetxe, Gorka Labaka, Eneko Agirre, and Kyunghyun Cho. 2018b. Unsupervised neural machine translation. In *ICLR 2018*.

Dzmitry Bahdanau, Philemon Brakel, Kelvin Xu, Anirudh Goyal, Ryan Lowe, Joelle Pineau, Aaron Courville, and Yoshua Bengio. 2017. An actor-critic algorithm for sequence prediction. In *ICLR 2017*.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *ICLR 2015*.

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

Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. 2017. Convolutional sequence to sequence learning. In *arXiv:1705.03122*.

Di He, Hanqing Lu, Yingce Xia, Tao Qin, Liwei Wang, and Tieyan Liu. 2017. Decoding with value networks for neural machine translation. In *Advances in Neural Information Processing Systems, NIPS 2017*, pages 178–187.

Felix Hill, Kyunghyun Cho, and Anna Korhonen. 2016. Learning distributed representations of sentences from unlabelled data. In *Proceedings of NAACL-HLT 2016*, pages 1367–1377.

Minghao Hu, Yuxing Peng, Zhen Huang, Xipeng Qiu, Furu Wei, and Ming Zhou. 2018. Reinforced mnemonic reader for machine reading comprehension. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018*, pages 4099–4106.

Diederik P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. 2014. Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems, NIPS 2014*, pages 3581–3589.

Diederik P Kingma and Max Welling. 2014. Auto–encoding variational bayes. In *ICLR 2014*.

Guillaume Lample and Alexis Conneau. 2019. Cross–lingual language model pretraining. In *arXiv:1901.07291*.

Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2018a. Unsupervised machine translation using monolingual corpora only. In *ICLR 2018*.

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

Jiwei Li, Will Monroe, and Dan Jurafsky. 2017. Learning to decode for future success. In *arXiv:1701.06549*.

Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, and Dan Jurafsky. 2016. Deep reinforcement learning for dialogue generation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, ENMLP 2016*, page 1192–1202.

Piji Li, Lidong Bing, and Wai Lam. 2018. Actor-critic based training framework for abstractive summarization. In *arXiv:1803.11070*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems, NIPS 2013*, pages 3111–3119.

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 on association for computational linguistics, ACL 2002*, pages 311–318.

Romain Paulus, Caiming Xiong, and Richard Socher. 2018. A deep reinforced model for abstractive summarization. In *ICLR 2018*.

Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. 2016. Sequence level training with recurrent neural networks. In *ICLR 2016*.

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

Steven J Rennie, Etienne Marcheret, Youssef Mroueh, Jarret Ross, and Vaibhava Goel. 2017. Self-critical sequence training for image captioning. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, pages 1179–1195.

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, ACL 2016*, 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, ACL 2016*, pages 1715–1725.

Shiqi Shen, Yong Cheng, Zhongjun He, Wei He, Hua Wu, Maosong Sun, and Yang Liu. 2016. Minimum risk training for neural machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016*, pages 1683–1692.

Kaitao Song, Xu Tan, Tao Qin, Jianfeng Lu, and Tie-Yan Liu. 2019. MASS: masked sequence to sequence pre-training for language generation. In *arXiv:1905.02450*.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems, NIPS 2014*, pages 3104–3112.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *arXiv:1706.03762*.

Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre Antoine Manzagol. 2008. Extracting and composing robust features with denoising autoencoders. In *Machine Learning, Proceedings of the Twenty-Fifth International Conference, ICML 2008*, page 1096–1103.

Li Wang, Junlin Yao, Yunzhe Tao, Li Zhong, Wei Liu, and Qiang Du. 2018. A reinforced topic-aware convolutional sequence-to-sequence model for abstractive text summarization. In *arXiv:1805.03616*.

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

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, and Klaus Macherey. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. In *arXiv:1609.08144*.

Yuxiang Wu and Baotian Hu. 2018. Learning to extract coherent summary via deep reinforcement learning. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI 2018*, pages 5602–5609.

Zhen Yang, Wei Chen, Feng Wang, and Bo Xu. 2018. Unsupervised neural machine translation with weight sharing. In *arXiv:1804.09057*.

Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun. 2017. Adversarial training for unsupervised bilingual lexicon induction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017*, pages 1959–1970.