Transfer Learning for Chinese-Lao Neural Machine Translation with Linguistic Similarity

Zhiqiang Yu[1,2], Zhengtao Yu[1,2], Yuxin Huang[1,2], Junjun Guo[1,2], Zhenhan Wang[1,2] and Zhibo Man[1,2]

[1] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China

[2] Yunnan Key Laboratory of Artifcial Intelligence, Kunming University of Science and Technology, Kunming, China

ztyu@hotmail.com

Abstract. As a typical low-resource language pair, besides severely limited by the scale of parallel corpus, Chinese-Lao language pair also has considerable linguistic differences, resulting in poor performance of Chinese-Lao neural machine translation (NMT) task. However, compared with the Chinese-Lao language pair, there are considerable cross-lingual similarities between Thai-Lao languages. According to these features, we propose a novel NMT approach. We first train Chinese-Thai and Thai-Lao NMT models wherein Thai is treated as pivot language. Then the transfer learning strategy is used to extract the encoder and decoder respectively from the two trained models. Finally, the encoder and decoder are combined into a new model and then fine-tuned based on a small-scale Chinese-Lao parallel corpus. We argue that the pivot language Thai can deliver more information to Lao decoder via linguistic similarity and help improve the translation quality of the proposed transfer-based approach. Experimental results demonstrate that our approach achieves 9.12 BLEU on Chinese-Lao translation task using a small parallel corpus, compared to the 7.37 BLEU of state-of-the-art Transformer baseline system using back-translation.

Keywords: Transfer Learning, Chinese-Lao, Neural machine translation, Linguistic similarity.

1 Introduction

Chinese-Lao NMT is a typical low-resource NMT, the research on which in the past decade is not widespread. Limited by the scale and domain of parallel corpus, the bulk of research on Chinese-Lao NMT has to focus on language model training and dictionary building [1,2], etc. However, with the introduction of the "the Belt and Road", the demand for translations of Chinese-Lao has been increasing. Therefore, it is important to investigate how to design an effective NMT model on a small scale of parallel corpus to improve translation performance on Chinese-Lao language pair.

To tackle the inefficiency problem in low-resource settings such as Chinese-Lao, some approaches have been proposed. Recent efforts [3,4,5,6] in NMT research have shown promising results when transfer learning techniques are applied to leverage existing rich-resource models to cope with the scarcity of training data in low-resource settings. However, these works mainly leverage the way that transfers the parameters of the rich-resource model to the low-resource model, barely adopt the strategy to extract the encoder or decoder from two pivot-relevant models separately. Even when the pivot strategy [6] is adopted, the similarity between the pivot and the target language is ignored.

Chinese and Lao have a mass of linguistic differences, the former belongs to Sino-Tibetan language family and the latter is from the Tai-Kadai language family. The tremendous cross-lingual different make Chinese and Lao are mutually unintelligible. Therefore, we choose a pivot language to overcome such cross-lingual different. Intuitively, a good pivot language for Chinese-Lao translation should have the following properties: (1) Adapt to unbalanced data set: the scale of Chinese-pivot parallel corpus could be larger than Chinese-Lao; and (2) considerable similarities with Lao: has high cross-lingual similarities, the best is in the same language family with Lao. Based on above considerations, we choose Thai as the pivot language for our transfer learning approach, and we elaborate the Language features in next section. Our main contributions are as follows:

- we investigate the cross-lingual similarities between Thai and Lao, and discuss the feasibility that chooses Thai as the pivot language for Chinese-Lao translation model construction.
- we propose a transfer Learning approach for Chinese-Lao NMT with pivot language. The central idea is to construct a new model by extracting encoder from the trained Chinese-Thai NMT model, and decoder from Thai-Lao NMT model which is trained on small scale parallel corpus of high similarity.

2 Linguistic Similarity between Thai and Lao

Thai and Lao are tonal languages and belong to Tai-Kadai language family, the speech and writing of the two languages are highly similar. Actually, spoken Thai and Lao are mutually intelligible. Moreover, the two languages share a large amount of correlative words on etymologically and have similar head-initial syntactic structures [8]. For writing, Thai and Lao are both written in abugida tokens, and in many cases the sentences composed of which are linguistically similar [9]. As the example illustrated in Figure 1, the similarity in the shape of certain tokens can be observed.

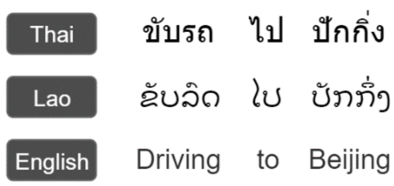
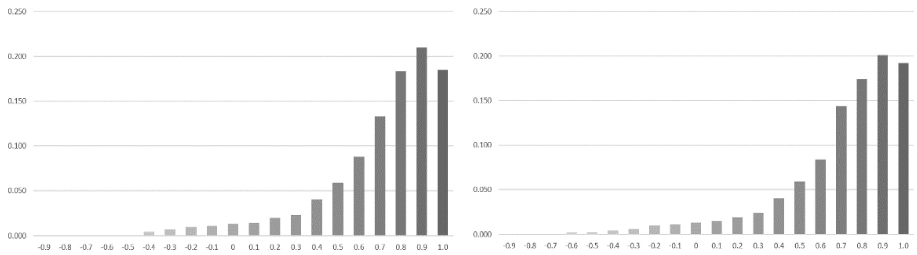


Fig. 1. Thai-Lao linguistic similarity

Besides the similarity investigation of token shape, we also discuss the similarity of syntactic structure. We leverage GIZA++ tool [10] to get word alignment over the 20K Thai-Lao portion of publicly ALT dataset. Then we use the approach proposed in Isozaki [11] to get the Kendall's τ according to the previous word alignment. Kendall's τ mainly indicate the cost of adjusting two parallel sentences to the same word order. As shown in Figure 2, Thai-Lao language pair shows a relative similar order with an average τ around 0.73. The result demonstrates the considerable similarity in syntactic structure of the two language.



(a) Thai-Lao (b) Lao-Thai

Fig. 2. Distribution of Kendall's τ on Thai-to-Lao (a) and Lao-to-Thai (b).

According to the above analysis, Thai-Lao language pair has considerable cross-lingual similarity in either token shape or syntactic structure. We argue that the similarity between two languages will bring more adequate information from Thai to Lao and improve the accuracy of Lao decoder. Therefore, choose Thai as the pivot language for Chinese-Lao translation task is positive. To the best of our knowledge, there is no existing work on transfer learning for Chinese-Lao NMT by choosing a target-similar pivot language Thai.

3 Our Approach

In this section, we will elaborate the detail of our proposed model. Our goal is to achieve a transfer-based NMT model which composed of trained Chinese encoder and Lao decoder. As illustrated in Figure 3, we first train Chinese-Thai and Thai-Lao translation model respectively. Then we compose new translation model using extracted Chinese encoder and Lao decoder. Lastly, we fine-tune the new Chinese-Lao model on small parallel corpus.

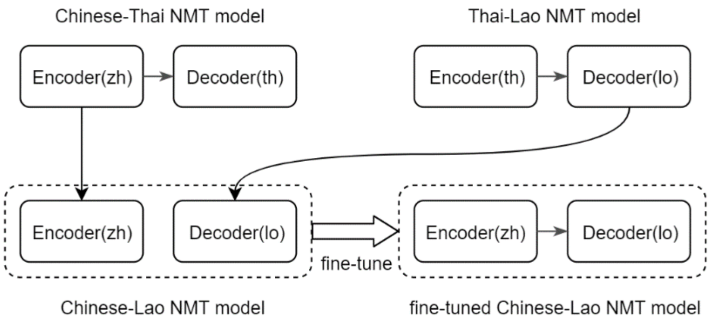


Fig. 3. Structure of transfer-based Chinese-Lao NMT model.

3.1 Chinese-Thai NMT model

As shown in the upper left of Figure 3. Given Chinese source sentence x and the Thai target sentence z. We denote the standard attention-based Chinese-Thai NMT model as $P(z|x;\theta_{x\to z})$, which can be trained on the Chinese-Thai parallel corpus $D_{x,z} = \{\langle x^{(m)}, z^{(m)}\rangle\}_{m=1}^{M}$ using maximum likelihood estimation:

$$\hat{\theta}_{x\to z} = \arg\max_{\theta_{x\to z}} \{\mathcal{L}(\theta_{x\to z})\} \tag{1}$$

where the log-likelihood is defined as:

$$\mathcal{L}(\theta_{x\to z}) = \sum_{m=1}^{M} \log P\left(z^{(m)} \mid x^{(m)}; \theta_{x\to z}\right) \tag{2}$$

3.2 Thai-Lao NMT Model

The standard attention-based Thai-Lao NMT model $P(y|z;\theta_{z\to y})$ with respect to the Thai-Lao parallel corpus $D_{z,y} = \{\langle z^{(n)}, y^{(n)}\rangle\}_{n=1}^{N}$ can be calculated similar with the Chinese-Thai NMT model, the model training procedure using maximum likelihood estimation is:

$$\hat{\theta}_{z\to y} = \arg\max_{\theta_{z\to y}} \{\mathcal{L}(\theta_{z\to y})\} \tag{3}$$

where the log-likelihood is defined as:

$$\mathcal{L}(\theta_{z\to y}) = \sum_{n=1}^{N} \log P\left(y^{(n)} \mid z^{(n)}; \theta_{z\to y}\right) \tag{4}$$

3.3 Chinese-Lao NMT Model

We compose a new translation model using the Chinese encoder $Enc_{x\to z}$ and the Lao decoder $Dec_{z\to y}$ that fetched from Chinese-Thai NMT model $P(z|x;\theta_{x\to z})$ and Thai-Lao NMT model $P(y|z;\theta_{z\to y})$ respectively. The process can be simply formulated as:

$$enc_{x\to z} = fetchEnc\left(P(z \mid x; \theta_{x\to z})\right) \tag{5}$$

$$dec_{z\to y} = fetchDec\left(P\left(y \mid z;\theta_{z\to y}\right)\right) \tag{6}$$

$$P\left(y \mid x;\theta_{x\to y}\right) = \left\{enc_{x\to z}, dec_{z\to y}\right\} \tag{7}$$

where $fetchEnc$ and $fetchDec$ are the functions that fetch encoder and decoder portion parameters from Chinese-Thai and Thai-Lao NMT model respectively. $P(y|x;\theta_{x\to y})$ is the composed Chinese-Lao NMT model.

Even in low-resource NMT settings, there often exist small-scale parallel corpus. In our approach, we first combine the extracted encoder and decoder into a new NMT model, but the model is not fine-tuned and suboptimal. Therefore, we use the small-scale parallel corpus of Chinese-Lao from ALT dataset to fine-tune the model parameters. In the fine-tuning process, we first try to fix some parameters. However, we observe that for the ALT dataset we used for fine-tuning, not fix parameters had a better effect.

4 Evaluation

4.1 Experimental Setup

Data. We conduct experiments on the publicly ALT dataset[1] and the in-house Chinese-Thai parallel corpus. For ALT dataset, we use the trilingual Chinese-Thai-Lao portion which comprise 20K sentences triples. Then we bin the ALT subset into three subsets: 19K for training, other two subsets of 500 sentences as the development and test datasets, respectively. For Chinese-Thai model training, we use the combined parallel data from the Chinese-Thai portion (19K) of ALT subset and the 50K in-house parallel corpus collected by ourselves. For Thai-Lao model training, we use the 19K parallel data from the Thai-Lao portion of ALT subset. For Chinese-Lao model fine-tuning training, we use the 19K parallel data from the Chinese-Lao portion of the ALT subset. We process the experiment corpus simply before applying our approach. For the Thai word segmentation, we use pythaipiece tool[2] which based on sentencepiece to segment Thai sentences, while for Lao word segmentation, we use LaoWordSegmentation tool[3] to segment Lao sentences. For Chinese we apply word segmentation by jieba tools[4]. We do not use BPE approach on the experimental parallel corpus.

Evaluation. We adopt the case insensitive 4-gram BLEU as the main evaluation metrics [12], and choose the multi-bleu.perl as scoring script. Significance tests are conducted based on the best BLEU results by using bootstrap resampling [13].

Baseline. We compare the proposed model against the state-of-the-art NMT system Transformer, which has obtained the state-of-the-art performance on machine translation and predicts target sentence from left to right relying on self-attention [14].

---

[1] http://www2.nict.go.jp/astrec-att/member/mutiyama/ALT/

[2] https://github.com/wannaphong/thai-word-segmentation-sentencepiece

[3] https://github.com/djkhz/LaoWordSegmentation

[4] https://github.com/fxsjy/jieba

Implement Detail. We adopt the prudent Transformer settings, uses a 2-layer encoder and 2-layer decoder, while each layer employs 4 parallel attention heads. The dimensions of word embeddings, hidden states and the filter sizes are set to 256, 256 and 512 respectively. The dropout is 0.2 for Chinese-Thai and 0.1 for Thai-Lao training. We train using the Adam optimizer [15] with a batch size of 256 words and evaluate the model every 1000 steps. The models are trained on 2 P100 GPUs. We implement our approach on Thumt [16], an efficient open source machine translation platform.

4.2 Experimental Results

Quantitative study. Table 1 shows the experimental results evaluated by BLUE score. We get 3.62 BLEU point improvement compared with transformer baseline which only use 19K tiny Chinese-Lao parallel corpus for training. Moreover, for a fair comparison, we back-translation [17] the Thai side sentences of 50K Chinese-Thai corpus collected by ourselves to corresponding Lao sentences on the Thai-Lao transformer model which is trained on original 19K ALT corpus. The Chinese sentences and corresponding translated Lao sentences are combined as new parallel corpus, which is fed into a new Chinese-Lao model together with Chinese-Lao ALT dataset for training again. Note that we do not back-translation the Lao side sentences of Thai-Lao ALT corpus to Chinese because ALT corpus is a Multilingual parallel corpus. As shown in Table 1, our approach still gains 1.75 BLEU point improvement compared with Transformer using back-translation.

Table 1. BLEU scores evaluated on test set (0.5K) compared with baseline. Parallel sentences for Transformer and transformer+back-translation training are 19K and 69K, respectively.

Models	BLEU
Transformer	5.50
Transformer+back-translation	7.37
Our approach	9.12

Table 2. Performance difference on our proposed approach when choosing different pivot language.

Pivot	BLEU
En	7.55
Th	9.12

The source side Chinese sentences for Chinese-Thai model training and Chinese-Lao model fine-tuning are identical. To dispel the concern that the improvement is brought by the same source training data, for a fair comparison, we also conduct the experiment that select different language as pivot. To ensure fair comparison, we choose English as the pivot, and for Chinese-English model training, we conduct the experiment on the trilingual Chinese-English-Lao portion of ALT dataset and the extracted 50K Chinese-English parallel sentences from IWSLT15 zh-en dataset. Table 2 reports the performance of choosing English and Thai as pivot language respectively. We observe that there is large gap when choosing English as pivot compared with Thai. The main possible reason is that English has few cross-lingual similarities with Lao compared with Thai.

Case study. Apart from the quantitative analysis, we illustrate an example of our propose approach. As we do not apply BPE to corpus, to avoid the UNK, we provide a relatively common sentence. As shown in Table 3, the Chinese word "*那人*" (*The man*), "*处理*" (*deal with*) and "*秘密的*" (*secret*) are translated correctly in our approach. We argue that one of the main reasons is that the pivot language Thai delivers more information in translation process. As shown in Table 4, for the three preceding Chinese words, the corresponding words in Lao are similar in morphology with the words in the pivot language Thai and all of them can be found in the training corpus.

Table 3. Example of Chinese-Lao translation.

Input:	那人 能 处理 秘密的 私人 问题
Golden:	ຜູ້ຊາຍ ສາມາດ ຈັດການກັບ ຄວາມລັບ ບັນຫາສ່ວນຕົວ (The man can deal with secret personal problems)
Baseline:	ຜູ້ຊາຍ ສາມາດ ໄດ້ຮັບ ບາງ ບັນຫາ (The man often gets some trouble)
Our approach:	ຜູ້ຊາຍ ສາມາດ ຈັດການກັບ ຄວາມລັບ ສິ່ງຂອງ (The man often deals with secret things)

Table 4. The morphological similarity between Thai and Lao words that translated correctly in Table 3.

Thai	Lao	Corresponding English
ผู้ชาย	ຜູ້ຊາຍ	the man
จัดการกับ	ຈັດການກັບ	deal with
ความลับ	ຄວາມລັບ	secret

5 Related Work

Many types of transfer learning approaches [3,4,5,6,7] have been proposed in the past few years. Since the advent of Transformer, To improve the quality of the translation, many authors have endeavored to adopt transfer-based method on Transformer framework. Lakew et al. propose a Transformer-adapted transfer learning approach [5] that extend an initial model for a given language pair to cover new languages by adapting its vocabulary as long as new data become available. Kim et al. propose three methods to increase the relation among source, pivot, and target languages in the pre-training and implement the models on Transformer [7]. While for Chinese-Lao translation task, limited by the scale of parallel corpus and the language processing tools of Lao, the research on Chinese-Lao NMT in the past decade is not widespread. The bulk of researches have to focus on the Analysis of Lao Language characteristics [1, 2]. Different from the above work, we endeavor to leverage the cross-lingual similarity between Thai and Lao to improve Chinese-Lao NMT performance based on Transformer framework.

6 Conclusions

We propose a new NMT approach focusing on language pair Chinese-Lao with an extremely limited amount of parallel corpus. Our proposed approach utilizes a transfer learning approach to reuse the encoder and decoder from two trained Chinese-Thai and Thai-Lao NMT models respectively. As the pivot language, Thai has considerable similarities with Lao, and we argue that it will bring significant improvement to entire framework. We conduct contrast experiments, as the results reported, our approach can achieve 9.12 BLEU on Chinese-Lao translation task using small parallel corpus, compared to the 7.37 BLEU of strong transformer baseline system using back-translation.

An interesting direction is to apply our approach to other low-resource NMT task, with the feature that the scale of source-pivot parallel corpus is obvious larger than pivot-target parallel corpus, and the pivot language is similar with target language, such as Chinese-Indonesian (Malay as pivot language) etc.

Acknowledgements

We would like to thank the anonymous reviewers for their constructive comments. The work is supported by National Natural Science Foundation of China (Grant Nos. 61732005, 61672271, 61761026, 61762056 and 61866020), National key research and development plan project (Grant No. 2019QY1800), Yunnan high-tech industry development project (Grant No. 201606), and Natural Science Foundation of Yunnan Province (Grant No. 2018FB104).

References

[1] A. Srithirath andP. Seresangtakul, “A Hybrid Approach to Lao Word Segmentation using Longest Syllable Level Matching with Named Entities Recognition,” in The 10th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Krabi, Thailand, pp. 1-5, May 15-17, 2013.

[2] M. Yang, L. Zhou, Z. Yu, S.Gao, and J. Guo, “Lao Named Entity Recognition based on conditional random fields with simple heuristic information,” in The 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Zhangjiajie, pp.1426-1431, 2015.

[3] Zoph B, Yuret D, May J, et al. Transfer learning for low-resource neural machine translation[J]. arXiv preprint,2016,arXiv:1604.02201

[4] T. Q. Nguyen and D. Chiang, Transfer learning across low-resource, related languages for neural machine translation, arXiv preprint arXiv:1708.09803, 2017.

[5] Lakew S M, Erofeeva A, Negri M, et al. Transfer Learning in Multilingual Neural Machine Translation with Dynamic Vocabulary[J]. arXiv preprint ,2018,arXiv:1811.01137

[6] Cheng Y, Yang Q, Liu Y, et al. Joint Training for Pivot-based Neural Machine Translation[C]//26th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 2017: 3974-3980.

[7] Kim, Yunsu et al. “Pivot-Based Transfer Learning for Neural Machine Translation Between Non-English Languages.” Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (2019): n. pag. Crossref. Web.

[8] Chenchen . Ding, Masao . Utiyama. , Eiichiro . Sumita. Similar Southeast Asian Languages: Corpus-Based Case Study on Thai-Laotian and Malay-Indonesian Proceedings of the 3rd Workshop on Asian Translation, pages 149–156, Osaka, Japan, December 11-17, 2016.

[9] Singvongsa, K. , & Seresangtakul, P. 2016. Lao-Thai machine translation using statistical model. International Joint Conference on Computer Science & Software Engineering. IEEE.

[10] Franz Josef Och, Hermann Ney. A Systematic Comparison of Various Statistical Alignment Models. Computational Linguistics, volume 29, number 1, pp. 19-51 March 2003.

[11] Hideki Isozaki, Katsuhito Sudoh, Hajime Tsukada, and Kevin Duh. 2012. Hpsg-based preprocessing for English-to-Japanese translation. ACM Transactions on Asian Language Information Processing, 11(3):1-16.

[12] Kishore Papineni, Salim Roukos, Todd Ward, and WeiJing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In Proceedings of ACL, pages 311–318.

[13] Philipp Koehn. 2004. Statistical significance tests for machine translation evaluation. In Proceedings of EMNLP 2004, pages 388–395.

[14] Ashish Vaswani, Noam Shazeer, Niki Parmar,Jakob Uszkoreit, Llion Jones, Aidan N Gomez,Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In I. Guyon, U. V.Luxburg, S. Bengio, H. Wallach, R. Fergus,S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems 30, pages 5998–6008.

[15] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In Proceedings of the 3th International Conference on Learning Representations.

[16] Jiacheng Zhang, Yanzhuo Ding, Shiqi Shen, Yong Cheng, Maosong Sun, Huanbo Luan, Yang Liu. 2017. THUMT: An Open Source Toolkit for Neural Machine Translation. arXiv:1706.06415.

[17] Sennrich, Rico & Haddow, Barry & Birch, Alexandra. (2015). Improving Neural Machine Translation Models with Monolingual Data.