Incorporating translation quality estimation into Chinese-Korean neural machine translation

Feiyu Li, Yahui Zhao,* Feiyang Yang, Rongyi Cui
Department of Computer Science and Technology,
Yanbian University, Yianbian, China
{2019010431,yhzhao,2018010454,cuirongyi}@ybu.edu.cn

Abstract

Exposure bias and poor translation diversity are two common problems in neural machine translation (NMT), which are caused by the general of the teacher forcing strategy for training in the NMT models. Moreover, the NMT models usually require the large-scale and high-quality parallel corpus. However, Korean is a low resource language, and there is no large-scale parallel corpus between Chinese and Korean, which is a challenging for the researchers. Therefore, we propose a method which is to incorporate translation quality estimation into the translation process and adopt reinforcement learning. The evaluation mechanism is used to guide the training of the model, so that the prediction cannot converge completely to the ground truth word. When the model predicts a sequence different from the ground truth word, the evaluation mechanism can give an appropriate evaluation and reward to the model. In addition, we alleviated the lack of Korean corpus resources by adding training data. In our experiment, we introduce a monolingual corpus of a certain scale to construct pseudo-parallel data. At the same time, we also preprocessed the Korean corpus with different granularities to overcome the data sparsity. Experimental results show that our work is superior to the baselines in Chinese-Korean and Korean-Chinese translation tasks, which fully certificates the effectiveness of our method.

1 Introduction

With the rapid development of deep learning, neural machine translation(NMT) has attracted much attention in recent years. This is because it has superior performance and does not require much manual intervention(Bahdanau D, 2015)(Junczys-Dowmunt M, 2016). Korean is the official language of the Korean ethnic group in China, and is also used in the Korean Peninsula, the United States, the Russian Far East and other areas where Koreans congregate. Transnational and transregional is a key feature of Korean. As the Korean ethnic group is one of the 24 ethnic minorities in China that have their own language(Yongshou, 2020), the research on Chinese-Korean neural machine translation plays an important role in promoting the language work of ethnic minorities and strengthening the communication and unity of ethnic groups.

At present, most NMT models adopt the teacher forcing strategy in training(Williams RJ, 1989). Teacher forcing is to minimize the difference between the source sentence and the reference and force the predicted translation to be infinitely close to the reference. First of all, as there is usually no reference for sentence prediction, exposure bias will be brought, which may affect the performance and robustness of the model (Ranzato M, 2016). Secondly, as there are a large number of synonyms and similar expressions in the language, the translation model cannot always generate the ground truth word even with teacher forcing in training. Using this strategy will greatly curb the diversity of translation and make many reasonable translations in an unreachable state(Zhang W, 2019). In addition, domestic research on Chinese-Korean machine translation started late, the foundation is poor, the lack of large-scale parallel corpus. Therefore, there are many challenges to improving the performance of Chinese-Korean machine translation under the condition of low resources.

---

*Corresponding author.

We attempt to introduce a sentence-level evaluation mechanism to guide the training of the model, so that the prediction can not converge completely to the ground truth word, to alleviate exposure bias and poor translation diversity. The evaluation mechanism is reference-free evaluation, which is called quality estimation(QE) in conference on machine translation. The main idea is that the evaluation model can estimate the quality of the unseen translation without reference after supervised training. The instruct mechanism is guided by the reinforcement learning(RL) of policy optimization, which enables the model to optimize the target sequence at the sentence level. In order to alleviate the acknowledged instability and variance difficulty fitting problems of reinforcement learning, we trained MLE and RL together and referred to the baseline reward method proposed by Weaver L (1999). Moreover, previous studies have directly used BLEU value as a reward(Wu L, 2018)(Yang Zhen, 2017), which would lead to serious bias in the model and exacerbate the problem of poor translation diversity. Therefore, we propose a reward function based on QE score. Meanwhile, monolingual corpus and Korean text preprocessing with different granularity are used in the training process overcome data sparsity and improve the quality of machine translation for low-resource languages.

2 Related work

Machine translation quality estimation: Machine translation quality estimation is different from the evaluation indexes of machine translation such as BLEU(Papineni K, 2002), TER(Snover M, 2006), METEOR(Lavie A, 2007), etc. It can automatically give the quality prediction of machine-generated translation without relying on any reference translation. The most commonly used quality score is the human translation edit rate (HTER). QuEst is a model proposed by Specia L (2013), which is used for quality estimation tasks. As the baseline model of machine translation quality estimation task, the model consists of feature extraction module and machine learning module. In order to solve the problem of machine translation quality estimation, Kim H (2017) firstly applied the machine translation model to the quality estimation task and proposed a translation quality estimation model based on RNN. Fan K (2019) replace the machine translation model based on RNN with Transformer model on the basis of Predictor-Estimator, and proposed a bilingual expert model, which improved the performance and interpretability of the model evaluation.

Reinforcement learning-based NMT: There is a lot of research showing that the advantages of reinforcement learning in sequence generation tasks. Ranzato M (2016) proposed a novel sequence level training algorithm that directly optimizes the metric used at the test time. Wu L (2018)conducted a systematic study on how to train better NMT models using reinforcement learning with several large-scale translations tasks. Keneshloo Y (2020) considered seq2seq problems from the RL point of view and provide a formulation combining the power of RL methods in decision-making with seq2seq models that enable remembering long-term memories.

3 Methodology

In this section, we describe the construction and training of a Chinese-Korean neural machine translation model that incorporates translation quality assessment. We introduce the model architecture, sentence-level translation quality estimation methods, reward function design and training of the whole model.

3.1 Model Overview

To alleviate exposure bias and poor translation diversity, we propose a Chinese-Korean machine translation model that incorporates translation quality estimation. The model introduces an evaluation mechanism at the sentence level to guide the model prediction not to converge completely to the ground truth word. The specific framework structure of the model is shown in Figure 1, which mainly includes two modules: machine translation and machine translation quality estimation. The translation module adopts the encoder-decoder architecture, and the framework is consistent with Transformer. The evaluation module adopts sentence-level machine translation quality estimation model Bilingual Expert. The training adopts reinforcement learning.
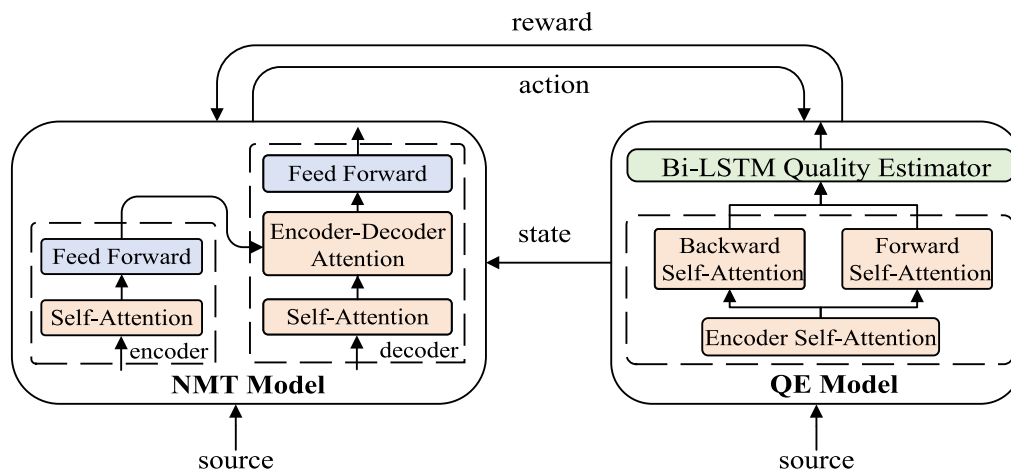
Figure 1: The architecture of the Chinese-Korean machine translation module

In the process of machine translation, NMT system, as an agent of reinforcement learning, obtains the environmental state information at the current moment through continuous interaction with the environment. The environmental state information is the source sentence x under the time step t and the above $P(y_t|x, \hat{y}_{<t})$ of the generated target sentence. Where $\hat{y}_{<t}$ represents the target sentence predicted by the model before the time step t. According to the state of the current environment, the agent decides to choose the next selected word, obtains the reward value of the word selection operation in the current state and enters the next state, and finally finds the optimal strategy of translation through reinforcement learning.

As shown in Figure 2, the machine translation task is described as: Training a machine translation model M_θ with parameter θ under the condition that the Chinese-Korean parallel corpus is given. The machine translation model M_θ translates the given source sentence sequence $\mathbf{x} = (x_1, x_2, ..., x_n)$ into a target sentence sequence $\mathbf{y} = (y_1, y_2, ..., y_m)$, where n, m are sequence lengths of source and target sentences respectively. In the time step t, the state S_t defines the target sentence $\mathbf{y} = (y_1, y_2, ..., y_m)$ generated by the machine translation model M_θ under the current time step, and the action a_t is defined as the selection of the next word y_{t+1} in the current environment. Training a machine translation quality estimation model Q_φ with parameter φ under the condition that the condition of the translation data and HTER scores is given. After supervised training, the quality estimation model Q_φ acts as a generator of reward functions to give quality scores to the unseen translations, and the machine translation model M_θ is guided to interact with the environment to produce the next word y_{t+1}.
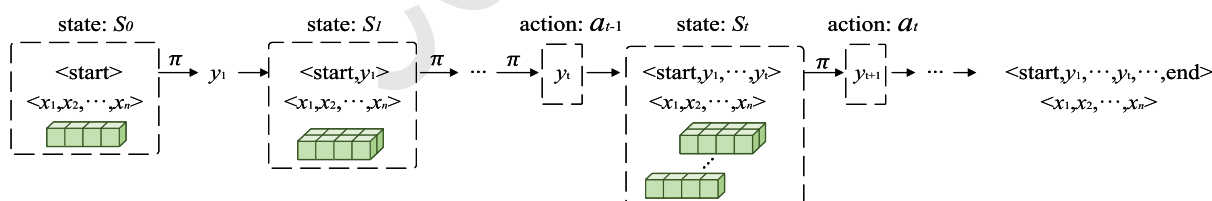

Figure 2: The schematic diagram of the decision process in translation

3.2 Generate Rewards through Sentence-Level Quality Estimation

An excellent translation usually includes multidimensional evaluation, such as fidelity and faithfulness, so it is difficult to abstract machine translation tasks into simple optimization problems. Therefore, instead of manually setting a single rule as the source of the reward function, we use the output of the machine translation quality estimation model as part of the reward. The model can generate a more comprehensive score through a relatively complex network structure, which is more relevant to human evaluation and more tolerant to the diversity of translation.

The model Q_φ in this paper uses the same network structure as Bilingual Expert. The model includes a word prediction module based on bidirectional Transformer and a regression prediction model based on Bi-LSTM. Bidirectional Transformer includes three parts encoder self attention, e forward and backward self-attentions and token reconstruction. It acquires hidden state features h by pre-training on large-scale parallel corpus. The encoder part corresponds to $q(h|x,y)$, and the decoder part corresponds to $p(y|h)$. The calculation formula is as follows:

$$q(h|x,y) = \prod_t q(\overrightarrow{h_t}|x, y_{<t})q(\overleftarrow{h_t}|x, y_{<t}) \tag{1}$$

$$p(y|h) = \prod_t p(y_t|\overrightarrow{h_t}, \overleftarrow{h_t}) \tag{2}$$

The hidden state $h = (h_1, ..., h_m)$ is a combination of the forward and backward hidden state, which captures the deep translation features of the sentence. The final features are as follows:

$$f = \text{Concat}\left(\overrightarrow{h}_t, \overleftarrow{h}_t, e_{t-1}, e_{t+1}, f^{mm}\right) \tag{3}$$

Where, e_{t-1}, e_{t+1} is the embedding concatenation of two neighbor tokens, and f^{mm} is the mis-matching features. Finally, the features are input to Bi-LSTM for training to get the predicted HTER score:

$$\text{HTER}' = \text{sigmoid}\left(w^T\left[\text{Bi} - \text{LSTM}(f)\right]\right) \tag{4}$$

The loss function in the training process is:

$$\text{argmin}||\text{HTER} - \text{HTER}'||_2^2 \tag{5}$$

The scalar value obtained in Equation (4) is the evaluation of the generated translation by the machine translation quality estimation module. Compared with BLEU, it has deeper translation characteristics. Therefore, our method uses this score to guide the machine translation module and can achieve the effect of the prediction can not converge completely to the ground truth word.

3.3 Reward Computation

It is critical to set up appropriate rewards for RL training. In previous researches on NMT, it is assumed that the effective predictive value of each word item in the generated target sentence is unique, that is to say, there is a fixed reference for each sentence. Therefore, both the minimum-risk training method (Shen Shiqi, 2016) and the reinforcement learning method (Yang Zhen, 2017)(Keneshloo Y, 2020) use the BLEU score of similarity between the generated sentence and the reference as the training target. However, in natural language, the same source sentence fragment can correspond to multiple reasonable translations, so the reward based on BLEU cannot give a reasonable reward or punishment for words other than the target language. As a result, most reasonable translations are denied, which greatly limits the improvement of translation effect by reinforcement learning and exacerbates the problem of poor diversity of machine translations. Thus we set the reward as:

$$R(\hat{y}_t) = \alpha Score_{BLEU}(\hat{y}_t) + \frac{1 - \alpha}{Score_{QE}(\hat{y}_t) + 1} \tag{6}$$

Where, $Score_{BLEU}(\hat{y}_t)$ is the normalized BLEU between the generated translation and the ground truth, and $Score_{QE}(\hat{y}_t)$ is the normalized QE evaluation score of the generated translation. The super parameter α is used to balance the weight between BLEU and QE scores, so as to avoid the problem that the QE score may aggravate the instability of training after introducing it. In this way, the training can be converged quickly and the diversity of translation can be fully considered.

In the machine translation task, the agent needs to take dozens of actions to generate a complete target sentence, but after generating a complete sequence, only one terminal reward can be obtained, and sequence-level reward cannot distinguish the contribution of each word item to the total reward. Therefore,

there is a problem of reward sparsity during the training, which will lead to slow convergence speed of the model or even failure to learn. Reward shaping can alleviate this problem. In reward shaping, the instant reward at each decoding step t is imposed, and the rewards correspond to word levels. The rewards are set as follows:

$$r_t\left(\hat{y}_t\right) = R\left(\hat{y}_t\right) - R\left(\hat{y}_{t-1}\right) \tag{7}$$

During the training, an accumulative reward is calculated as the current sequence reward after each sampling action is completed, and the reward difference between two continuous time steps is the word level reward. In this way, the model can get an instant reward for the current time step after each action, thus alleviating the problem of reward sparse.

$$R(\hat{y}_t) = \sum_{t=1}^{T} r_t\left(\hat{y}_t\right) \tag{8}$$

Experiments have shown that using reward shaping does not change the optimal strategy. Since the reward of the whole sequence is the sum of the reward value of each word item level, which is consistent with the reward of the sequence level, the total reward of the sequence will not be affected.

3.4 The Training of Reinforcement Learning

The idea of reinforcement learning is that the agent selects an action to execute according to the current environment, and then the environment shifts with a certain probability and gives a reward to the agent, and the agent repeats the above process for the purpose of maximizing the reward. Specifically, in the translation task, the NMT model is regarded as the agent making decisions, and the random strategy $\pi\left(a_t|\mathbf{s_t};\Theta\right)$ is adopted to select candidate words from the word list as an action. During the training, the agent learns better translation through the reward given by the environment after the target sentence is generated by the decoder.

$$\pi\left(a_t|\mathbf{s_t};\Theta\right) = \sigma\left(\mathbf{W} * \mathbf{s}_t + \mathbf{b}\right) \tag{9}$$

Where $\pi\left(a_t|\mathbf{s_t};\Theta\right)$ is the probability of choosing an action, and σ is the sigmoid function, $\Theta = \{\mathbf{W}, \mathbf{b}\}$ is a parameter of the policy network. During the training, the action from all the conditional aim-listed probability $p(y_t|x, y_{<t})$ of the given source sentence and the word selected below, and the goal is to pursue the maximum expected reward, i.e:

$$a_t^* = \mathrm{argmax}_a \pi\left(a|\mathbf{s_t};\Theta\right) \tag{10}$$

When the complete target sentence is generated, the quality estimation score of the sentence is used as the label information to calculate the reward. Then the Policy Gradient method (Silver D, 2014) of reinforcement learning algorithm is used to maximize the expected revenue, as shown follows:

$$
\begin{aligned}
J(\Theta) &= \sum_{i=1}^{N} E_{\hat{y} \sim p(\hat{y}|x^i)} R\left(\hat{y}\right) \\
&= \sum_{i=1}^{N} \sum_{\hat{y} \in Y} p\left(\hat{y} \mid x^i\right) R\left(\hat{y}\right)
\end{aligned} \tag{11}
$$

Where Y is the space composed of candidate translated sentences and $R(\hat{y})$ is the sentence-level reward of the translation. Because the state at time step $t + 1$ is completely determined by the state at time step t, the probabilities $p\left(\mathbf{s_1}\right)$ and $p\left(\mathbf{S_{t+1}}|\mathbf{S_t}, a_t\right)$ are equal to 1. The gradient update strategy as shown follows:

$$\nabla_\Theta J(\Theta) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{t=1}^{L} (R_L - b) \nabla_\Theta \log \pi_\Theta\left(a_t|\mathbf{s_t}\right) \tag{12}$$

Where N is the number of turns, $b \approx E[R_L]$.

The action space of reinforcement learning-based machine translation is considerable and discrete. Its size is the capacity of the entire word list. So we use beam search to sample the actions. It reduces the
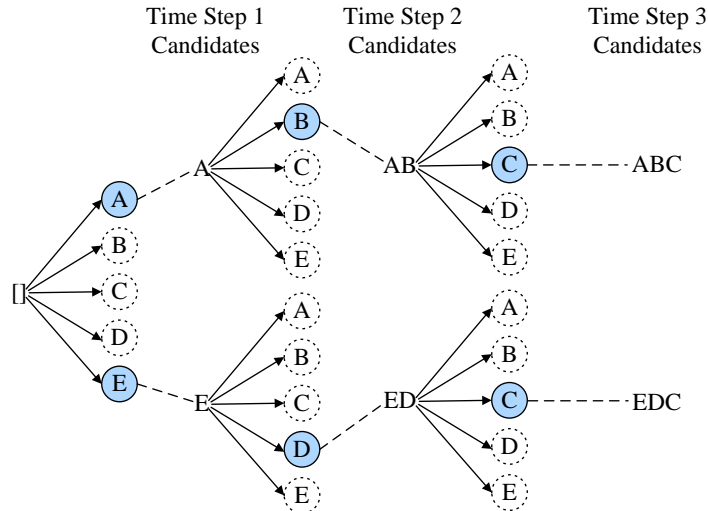
Figure 3: The schematic diagram of beam search

computational cost and increases the probability of high quality translation results in the decoding stage. The principle of beam search is shown in Figure 3.

In order to stabilize the process of reinforcement training and alleviate the large variance that may be brought by reinforcement learning, we combined MLE training goals with RL goal. The specific step is to retain the cross-entropy loss function of traditional machine translation in the loss function, and then combine it with the reinforcement learning training objective linearly. The loss function after mixing is shown below:

$$L_{\text{combine}} = \gamma \times L_{\text{mle}} + (1 - \gamma) L_{\text{rl}} \tag{13}$$

Where, L_{combine} is the binding loss function, L_{mle} is the cross entropy loss function, L_{rl} is the reinforcement learning reward, and γ is the super parameter controlling the weight between L_{mle} and L_{rl}. Different γ will affect the performance of the final translation results.

4 Experiments

4.1 Datasets

The data set used in the experiments comes from the corpus constructed by the laboratory to undertake the "China-Korea Science and Technology Information Processing Comprehensive Platform" project(Tian Mingjie, 2018). The original corpus includes more than 30,000 documents and more than 160,000 parallel sentence pairs, covering 13 fields such as biotechnology, Marine environment and aerospace. To alleviate the problem about data sparsity, we also used additional monolingual corpus in the experiment. The detailed data information obtained after preprocessing according to the task in this paper is shown in Table 1. The HTER score for quality estimation tasks is automatically calculated by the TERCOM tool.

Category	Languages	Size(sentence)
Parallel corpora		163,449
Monolingual corpora	Chinese	115,000
	Korean	115,000
QE corpora	Chinese	163,449
	Korean	163,449

Table 1: Description of data in Chinese-Korean machine translation

4.2 Preprocessing

Large-scale corpus word embedding can provide sufficient priori information for the model, accelerate the convergence rate of the model, and effectively improve the effect of downstream tasks. However, as a

low-resource language, Korean lacks a large corpus, so there will be a large number of low-frequency words in the corpus, which will lead to low quality of word vectors. To solve this problem, we use more flexible Korean language granularity for word embedding to alleviate the data sparsity problem.

Korean is a phonetic alphabet. From a phonetic point of view, the Korean language consists of phonemes that form syllables according to rules, and syllables that form sentences. Since the number of phonemes and syllables is relatively fixed (67 phonemes and 11172 syllables), the scale of dictionary construction using such granularity is very small, and the existence of low-frequency words can be significantly reduced compared with other granularities. And from a semantic point of view, word may have clearer morphological and linguistic features. Therefore, we use phoneme, syllable and word to preprocess the corpus of the Korean text. Phonemes are obtained by the open source phoneme decomposition tool hgtk, syllables are obtained by reading characters directly, and word is obtained by the word segmentation tool Kkam.

4.3 Setting

Our translation module is implemented on an encoder-decoder framework based on self-attention, and the Transformer system adopts the same model configuration as described by (Vaswani A, 2017), which is implemented using Tensor2Tensor, an open source tool built on Google Brain. We set dropout to 0.1, and the word vector dimension to 512. The MLE training gradient optimization algorithm uses the Adam (Kingma DP, 2015) algorithm and learning rate decay. In the feature extraction part of our machine translation quality estimation module, the number of encoder and decoder layer is 2, the number of hidden units of feedforward sublayer is 1024, and the number of head of attention is 4. In our quality estimation part, the network structure is single-layer Bi-LSTM, the hidden layer unit is set as 512, the gradient optimization algorithm uses Adam, and the learning rate is set as 0.001. During reinforcement learning training, the MLE model was used for parameter initialization, the learning rate was set as 0.0001, and the beam search width was set as 6.

4.4 Main results and Analysis

In order to test the translation performance of the model, we conducted Chinese-Korean machine translation experiments under the same hardware and corpus environment and calculated the BLEU and QE of the test set respectively. The results are shown in Table 2.

Model	zh-ko		ko-zh	
	BLEU ↑	QE ↓	BLEU ↑	QE ↓
LSTM+attention	11.52	83.95	12.61	83.20
Transformer	15.82	30.91	20.20	27.41
Our Method	**21.67**	**25.58**	**23.46**	**25.16**

Table 2: The score of translation performance

As can be seen from Table 2, our method can exceed the baseline model for both the Chinese-Korean and Korean-Chinese translation tasks. Compared with LSTM+attention, BLEU increases by 10.15 and QE score decreases by 58.37 in Chinese-Korean direction, and BLEU increases by 10.85 and QE score by 58.04 in Korean-Chinese direction. Compared with Transformer, BLEU increases by 5.85 and QE score decreases by 5.33 in Chinese-Korean direction, and BLEU increases by 3.26 and QE score by 2.25 in Korean-Chinese direction. Therefore, the introduction of evaluation module effectively improves the performance of Chinese-Korean machine translation.

4.5 Performance verification about QE

To ensure the rationality and effectiveness of our method, we verify the performance of the machine translation quality estimation module. Pearson's Correlation Coefficient, Mean Average Error (MAE) and Root Mean Squared Error (RMSE) used in WMT competition were used as verification indexes. Pearson's Correlation Coefficient is used to measure the correlation between the predicted value and the ground truth. The higher the positive correlation, the better the performance of QE module. MAE and RMSE

represent the mean and square root of the absolute error between the predicted value and the true value respectively, the smaller the value, the better. The baseline system adopts the open source system QuEst ++(Specia L, 2013), which is the official baseline system of WMT2013-2019. The specific experimental results are shown in Table 3.

Model	Bilingual Expert	QuEst++
Pearson's ↑	**0.476**	0.397
MAE ↓	**0.118**	0.136
RMSE ↓	**0.166**	0.173

Table 3: Verification results for QE module performance

As can be seen from the experimental results in Table 3, compared with the baseline system of QE task, the Bilingual Expert used in the experiment has a better performance improvement, with Pearson correlation coefficient increased by 0.079, MAE decreased by 0.018, and RMSE decreased by 0.007. Our results have a high correlation with manual evaluation, thus proving the effectiveness of using the machine translation quality assessment module in this experiment. In conclusion, it is reasonable to use the machine translation quality estimation module to optimize the translation module.

4.6 Example of translation results

Examples of translation results for different models are shown in Table 4.

Model	Sentence
Source	스마트폰을 이용해서 정밀한 길이 측정이 가능해요.
LSTM+attention	智能手机以测量的。
Transformer	通过使用智能手机的方式准确测量。
Our Method	使用智能手机精确测量长度成为可能。
Source	이미지가 저장이 안되고 보기만 가능해요.
LSTM+attention	图像查看因为他不能保存。
Transformer	图像只可以查看并不能保存。
Our Method	图像不能保存只能查看。
Source	而其余的英雄们回到了他们的日常生活。
LSTM+attention	내 가 다음 날 들은 일상 생활 로 돌아 갔어 .
Transformer	그리고 나머지 영웅 들은 일상 생활 로 돌아 갔어 .
Our Method	그리고 나머지 영웅 들은 일상 생활로 돌아 갔어.
Source	这些都与织物在螺杆或机头中的流动有关。
LSTM+attention	스트레스 없이 모두 외부 에 의해 바꿀 수 있어.
Transformer	이들은 나사 또는 머리의 직물 흐름과 관련이 있습니다.
Our Method	이러한 것들은 스크류나 헤드 내에서 생지의 유동과 관계가 있어.

Table 4: Examples of translation results

As can be seen from the translation examples in Table 4, the translation obtained by our method is more accurate in both directions, the fluency and fidelity of the translation are in line with the target language specification, and the quality of the translation is significantly better than that of the other baseline models, which proves that our method can effectively improve the performance of the Chinese-Korean neural machine translation model.

5 Conclusion

In order to alleviate the exposure bias and poor translation diversity problems caused by teacher forcing in machine translation, we propose a Chinese-Korean neural machine translation model that incorporates machine translation quality estimation. The model introduces an evaluation mechanism at the sentence level to guide the training of the model, so that the prediction cannot converge completely to the ground truth word. The evaluation mechanism is reference-free evaluation. The instruct mechanism is guided by the reinforcement learning. The experimental results clearly show that our approach can effectively improve the performance of Chinese-Korean neural machine translation.

Acknowledgements

References

Bengio Y Bahdanau D, Cho KH. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–15.

Li B Fan K, Wang J. 2019. Bilingual expert1 can find translation errors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 6367–6374.

Hoang H Junczys-Dowmunt M, Dwojak T. 2016. Is neural machine translation ready for deployment? In *A Case Study on 30 Translation Directions*.

Ramakrishnan N Keneshloo Y, Shi T. 2020. Deep reinforcement learning for sequence-to-sequence models. page 2469–2489.

Na S-H Kim H, Lee J-H. 2017. Predictor-estimator using multilevel task learning with stack propagation for neural quality estimation. In *Proceedings of the 2nd conference on machine translation*, pages 562–568.

Ba JL Kingma DP. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR*, pages 1–15.

Agarwal A Lavie A. 2007. Meteor: An automatic metric for mt evaluation with high levels of correlation with human judgments. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 228–231.

Ward T Papineni K, Roukos S. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 2002*, pages 311–318.

Auli M Ranzato M, Chopra S. 2016. Sequence level training with recurrent neural networks. In *4th International Conference on Learning Representations, ICLR 2016*, pages 1–16.

He Zhongjun Shen Shiqi, Cheng Yong. 2016. Minimum risk training for neural machine translation. In *Proc of the 54th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA: Association for Computational Linguistics*, page 1683–1692.

Heess N Silver D, Lever G. 2014. Deterministic policy gradient algorithms. In *31st International Conference on Machine Learning, ICML*, pages 605–619.

Schwartz R Snover M, Dorr B. 2006. A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation of the Americas: Visions for the Future of Machine Translation, AMTA*, pages 223–231.

De Souza JGC Specia L, Shah K. 2013. Quest++a translation quality estimation framework. In *Proceedings of the 51st ACL: System Demonstrations*, pages 79–84.

Cui Rongyi Tian Mingjie, Zhao Yahui. 2018. Identifying word translationsin scientific literature based on labeled bilingual topic model and co-occurrence features. In *Proceedings of Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data*, page 76–87.

Parmar N Vaswani A, Shazeer N. 2017. Attention is all you need. In *The proc of Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*, page 5998–6008.

Tao N Weaver L. 1999. The optimal reward baseline for gradient-based reinforcement learning. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 538–545.

Zipser D Williams RJ. 1989. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, pages 270–280.

Qin T Wu L, Tian F. 2018. A study of reinforcement learning for neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP*, pages 3612–3621.

Wang Feng Yang Zhen, Chen Wei. 2017. Improving neural machine translateon with conditional sequence generative adversarial nets. page arXiv preprint arXiv:1703.04887.

Jin Yongshou. 2020. Current situation and future research direction of chinese-korean translation theory. *Korean Language in China*, pages 66–73.

Meng F Zhang W, Feng Y. 2019. Bridging the gap between training and inference for neural machine translation. In *57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, ACL*, page arXiv preprint arXiv:1906.02448.