www.cadanda.com Computer-Aided Design AND APPLICATIONS

Application of Artificial Intelligence in Computer-Assisted English Vocabulary Translation

Yuanyuan Wang[1] and Wei Dong[2,*]

[1]Department of Foreign Languages, Qinhuangdao Vocational and Technical College, Hebei 066100, China, wangyy@qvc.edu.cn
[2]Department of Foreign Languages, Qinhuangdao Vocational and Technical College, Hebei 066100, China, dw@qvc.edu.cn

Corresponding author: Wei Dong, dw@qvc.edu.cn

Abstract. In the era of artificial intelligence, machine translation based on neural networks and deep learning is an important tool for computer-aided English translation. However, the accuracy of existing word-based domain feature learning methods in domain recognition is low, which reduces the efficiency and accuracy of English word translation. As a result, this paper proposes a multi-domain neural machine translation method based on word domain feature sensitivity to address the problem of translation models considering word domain features in isolation when there is no apparent domain tendency in a sentence and low domain discrimination accuracy. Compared with other models, the proposed model can extract contextual features on top of domain features of words and calculate enhanced domain proportions for each word to guide translation generation. A significant improvement in word translation accuracy is also observed in the proposed model compared with the baseline model, as well as a stronger learning ability, which has significant potential for use in English word translation.

Keywords: English vocabulary translation, artificial intelligence, machine learning, domain feature-sensitive.
DOI: https://doi.org/10.14733/cadaps.2023.S5.32-41

1 INTRODUCTION

With the rapid advancement of deep learning, Neural Machine Translation (NMT) models have demonstrated excellent performance in a range of translation tasks. Consequently, Multi-Domain NMT has become a new research hotspot because of the increasing demand for high-quality machine translation in professional fields [1].

Multi-Domain NMT aims to design and build a unified cross-domain NMT model that can automatically discriminate the domain of input sentences and translate them accordingly. The most basic approach involves combining bilingual data from multiple domains to train a unified translation model, which can then be enhanced by learning linguistic features. The data, however, come from different sources. Based on the findings of Cheng et al. [2], it is possible to train a model in a domain with good performance if sufficient high-quality training data are available. However, heterogeneity of data from multiple domains may affect the performance of a unified model. Due to the heterogeneity of multi-domain data, a uniform model may perform worse than one trained on domain data alone. Thus, building multi-domain neural machine translation is still a challenge.

Researchers model domain discrimination using context as the domain feature. In addition, multi-domain NMT usually implements a joint learning framework based on domain discrimination and translation [3]. With multi-domain NMT based on word domain feature encoding, domain features are learned from each word, which results in a significant performance improvement compared with multi-domain NMT based on sentence domain feature encoding; however, these models have difficulty providing accurate translations when there are no obvious domain tendencies in the sentence. The accuracy of existing word-based domain feature learning methods in domain recognition is low, and there is still a lot to learn about domain feature learning.

Therefore, we propose a word-level domain feature-sensitive learning mechanism, including 1) encoding of contextual features at the encoder side, to extend the range of word-level domain feature learning by introducing convolutional neural networks to extract word strings with different window sizes as word contextual features in parallel; 2) enhanced domain feature learning, designing a domain discriminator module based on a multilayer perceptron to enhance the domain proportion from word contextual features. The domain discriminator module based on multilayer perceptron is designed to enhance the learning ability to acquire more accurate domain proportions from word context features and improve the accuracy of word domain discrimination.

2 RELATED WORK

NMT multi-domain models are classified into two types based on how they learn domain features: word-level models and sentence-level models. In Kumari et al. [4], domain labels are attached to each source sentence in order to incorporate domain information into the model. Jamel et al. [5] added a domain discriminator outside the encoder to discriminate the source sentence's domain from the encoder output's embedded representation and integrated domain knowledge into the model representation by learning both the translation model and the domain discriminator in tandem. This method uses sentence-level fusion of domain features and is simple and intuitive. The model, however, is unable to learn the domain feature representation of words, so it can't provide more accurate word translations.

The word-level multi-domain NMT approach focuses more on the domain features carried by words. Su et al. [6] divided the word embedding representation into the generic domain part and the domain-specific part. Domain-specific data were activated only during training, and domain features were then incorporated into the word embedding representation by using the generic domain data. By using two different attention mechanisms, Zhao et al. [7] developed a domain-specific representation of the source sentence as well as a generic representation for the encoder. Additionally, they adjusted the objective function to refine the training by adjusting the attention weights at the target side. After that, Chen et al. [8] proposed the concept of Domain Proportion, which sets a domain proportion vector for each word and obtains the weights of various domains through training; then we redesigned the Transformer model by setting different attention mechanisms for different domains in each layer, and adding model parameters layer by layer based on domain proportion. In order to achieve domain mixing, the model parameters of the domain scale are weighed and averaged. This approach allows effective domain knowledge sharing and captures fine-grained domain-specific knowledge, which solves the problem of lack of adaptation to each domain caused by the previous approach forcing one encoder to learn shared embedding representation across all domains. However, when the number of target domains increases, the model degrades in translation quality. Thus, Dabre et al. [9] proposed a domainaware self-attentive mechanism based on the fact that each word has a domain vector in addition to a semantic vector. The learned domain vector representation is incorporated into the encoder and decoder of NMT together with the semantic vector representation, forcing the NMT model to encode and decode both semantic and domain information.

In sum, the purpose of this paper is to extract word contextual features and encode word-level domain scaling. Firstly, we propose a mechanism for learning domain features at the word level by extending the domain feature learning range to the word level and incorporating multiple convolutional neural networks into the encoder to extract context with varying window sizes as additional features of words at the same time. By replacing the original domain proportion with reinforced domain proportion, the domain discriminative accuracy of words is further enhanced.

3 MULTI-DOMAIN NEURAL MACHINE TRANSLATION MODEL

We design a multi-domain translation model based on word domain feature sensitivity in order to make the model learn domain features effectively. As a means of extending the domain feature learning scope at the word level, we introduce a domain context sensitive mechanism (DCSM) to learn single word contextual features from word strings with different window sizes; on the other hand, to enhance the learning of word domain features, we designed the enhanced domain discrimination mechanism (RDDM) to enhance word domain feature learning by improving the domain discrimination accuracy of words through joint learning. According to Figure 1, the overall model framework is based on a Transformer and the DCSM and RDDM modules are located in the purple and green parts, respectively, of the encoders and decoders.

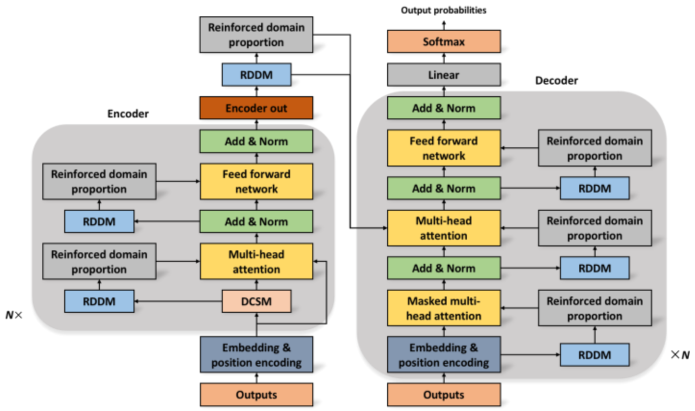

Figure 1: Architecture of proposed NMT model.

3.1 Domain Context Sensitive Mechanisms

Words that do not have significant domain properties, but surround them with words that have strong domain properties, can be disambiguated by combining them; if a word has no substantial domain properties, and the words around it do not have strong domain properties, but they can be combined into word strings to provide clear domain attribution. Figure 2 illustrates how DCSM expands the domain feature learning scope of the current word. In contrast to existing models that focus solely on individual isolated word domain features, this model further learns word context representations to provide a better understanding of word domain features.

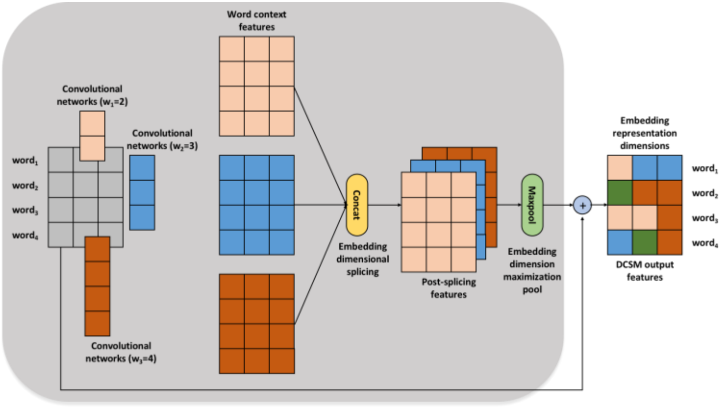

Figure 2: Architecture of domain context sensitive mechanism.

In order to give words with more definite domain features, three convolutional networks are introduced based on the theory that strings of words with a larger granularity have more definite semantics. This is based on the concept of Wang et al. [10], who integrated convolutional neural networks into a translation model to extract character features of varying granularities. We use the same number of convolutional kernels as the embedding representation size for each convolutional network and use the same padding to extract multi-channel domain features from word contexts while maintaining the same shape of the output. After extracting the word context domain features from three convolutional networks in parallel, we stitch these features in the embedding dimension and perform maximum pooling. By stitching and max-pooling in the embedding dimension, we can learn which domain features represent the most salient context. After extracting the contextual domain features of words, we finally sum the input and output of DCSM using residual concatenation in order to preserve the semantic features of individual words.

3.2 Enhanced Domain Discrimination Mechanism

The domain scale, as a learnable weight vector of words in each domain, can incorporate word-level domain features for the modules in the Transformer. In our initial experiments with DCSM, we found that the domain discrimination accuracy of the model improved to a certain extent, but due to the simple structure of the domain scale learning mechanism, the model ended up with a learning bottleneck, and the domain discrimination accuracy of words improved but only remained at a low level; In addition, due to the simple structure of the domain scale learning mechanism, the model does not take full advantage of these additional domain features, resulting in the translation performance of the model is not effectively improved. For this reason, we propose a reinforced domain discrimination mechanism (RDDM), which is based on the reinforced domain proportion of each word rather than the domain proportion in the previously examined models, in order to improve the domain recognition accuracy of words and make full use of the contextual features learned by DCSM. RDDM, on the other hand, can provide a more accurate representation of domain features and enhance word domain recognition accuracy.

As part of the model representation, the domain proportions of each word reinforcement learned by RDDM can be incorporated as domain features. Transformer's FFN module, for example, outputs the proportion of the reinforcement domain P corresponding to an arbitrary input sequence X. The input sequence X first passes through the first fully-connected network of the FFN, which projects the input representation to a representation of a multiple of the domain size, and then uses P as the weight, which is calculated by weighting the embedded representation, and another fully-connected network, which performs a similar operation to obtain the final model representation incorporating the domain features, while doubling the weight of the entire FFN.

Similarly, suppose the enhanced domain scaling is incorporated into the Transformer's multi-headed attention mechanism. In that case, the query sequence Q, the key sequence K, and the value sequence V of the multi-headed attention mechanism are first projected into the space of multiples of the number of domains of the original representation size before the computation of the self-attentive function of the deflated dot product operation, and the enhanced domain scaling P is used as the weight-to-weight Q, K, and V, respectively. The weighted sum of Q, K, and V is reprojected, and the attention calculation is finally performed.

4 RESULT ANALYSIS

4.1 Data Sources and Experimental Settings

We validate the effectiveness of the proposed approach on a multi-domain (English-Chinese) translation task. The training set is 445K, 208K, 444K, 263K, and 216K, the validation set is 2K, and the test set is 462, 456, 1500, 503, and 455. We preprocessed the datasets, first, we de-duplicated the data and removed the non-printing characters. Second, we used the open-source toolkit MOSES to normalize the sentences for punctuation and tokenize the English and French sentences; meanwhile, we used the Jieba1 tool to split the Chinese sentences. Finally, the pre-processed parallel data were sub-phrase sliced using BPE. In addition, we use domain classification accuracy curves to evaluate the model in order to evaluate the learning effect of this model and traditional machine translation models on word domain ratio. In the inference stage, we use beam search for the model and set the beam size to 5.

Multi-domain neural machine translation uses all domain data to train translation models, but it is also possible to obtain translation models for a domain by training with only one domain's data. As well, we classify multi-domain translation models into sentences and words based on how they learn domain features. Based on the above considerations, we design and reproduce the corresponding comparative models as follows: 1) Single: We train the translation model on a single domain using Transformer data from a single domain. 2) Mixed: We train the translation model using a unified multi-domain neural machine based on Transformer with a mixture of data from all domains. 3) Discriminator (Disc): This method encodes domain features at the sentence level and adds a domain discriminator outside the encoder for domain feature representation learning. 4) Adversarial Discriminator (AdvL): Based on Disc, this method takes the opposite of the gradient of the back-propagation process from discriminator to encoder, thus weakening the gradient of the encoder. 5) Partial Adversarial Discriminator (PAdvL): At the sentence level, this method performs domain feature representation learning based on the combination of Disc and AdvL. 6) Word-level Adaptive Layer-wise Domain Mixing (LWDM): This method is based on the combination of Disc and AdvL. With Layer-wise Domain Mixing (WALDM), a domain discriminator is included as part of the encoder-decoder in order to learn the domain proportion of words involved in the translation model. Additionally, all models in this study are implemented using Fairseq v0.10.2, and hyperparameters are set by default. The Adam optimizer is used for all models, and the inverse square root algorithm is used to dynamically adjust the learning rate.

4.2 Experimental Results

Figure 3 shows the experimental results of this model versus the comparison model in the direction of UM-Corpus English-Chinese translation. Among them, Single achieves 71.83 and 33.39 BLEU values for Law and News, which are the highest compared with other models because training one domain data alone does not introduce additional noise. In contrast, the performance of training with only single domain data is not satisfactory for Science and Spoken due to the small amount of data in each domain and its complexity. While the volume of data in Law and these two domains is comparable, the form of the data in Law is more standardized, so training the translation model alone performs better. Training with mixed data for these domains can improve the model's performance.

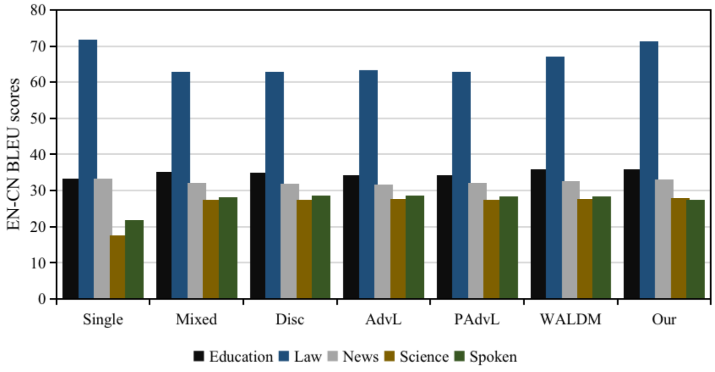

Figure 3: EN-CN BLEU scores of different models.

In the multi-domain approach based on the sentence level, Disc, AdvL and PAdvL have improved in Law, Science, and Spoken compared to Mixed. Moreover, Disc and AdvL achieve a BLEU of 28.70 on Spoken, which is the highest level compared to other models. Among the word-level domain coding methods, WALDM outperforms Mixed in all domains, with an average improvement of 1.2 BLEU values compared to Mixed, which is significant. Compared with WALDM, our model outperforms WALDM in all four domains except the Spoken domain, and the average BLEU value of all domains is improved by 0.82, which is the highest average performance among all models. In addition, our model is close to the highest level of Single performance in Law, with 4.1 BLEU higher than WALDM, which is a significant improvement.

We note that the results of this model on News and Spoken still need to be improved compared with the highest values because the domain characteristics of these two domains are weaker than the other three domains. The model has the ability to extract more domain features to learn and generate accurate translations for specialized domains, while the model has difficulty in extracting domain-specific information for "mixed" domains such as News and Spoken, resulting in poor performance in these two domains.

To further verify the effectiveness of this model on other language pairs, we also conducted experiments on the OPUS English-French dataset, and the experimental results are shown in Figure 4. The WALDM based on word-level domain coding improves the BLEU value in all three domains compared with Mixed, with an average improvement of 1.35 BLEU values, among which, the JRC improves 0.52 BLEU values compared with Single. The model achieves the highest level in all three domains, with an average performance improvement of 0.91 BLEU values compared to Single and 1.06 BLEU values compared to WALDM, indicating that the model can learn more contextual domain knowledge in the specialized domains and improve the performance of the model in each domain.

In addition to evaluating the translation performance of the model using BLEU values, we also compare the domain discrimination accuracy of this model with that of the comparison model WALDM in different translation directions, as shown in Figure 5 and Figure 6.

Figure 5 and Figure 6 show the word domain discrimination accuracy curves of the two models in the validation set for the English-Chinese and English-French translation tasks during the training process. As the number of iterations increases, the accuracy curves of WALDM in the two translation tasks fluctuate significantly in the process of increasing, but the model can effectively learn more domain features so that the curves fluctuate less and have some stability. In task 1, the highest accuracy rate of WALDM is 53.22%, while the accuracy rate of this model is 63.29%; in task 2, the highest accuracy rate of WALDM is 64.96%, while the accuracy rate of this model is

83.02%, which is a more remarkable improvement than that in task 1. It is proved that this method enhances the domain feature learning of words by improving the accuracy of word domain discrimination in the English-French task and finally improves the translation performance of the model.

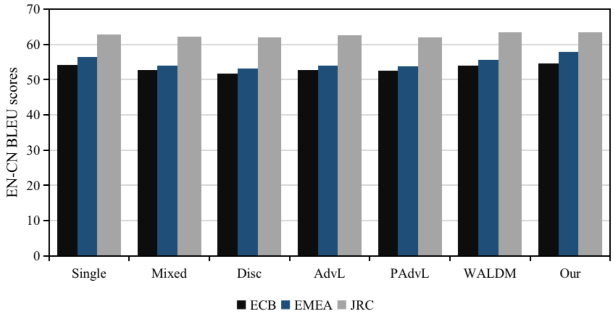

Figure 4: EN-CN BLEU scores of different models.

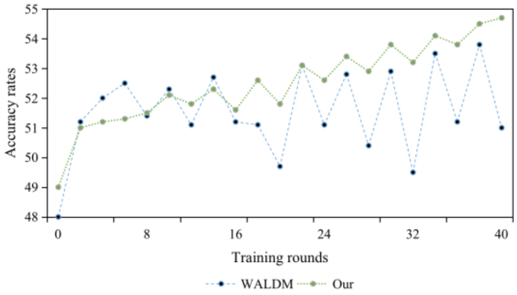

Figure 5: Comparison of accuracy of domain discrimination (Task 1).

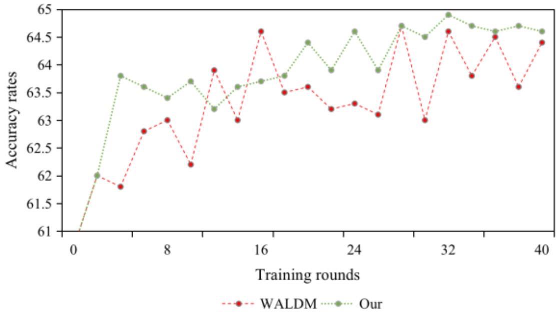

Figure 6: Comparison of accuracy of domain discrimination (Task 2).

4.3 Ablation Experiments

Considering that this model introduces additional model parameters to a certain extent, we propose an ablation experiment to verify the effect of different model parametric quantities on experimental results. Figure 7 shows the experimental results. When Mixed (big) is layered over

Mixed, the performance can be improved to some extent by increasing the number of model parameters. In comparison with the word-level-based translation model WALDM and the present model, the BLEU values are 0.58 and 1.40 lower. This is because increasing the number of Transformer model parameters to a certain extent is not as effective as a model designed for multi-domain data characteristics. Further, with only a few parameters introduced from WALDM, this model's BLEU value increases by 0.82 compared with the number of parameters introduced from Mixed to Mixed (big), suggesting that this model can achieve higher performance with fewer parameters.

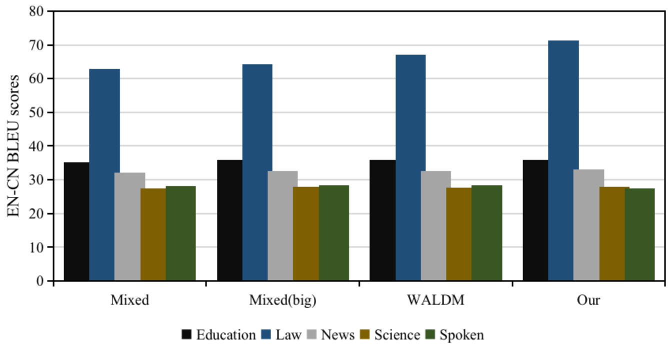

Figure 7: EN-CN BLEU scores of different models with similar parameters.

The present paper conducts ablation experiments on DCSM and RDDM in order to further investigate their role as a whole mechanism. Figure 8 shows the results. As a result of removing RDDM from the model, the domain discrimination accuracy decreases by 8.76% compared to the model in this paper, but improves by 1.31% compared to 53.22% in the WALDM model. In contrast to the DCSM model in this paper, the BLEU value decreases by 1.07, indicating that the model learns different contexts as domain features. Nevertheless, it is limited by not using the domain discriminator module, which does not fully incorporate these additional features into the model and introduces a lot of irrelevant noise. The RDDM can learn the domain ratio more effectively and improve the domain discrimination accuracy of words when the DCSM is removed from the model. Since the model lacks additional word context features extracted from the DCSM, its BLEU value decreases severely, leading to overfitting of the reinforced domain discriminator. During training, the model has high accuracy in word domain discrimination due to the fact that it lacks further word context features extracted by DCSM, resulting in overfitting of the enhanced domain discriminator. The model lacks domain-specific features learned by DCSM to complement it, and finally, the translation performance in the test set is poor.

As opposed to using DCSM alone, RDDM can effectively utilize this model after learning additional domain features with DCSM without affecting the model performance as noise; when compared with using RDDM alone, this model provides DCSM with additional extracted word context features so that RDDM does not overfit. When DCSM and RDDM are selectively removed from the model, it is found that they complement each other, and that they should be used together when building the model.

5 CONCLUSION

To address the problems of traditional machine translation models interpreting domain features of words in isolation and the low accuracy of domain discrimination, we improve the following aspects: 1) we propose an efficient mechanism for encoding contextual features: based on focusing on word-level domain features, we learn domain features of multiple granularities from the context of words, extend the scope of domain feature learning at the word level, and enhance the ability of the model to encode domain features of words. (2) We propose an enhanced domain discrimination mechanism: we further enhance the domain feature learning of words and improve the domain discrimination accuracy of words.

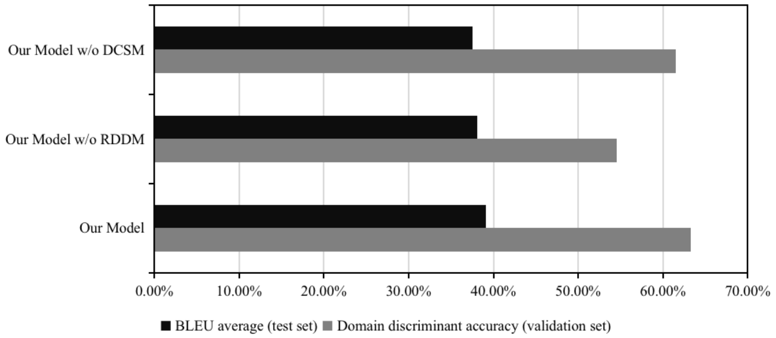

Figure 8: Results of ablation experiment.

The experimental results show that our model improves the domain discrimination accuracy by 10.07% and 18.06% in different tasks compared with the baseline model. 3) We conduct experiments on the public dataset. The experimental results show that the proposed multi-domain NMT model improves the accuracy of English-Chinese translation in four out of five domains compared with the baseline model. The average BLEU value of all five domains exceeds 0.82%. In conclusion, the proposed model has significantly improved the word recognition rate and accuracy in all three domains. Future research can further develop and improve the domain feature learning mechanism from the comprehension of the context in English translation tasks to improve the translation performance of the model in different task environments.

Yuanyuan Wang, https://orcid.org/0000-0002-7502-7748
Wei Dong, https://orcid.org/0000-0002-9355-6811

REFERENCES

[1] Krüger, R.: Explicitation in neural machine translation, Across Languages and Cultures, 21(2), 2020, 195-216. https://doi.org/10.1556/084.2020.00012

[2] Cheng, Z.; Zhang, H.; Tan, Y.; & Lim, Y.: SMT-based scheduling for overloaded real-time systems, IEICE Transactions on Information and Systems, 100(5), 2017, 1055-1066. https://doi.org/10.1587/transinf.2016EDP7374

[3] Singh, S.-M.; & Singh, T.-D.: An empirical study of low-resource neural machine translation of manipuri in multilingual settings, Neural Computing and Applications, 31(17), 2022, 14823-14844. https://doi.org/10.1007/s00521-022-07337-8

[4] Kumari, D.; Ekbal, A.; Haque, R.; Bhattacharyya, P.; & Way, A.: Reinforced NMT for sentiment and content preservation in low-resource scenario, Transactions on Asian and Low-Resource Language Information Processing, 20(4), 2021, 1-27. https://doi.org/10.1145/3450970

[5] Jamal, A.; Deodhare, D.; Namboodiri, V.; & Venkatesh, K.-S.: Eclectic domain mixing for effective adaptation in action spaces, Multimedia Tools and Applications, 77(22), 2018, 29949-29969. https://doi.org/10.1007/s11042-018-6179-y

[6] Su, J.; Zeng, J.; Xie, J.; Wen, H.; Yin, Y.; & Liu, Y.: Exploring discriminative word-level domain contexts for multi-domain neural machine translation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(5), 2019, 1530-1545. https://doi.org/10.1109/TPAMI.2019. 2954406

[7] Zhao, L.; Gao, W.; & Fang, J.: High-performance English–Chinese machine translation based on GPU-enabled deep neural networks with domain corpus, Applied Sciences, 11(22), 2021, 10915. https://doi.org/10.3390/app112210915

[8] Chen, K.; Wang, R.; Utiyama, M.; Sumita, E.; & Zhao, T.: Neural machine translation with sentence-level topic context, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27(12), 2019, 1970-1984. https://doi.org/10.1109/TASLP.2019.2937190

[9] Dabre, R.; Chu, C.; & Kunchukuttan, A.: A survey of multilingual neural machine translation, ACM Computing Surveys (CSUR), 53(5), 2020, 1-38. https://doi.org/10.1145/3406095

[10] Wang, X.; Tu, Z.; & Zhang, M.: Incorporating statistical machine translation word knowledge into neural machine translation, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(12), 2018, 2255-2266. https://doi.org/10.1109/TASLP.2018. 2860287