Neural Machine Translation with Diversity-Enabled Translation Memory

Quang Chieu Nguyen[1,2], Xuan Dung Doan[1], Van-Vinh Nguyen[2],
and Khac-Hoai Nam Bui[1](✉)

[1] Viettel Cyberspace Center, Viettel Group, Hanoi, Vietnam
{chieunq,dungdx4,nambkh}@viettel.com.vn
[2] Vietnam National University of Hanoi, Hanoi, Vietnam
vinhvn@vnu.edu.vn

Abstract. Neural machine translation (NMT) using translation memory (TM) has been introduced as an emergent technique for improving machine translation systems (MTS). In this study, we propose an end-to-end NMT model with TM by exploiting the diversity of the retrieval-augmented phase using maximal marginal relevance (MMR). In particular, the proposed model is designed with monolingual TM, which is able to support low-resource scenarios. Furthermore, the memory retriever and translation models are jointly trained to improve translation performance. For the experiment, we use IWSLT15 (En $\longleftrightarrow$ Vi) as a benchmark dataset to evaluate the performance of the proposed method. Accordingly, the experiential results show the effectiveness of the proposed method compared with strong baselines in this research field.

Keywords: Neural Machine Translation · Translation Memory · Maximal Marginal Relevance · Low Resource Language

1 Introduction

Translation Memory (TM) is conceptually regarded as a database of sentence pairs (source and target texts), which is utilized to reuse previously translated content when working on new texts. Recent works have focused on memory augmentation to improve the performance of neural machine translation (TM-augmented NMT) [12].

Technically, a typical TM-augmented NMT model performs the translation process in two phases, as shown in Fig. 1: i) *Retrieval Stage* extracts the candidate sentence memories from training corpus based on calculating similarity; and ii) *Generation Stage* integrates the candidate sentences into translation model for the translation. Subsequently, the trend research focuses on jointly learning models of two phases (retriever and translation models) with remarkable results [4].

N. T. Nguyen et al. (Eds.): ACIIDS 2023, LNAI 13995, pp. 322–333, 2023.
https://doi.org/10.1007/978-981-99-5834-4_26

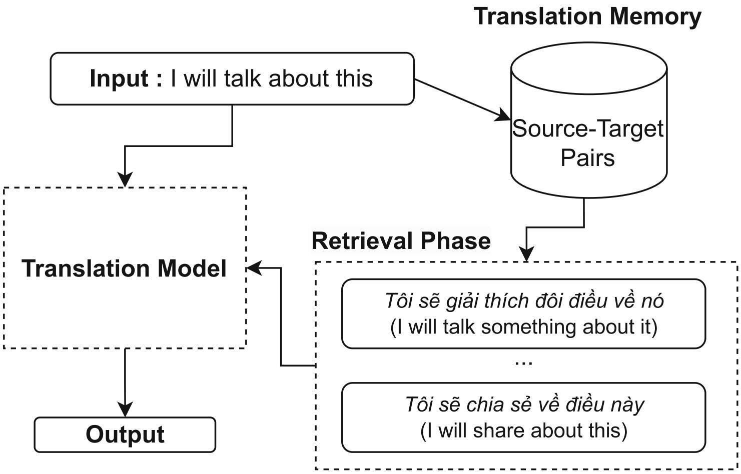
Translation Memory
Input : I will talk about this
Source-Target Pairs
Retrieval Phase
Translation Model
Tôi sẽ giải thích đôi điều về nó
(I will talk something about it)
...
Tôi sẽ chia sẻ về điều này
(I will share about this)
Output

Fig. 1. An example of the neural machine translation using translation memory.

Despite the success of the recent TM-augmented NMT models, there are still two remaining research issues that need take into account: i) The retrieval stage mainly uses a greedy method to extract the top-r nearest sentence pairs, which results in redundant information because the top-r sentence pairs are highly similar to each other [4]. ii) Most previous works use TM with sentence pairs (source-target pairs), which is not able to take advantage of abundant monolingual data [9]. In this regard, this study proposes a new method for TM-based NMT to deal with the aforementioned issues. Specifically, for the retrieval phase, we adopt Maximal Marginal Relevance (MMR) [7] to enable the diversity, guaranteeing the two most challenging properties of candidates: *informativeness* obtained by the distance between query and candidates; *diversity* expressed by the distances among candidates themselves. For the monolingual TM, following the work in [4], a simple dual-encoder framework is adopted for selecting the most relevant sentences. Generally, the main contributions of this study are two folds as follows:

- We present a novel end-to-end model for TM-augmented NMT, which aims to leverage two emergent issues of TM-augmented NMT such as balancing the relevance and diversity of the retrieval phase and using monolingual data. To the best of our knowledge, the proposed method is the first study that integrates two aforementioned issues in a unified framework.
- We execute our approach on IWSLT15, a benchmark English-Vietnamese dataset [8] to demonstrate the effectiveness of the model in low-resource language pair scenarios. Specifically, the reported results show that our model outperforms strong baseline models in this research field.

The rest of the paper is organized as: Sect. 2 presents a brief review of previous works regarding this study. The proposed method is presented in Sect. 3. We report and analyze the evaluated results in Sect. 4. Section 5 is the conclusion and discussion of this study.

2 Related Work

Recent work tries to jointly train the retrieval model and a translation model with monolingual TM and achieve impressive results [4]. The proposed method in this study is the orthogonality of recent works of this research line. Specifically, we present an end-to-end monolingual TM-augmented NMT model that includes a retrieval stage with a special focus on extracting both relevant and diverse sentences. In this section, we take a brief review of those aforementioned techniques for improving the performance of the TM-augmented NMT approach.

2.1 Neural Machine Translation for Low-Resource Languages

In recent years, Neural Machine Translation (NMT) [19] has emerged as a state-of-the-art approach to machine translation, gaining widespread popularity. Specifically, the Transformer architecture [25] has revolutionized the field of NMT by achieving success in multiple language pairs. However, supervised NMT requires large datasets, which are often limited in low-resource languages. To address this issue, several data augmentation techniques have emerged, including back-translation [23], and self-training [15]. Additionally, transfer learning techniques [20,29] show promise in leveraging pre-trained models for improved performance. In cases where parallel data is not available, unsupervised technique NMT [2], pivot-based [10] or multi-NMT-based solution [11] can be employed. Recent studies focus on using TM with monolingual data, as an emergent technique, to improve the translation quality of NMT.

2.2 Translation Memory-Augmented Neural Machine Translation

Augmenting TM has become an emerging research topic for improving NMT. There are primarily two approaches for incorporating translation memory (TM) into neural machine translation (NMT): constraining the decoding process with TM and using TM to train a more powerful NMT model.

The main idea of the first research line is to increase the generation probability of some target words based on TM. Zhang et al. [28] increased the generation probability of target words aligned with the TM. In [13] a bilingual dictionary is used as auxiliary information to tackle infrequent word translation. Khandelwal et al. [17] used kNN-MT to retrieve TMs from dense vectors by creating a key-value datastore and interpolating the generation probability of the NMT model with similar target distributions from the datastore at each time step.

The second research line aims to train the translation model to learn how to deal with the retrieved TMs. A data augmentation way was used by Pham et al. [22] to concatenate the retrieved TMs with input sentences during training. Several studies have explored modifying the architecture of the NMT model to improve integration with TM. Cao et al. [6] introduced a gating mechanism to control the signal from the retrieved TM, and following this, in [5] an additional transformer encoder is designed to incorporate the target sentence of the TM through attention. In Xia et al. [26] work, multiple retrieved TMs are compressed into a graph structure to enhance efficiency and space usage and then integrated into the model through attention.

2.3 Retrieval for Translation Memory-Augmented Neural Machine Translation

Previous works focus on the TM with bilingual sentence pairs [26,27], which used fuzzy matching to retrieve the most similar sentences from the corpus with a query. In TM with monolingual, retrieval task is more challenging due to the cross-lingual setting. To address this challenge, Cai et al. [4] proposed a simple dual-encoder framework pre-trained on two tasks: sentence-level cross-alignment and token-level cross-alignment.

Regarding diversity for the retrieval results, authors in [9] have proven that diverse translation memories are able to improve the performance of the NMT, making it important to ensure diversity in the retrieval stage. There are several methods to enable diversity, including MMR [7], IA-Select [1] or MaxSum Diversification [3]. We employed MMR in this study due to its straightforward implementation and, more importantly, its ease of interpretation.

3 Methodology

3.1 Overview System

Figure 2 depicts the overview structure of the proposed method. In particular, the main contribution of this study focuses on the retrieval stage, which selects the most relevant and diverse sentences from a large monolingual TM in the target language. Specifically, given an input sentence x in the source language and a large monolingual TM $M = \{m_1, m_2, .., m_n\}$, the output of the retrieval stage is a subset (top k) TM and relevance scores $\{f(x, m_i)\}_{i=1}^{k}$ Then, the translation model conditions on both the input x, the retrieved set, and their scores x to generate the output y.

3.2 Retrieval Model

3.2.1 Relevant Monolingual TM: The input sentence x (source sentence) and monolingual TM M of target language are encoded by using two independent Transformers [25], which are sequentially formulated as follows:

$$\begin{aligned} z_x &= W_1 Trans_x(< bos >, x^1, x^2, ..x^{|x|}) \\ z_{m_i} &= W_2 Trans_m(< bos >, m_i^1, m_i^2, ..m_i^{|m_i|}) \end{aligned} \tag{1}$$

where $m_i \in M$ denotes the memory target sentence. W_1 and W_2 are learning parameters. In this regard, the relevance score $f(x, m_i)$ between the source sentence and the candidate sentence can be calculated using the dot product:

$$f(x, m_i) = z_x^T z_{m_i} \tag{2}$$

Subsequently, the top r relevant sentences are extracted using Maximum Inner Product Search (MIPS).

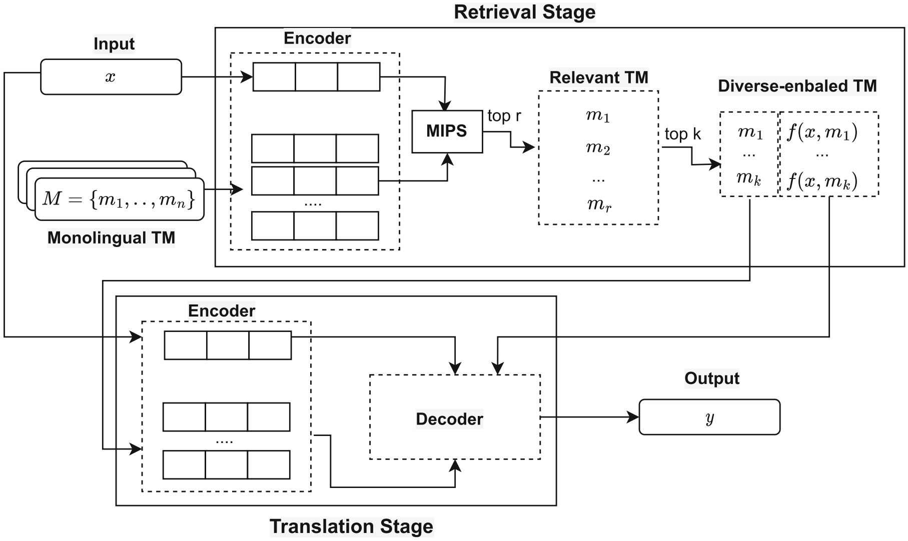

Fig. 2. Overall structure.

3.2.2 Diversity-Enabled TM: After obtaining $R = \{m_1, .., m_r\}$ as the relevant sentences, a subset of size k is selected from R is selected to increase the diversity by using MMR [7], the MMR function can be formulated as follows:

$$MMR(x, R, S) = \underset{m_i \in R \setminus S}{argmax}[\lambda.cosine(x, m_i) - (1 - \lambda).\underset{m_j \in S}{max}(cosine(m_i, m_j))] \quad (3)$$

where S is the current set of chosen candidates. $R \setminus S$ is a set of unselected sentences. The hyperparameter λ, which takes values in the range $[0, 1]$, is used to trade off accuracy and diversity. A high value of lambda corresponds to high accuracy, whereas a low value corresponds to high diversity.

Algorithm 1: Maximal Marginal Relevance (MMR)

```
input  : top-$r$ most relevant sentences $R = \{m_i\}_{i=1}^{r}$ and result size $k$
output: result set $S \subseteq R, |S| = k$
  $S \leftarrow \emptyset$
  while $|S| < k$ do
    $s_s = MMR(x, R, S)$
    $S \leftarrow S \cup s_s$
    $R = R \setminus s_s$
  end
```

The diverse-enabled TM processed can be described in the Algorithm 1. Specifically, the output of this process is a set of translation memories $S = \{m_1, .., m_k\}$ and its retrieval score $f(x, S) = \{f(x, m_1), .., f(x, m_k)\}$.

3.3 Translation Model

For the translation stage, we follow the work in [4] for the end-to-end model, which is built based on the standard encoder-decoder NMT model [25]. Specifically, given source sentence x, a set of retrieval TM $S = \{m_i\}_{i=1}^{k}$ and its scores $\{f(x, m_i)\}_{i=1}^{k}$ in the previous step, the objective of the translation model is to define the conditional probability as follows:

$$p(y|x, m_1, f(x, m_1), ..., m_k, f(x, m_k)) \tag{4}$$

To incorporate the information of TM contextualized token embeddings $\{z_{m_i,j}\}_{j=1}^{|m_i|}$ $(1 \le i \le k)$, the cross attention is calculated as follows:

$$\alpha_{ij} = \frac{exp(h_t^T W_3 z_{m_i,j} + \beta f(x, m_i))}{\sum_{i=1}^{i=k} \sum_{l=1}^{L_i} exp(h_t^T W_3 z_{m_i,l} + \beta f(x, m_i))} \tag{5}$$

Here, α_{ij} and L_i denote the attention score of the j-th token in z_{m_i} and the length of the sentence z_{m_i}, respectively. W_3 represents the learning parameter. h_t is the decoder's hidden state at time step t. The weighted sum of memory information can be updated as follows:

$$c_t = W_4 \sum_{i=1}^{k} \sum_{j=1}^{L_i} \alpha_{i,j} z_{i,j} \tag{6}$$

where W_4 denotes the learning parameter. Following this, h_t is updated with c_t, i.e., $h_t = h_t + c_t$. In this regard, the next-token probabilities can be computed as follows:

$$p(y_t|x, m_1, f(x, m_1), ..., m_k, f(x, m_k)) = (1 - \lambda_t) P_v(y_t) + \lambda_t \sum_{i=1}^{k} \sum_{j=1}^{L_i} \alpha_{i,j} \mathbb{1}_{z_{m_i,j} = y_t} \tag{7}$$

where $\lambda_t = g(h_t, c_t)$ denotes the feed-forward network, $\mathbb{1}$ is the indicator function, the next-token probabilities P_v are obtained by converting the hidden state h_t using a linear projection and then applying the softmax function, which can be formulated as follows:

$$P_v = softmax(W_v h_t + b_v) \tag{8}$$

4 Experiment

4.1 Experiment Setup

4.1.1 Dataset and Evaluation: We use the English-Vietnamese as the evaluated dataset of this study (publicly available in the MT track of the IWSLT 2015 corpus [8]). Specifically, this dataset comprises a collection of parallel sentences in spoken language domains. The detailed data statistic is illustrated in Table 1.

Table 1. Statistics of the evaluated dataset.

# Train Pairs	# Dev Pairs	# Test Pairs
133317	1553	1268

In all experiments, the target language in the training set is utilized as monolingual translation memory data M. Subsequently, different bilingual datasets are generated for later experiments by randomly selecting 60%, 80%, 100% of the training dataset, referred to as D60, D80, and D100 datasets, respectively. For evaluation, we use the BLEU score [21].

4.1.2 Baseline Models: We compare the proposed model with the following baselines:

- **NMT wo TM**: the original NMT model without TM [25].
- **NMT + TM-BM25**: source similarity search method based on BM25, which is used in many recent TM-augmented NMT models [14,26].
- **NMT + Monolingual TM**: The joint training retrieval and translation models by adopting a dual encoder architecture [4].

4.1.3 Implementation Details: Our model utilizes Transformer blocks with the same setup as Transformer Base [25], which includes 8 attention heads, a hidden state with 512 dimensions, and a feed-forward state with 2048 dimensions. We employ 3 Transformer blocks for the retrieval model, 4 blocks for the memory encoder in the translation model, and 6 blocks for the encoder-decoder architecture in the translation model. We set trade-off hyperparameter in MMR $\lambda = 0.5$. The FAISS [16] has been used for indexing the dense representations. The learning rate schedule, dropout, and label smoothing are set following the default settings in [25]. We use Adam optimizer [18] and train models with up to $30K$ steps throughout all experiments. The number of tokens in every batch is 4096. BPE [24] tokenizer is employed with a vocabulary size of 4000. In order to execute the BM25-based method, we used a BM25 search engine[1] to obtain a preliminary set of TM sentences.

4.2 Main Results

Table 2 shows the results of the evaluation on different sizes of the training dataset. Particularly, the reported results are conducted with 3 and 5 retrieval sentences for the TM, respectively. As reported results, we make the following observations: i) NMT by using greedy retrieval (e.g., BM25) does not outperform the original NMT model, which indicates that joint training is an important method for the MT-augmented NMT approach; ii) Our model, which focuses

[1] https://github.com/dorianbrown/rank_bm25.

Table 2. Report results (BLEU scores) with two different values of top k retrieval TM for English ⟶ Vietnamese. Bold texts are the best results of each column.

Dataset	Model	k = 3		k = 5	
		Dev	Test	Dev	Test
D60	NMT wo TM	24.31	26.14	24.31	26.14
	NMT+TM-BM25	23.8	25.1	24.03	25.39
	NMT+Monolingual TM	24.22	26.51	24.38	26.72
	Our Model	**24.45**	**27.3**	**24.59**	**26.87**
D80	NMT wo TM	25.61	28.14	25.61	28.14
	NMT+TM-BM25	25.65	27.5	25.38	28.21
	NMT+Monolingual TM	25.61	28.32	25.7	**28.69**
	Our Model	**25.8**	**28.71**	**25.84**	28.67
D100	NMT wo MT	26.53	29.56	26.53	29.56
	NMT+TM-BM25	26.31	28.99	26.56	29.37
	NMT+Monolingual TM	**26.7**	**30.03**	26.74	29.28
	Our Model	**26.7**	29.6	**26.88**	**29.84**

on improving the retrieval sentence in terms of enabling the diversity for TM, achieves the best results compared with strong baseline models. The reported results indicate that diverse TM is able to improve the performance of NMT, especially with the low-resource scenarios; iii) The results between the number k sentences (k = 3 and k = 5) of the retrieval models are not too different. However, in our opinion, selecting the number of k sentences should be regarded as a hyperparameter and tuned during the training process. We leave this issue as future work regarding this study.

Table 3. Report the BLEU scores obtained when comparing monolingual translation memory (TM) and bilingual TM for English ⟶ Vietnamese.

Dataset	Model	Dev	Test
D60	Cheng et al., 2022 [9]	24.22	26.48
	Our Model+Bilingual TM	24.4	27.17
	Our Model+Monolingual TM	**24.45**	**27.3**
D80	Cheng et al., 2022 [9]	25.53	28.39
	Our Model+Bilingual TM	25.78	28.54
	Our Model+Monolingual TM	**25.8**	**28.71**
D100	Cheng et al., 2022 [9]	26.02	29.34
	Our Model+Bilingual TM	26.62	29.1
	Our Model+Monolingual TM	**26.7**	**29.6**

Furthermore, we also try to evaluate the performance between monolingual TM and bilingual TM. Accordingly, we re-implement the most recent work [9] for bilingual TM and comparing with our method in terms of both monolingual and bilingual, respectively. Table 3 shows the results of the variant of our model and Cheng et al., [9] with $k = 3$. An interesting observation is that the performance of monolingual TM is slightly better than bilingual TM. The evaluated results indicate that taking advantage of abundant monolingual data is able to improve the performance of NMT tasks, especially for low-resource scenarios.

5 Conclusion

In this paper, we propose a new framework for TM-augmented NMT by enabling the diversity of monolingual TM. To the best of our knowledge, the proposed method is the first study of end-to-end TM-augmented NMT that takes both monolingual and diversity-enabled TM into account. Specifically, by adding a non-heuristic module using the MMR algorithm, our proposed framework is able to enable diversity for the retrieval stage. Furthermore, instead of utilizing bilingual sentence pairs for the retrieval stage, we adopt two transformer encoders to exploit the capability of abundant information by monolingual data. Experiments show the effectiveness of the proposed method. Specifically, with varying the number of training datasets, our method is able to increase the performance from 0.5 to 1 Bleu score compared with strong baseline models in this research field. Regarding the future work of this study, we plan to exploit the size of translation memory by integrating this hyperparameter into the learning process in order to improve the performance of TM-augmented NMT tasks.

References

1. Agrawal, R., Gollapudi, S., Halverson, A., Ieong, S.: Diversifying search results. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining, WSDM 2009, pp. 5–14. Association for Computing Machinery, New York (2009). https://doi.org/10.1145/1498759.1498766
2. Artetxe, M., Labaka, G., Agirre, E., Cho, K.: Unsupervised neural machine translation (2018)
3. Borodin, A., Lee, H.C., Ye, Y.: Max-sum diversification, monotone submodular functions and dynamic updates. In: Proceedings of the 31st ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems. PODS 2012, pp. 155–166. Association for Computing Machinery, New York (2012). https://doi.org/10.1145/2213556.2213580
4. Cai, D., Wang, Y., Li, H., Lam, W., Liu, L.: Neural machine translation with monolingual translation memory. In: Zong, C., Xia, F., Li, W., Navigli, R. (eds.) Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, 1–6 August 2021, pp. 7307–7318. Association for Computational Linguistics (2021). https://doi.org/10.18653/v1/2021.acl-long.567

5. Cao, Q., Kuang, S., Xiong, D.: Learning to reuse translations: guiding neural machine translation with examples. In: Giacomo, G.D., et al. (eds.) 24th European Conference on Artificial Intelligence, ECAI 2020, Santiago de Compostela, Spain, 29 August–8 September 2020, Including 10th Conference on Prestigious Applications of Artificial Intelligence, PAIS 2020. Frontiers in Artificial Intelligence and Applications, vol. 325, pp. 1982–1989. IOS Press (2020). https://doi.org/10.3233/FAIA200318
6. Cao, Q., Xiong, D.: Encoding gated translation memory into neural machine translation. In: Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J. (eds.) Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018, pp. 3042–3047. Association for Computational Linguistics (2018). https://doi.org/10.18653/v1/d18-1340
7. Carbonell, J., Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 1998, pp. 335–336. Association for Computing Machinery, New York (1998). https://doi.org/10.1145/290941.291025
8. Cettolo, M., Niehues, J., Stüker, S., Bentivogli, L., Cattoni, R., Federico, M.: The IWSLT 2015 evaluation campaign. In: Proceedings of the 12th International Workshop on Spoken Language Translation: Evaluation Campaign, Da Nang, Vietnam, 3–4 December 2015, pp. 2–14 (2015). https://aclanthology.org/2015.iwslt-evaluation.1
9. Cheng, X., Gao, S., Liu, L., Zhao, D., Yan, R.: Neural machine translation with contrastive translation memories. In: Goldberg, Y., Kozareva, Z., Zhang, Y. (eds.) Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022, Abu Dhabi, United Arab Emirates, 7–11 December 2022, pp. 3591–3601. Association for Computational Linguistics (2022). https://aclanthology.org/2022.emnlp-main.235
10. Cheng, Y., Yang, Q., Liu, Y., Sun, M., Xu, W.: Joint training for pivot-based neural machine translation. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp. 3974–3980 (2017). https://doi.org/10.24963/ijcai.2017/555
11. Fan, A., et al.: Beyond English-centric multilingual machine translation. J. Mach. Learn. Res. **22**, 1–48 (2020)
12. Feng, Y., Zhang, S., Zhang, A., Wang, D., Abel, A.: Memory-augmented neural machine translation. In: Palmer, M., Hwa, R., Riedel, S. (eds.) Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, 9–11 September 2017, pp. 1390–1399. Association for Computational Linguistics (2017). https://doi.org/10.18653/v1/d17-1146
13. Feng, Y., Zhang, S., Zhang, A., Wang, D., Abel, A.: Memory-augmented neural machine translation. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, September 2017, pp. 1390–1399. Association for Computational Linguistics (2017). https://doi.org/10.18653/v1/D17-1146

14. Gu, J., Wang, Y., Cho, K., Li, V.: Search engine guided neural machine translation. In: McIlraith, S.A., Weinberger, K.Q. (eds.) Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI 2018, The 30th Innovative Applications of Artificial Intelligence, IAAI 2018, and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2018, New Orleans, Louisiana, USA, 2–7 February 2018, pp. 5133–5140. AAAI Press (2018). https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17282
15. He, J., Gu, J., Shen, J., Ranzato, M.: Revisiting self-training for neural sequence generation (2020)
16. Johnson, J., Douze, M., Jégou, H.: Billion-scale similarity search with GPUs. IEEE Trans. Big Data **7**(3), 535–547 (2021). https://doi.org/10.1109/TBDATA.2019.2921572
17. Khandelwal, U., Fan, A., Jurafsky, D., Zettlemoyer, L., Lewis, M.: Nearest neighbor machine translation. CoRR abs/2010.00710 (2020). https://arxiv.org/abs/2010.00710
18. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015, Conference Track Proceedings (2015). http://arxiv.org/abs/1412.6980
19. Luong, M.T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation (2015)
20. Neubig, G., Hu, J.: Rapid adaptation of neural machine translation to new languages. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium October–November 2018, pp. 875–880. Association for Computational Linguistics (2018). https://doi.org/10.18653/v1/D18-1103. https://aclanthology.org/D18-1103
21. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a method for automatic evaluation of machine translation. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 311–318. Association for Computational Linguistics (2002)
22. Pham, M.Q., Xu, J., Crego, J., Yvon, F., Senellart, J.: Priming neural machine translation. In: Proceedings of the Fifth Conference on Machine Translation, November 2020, Online, pp. 516–527. Association for Computational Linguistics (2020). https://aclanthology.org/2020.wmt-1.63
23. Sennrich, R., Haddow, B., Birch, A.: Improving neural machine translation models with monolingual data (2016)
24. Sennrich, R., Haddow, B., Birch, A.: Neural machine translation of rare words with subword units. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Berlin, Germany, August 2016, pp. 1715–1725. Association for Computational Linguistics (2016). https://doi.org/10.18653/v1/P16-1162. https://aclanthology.org/P16-1162
25. Vaswani, A., et al.: Attention is all you need. CoRR abs/1706.03762 (2017). http://arxiv.org/abs/1706.03762
26. Xia, M., Huang, G., Liu, L., Shi, S.: Graph based translation memory for neural machine translation. In: The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, 27 January–1 February 2019, pp. 7297–7304. AAAI Press (2019). https://doi.org/10.1609/aaai.v33i01.33017297

27. Xu, J., Crego, J.M., Senellart, J.: Boosting neural machine translation with similar translations. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R. (eds.) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, 5–10 July 2020, pp. 1580–1590. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.acl-main.144
28. Zhang, J., Utiyama, M., Sumita, E., Neubig, G., Nakamura, S.: Guiding neural machine translation with retrieved translation pieces (2018). https://doi.org/10.48550/ARXIV.1804.02559. https://arxiv.org/abs/1804.02559
29. Zoph, B., Yuret, D., May, J., Knight, K.: Transfer learning for low-resource neural machine translation (2016)