Machine Translation System Based on Deep Learning

Ting Yang¹, Shinan Zhao¹, He Chen² and Bo Chen³*

¹ School of Foreign Language Studies, Shaanxi University of Chinese Medicine, Xianyang, Shaanxi, 712046, China
² School of Food and Biologil Engineer, Shaanxi University of Science & Technology, Xi’an, Shaanxi, 710021, China
³ School of Physics & Electronic Engineering, Xianyang Normal University, Xianyang, Shaanxi, 712000, China
*Corresponding author’s e-mail: xuxiaoxiaqing0601@dingtalk.com

Abstract. The translation fusion of multi-machine translation is a strategy that has been proposed very early to improve the quality of machine translation. It has been fully discussed under the framework of statistical machine translation, and further breakthroughs in performance have entered a bottleneck period. Since the fusion methods of existing methods are mostly limited to surface features and lack effective deep fusion strategies, this paper encodes source language sentences and multiple system translations separately to achieve fusion at the coding level. In the specific decoding, we limit the decoding space, so that the translation model can obtain a higher performance improvement on less data. When increasing the values of the two hyper-parameters \( \alpha \) and \( \beta \), the translation effect can be gradually improved. The performance of the model reaches the maximum when \( a \) and \( b \) are respectively set to 0.4, and the translation performance will decrease if it continues to increase.

1. Introduction

Deep neural network models have achieved great success in the research of machine translation, and are gradually introduced into the research of machine translation translation fusion[1,2]. However, the neural network model usually requires a lot of data for training, in order to exceed the classic fusion model in performance. In the task of actual machine translation translation fusion, affected by the scale of training data, direct use of small-scale data training usually does not achieve ideal performance. In order to solve the problems listed above, this chapter proposes a deep fusion method of multi-system translation based on neural network to optimize translation quality[3]. This method uses the codec model framework to encode the source language sentence and the machine translation, and then merges and decodes the new translation on this basis[4]. The advantage of this method is that not only the source language sentences are encoded, the information contained in it can better guide the generation of translations when fused, and the machine translations are uniformly coded without distinction[5]. It also effectively alleviates the dependence of the neural network fusion model training on the machine translation system that needs to collect a large number of machine translations, so as to improve the adaptability of the fusion model to the fusion translation. At the same time, we also limit the space for generating translations in the final fusion decoding to reduce the uncertainty of the generated translations.
2. Principle

We propose a multi-translation optimization deep fusion model based on neural network. This model uses the classic encoding-decoding model framework in deep learning, and its structure is shown in Figure 1.

In the encoding process, we use a source language sentence encoder and a machine translation encoder, which are used to encode the source language sentence and the machine translation into a sequence of vectors, respectively. Define the source language sentence as $X$ and the machine translation as $Z(i) \ (i \in \{1, \ldots, n\})$, Two encoders respectively encode them into corresponding vector sequences $H$ and $Q(i) \ (i \in \{1, \ldots, n\})$, and respectively denote as:

$$H = \text{Source Encoder}(X)$$
$$Q(i) = \text{arg max}_{\text{Encoder}(H, Z(i))} \ (i \in \{1, \ldots, n\})$$

We define the encoding $Q(i)$ of the machine translation depends on the encoding of the source language sentence. The rationality of this definition is that in general, the machine translation cannot perfectly express the meaning of the source language sentence, so it needs source language information to supplement. In the decoding stage, we use a fusion decoder to fuse the source language sentence encoding and machine translation encoding to generate a new translation. The fusion decoding process can be described as:

$$Y = \text{arg max} \ P(\hat{Y}, Q(1), Q(2), \ldots Q(n))$$

Where $H$ is the encoding of the source language sentence, and $Q(i)$ is the code of the $i$th machine translation.

Compared with the traditional neural machine translation framework based on the encoder and decoder, our method draws on the idea of teacher-forcing, allowing the input of the source language sentence and the previous word of the target sentence, and generating the next word based on the input.
The advantage is that it can not only eliminate the hidden layer to the hidden layer loop, but also decouple all time steps in the loss function based on the prediction of a certain moment and the training target at that moment, which is conducive to the parallel training of the model. More importantly, for fusion tasks, translation candidates can usually be obtained in advance, so this mechanism can be better consistent with the fusion process in the decoding process. Different from the classic fusion method based on voting, the deep fusion method of multi-system translation based on neural network proposed by us does not directly reassemble segments of machine translation, but encodes the source language sentence and the information of multiple machine translations. And fusion is carried out in the coded space, so the source language information can be used more effectively to influence the generation of the final fusion translation.

For a machine translation $Z$ with a length of $L$ machine translation systems, it can be expressed as

$$Z = (z_1, z_2, ..., z_L) \in \mathbb{R}^K$$

The corresponding source language sentence $X$ can be expressed as

$$X = (x_1, x_2, ..., x_L) \in \mathbb{R}^K$$

Then the sequence of the codeable vectors of $Z$ with respect to $X$ can be expressed as

$$Q = (q_1, q_2, ..., q_L)$$

We use the attention mechanism to solve the source language dependence problem, and use the conditional gated recurrent network unit to calculate each vector $q_j (j \in \{1, ..., L\})$ in the sequence $Q$, and the specific process is shown in Figure 2.

![Source-aware Encoder of Machine Translation Outputs](image)

**Figure 2. The Source-aware Encoder of Machine Translation Outputs.**

### 3. Experiment and result

Figure 3 shows a translation example, which comes from the test set of the NIST Chinese-English translation task, where "Source" represents the input source language, and "Reference" represents the corresponding correct target language research translation.
The experiment also observes the influence of different values of the hyperparameters on the experimental results by changing the a and b of the hyperparameters. We adopt the form of control variables to analyze, that is, when the value of one hyperparameter is changed, the value of the other hyperparameter is set to 0 to eliminate interference. It can be seen from Figure 4 that with the increase of the two hyperparameter values, the translation effect gradually becomes better. When $\alpha$ and $\beta$ are respectively set to 0.4, the performance of the model reaches the maximum. If it continues to increase, the translation performance will decrease instead. When $\alpha$ and $\beta$ are relatively small, the influence of syntactic structure on the model is very weak, so syntactic knowledge is also very limited in helping translation. As the value increases, the role played by the syntax also becomes larger. When the value reaches 0.4, the help to the translation model is greatest. However, if it is further increased, the supervision signal will be too strong, which will interfere with the modeling of other aspects of the sentence by the translation model, so the translation performance will begin to decline.

4. Conclusion
We used a recurrent neural network for encoding and decoding, which was the best performing neural machine translation framework at the time. The recently proposed self-attention has attracted more and
more attention due to its training efficiency and its excellent performance. Future research is necessary to further test the performance of the deep fusion system based on the self-focused network. In addition, how to design a more efficient fusion model based on the self-focused network is also a direction worthy of attention.

Acknowledgments
This work was supported in part by Shaanxi Provincial Key Research and Development Program 2021NY-213, Shaanxi Provincial Social Science Foundation Program 2020K014.

Social Science Foundation Program of Shaanxi Province

References
[1] Shen S, Cheng Y, He Z, et al. (2016) Minimum Risk Training for Neural Machine Translation. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Brilon, Germany. pp. 1683-1692.
[2] Zhang X, Lapata M, Wei F, et al. (2018) Neural Latent Extractive Document Summarization. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium. pp. 779-784.
[3] Mi H, Sankaran B, Wang Z, et al. (2016) Coverage Embedding Models for Neural Machine Translation. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, Texas. pp. 955-960.
[4] Sennrich R, Haddow B, Birch A. Neural Machine Translation of Rare Words with Subword Units. In: Proceedings of the the 54th Annual Meeting of the Association for Computational Linguistics. Brilon, Germany. pp. 86-96.
[5] Xu J, Durrett G. (2019) Neural Extractive Text Summarization with Syntactic Compression. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing. Hong Kong, China. pp. 3290-3301.