Research Statement: Toward Intelligent Machine Translation

Rui Wang
National Institute of Information and Communications Technology (NICT)
Kyoto, Japan
wangrui@nict.go.jp

1 Overview

Natural language, or human language, is understood as a cultural specific communication system in informal usage. The language barrier of human beings has led to a cultural barrier to human communication. To overcome this barrier, a translator who understands two or more languages can translate one language into another one. The monolingual texts in the world are countless; however, the translators, especially the professional translators, are very few in comparison. Therefore, machine, instead of a human, is used for translation.

The history of machine translation (MT) can date back to 1940s, which is almost immediately after the first computer ENIAC. In the beginning, MT played a role as the translation memory tools for translators. One of the basic ideas behind MT is that a foreign language is considered as encrypted English (Weaver, 1947). So *noisy-channel model* was used for decoding foreign language into English. From the 1980s, more and more bilingual corpora have become available, and data-Driven methods have been proposed. Example-based MT systems have been built especially in Japan since the 1980s. In the 1990s, the emergence of statistical machine translation (SMT) was groundbreaking, especially IBM proposed a series of statistic models for SMT. Phrase-based SMT was widely considered as the state-of-the-art system before 2010s (Koehn, 2009). From 2000s, continuous-space methods, especially Neural network (NN) based methods become popular, as the upgrading of computer performance (Bengio et al., 2003). NN methods have been used to improve translation model, language model or directly integrated into end-to-end neural machine translation (NMT) systems (Bahdanau et al., 2015).

MT becomes a classic topic of artificial intelligence. The key point of MT is to transfer the linguistic knowledge from human to machine. The knowledge can be learned from the human-annotated data or human-designed model. My claim of the trend of MT is that MT is becoming more and more intelligent. That is, less knowledge is provide by human and MT has more independent ability. From this point, I think that three aspects are the research trends of MT.

Linguistic-motivated modeling and learning: From the model aspect, a common idea is that human designs a machine-learning based model and let the machine to learn its parameter. In this way, machine has little chance to learn the linguistic knowledge. I have introduced several linguistic-motivated mechanisms into statistical (Section 2) and neural (part of Section 3) machine translation. In this way, machine can obtain the ability to model and learn linguistic knowledge.

Intelligent data processing: From the data aspect, parallel data (annotated translation) is basically necessary for MT. Traditional MT just simply makes use of the data and ignores exploring the knowledge implied in the data. I have proposed several methods to advance the MT systems by making use of the knowledge implied in the data (part of Section 3 and 4).

Unsupervised machine translation: Despite the success of supervised NMT, it strongly relies on the availability of large amounts of parallel corpora, which hinders its applicability to language pairs with only large monolingual corpora in each language. I have proposed a universal unsupervised approach which train the translation model without using any parallel data. Compared with the existing unsupervised NMT (UNMT) methods, which has only been applied to similar or rich-resource language pairs, our method can be adapted to universal scenarios (Section 4).

2 Research in the Ph.D. Period (SJTU, 2012-2016)

In early 2010s, neural network based methods, which provide a way to represent language into a continuous-space, become popular. During my Ph.D. study, my researchs focus on the continuous-space language representation for SMT.

Neural Network based Language Modeling, 2013-2014

Language model (LM), which is a probability distribution over sequences of words, is a crucial component in phrase-based SMT. Before 2014, PBSMT is still the state-of-the-art machine translation system. Neural network LM (NNLM), which is also called continuous-space LM (CSLM) (Schwenk et al., 2006) in Figure 1, outperforms back-off n-gram LM (BNLM) significantly in perplexity. However, due to the non-linear neural network structure, NNLM is so time-consuming that can be hardly integrated into SMT efficiently. To maintain the performance of CSLM as well as the efficiency of BNLM, we proposed a method to converting NNLM into BNLM. In this way, the converted CSLM can predict the probability of words as accurate as CSLM and can be integrated into PBSMT the same as BNLM. These works have been published in EMNLP-2013 (Wang et al., 2013) and TALLIP-2016 (Wang et al., 2016a).

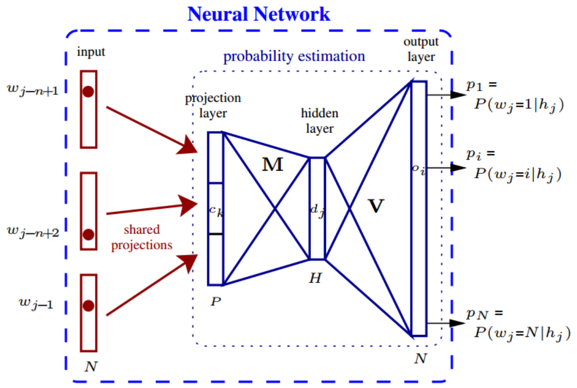

Figure 1: Monolingual NNLM (CSLM).

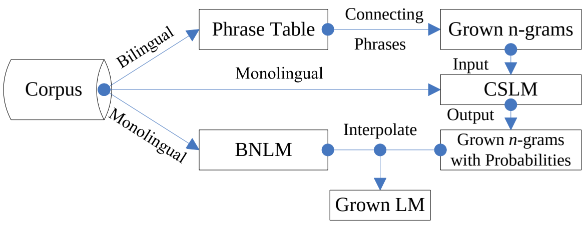

Figure 2: Bilingual NNLM (Wang et al., 2014).

In addition, to introduce the bilingual knowledge into NNLM, as shown in Figure 2, we adopted the connecting phrases from bilingual phrase table as a rule to generate the potential monolingual n-grams. The probabilities of these n-grams were calculated by CSLM. These generated n-grams, together with their probabilities were added to original LM as a new grown LM in EMNLP-2014 (Wang et al., 2014) and TASLP-2015 (Wang et al., 2015).

Graph-based Bilingual Word Embedding, 2015-2016

Continuous representations of words onto multi-dimensional vectors enhance natural language processing, especially SMT, by measuring similarities of words using distances of corresponding vectors. Most of the bilingual or monolingual word embedding methods at that time are the neural network based. However, most of these methods suffer from two obvious drawbacks. First, they only focus on simple contexts such as an entire document or a fixed sized sliding window to build word embedding and ignore latent useful information from the selected context. Second, the word sense but not the word should be the minimal semantic unit; however, most existing methods still use word representation. As we know, sense gives more exact meaning formulization than word itself. Motivated by this, we propose bilingual contexonym cliques as shown in Figure 3, which are extracted from bilingual point-wise mutual information based word co-occurrence graph. BCC plays a role of a minimal unit for bilingual sense representation. Correspondence analysis or neural network methods are used for summarizing BCC-word matrix into lower dimensions vectors for word representation. The whole pipeline is shown. The proposed bilingual word embedding method is applied to SMT to help predict word translation and generate more translation candidates in IJCAI-2016 (Wang et al., 2016b) and TALLIP-2018 (Wang et al., 2018c).

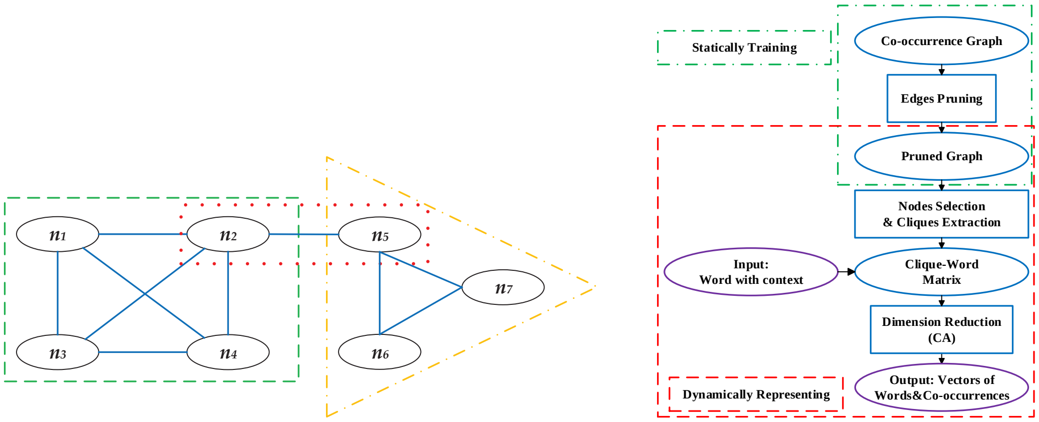

Figure 3: Bilingual contexonym cliques (left) (Wang et al., 2016b) and pipeline for graph-based bilingual word embedding (right) (Wang et al., 2018c).

3 Research in PostDoc (NICT, 2016-2019) and Tenure-track Period (NICT, 2019-now)

From 2015, NMT has set new state-of-the-art benchmarks (Bahdanau et al., 2015). Unlike PBSMT, which is a linear combination of several independent models, NMT jointly learns all of the parameters simultaneously. Therefore, I primarily studied NMT from data, modeling, and decoding aspects.

3.1 Data Distribution

Training and testing data mismatching. In a real-world NMT scenario, the domain distributions between training and testing data are sometimes different, because the testing data primarily focus on one domain; however, the training data are mostly collected from several domains described in my COLING-2018 survey paper (Chu and Wang, 2018). We proposed two methods to solve this problem: 1) We adopt the sentence embedding similarity as the criteria to select pseudo in-domain data in ACL-2017 (Wang et al., 2017a). 2) We assign a higher weight to the sentences from low-resource domains, in order to balance the domain distribution in the training data in EMNLP-2017 (Wang et al., 2017b). The proposed method can also be used to solve the multi-domain translation problem where the target domain is unknown in TASLP-2018 (Wang et al., 2018a). In addition, we propose an efficient method to dynamically sample the sentences in order to accelerate the NMT training in ACL-2018 (Wang et al., 2018b). The flowchart of these methods is illustrated in Figure 5.

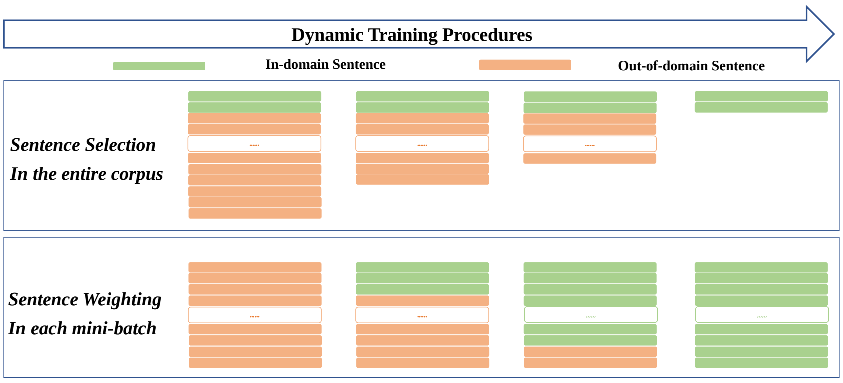

Figure 4: Domain adaptation for NMT (Wang et al., 2018a).

Exploring the knowledge implied in the data. For typical sequence prediction problems such as MT, maximum likelihood estimation (MLE) has commonly been adopted as it encourages the predicted sequence most consistent with the ground-truth sequence to have the highest probability of occurring.

However, MLE focuses on once-to-all matching between the predicted sequence and gold-standard, consequently treating all incorrect predictions as being equally incorrect. To counteract this, we augment the MLE loss by introducing an extra Kullback–Leibler divergence term derived by comparing a data-dependent Gaussian prior and the detailed training prediction in ICLR-2020 (Li et al., 2020) (This paper is finished by intern Zuchao Li and **it is one of the full review score paper among entire 3000+ submissions**). The proposed data-dependent Gaussian prior objective (D2GPo) is defined over a prior topological order of tokens and is poles apart from the data-independent Gaussian prior (L2 regularization) commonly adopted in smoothing the training of MLE.

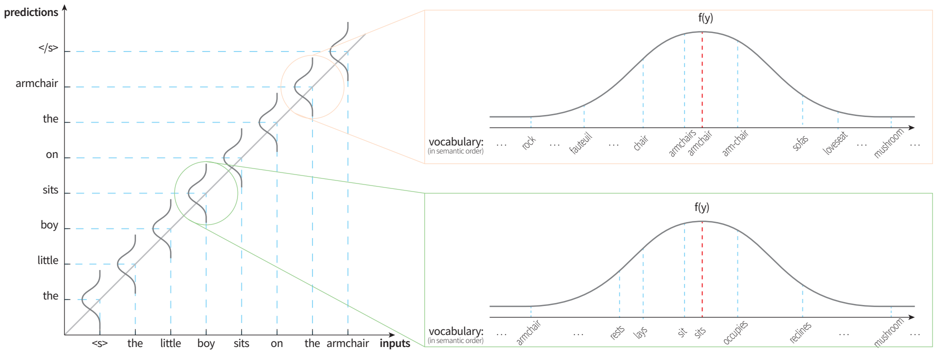

Figure 5: Data-dependent Training Objective for NMT (Li et al., 2020).

3.2 Syntax-directed Context Modeling

One crucial problem in NMT is how to represent the contextual information of a given source word. Motivated by our previous work on bilingual syntax-based context representation for PBSMT in TASLP (Chen et al., 2018), we proposed a source context representation by using the source dependency information in EMNLP-2017 (Chen et al., 2017a) which is illustrated in Figure 6 (left). Using this method, NMT performance was improved, especially the long sentences. Besides word-level contextual information, we also investigate the larger granularity-based context representation. For local attention, we introduce the syntax information into the attention mechanism in AAAI-2018 (Chen et al., 2017b) as illustrated in Figure 6 (right). Using this approach, we can achieve a better context by a better local attention construction. More recently, we also introduce a linguistic feature, i.e., content word information into NMT (Chen et al., 2020). (Kehai Chen is an intern supervised by me when conducting these works.)

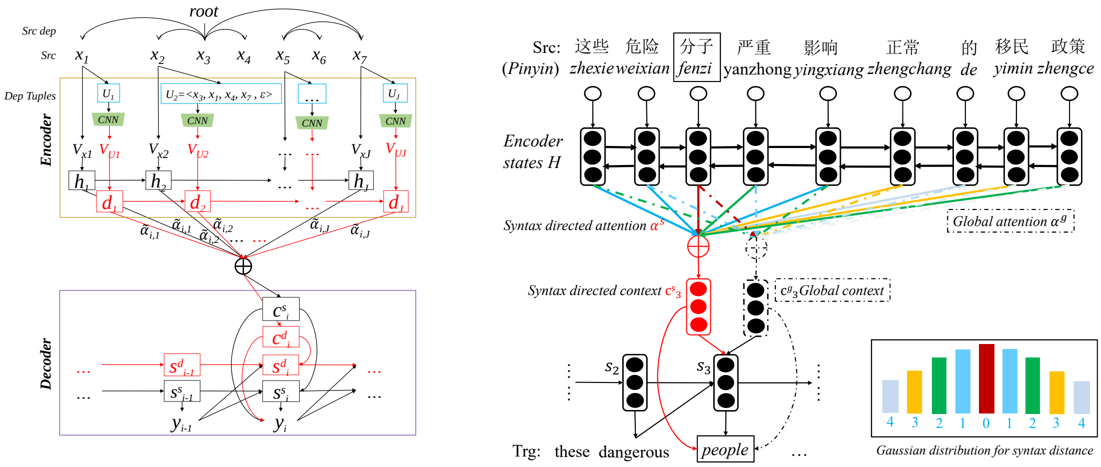

Figure 6: Source dependency (Chen et al., 2017a) (left) and Syntax-directed attention for NMT (Chen et al., 2017b) (right).

3.3 Decoding and Searching

In NMT, the decoder can capture the features of the entire prediction history through hidden layers. This enhances the translation model significantly; however, it restricts the explorable space in decoding. To enlarge the reachable space beyond that of a typical beam search and to search more efficiently, we revisit the idea of recombination from PBSMT and adapt it as an n-gram suffix based merger in the NMT beam search decoder as illustrated in Figure 7 in EMNLP-2018 (Zhang et al., 2018). In addition, we propose a sentence-level agreement module to directly minimize the difference between the representation of source and target sentence. In addition, we propose a training objective function and can also be used to enhance the representation of the source sentences in ACL-2019 (Yang et al., 2019). (Zhang and Yang are interns supervised by me when conducting these works.)

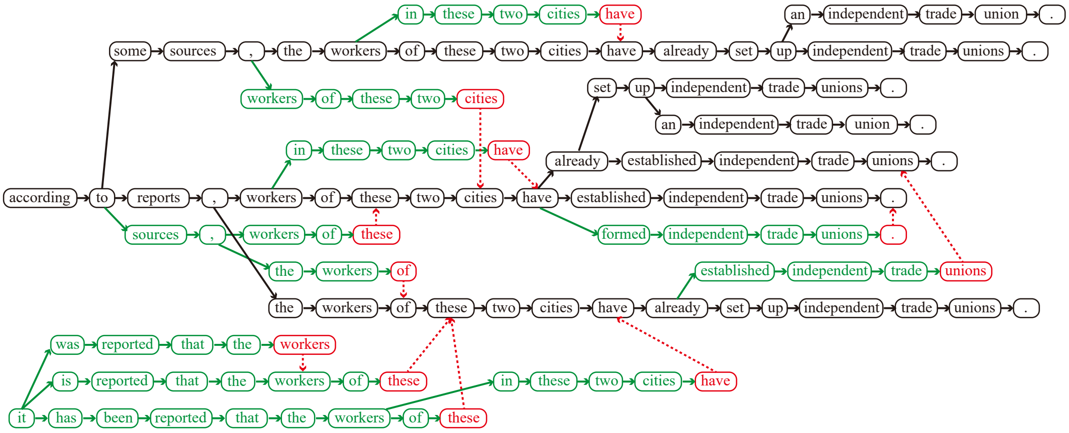
in these two cities have
an independent trade union .
some sources , the workers of these two cities have already set up independent trade unions .
workers of these two cities
set up independent trade unions .
an independent trade union .
in these two cities have
already established independent trade unions .
according to reports , workers of these two cities have established independent trade unions .
sources , workers of these
formed independent trade unions .
the workers of
established independent trade unions
the workers of these two cities have already set up independent trade unions .
was reported that the workers
is reported that the workers of these
in these two cities have
it has been reported that the workers of these

Figure 7: Illustration of state merging for NMT (Zhang et al., 2018).

4 Long-Term Research Goal: Intelligent Machine Translation

4.1 Unsupervised Learning

In spite of the recent success of NMT in standard benchmarks, the lack of large parallel corpora poses a major practical problem for many language pairs. There have been several proposals to alleviate this issue with, for instance, triangulation and semi-supervised learning techniques, but they still require a strong cross-lingual signal. Recently, some preliminary unsupervised methods are proposed for NMT (Artetxe et al., 2017; Lample et al., 2017). They adopt an unsupervised bilingual word embedding (UBWE) or cross-lingual masked language model (CMLM) as the pre-training word embedding. A shared encoder works together with a source language reconstruction (de-noising) decoder and a target language back-translation decoder. Results indicate that the proposed unsupervised neural machine translation (UNMT) can achieve a reasonable MT performance.

However, UBWE/CMLM training and UNMT training are independent, which makes it difficult to assess how the quality of UBWE/CMLM affects the performance of UNMT during UNMT training. We empirically explored relationships between UNMT and UBWE/CMLM. The empirical results demonstrate that the performance of UBWE and CMLM has a significant influence on the performance of UNMT. Motivated by this, we proposed a novel UNMT structure with cross-lingual language representation agreement to capture the interaction between UBWE/CMLM and UNMT during UNMT training in ACL-2019 (Sun et al., 2019) and TASLP-2020 (Sun et al., 2020a).

In the future, I will propose a universal unsupervised approach which train the translation model without using any parallel data. Compared with the existing unsupervised NMT (UNMT) methods, which has only been applied to similar or rich-resource language pairs, our method can be adapted to universal scenarios. Our main goals are: 1) Narrow the gap of performances between UNMT and SNMT in rich-resource languages. 2) Let UNMT achieve appropriate performances in low-resource and zero-resource language

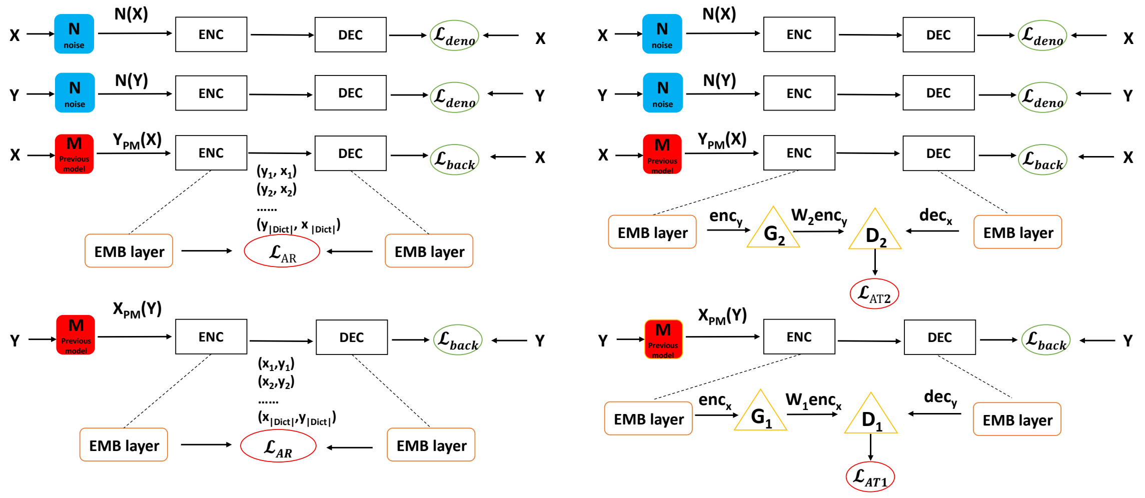

Figure 8: (a) Architecture of UNMT with UBWE Agreement Regularization (Sun et al., 2019); (b) Architecture of UNMT with UBWE Adversarial Training TASLP-2020 (Sun et al., 2020a).

pairs, where SNMT is almost impossible. 3) Adapt UNMT in real translation scenario, where the domain is difficult to be predicted and is often mismatched with the training corpus. 4) Multi-lingual UNMT: to translate several languages using a single model. Partial work has been accepted by ACL-2020 (Sun et al., 2020b). Haipeng Sun finished these works during his internship in NICT.

4.2 Intelligent Ordering

The two key components of MT are: word translation and reordering. NMT is good at solving the problem of translation candidate selection. Meanwhile, the reordering problem is quite difficult for NMT. The reason is that there are order differences between most of the languages. In phrase-based SMT, which is a combination of various features, there is a statistical reordering table learned from parallel data. In NMT, which is an entire end-to-end structure, there is no explicit mechanism to help the reordering.

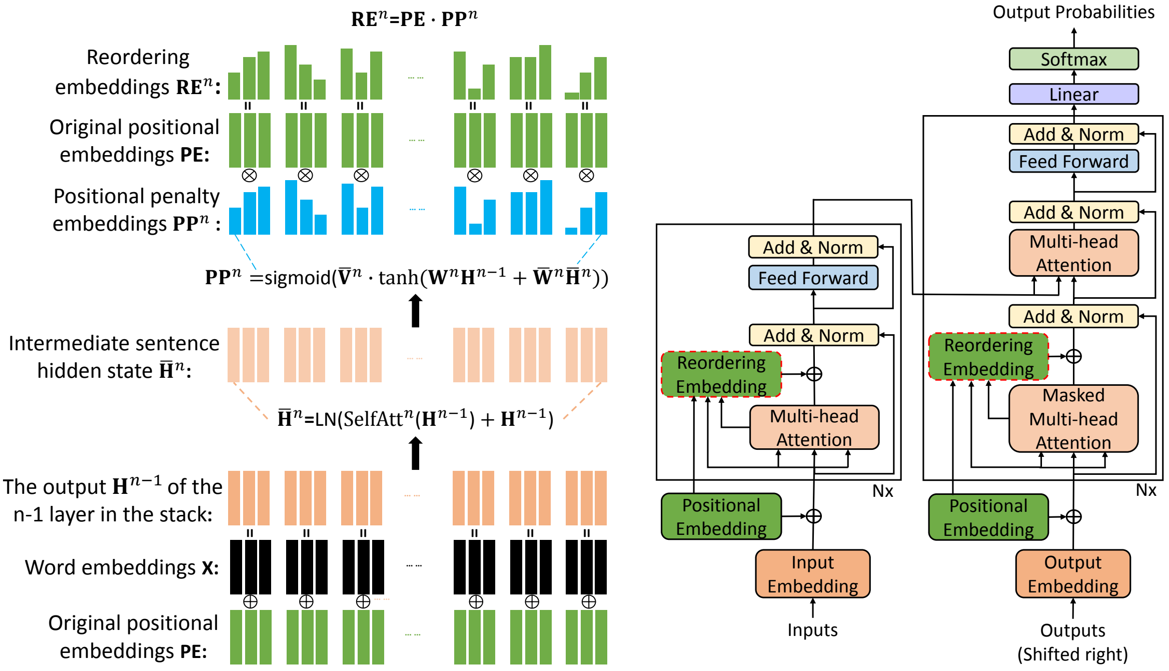
$\mathbf{RE}^n$=$\mathbf{PE} \cdot \mathbf{PP}^n$
Reordering embeddings $\mathbf{RE}^n$:
Original positional embeddings $\mathbf{PE}$:
Positional penalty embeddings $\mathbf{PP}^n$:
$\mathbf{PP}^n$ =sigmoid($\bar{\mathbf{V}}^n \cdot \tanh(\mathbf{W}^n\mathbf{H}^{n-1} + \bar{\mathbf{W}}^n\bar{\mathbf{H}}^n)$)
Intermediate sentence hidden state $\bar{\mathbf{H}}^n$:
$\bar{\mathbf{H}}^n$=LN(SelfAtt$^n(\mathbf{H}^{n-1}) + \mathbf{H}^{n-1}$)
The output $\mathbf{H}^{n-1}$ of the n-1 layer in the stack:
Word embeddings $\mathbf{X}$:
Original positional embeddings $\mathbf{PE}$:
Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Multi-head Attention
Masked Multi-head Attention
Reordering Embedding
Positional Embedding
Input Embedding
Inputs
Output Embedding
Outputs (Shifted right)
Nx

Figure 9: Learning reordering embeddings (left) and NMT with reordering mechanism (right) (Chen et al., 2019).

In Transformer-based NMT, the positional encoding mechanism helps the self-attention networks to learn the source representation with order dependency, which makes the Transformer-based NMT achieve state-of-the-art results for various translation tasks. However, Transformer-based NMT only adds representations of positions sequentially to word vectors in the input sentence and does not consider reordering information in this sentence. We propose a novel reordering method to model this reordering information for the Transformer-based NMT (Chen et al., 2019).

In the future, we will extent the latent ordering information into explicit order modeling. In practice, we will design a reordering model to learn the explicit order information. This will directly reorder the original source input to a reorder sequence whose order is similar to the target language.

4.3 Multi-signal Machine Translation

Theoretically, machine translation only considers text information. However, in human translation, human also take the speech, image, etc, into consideration. Motivated by this, multi-signal or multi-domain based neural machine translation is a way to simulate the human intelligence.

Though visual information has been introduced for enhancing neural machine translation (NMT), its effectiveness strongly relies on the availability of large amounts of bilingual parallel sentence pairs with manual image annotations. We present a universal visual representation learned over the monolingual corpora with image annotations, which overcomes the lack of large-scale bilingual sentence-image pairs, thereby extending image applicability in NMT. In detail, a group of images with similar topics to the source sentence will be retrieved from a light topic-image lookup table learned over the existing sentence-image pairs, and then is encoded as image representations by a pre-trained ResNet. An attention layer with a gated weighting is to fuse the visual information and text information as input to the decoder for predicting target translations. This work has been finished by an intern in ICLR-2020 (Zhang et al., 2020).

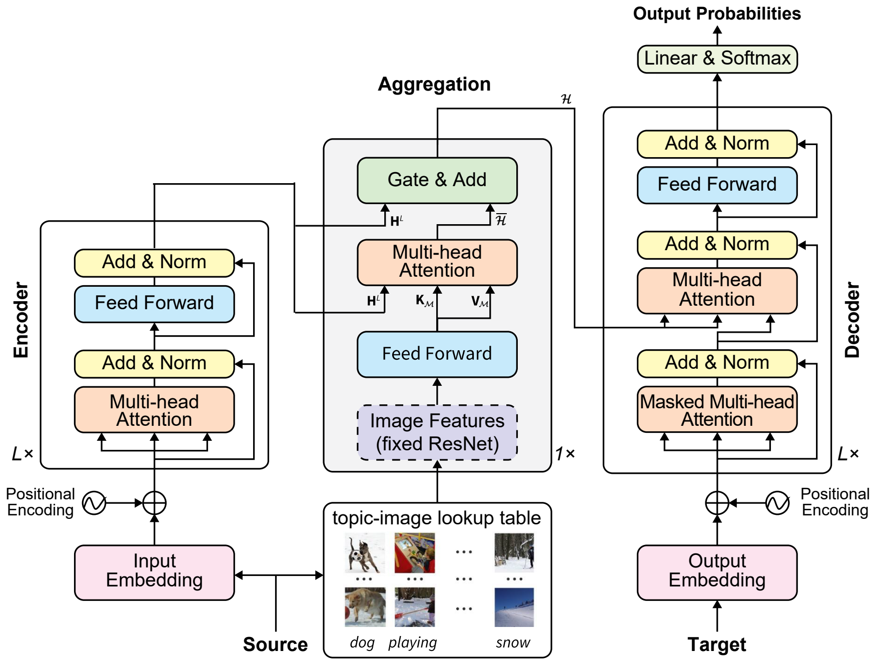

Figure 10: Overview of the framework of our proposed method.

In the future, we will implore more signals, such as speech, image, video, etc, into NMT. In addition, we will also deeply analyze the role of image in NMT.

4.4 Competition and Industry

Due to the importance of MT, it does not only attract the attention from academic researchers, but also the industry. Especially after NMT setup a series of state-of-the-art benchmarks in several translation shared-tasks (Luong and Manning, 2015; Bojar et al., 2017), several translation software has become popular in people's daily lives. Much famous translation software and products are widely used around the world, including Google and Bing of USA, VoiceTra of Japan, and iFlytek of China. Therefore, the machine translation study does not only have an effect on academic research but also directly help improve the products in the industry.

My researches, such as multi-domain NMT and reordering for NMT, have obtained the patent authorization. These method is applied to the NICT product VoiceTra, which is the most popular translation machine in Japan and the official machine translation system for Tokyo Olympic-2020. In addition, I have attended several top-tier machine translation shared task and achieve several first places.

- CONLL-2019 (Corresponding author): 1st in the DM sub-task and the 2nd overall.
- WMT-2019 (Corresponding author): 1st in the only unsupervised MT task (German–Czech) by BLEU and human evaluation
- WMT-2018 (Second author): 1st places in four tasks (English–Estonian and English–Finnish).
- WAT-2018 (First author): 1st places in Myanmar (Burmese) – English by BLEU.

References

Mikel Artetxe, Gorka Labaka, Eneko Agirre, and Kyunghyun Cho. 2017. Unsupervised neural machine translation. *CoRR*, abs/1710.11041.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *International Conference on Learning Representations*, San Diego.

Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A neural probabilistic language model. *Journal of Machine Learning Research*, 3:1137–1155, March.

Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Shujian Huang, Matthias Huck, Philipp Koehn, Qun Liu, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Raphael Rubino, Lucia Specia, and Marco Turchi. 2017. Findings of the 2017 conference on machine translation (WMT17). In *Proceedings of the Second Conference on Machine Translation*, pages 169–214, Copenhagen, Denmark, September. Association for Computational Linguistics.

Kehai Chen, Rui Wang, Masao Utiyama, Lemao Liu, Akihiro Tamura, Eiichiro Sumita, and Tiejun Zhao. 2017a. Neural machine translation with source dependency representation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2846–2852, Copenhagen, Denmark, September. Association for Computational Linguistics.

Kehai Chen, Rui Wang, Masao Utiyama, Eiichiro Sumita, and Tiejun Zhao. 2017b. Syntax-directed attention for neural machine translation. In *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*, New Orleans, USA.

K. Chen, T. Zhao, M. Yang, L. Liu, A. Tamura, R. Wang, M. Utiyama, and E. Sumita. 2018. A neural approach to source dependence based context model for statistical machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(2):266–280, Feb.

Kehai Chen, Rui Wang, Masao Utiyama, and Eiichiro Sumita. 2019. Neural machine translation with reordering embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1787–1799, Florence, Italy, July. Association for Computational Linguistics.

Kehai Chen, Rui Wang, Masao Utiyama, and Eiichiro Sumita. 2020. Content word aware neural machine translation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July. Association for Computational Linguistics.

Chenhui Chu and Rui Wang. 2018. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1304–1319, Santa Fe, New Mexico, USA, August. Association for Computational Linguistics.

Philipp Koehn. 2009. *Statistical machine translation.* Cambridge University Press.

Guillaume Lample, Ludovic Denoyer, and Marc'Aurelio Ranzato. 2017. Unsupervised machine translation using monolingual corpora only. *CoRR*, abs/1711.00043.

Zuchao Li, Rui Wang, Kehai Chen, Masso Utiyama, Eiichiro Sumita, Zhuosheng Zhang, and Hai Zhao. 2020. Data-dependent gaussian prior objective for language generation. In *International Conference on Learning Representations*.

Minh-Thang Luong and Christopher D Manning. 2015. Stanford neural machine translation systems for spoken language domains. In *Proceedings of the International Workshop on Spoken Language Translation*, pages 76–79, Da Nang, Vietnam.

Holger Schwenk, Daniel Dchelotte, and Jean-Luc Gauvain. 2006. Continuous space language models for statistical machine translation. In *Proceedings of the COLING/ACL on Main conference poster sessions*, COLING-ACL '06, pages 723–730, Sydney, Australia, July. Association for Computational Linguistics.

Haipeng Sun, Rui Wang, Kehai Chen, Masao Utiyama, Eiichiro Sumita, and Tiejun Zhao. 2019. Unsupervised bilingual word embedding agreement for unsupervised neural machine translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1235–1245, Florence, Italy, July. Association for Computational Linguistics.

H. Sun, R. Wang, K. Chen, M. Utiyama, E. Sumita, and T. Zhao. 2020a. Unsupervised neural machine translation with cross-lingual language representation agreement. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, pages 1–1.

Haipeng Sun, Rui Wang, Kehai Chen, Masao Utiyama, Eiichiro Sumita, and Tiejun Zhao. 2020b. Knowledge distillation for multilingual unsupervised neural machine translation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July. Association for Computational Linguistics.

Rui Wang, Masao Utiyama, Isao Goto, Eiichro Sumita, Hai Zhao, and Bao-Liang Lu. 2013. Converting continuous-space language models into n-gram language models for statistical machine translation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 845–850, Seattle, Washington, USA, October. Association for Computational Linguistics.

Rui Wang, Hai Zhao, Bao-Liang Lu, Masao Utiyama, and Eiichiro Sumita. 2014. Neural network based bilingual language model growing for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 189–195, Doha, Qatar.

Rui Wang, Hai Zhao, Bao-Liang Lu, Masao Utiyama, and Eiichiro Sumita. 2015. Bilingual continuous-space language model growing for statistical machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 23(7):1209–1220.

Rui Wang, Isao Goto, Masao Utiyama, Eiichiro Sumita, Hai Zhao, and Bao-Liang Lu. 2016a. Converting continuous-space language models into n-gram language models with efficient bilingual pruning for statistical machine translation. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 15(3).

Rui Wang, Hai Zhao, Sabine Ploux, Bao-Liang Lu, and Masao Utiyama. 2016b. A bilingual graph-based semantic model for statistical machine translation. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pages 2950–2956, New York, July.

Rui Wang, Andrew Finch, Masao Utiyama, and Eiichiro Sumita. 2017a. Sentence embedding for neural machine translation domain adaptation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 560–566, Vancouver, Canada.

Rui Wang, Masao Utiyama, Lemao Liu, Kehai Chen, and Eiichiro Sumita. 2017b. Instance weighting for neural machine translation domain adaptation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1483–1489, Copenhagen, Denmark.

Rui Wang, Masao Utiyama, Andrew Finch, lemao Liu, Kehai Chen, and Eiichiro Sumita. 2018a. Sentence selection and weighting for neural machine translation domain adaptation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(10):1727–1741.

Rui Wang, Masao Utiyama, and Eiichiro Sumita. 2018b. Dynamic sentence sampling for efficient training of neural machine translation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 298–304, Melbourne, Australia, July. Association for Computational Linguistics.

Rui Wang, Hai Zhao, Sabine Ploux, Bao-Liang Lu, Masao Utiyama, and Eiichiro Sumita. 2018c. Graph-based bilingual word embedding for statistical machine translation. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 17(4), July.

Warren Weaver. 1947. Letter to norbert wiener, march 4, 1947. *Rockefeller Foundation Archives*.

Mingming Yang, Rui Wang, Kehai Chen, Masao Utiyama, Eiichiro Sumita, Min Zhang, and Tiejun Zhao. 2019. Sentence-level agreement for neural machine translation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3076–3082, Florence, Italy, July. Association for Computational Linguistics.

Zhisong Zhang, Rui Wang, Masao Utiyama, Eiichiro Sumita, and Hai Zhao. 2018. Exploring recombination for efficient decoding of neural machine translation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4785–4790, Brussels, Belgium, October-November. Association for Computational Linguistics.

Zhuosheng Zhang, Kehai Chen, Rui Wang, Masao Utiyama, Eiichiro Sumita, Zuchao Li, and Hai Zhao. 2020. Neural machine translation with universal visual representation. In *International Conference on Learning Representations*.