ISSN: 2688-8645

Study of Neural Machine Translation with Fused Monolingual Data

Juanjuan Gao

School of Control and Computer Engineering, North China Electric Power University, Baoding 071000, China

Abstract

A data enhancement method based on dynamic data expansion is proposed to alleviate the data sparsity problem that machine translation is facing by minimising the difference between the performance of neural machine translation and human translation and the problem of insufficient training corpus. The translation model is trained by modifying the target-side sentences to add noise and creating new pseudo-parallel sentence pairs in combination with the source-side utterances. In order to generate a variety of final translated translations, a data augmentation approach is used to construct pseudo-sentence pairs using a sampling decoding strategy in conjunction with the decoding stage, and different baseline model approaches and decoding strategies are selected for comparison experiments. The experimental results verify that the data augmentation method proposed in this paper can effectively alleviate the problem of insufficient generalisation ability of the neural machine translation model, thus improving the sentence representation ability and the performance of neural machine translation.

Keywords

Neural Machine Translation; Data Expansion; Sampling Decoding.

1. Introduction

Currently, neural machine translation models have shown good translation results on a variety of benchmarks. However, these models often rely on massively parallel preconditioning for training and show degraded performance and difficulties in machine translation on low-resource languages[1]. In addition, today's neural machine translation models are often fragile, as noise (e.g., grammatical errors) can lead to serious mistranslations.

Monolingual corpora combined with data augmentation methods allow for data augmentation, which works by expanding the number of data points used for training without manually collecting new data, and has been widely used in computing to improve diversity and robustness and to avoid overfitting on small data sets[3,4]. Although data enhancement (e.g. image inversion, cropping and blurring) has become a standard technique for training deep networks in the field of computer vision, applying it to machine translation is not an easy task[5]. Currently, in the field of machine translation, there are 2 main types of data augmentation methods, one is reverse translation and the other is lexical replacement.

Reverse translation is run in a semi-supervised environment. Zhang et al[8].proposed an adversarial method based on Self-encoder and reverse translation, which adds noise at the target end and then performs reverse translation to obtain the final language mapping. Experimental results demonstrate the effectiveness of the method for data expansion. However, reverse translation requires additional training of a translation system, which not only leads to increased computing costs, but also reverse translation can lead to degradation of translation quality.

ISSN: 2688-8645

In order to solve the problem of insufficient generalization ability of neural machine translation caused by data sparsity, this study adopts a new method, i.e. adding certain noise to the target end of the original parallel corpus, by modifying and replacing the data at the target end according to certain strategies, and at the same time using the principle of noise-reducing Self-encoder, constraining the encoder to restore the sentence at the target end before adding noise, in order to achieve the effect of data expansion, and thus The target language model's ability to express sentences is improved. In addition, in order to make the generated translations sufficiently diverse, we use the sampling decoding strategy to generate the final translations of the model in the decoding stage, and compare them with other baseline models, and demonstrate through experiments that the translations generated by the sampling decoding strategy are diverse, thus improving the generalization ability of the model.

2. Related Work

2.1. Neural Machine Translation

End-to-end neural machine translation (NMT) has been a hot topic of research in the community, and it has achieved rapid development until now. Neural machine translation has surpassed phrasal statistical machine translation in multiple language pairs [9], and it has shown great potential to become the frontier of research on machine translation under the condition of large-scale corpus and computing power [11].

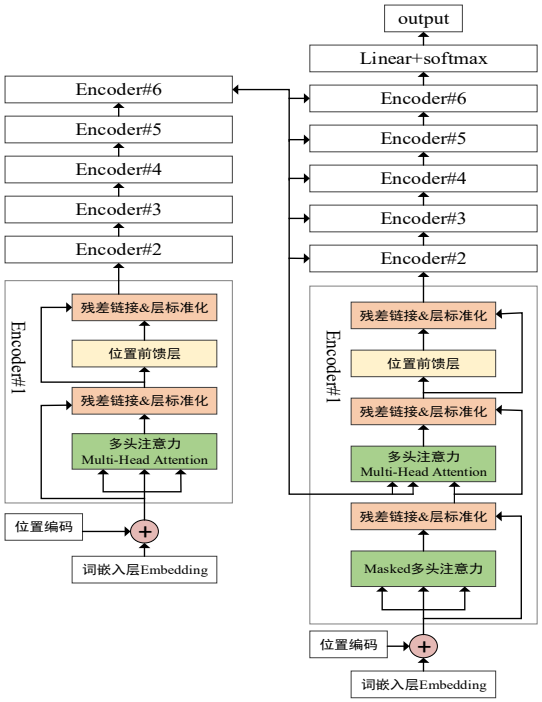

Figure 1. Transformer model structure diagram

Neural machine translation based on encoder-decoder architecture is a general model [12] and is not entirely specific to the machine translation task itself. Therefore, neural machine translation also faces problems such as limited translation sentence length, difficulty in fully utilizing the fused large-scale monolingual corpus, over-translation and under-translation. To address these problems, improving the attention mechanism is a major breakthrough in neural machine translation [13,14]. Compared with statistical machine translation, neural machine translation can handle complex structures and short sentences more effectively, but it still has

ISSN: 2688-8645

some problems and challenges in terms of translation fidelity and translation quality of long sentences. Based on this, as shown in Figure 1, this paper investigates an attention-based machine translation model, which uses the Transformer structure for neural machine translation tasks. The training goal is to maximize the likelihood of the model parameters over a parallel corpus S (|S| denotes the number of sentences in the parallel corpus).

2.2. Noise Reduction Self-encoder Model

Noise-reducing Self-encoders are an improvement on Self-encoders. Unlike Self-encoders where the weights obtained after training are chaotic and noisy, noise-reducing Self-encoders can learn features from the original data with noise, thus allowing them to show better robustness and expressiveness in data extraction.

A typical noise-reducing Self-encoder training process is shown in Figure 2, given a source-side sequence $x=(x_1,x_2,...,x_n)$, The source sequence is first noisified by introducing noise f(x) to obtain a sequence with noise $x^{'}=(x_1^{'},x_2^{'},...,x_n^{'})$, The model is remapped back to x from x' with noise by minimizing the loss function, where the loss function is:

$$Loss(x\,|\,x')=-\log P_{dec}(x\,|\,f(x')) \tag{1}$$

where, $f(x')$ denotes the output of the sequence x' with noise after input to the encoder, $P_{dec}(x\,|\,f(x'))$ denotes the noise addition probability of the model.

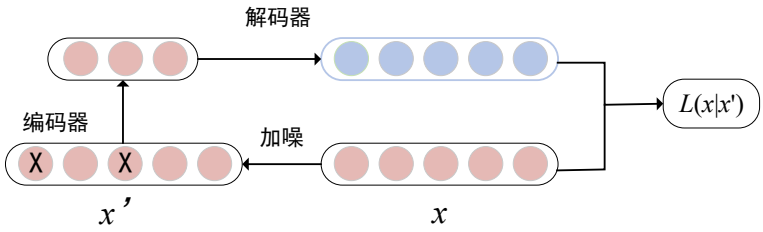

Figure 2. Noise-cancelling Self-encoder construction diagram

3. Model Structure

Similar to noise reduction self-coding, given a target-end sequence $y=(y_1,y_2,...,y_m)$, this study first adds noise-ification to the target input sentences, and for each input target-end sequence, 15% of the words are randomly selected for overwriting to obtain the sequence with noise $y^{'}=(y_1^{'},y_2^{'},...,y_m^{'})$, In particular, this study uses three types of noise strategies to obtain the target end sequences: 1) replacing selected words with 10% probability; 2) covering selected words with [MASK] with 80% probability; 3) keeping words unchanged with 10% probability. DA (dynamic expansion of data) is like methods such as [15] in the literature. The framework diagram is shown in Figure 3, where the method improves the ability of the target language model to represent the sentence by randomly noise-ifying the sentence according to some strategy each time when loading the target-side sentence. By randomly selecting some words in the target-side sentences and adding noise to them, the encoder predicts the covered words, thus enhancing the data training effect of NMT. The original sentence is restored by constraining the encoder to improve the ability of the model to represent the sentence. First construct the target sequence with noise $y=\{y_i\}_{i=1}^{n}$, The same training strategy as in [11], followed by random overwriting of the words in each sequence to obtain the target sequence with noise $y^{'}=\{y_i^{'}\}_{i=1}^{n}$, afterwards, the decoder combines the output of the encoder y' reduced

ISSN: 2688-8645

to y. The process of recombining the target sequence obtained by the decoder can be thought of as maximizing the conditional probability $p(y \mid h, y^{'}; \theta^{dec})$.

$$P(y \mid h, y^{'}; \theta_{dec}) = \prod_{i=1}^{m} P(y_i \mid h, y^{'} < i; \theta_{dec}) \tag{2}$$

The decoder generates the target words one by one from left to right, resulting in a complete translation y, the model is used in each pseudo-parallel sentence pair(x,y') ,the defined loss function is.

$$Loss(\theta_{mt}) = \sum_{i=1}^{m} -\log P(y_i \mid x, y^{'} < i; \theta_{mt}) \tag{3}$$

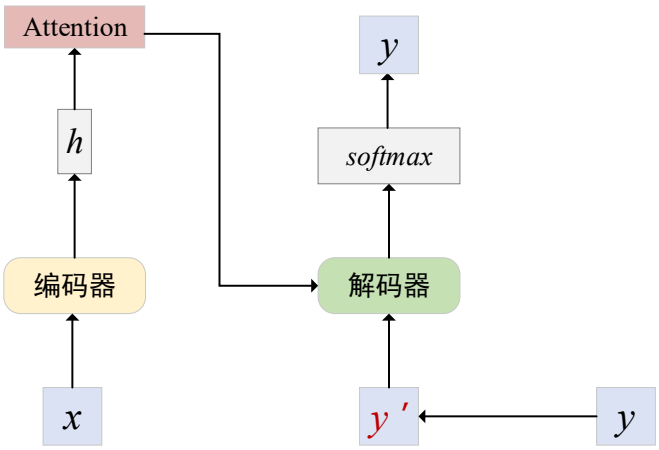

Figure 3. Neural machine translation based on data expansion

4. Experiment and Analysis

Experiments were conducted on the WMT14 English-German dataset, with the baseline model based on two network architectures, RNNSearch, Transformer, and the comparison methods base (no data augmentation), Edunov (reverse translation method + beam), and our method (dynamic noise addition at the target end + sampling decoding strategy), with each model method The experimental results on different datasets are shown in Table 1.

Table 1. English-German translation experimental results BLEU values (%)

Network Architecture	model method	Tst2013	Tst2014	average value
RNNSearch	base	20.72	22.90	21.81
	DA+beam	22.59	23.13	22.86
	Our method	23.14	24.32	23.73
Transformer	Base	26.52	27.22	26.87
	DA+beam	26.98	27.84	27.41
	Our method	27.44	28.36	27.90

Comparing the results of the base, DA+beam, and our method data enhancement methods on the English-German translation dataset, it can be seen that the pseudo-parallel sentence pairs generated using the target-side dynamic data expansion method combined with the sampling decoding data expansion method proposed in this study achieved, on the RNNSearch model and

the transformer model, respectively BLEU evaluation metrics of 23.73 and 27.90, respectively, outperforming those without any data augmentation method (base). Based on using data augmentation, we find that the data augmentation method using the sampling decoding strategy algorithm improves the BLEU by an average of +0.87 and +0.49 BLEU points over different backbone networks than the data augmentation method using the beam search algorithm, respectively.

5. Summary

The focus of this paper is to address the data sparsity problem faced by neural machine translation by using a data expansion approach to enhance the generalization ability of the model and its resistance to noise. Also, in order to improve the diversity of the final translation generated, a sampling decoding strategy is used in decoding, allowing it to generate more and more realistic translations at the sentence level. The experimental results show that the data expansion combined with sampling decoding can effectively improve the robustness of the neural machine translation model on the English-German dataset of WMT14, and has a good effect in improving the performance of machine translation. In addition, the data expansion method proposed in this study has a certain impact on the semantic information of the sentences in the lexical replacement and modification, and it can be considered in the future to add syntactic information while expanding the utterances to ensure the integrity of the sentences.

References

[1] Lin Qian, Liu Qing, Su Jin-song, et al. Focuses and frontiers tendency in neural machine translation research[J]. Journal of Chinese Information Processing, 2019,33 (11):1-14. (in Chinese).

[2] Wang Kun, Yin Ming-ming, Yu Hong-fei, et al. The study on low-resource Uygur-Chinese neural machine translation [J].Journal of Jiangxi Normal University (Natural Science), 2019,43(6):638-642. (in Chinese).

[3] Du Dong-yang, Lu Li-jun, Fu Rui-yang ,et al. Palm vein recognition based on end-to-end convolutional neural network[J]. Journal of Southern Medical University, 2019 , 39 (2):207-214.(in Chinese).

[4] Asaoka R, Tanito M ,Shibata N ,et al. Validation of a deep learning model to screen for glaucoma using images from different fundus cameras and data augmentation[J]. Ophthalmology Glaucoma, 2019, 2(4):224-231.

[5] Tustison N J, Avants B B, Lin Z, et al. Convolutional neural networks with template-based data augmentation for functional lung image quantification[J]. Academic Radiology,2019,26(3):412-423.

[6] Huang Guo, Xu Li, Chen Qing-li, et al. Research on non-local multi-scale fractional differential image enhancement algorithm [J]. Journal of Electronics and Information Technology,2019,41(12) : 2972-2979.(in Chinese).

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton.2020.A simple framework for contrastive learning of visual representations. arXiv preprint arXiv:2002.05709.

[8] Zhang Y, Li Y, Zhu Y, et al. Wasserstein GAN based on autoencoder with back-translation for cross-lingual embedding mappings[J]. Pattern Recognition Letters,2019,129:311-316.

[9] Bentivogli L, Bisazza A, Cettolo M, et al. Neural versus phrase based machine translation quality// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016). Austin, USA, 2016: 257-267.

[10] Y. Nishimura, K. Sudoh, G. Neubig and S. Nakamura, "Multi-Source Neural Machine Translation with Missing Data," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 569 -580, 2020, doi: 10.1109/TASLP.2019.2959224.

ISSN: 2688-8645

[11] Hou Qiang, HOU Ruili. A Review of Research and Development of Machine Translation Methods [J]. Computer Engineering and Applications,2019,55(10):30-35+66.

[12] Xiaoqian Sun, Yila Su, Yaping Zhao, Yufei Wang, RenQing Daoerji. Mongolian-chinese Neural machine Translation based on Encoder-Decoder Reconstruction Framework [J]. Computer Application and Software,2020,37(04):150-155+163.

[13] B. Zhang, D. Xiong and J. Su, "Neural Machine Translation with Deep Attention," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 1, pp. 154-163, 1 Jan. 2020, DOI: 10.1109/TPAMI.2018.2876404.

[14] X. Li, L. Liu, Z. Tu, G. Li, S. Shi and M. Q, H. Meng, "Attending from Foresight: A Novel Attention Mechanism for Neural Machine Translation," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 29, pp. 2606-2616, 2021, doi: 10.1109/TASLP.2021.3097939.

[15] Zhidong Liu, Junhui Li, Zhengxian Gong. A simple dynamic Data Expansion method for neural machine Translation [J]. Journal of Xiamen University (Natural Science edition),2021,60(04):680-686.