Corpus Augmentation for Improving Neural Machine Translation

Zijian Li[1], Chengying Chi[1, *] and Yunyun Zhan[2, *]

Abstract: The translation quality of neural machine translation (NMT) systems depends largely on the quality of large-scale bilingual parallel corpora available. Research shows that under the condition of limited resources, the performance of NMT is greatly reduced, and a large amount of high-quality bilingual parallel data is needed to train a competitive translation model. However, not all languages have large-scale and high-quality bilingual corpus resources available. In these cases, improving the quality of the corpora has become the main focus to increase the accuracy of the NMT results. This paper proposes a new method to improve the quality of data by using data cleaning, data expansion, and other measures to expand the data at the word and sentence-level, thus improving the richness of the bilingual data. The long short-term memory (LSTM) language model is also used to ensure the smoothness of sentence construction in the process of sentence construction. At the same time, it uses a variety of processing methods to improve the quality of the bilingual data. Experiments using three standard test sets are conducted to validate the proposed method; the most advanced fairseq-transformer NMT system is used in the training. The results show that the proposed method has worked well on improving the translation results. Compared with the state-of-the-art methods, the BLEU value of our method is increased by 2.34 compared with that of the baseline.

Keywords: Neural machine translation, corpus argumentation, model improvement, deep learning, data cleaning.

1 Introduction

At present, significant achievements have been made in various tasks in the field of natural language processing (NLP) based on deep neural networks, such as machine translation [Su, Zeng, Xiong et al. (2018); Bahdanay, Cho and Bengio (2014); Sutskever, Vinyals and Le (2014)], text classification [Socher, Perelygin, Wu et al. (2013); Kim (2014)] and recommendation algorithms [Xin, Zhang, Li et al. (2017); Hu, Liu, Zhang et al. (2018)]. However, if a large-scale high-quality bilingual parallel corpus for training is not available, a model may over-fit the data when using a small-scale dataset [Koehn and

[1] University of Science and Technology Liaoning, Anshan, 114031, China.

[2] College of Science & Health, Technological University Dublin, Dublin, D08 X622, Ireland.

[*] Corresponding Authors: Chengying Chi. Email: chichengying@ustl.edu.cn;
Yunyun Zhan. Email: D16123420@mytudublin.ie.

Received: 21 February 2020; Accepted: 11 April 2020.

Knowles (2017)]. Therefore, the effectiveness of model learning is closely related to the scale and quality of the bilingual parallel sentence pairs available. However, for low-resource languages, such as Thai, Lao, etc., the lack of standard bilingual parallel corpora and the small size of open corpora have greatly limited the learning ability of neural machine translation (NMT). It takes significant amount of time, money, and expertise to translate a large number of texts in the construction of a large-scale parallel corpus of high quality; thus, the practical cost is too high. Therefore, using computers to realize the automatic corpus construction under the premise of a basic bilingual corpus has a significant practical value for improving the translation accuracy of NMT models.

In the field of computer vision, image data enhancement technology has received significant attention and been widely adopted [Chatfield, Simonyan, Vedaldi et al. (2014); Shorten and Khoshgoftaar (2019)]. These approaches re-move noise from training data and expand the data scale through horizontal inversion, random clipping, tilting and changing the RGB channels of the original image. These techniques enhance the robustness of the model and improve its learning effect. For the field of NMT, the data enhancement technology can also be used. Here, the data can be expanded on the premise of ensuring the mutual translation of bilingual data and the rationality of sentences in the field of NLP in order to improve the final translation accuracy of the model. Fig. 1 shows the comparison of the main methods of image data enhancement and natural language data enhancement. The upper part of the figure shows the image processing method of image data enhancement technology. The lower part shows how NLP data enhancement techniques the sentences.

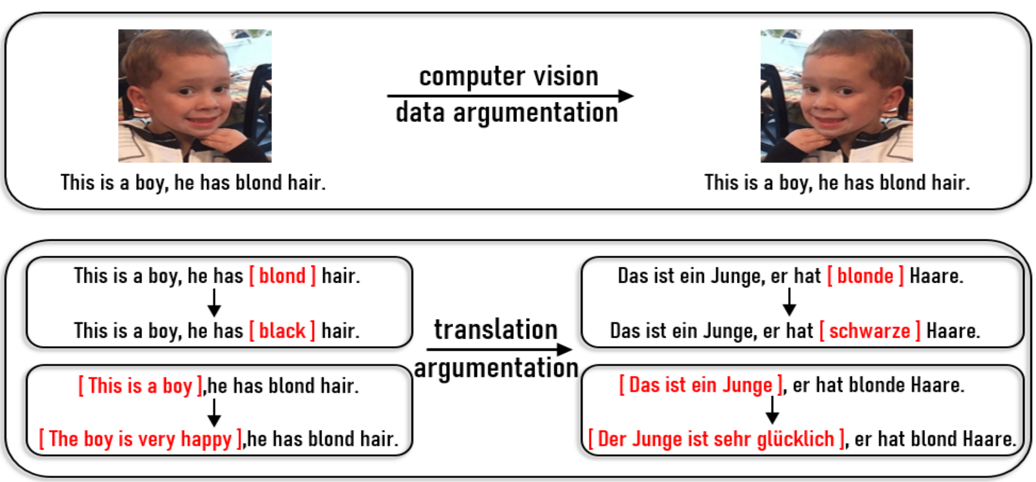


Figure 1: Comparison of data enhancement methods between computer vision and machine translation

In the prior research on data enhancement methods, many ways were proposed to improve data quality [He, Wang, Zhou et al. (2018); Duong, Cohn, Bird et al. (2015); Wan and Tao (2018)]. Among them, Bertoldi et al. solved the problem of adaptation in the field of machine translation through a pseudo-parallel corpus generated from a monolingual domain corpus [Bertoldi and Federico (2009)]. Hsieh et al. established a pseudo-parallel corpus based on the patterns learned from the source language and the monolingual target language to solve the cross-domain adaptation problem in machine translation [Hsieh, Huang and Chen (2013)]. Zhang et al. proposed a method to generate a pseudo-parallel corpus based on a source monolingual corpus and its automatic translation [Zhang and Zong (2016)]. In addition, since the resources of the monolingual data were extremely abundant, bilingual parallel corpora were also quickly constructed by the reverse translation of the monolingual target language data [Sennrich, Haddow and Birch (2016)]. Sennrich et al. proved that even if the monolingual target language data was only copied to the source language, it provided certain benefits to the training of NMT models [Sennrich, Birch, Currey et al. (2017)]. In addition, the expansion of low-frequency words was also proven to be an effective data enhancement method [Fadaee, Bisazza and Monz (2017)]. The method of back-translation, traced back to the era of statistical methods, was often applied in semi-supervised learning [Bojar and Tamchyna (2011)]. Lample et al. removed noise through the generated reverse translation data and the language model in the target language [Lample, Ott, Conneau et al. (2018)]. In addition, they also proposed a more effective reverse translation method [He, Xia, Qin et al. (2016)].

In this paper, we first filter the bilingual sentences to remove data noise and obtain relatively high quality basic data. Then, the sentence structure in the basic data is changed with different granularity data enhancement methods and a batch of new data is constructed. The new batch is merged with the basic data according to a certain proportion. The result is a large-scale training set with rich data information. In the process of data enhancement, this method uses the fast_align[3] technology to obtain the word alignment information between sentence pairs [Dyer, Victor and Smith (2013)].

2 Corpus augmentation methods

2.1 Data cleaning

Due to the various sources of bilingual parallel sentence pairs, there may be a large amount of noise in the data corpus, which can adversely affect the results of the model. If the noise data contained in the corpus are removed before the training starts, the negative effects of data noise on the model can be shielded to the maximum extent. In addition, using a high-quality bilingual parallel corpus as the basic data for data enhancement can obtain better training data during data expansion. This is because data noise is readily expanded during data enhancement; therefore, the expansion may generate excessive noise. Six processing methods are utilized to clean the data; these are described below.

2.1.1 Sentence coding filtering

The coded sentences that do not conform to the specifications in the corpus are cleared or normalized to ensure that all of the parallel sentence pairs in the corpus maintain the same coding format.

[3] https://github.com/clab/fast_align.

2.1.2 Punctuation and number filtering

Sentences are filtered that satisfy the following two points:

1. In a sentence pair, less than half of the source or target sentences are letters or numbers.
2. The source language or target sentence consists of numbers and symbols and does not contain any letters.

2.1.3 Language filtering

Language consistency is ensured in the source and target languages by filtering. The language filtering can be completed by langid.py[4].

2.1.4 Repeated filtration

Repeated sentence pairs in the corpus are removed. The repetition can result from the variety of sources used to acquire the data.

2.1.5 Length ratio filtering

The average length ratio value is filtered. This is obtained by counting the sentence length ratio in the whole corpus as a threshold.

2.1.6 Physical proofreading

Entity sentence pairs that do not conform to the corresponding specifications are manually filtered. This is accomplished by experts who compare the untranslated entities in the sentence pairs.

2.2 Word level corpus augmentation methods

Using word-level data enhancement methods, we aim to solve the problem of rare word translation in NMT training on the premise of ensuring the mutual translation of bilingual parallel sentence pairs. In the process of learning an NMT model, abundant context information of rare words cannot be obtained. As a result, the translation of rare words can-not achieve a very good effect, thus reducing the final overall translation effect of the model. Therefore, we plan to enhance the data by replacing some common words with rare words in bilingual parallel sentence pairs. In addition, in order to ensure the translation of bilingual parallel data, we also use the fast_align technology to learn the word alignment relationship between sentences, thus performing the same replacement operation on the designated word part of the target sentence sub sentence.

However, for the above mentioned alternatives, many sentence pairs with poor effects may also be produced. The reason is that it is difficult to ensure the correctness of the sentence in the process of replacing the content of the sentence. In addition, many replaced sentences are not ensured to conform to the standard grammatical structure.

---

[4] langid.py: https://github.com/saffsd/langid.py.

Table 1: Word-level bilingual substitution sample table

Source Language	Target Language	Normalization
Is this the right **bus**, please?	Is this the right **electromobile**, please?	yes
This **bus** goes downtown.	This **electromobile** goes downtown.	**yes**
It's one **block** west of here.	It's one **electromobile** west of here.	no
But where is the **elevator**?	But where is the **electromobile**?	no

As can be seen from Tab. 1, errors often occur in the normal word level replacement process. After replacing the specified words, the normalization of the replaced sentence cannot be guaranteed. In order to avoid this problem, we consider using language models to assist our work. Fig. 2 shows that before replacing the words in the sentence, the LSTM language model is used to evaluate the context fluency after replacing the specified words for evaluating the replaceability of the current words.

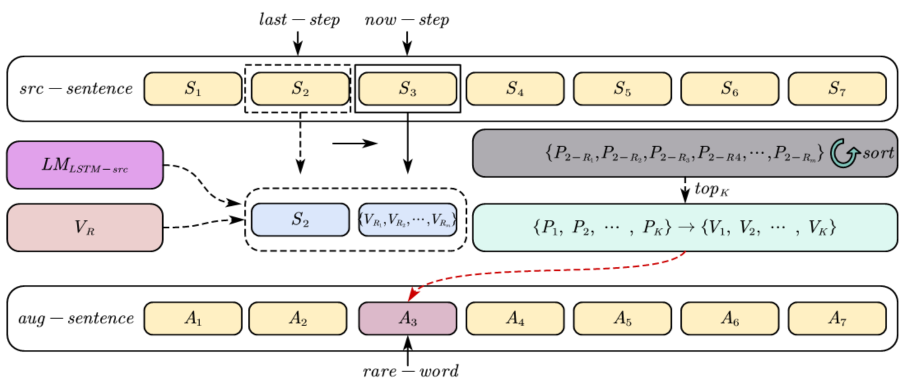


Figure 2: Example of extended sentence generation

Our work mainly consists of the following two steps:

2.2.1 One step

During the training of the NMT model, a vocabulary is generated for the words existing in the corpus. This vocabulary consists of the words that appear relatively frequently in the corpus. We also select words that appear less frequently in the vocabulary to obtain a rare vocabulary. Fig. 3 shows a word-level data enhancement method by using rare words in the data set to replace the specified words in the sentence and using LSTM language model to evaluate the fluency of the replaced sentence.

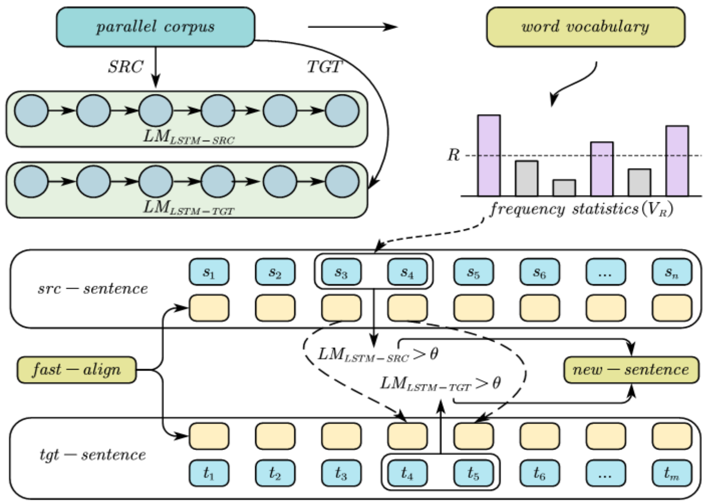


Figure 3: Example of word level data enhancement technology

2.2.2 Two step

In this step, we replace the rare words in the source statement. The replacement is determined by the trained LSTM language model. In order to ensure the substitutability of rare words, we use the trained language model to calculate the probability distribution for the words in the specified position in the sentences. The rare replacement word is obtained, and the specific formula is as follows:

$$P\left(t_j\right) = \underset{t \in trans(s_i)}{\arg\max}\, P_{LM}(t \mid t_{j-1}) \tag{1}$$

If the probability of the target replacement word t_j is underestimated by the target language model, the current reinforcement sentence pair is discarded. This approach greatly reduces the probability of sentences with incorrect semantics or syntax.

2.3 Sentence level corpus augmentation methods

The extended data set is obtained in this work by replacing clauses in the data set to be enhanced. In the training process of the NMT model, the combination of different clauses also brings many beneficial effects to the training of NMT models. In the process of model training, the translation of short sentences often depends on context information. Therefore, making the data have richer sentence context information is also an effective means to expand high-quality data. Here, we mainly use the following seven methods to expand sentences:

2.3.1 Get word alignment information

Before sentence-level data enhancement starts, it is necessary to obtain the corresponding information between sentence pair positions to ensure that the same positions of the source language and the target language can be replaced and the mutual translation of bilingual data can be ensured. The word vector information is obtained mainly by using the fast_align word alignment technology. According to the existing bilingual parallel sentence pairs, the corresponding relation of the designated positions between learning sentences is used.

2.3.2 Obtaining clause correspondence information

In this step, according to the commonly used segmentation punctuation marks in language sentences, such as "," , "?" and "." etc., long sentences are cut into different numbers of clause combinations. The source language and the target language are segmented in the same way to obtain sub sentence sets corresponding to the language direction.

2.3.3 Get clause correspondence

Assuming the active sentence S and the target sentence T , in Step 2, the sentence S is divided into n clauses according to punctuation marks, and is denoted as $\{S_1, S_2, \mathrm{S}_3, \ldots, \mathrm{S}_n\}$. The target statement T is divided into m clauses, and is denoted as $\{T_1, T_2, T_3, \ldots, \mathrm{T}_m\}$. According to the sentence position correspondence information obtained in Step 1, the correspondence probabilities between different source language clauses and target language clauses is inferred according to the correspondence relationship between words. If the probability of a generation between a pair of clauses is higher than a certain threshold θ , it is regarded as a translation sentence pair. The corresponding probability inference between the source clause and the target clause is as follows:

$$\theta = \frac{2 \times N_m}{N_s + N_t} \tag{2}$$

Here, N_m is the number of words corresponding to each other between the source language and the target language. N_s and N_t are the number of words in the source language and the target sentence, respectively.

2.3.4 Acquiring a short sentence set

From the content obtained in Step 3, an appropriate threshold θ is set to obtain the mutual translation correspondence between the source language clause and the target language clause. Then, the source language short sentence set S_k, the target language short sentence set with T_k, and the target language short sentence set k are obtained. In the process of determining the clause correspondence, if the source sub sentences S_i and S_j correspond to the target clause T_i then the source clauses S_i and S_j are combined to generate a new mutual translation correspondence with T_i.

2.3.5 Back-translation

For the corresponding sets of source language and target language clauses generated in Step 4, we use the openNMT[5] open source machine translation system to carry out a

[5] OpenNMT: https://github.com/OpenNMT/OpenNMT.

back-translation operation on the target language clause sets to generate the target language clause translation set $T_{k-trans}$.

2.3.6 Sentence generation in pseudo-source language

In this step, for each long sentence, we use a translation clause in $T_{k-trans}$ to replace the corresponding clause in the long sentence. A prerequisite for this step is that each clause in a long sentence has a corresponding target clause. If there is an unknown corresponding clause in the source sentence, then we discard it. The sentence is not used for the data enhancement because the sentence pairs in this situation may be poorly suited to augment the corpus.

2.3.7 Generation of pseudo-parallel sentence pairs

For the pseudo-source language sentence generated in Step 6, we copy the target language sentence corresponding to the original sentence and directly use it as the target language sentence corresponding to the newly generated pseudo-source language sentence.

Fig. 4 shows a method of using word alignment information to obtain clause alignment information in a sentence to facilitate sentence-level data enhancement.

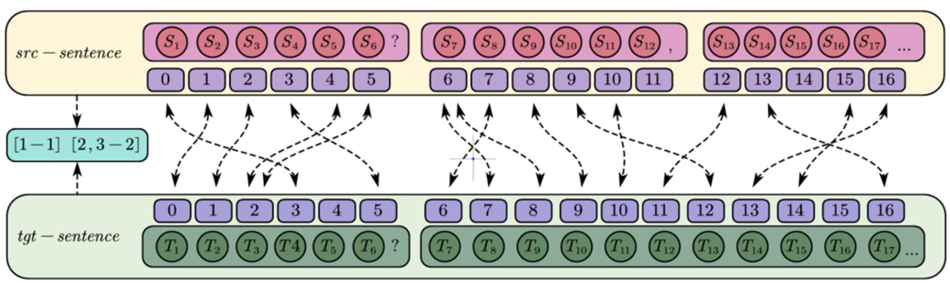


Figure 4: Sentence-level data enhancement clause correspondence

3 Experimental results

The effectiveness of the method proposed in this paper is experimentally validated. During the experiment, we use three standard data sets to test the NMT model to ensure the fairness of the test. The details are as follows:

3.1 Dataset and experimental setting

In the experiment, the English-German data set of WMT14 is used as the training set, and newtest2014 and newtest2015 are used as the test sets (refer to Tab. 2). The English-German original training data is 4.27 M, and the basic data size after data cleaning is 3.67 M. For the NMT, the fairseq-transformer model is used as the basic model for training [Ott, Edunov, Baevski et al. (2019)], and 20 rounds of training are conducted based on the existing training set to establish the system baseline. In the training process, we set the batch size to 4096 and use the 32 K Byte Pair Encoding (BPE) vocabulary to control the common 35 K vocabulary for NMT. We carry out a BPE operation on the data after data expansion to control the size of the model vocabulary.

Table 2: Data scale reference table

Base Dataset	Amount of data
Train-Dataset	3.6 M
Valid-Dataset	36 K
Test-2014-Dataset	2.7 K
Test-2015-Dataset	2.1 K

The language model uses a single-layer LSTM neural network model to estimate the probability of the simultaneous occurrence of a word context. We set the threshold R for the occurrence of rare words to 50; we set the number of words $topK$ selected from the rare word list V_R to 500 during each sentence word replacement, because these rare words can already generate enough new sentences. In the word-level data expansion experiment, rare word replacement methods conforming to syntax and grammar rules are used to replace the words in the specified positions in the source language and the target language. This expands the basic data while ensuring the mutual translation of bilingual sentences. For the sentence-level data expansion method, the threshold θ=0.5 is set, that is, a sentence pair with the correspondence probability greater than θ is treated as a translation sentence pair. We believe that when θ=0.5, sufficient expanded sentence pairs can be generated. If θ is too large, it is difficult to obtain the corresponding clause. If θ is too small, it leads to a one-to-many situation between clauses. Both cases have adverse effects on the generation of sentence pairs. Subsequent experiments have also proved this view. In addition, the back-translation technology and openNMT are used to implement the German-English NMT model to translate the German short sentence set. Using the obtained translation results, a sentence-level data expansion method is implemented.

In addition, we also compared our method with the Sennrich et al. [Sennrich, Haddow and Birch (2016)] back-translation methods. Specifically, we construct bilingual data from back-translation monolingual data and mix them into the basic data set. The mixing ratio of 1:1 is used to maintain consistency with our method.

The BLEU score [Papineni, Roukos, Ward et al. (2002)] calculated from the final translation results of the model is used to evaluate the final effect of the NMT model. We use the multi-bleu.perl[6] script in Moses to calculate the BLEU score.

3.2 Experimental process

Before starting the experiments, the quality of the basic data needs to be ensured to obtain better expanded data. Therefore, the basic data are cleaned. In the experiment, we carry

[6] https://github.com/moses-smt/mosesdecoder/tree/master/scripts/generic/multi-bleu.perl.

out data cleaning on the complete set of basic data to remove the noise data that may exist. The obtained high-quality training data are used for baseline training of the NMT model, and the experimental results are slightly better than that of the BLEU value complete dataset.

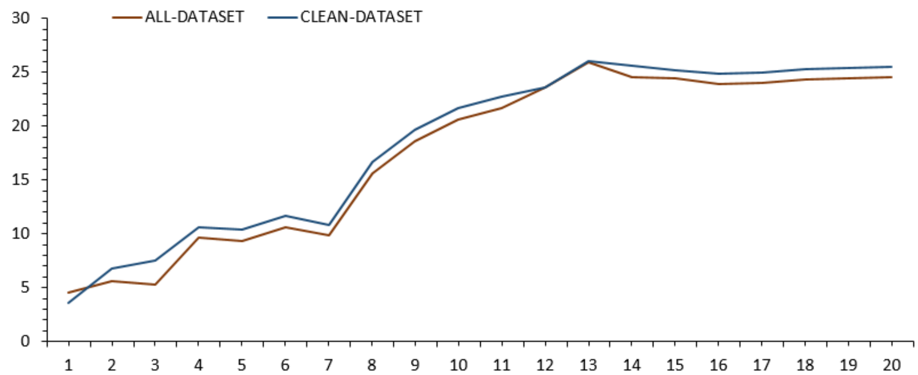


Figure 5: Comparison chart of data cleaning effect

As can be seen from Fig 5，we have removed about 600 K of bilingual data from the basic data set and obtained better experimental results. According to the BLEU value in the experimental results, we can obtain higher quality NMT model by using a smaller bilingual parallel data set. The experiments prove that the noise data in the bilingual parallel data set can adversely affect the NMT model, which relies more on high-quality bilingual parallel data set for training support. After obtaining high-quality bilingual data, we first construct the sentence-level expanded data. In the process of construction, we set the threshold θ to 0.5. This value has been determined experimentally to maximize the expanded data.

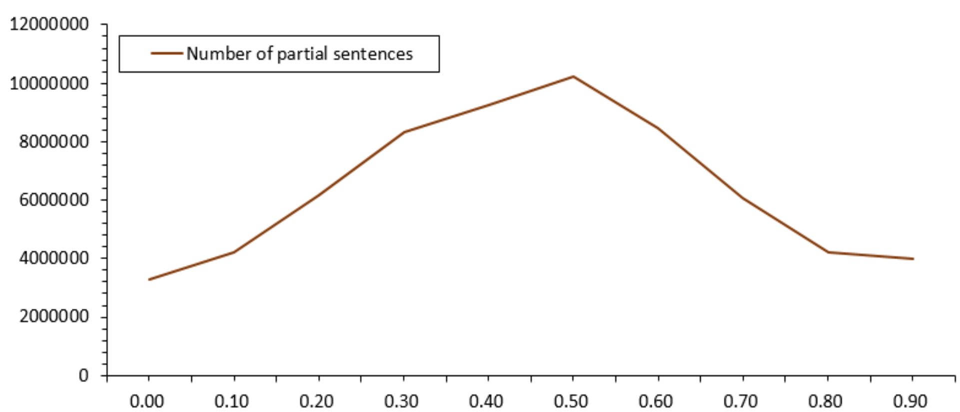


Figure 6: Number and effect of sentence-level data enhancement

As can be seen from Fig. 6, when θ is 0.5, the maximum data expansion can be constructed. This experimental result is consistent with our inference. If the parameter θ

is too large, it is difficult to obtain the corresponding relation between clauses; otherwise it is easy to make clauses appear one-to-many situation. Both of the above situations have adverse effects on the process of data construction. In addition, Tab. 3 shows the BLEU values of sentence-level data enhancement methods and back-translation methods under two different test sets.

Table 3: The Implementation effect of sentence-level data enhancement

Model	Amount of data	newtest2014	newtest2015
Base-Dataset	3.6 M	25.68	24.92
Back-translation$_{1:1}$	7.2 M	26.02(+0.34)	25.23(+0.31)
Sent-Argu-Dataset	6.0 M	**26.24(+0.56)**	**25.80(+0.88)**

As shown in Tab. 3, we use the NMT model trained by the basic data set as the baseline model of the method. In addition, the data constructed by the back-translation method is mixed with the original data set according to a ratio of 1:1, and the NMT model obtained by training is taken as a comparison model. The BLEU values of the above two models are compared with NMT model trained by data set processed by sentence-level data enhancement methods.

According to the experimental results, sentence-level data enhancement method can effectively improve the performance of NMT model and achieve better results than previous research results.

In addition, we also carry out word-level data expansion; the basic data is also used for this study. In our experiment, we use the language model corresponding to the language direction to infer the feasibility of substitution for both source and target sentences. Only when the evaluation probability of the substitution context is higher than the corresponding threshold value do we carry out the data enhancement operation to obtain the latest data. Tab. 4 shows the BLEU value performance of word-level data enhancement methods under two different data sets.

Table 4: The implementation effect of word-level data enhancement

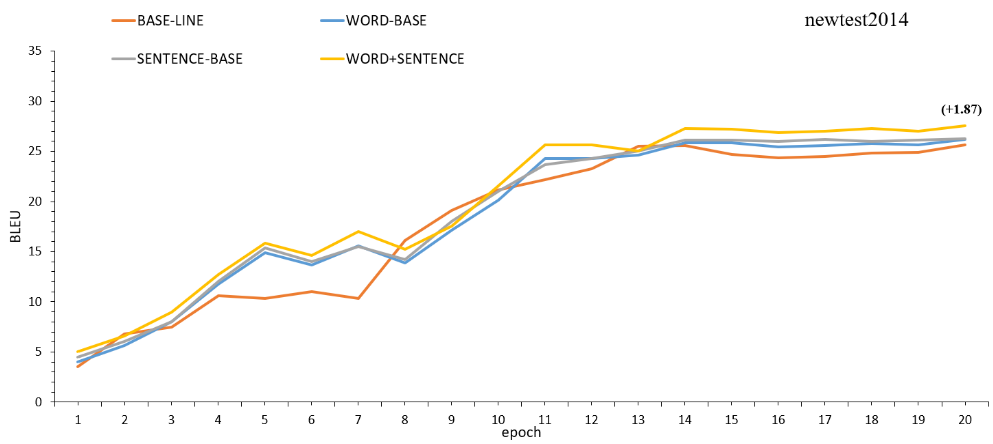
Model	Amount of data	newtest2014	newtest2015
Base-Dataset	3.6 M	25.68	24.92
Back-translation$_{1:1}$	7.2 M	26.02(+0.34)	25.23(+0.31)
Word-Argu-Dataset	6.0 M	**26.19(+0.51)**	**25.30(+0.38)**

As shown in Tab. 4, we use the data set processed by the word-level data enhancement method as the training set for NMT model training. The obtained model to compare the BLEU value results of baseline and back-translation models to prove the effectiveness of the word-level data enhancement method.

According to the experimental results, the word-level data enhancement method can construct more sentences with rich word combination information, and can ensure the rationality of the data. This method can effectively improve the translation quality of

NMT model.

In addition, after expanding the basic corpus in two ways (sentence and word-level), we add the word-level expanded sentence pairs and the sentence-level expanded sentence pairs to the basic data according to a certain proportion. NMT models often rely on word and sentence-level context in the translation process. It is generally believed that the model is equally important for modeling rare words and the learning context information of clauses, so we expand the basic data according to the ratio of 1:1 in the experiment. Fig. 7 shows the BLEU value results of the method on the newtest2014 and newtest2015 test set.



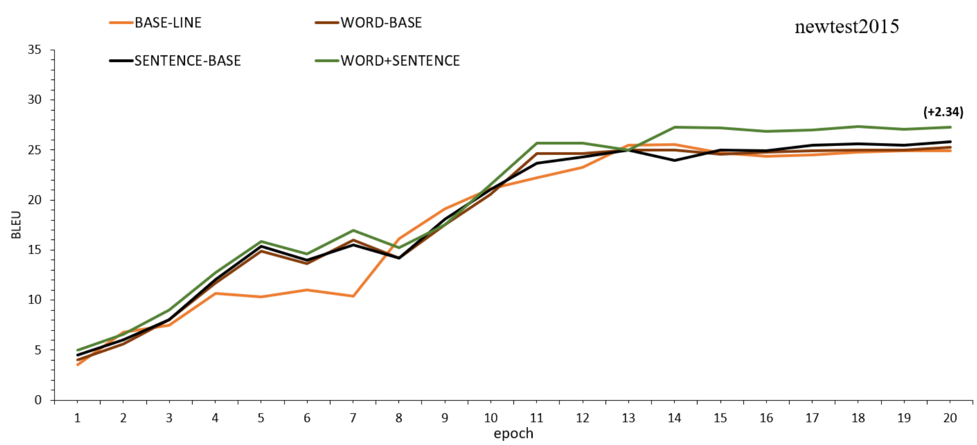


Figure 7: Performance comparison of different methods on different data sets

In this paper, a new data enhancement method is implemented by merging the two methods, and the performance of BLEU value is tested on two test sets for different methods.

The experimental results show that the NMT model trained with the bilingual data set constructed by the hybrid method can obtain significantly higher BLEU value results than those of the other methods. Using this method to construct sentences can effectively enhance the basic data set and bring the greatest performance improvement to NMT model.

4 Conclusion

In recent years, the main trend of low-resource NMT research has focused on how to better develop basic data resources. This paper proposes an effective method to improve the quality of NMT model training corpora. The experimental results indicate that using the rare words and clause combination information in an enhanced data set to realize a more powerful NMT model is simpler and more effective.

By using word alignment, language model, back-translation, and other technologies, we have realized the generation of new parallel sentence pairs on the basis of the original basic data set. According to the experimental results, the method produces more rare word contexts and more clause combinations in the new data corpus, which enable the model to obtain more abundant data sample information during the training process. Therefore, we can train a more competitive NMT model under the condition of limited data resources. This method is of great significance when it is difficult to obtain a large number of bilingual parallel data sets.

Funding Statement: This research was supported by the National Natural Science Foundation of China (NSFC) under the grant (No. 61672138).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Bahdanay, D.; Cho, K.; Bengio, Y. (2015): Neural machine translation by jointly learning to align and translate. arXiv:1409.0473.

Bertoldi, N.; Federico, M. (2009): Domain adaptation for statistical machine translation with monolingual resource. arXiv:1511.06709.

Bojar, O.; Tamchyna, A. (2011): Improving translation model by monolingual data. arXiv:1511.06709.

Chatfield, K.; Simonyan, K.; Vedaldi, A.; Zisserman, A. (2014): Return of the devil in the details: delving deep into convolutional nets. https://doi.org/http://dx.doi.org/10.5244/C.28.6.

Duong, L.; Cohn, T.; Bird, S.; Cook, P. (2015): A neural network model for low-resource universal dependency parsing. https://www.aclweb.org/anthology/D15-1040.

Dyer, C.; Chahuneau, V.; Smith, N. A. (2013): A simple, fast and effective reparameterization of IBM model 2. https://www.aclweb.org/anthology/N13-1073.

Fadaee, M.; Bisazza, A.; Monz, C. (2017): Data augmentation for low-resource neural machine translation. arXiv:1705.00440.

Hsieh, A. C.; Huang, H. H.; Chen, H. H. (2013): Use of monolingual in-domain corpor a for cross-domain adaptation with hybrid MT approaches. https://www.aclweb.org/anthology/W13-2817.

He, M.; Wang, H.; Zhou, H.; Wang, P.; Ju, A. (2018): Symmetric learning data augmentation model for underwater target noise data expansion. *Computers, Materials & Continua*, vol. 57, no. 3, pp. 521-532.

He, D.; Xia, Y.; Qin, T.; Wang, L.; Yu, N. et al. (2016): Dual learning for machine translation. http://papers.nips.cc/paper/6469-dual-learning-for-machine-translation.

Hu, J.; Liu, L.; Zhang, C.; He, J.; Hu, C. (2018): Hybrid recommendation algorithm based on latent factor model and personal rank. *Journal of Internet Technology*, vol. 19, no. 3, pp. 919-926.

Kim, Y. (2014): Convolutional neural networks for sentence classification. arXiv:1408.5882.

Koehn, P.; Knowles, R. (2017): Six challenges for neural machine translation. arXiv:1706.03872.

Lample, G.; Ott, M.; Conneau, A.; Denoyer, L.; Ranzato, M. (2018): Phrase-based & neural unsupervised machine translation. arXiv:1804.07755.

Ott, M.; Edunov, S.; Baevski, A.; Fan, A.; Gross, S. et al. (2019): Fairseq: a fast, extensible toolkit for sequence modeling. arXiv:1904.01038.

Papineni, K.; Roukos, S.; Ward, T.; Zhu, W. J. (2002): BLEU: a method for automatic evaluation of machine translation. https://www.aclweb.org/anthology/P02-1040.

Sutskever, I.; Vinyals, O.; Le Q. V. (2014): Sequence to sequence learning with neural networks. arXiv:1409.3215.

Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C. D. et al. (2013): Recursive deep models for semantic compositionality over a sentiment treebank. https://www.aclweb.org/anthology/D13-1170.

Sennrich, R.; Birch, A.; Currey, A.; Germann, U.; Haddow, B. et al. (2017): The university of edinburgh's neural MT systems for WMT17. https://www.aclweb.org/anthology/W17-4739.

Sennrich, R.; Haddow, B.; Birch, A. (2016): Improving neural machine translation models with monolingual data. arXiv:1511.06709.

Su, J.; Zeng, J.; Xiong, D.; Liu, Y.; Wang, M. et al. (2018): A hierarchy-to-sequence attention neural machine translation model. *IEEE/ACM Translations on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 623-632.

Shorten, C.; Khoshgoftaar, T. M. (2019): A survey on image data argumentation for deep learning. *Journal of Big Data*, vol. 6, no. 1, pp. 60.

Wan, T.; Tao, L. (2018): Building ontology for different emotional contexts and multilingual environment in opinion mining. *Intelligent Automation and Soft Computing*, vol. 24, no. 1, pp. 65-71.

Xin, M.; Zhang, Y.; Li, S.; Zhou, L.; Li, W. (2017): A location-context awareness mobile services collaborative recommendation algorithm based on user behavior prediction. *International Journal of Web Services Research*, vol. 14, no. 2, pp. 45-46.

Zhang, J.; Zong, C. (2016): Exploiting source-side monolingual data in neural machine translation. https://www.aclweb.org/anthology/D16-1160.