TransLingua: A Transfer Learning Approach to Enhancing Myanmar-Wa Neural Machine Translation

Florance Yune
Natural Language Processing Lab
(Student)
University of Computer Studies
Yangon, Myanmar
floranceyune@ucsy.edu.mm

Khin Mar Soe
Natural Language Processing Lab
(Professor, HOD)
University of Computer Studies
Yangon, Myanmar
khinmarsoe@ucsy.edu.mm

Abstract— In the field of Neural Machine Translation (NMT), transfer learning emerges as a pivotal technique, significantly enhancing NMT model performance, especially in low-resource language settings. Transfer learning methodologies in NMT typically involve the transfer of knowledge from a well-established parent model to a target model during the parameter initialization phase. While conventional neural network training relies on vast datasets, often comprising millions of data points, this approach faces significant challenges when applied to under-resourced languages like Myanmar (Standard Burmese Language) and Wa (a niche ethnolinguistic language spoken by the 'Wa' ethnic group in Myanmar), where access to parallel corpora has been exceptionally limited until recent times. This study builds upon established NMT architectures, encompassing both traditional transformer-based and LSTM-based models, as its foundational underpinning. Specifically, it explores the domain of Transfer Learning-based Machine Translation from Myanmar (Standard Burmese Language) to Wa (Ethnolinguistic Language). The proposed Transfer Learning approaches demonstrate substantial improvements compared to conventional direct translation methods.

Keywords— Neural Machine Translation, Transfer Learning, Transformer, LSTM.

I. Introduction

The In the realm of Neural Machine Translation (NMT), the efficacy of various techniques has been a subject of continual exploration. Traditional methods, such as Long Short-Term Memory (LSTM) and Transformer-based approaches, have paved the way for substantial progress. However, they often encounter significant challenges when resources are limited, both in terms of data availability and computational costs. Transfer learning, in this context, emerges as a compelling and innovative solution. It offers a simpler and more potent approach to enhancing NMT models. Unlike traditional neural network learning, which heavily relies on extensive, resource-intensive training data, transfer learning leverages pre-existing knowledge from large-scale datasets. This knowledge can be transferred to specific translation tasks, even when limited data is available. In this article, we delve into a comparative exploration of Transfer Learning against traditional NMT techniques, namely LSTM and Transformer. We aim to shed light on why Transfer Learning excels in scenarios characterized by resource constraints. Through empirical analysis and experimentation, we will demonstrate how Transfer Learning outperforms these conventional techniques in terms of translation accuracy and efficiency, providing a novel perspective on the evolving landscape of NMT research and application.

Transfer learning-based Myanmar-Wa language machine translation refers to the process of creating pre-trained models and knowledge from another related language pairs (such as Myanmar to English) to get the better performance of machine translation particularly for the Myanmar - Wa language. This approach is especially useful when there is limited parallel data available for direct training translation model. The Myanmar language, also known as Burmese, is the conclusive Myanmar language and is spoken by the generality of the population in Myanmar. Wa is an Austroasiatic language spoken by the Wa people, an ethnic group living in Myanmar and China. Since Wa is a less-resourced and lesser-studied language compared to Myanmar. So, the developing a productive machine translation system directly translation from Myanmar language to Wa language may be complicated due to the deficiency of parallel data. Transfer learning comes into play in such scenarios. Instead of building a Myanmar-Wa translation model from scratch, transfer learning allows to benefit from the representation and knowledge learned using a pre-trained model on a pair of related language (such as Translation Myanmar to English).

In conclusion, transfer learning-based Myanmar-Wa language machine translation offers a promising approach to overcome the scarcity of parallel data and improve quality of translation for low-resource language pairs. By leveraging knowledge from related language pairs, proposed system can develop more robust and accurate machine translation systems for the Myanmar and Wa languages. In this proposed system, mt5 pre-trained models of huggingface is used to create the Transfer Learning based Neural Machine Translation System for Myanmar-Wa language to overcome the low-resource and to get the better performance of the Machine Translation System. The experimental result of the system is analyzed based on the two based line Methods of Traditional Neural Machine Translation: LSTM and TRANSFORMER based Machine Translation Systems.

II. Related Works

Before NMT has effectively implemented transfer learning in low-resource situations. In [7], writers transfer 5 BLEU points from French-English to Uzbek-English. Their method of transfer learning ignores the vocabulary gap between parent and kid and duplicates the full model, including word embedding. They employed separate embedding for words in the target and source languages, but tied embedding is now the accepted practice in NMT for low-resource environments. We have the chance to review some of their finding’s thanks to tied embedding. While they found no improvements from a copy model with an untied embedding.

Since Zoph et al., strategies for overcoming vocabulary mismatch have evolved. Kocmi and Bojar make the case in [8] that a common vocabulary between the kid's language and the parent language is advantageous, however doing so necessitates that the parent be taught in the child language. Gheini and May developed a general vocabulary for transfer learning to address this problem. By simultaneously the sub-word tokens training across several languages and applying Romanization to languages written in scripts other than Latin, they were able to acquire a global vocabulary. However, this global vocabulary could only be able to represent unknown languages with an extremely aggressive and perhaps ineffective sub-word segmentation [9].

Multilingual models are another strategy for low-resource (or even zero-hot) NMT, which is comparable to teaching the parent and kid at the same time [10].

III. LANGUAGES

The fundamental operations of machine translation encompass the following two primary steps. Firstly, decoding the underlying meaning of the original text in the source language. Secondly, encoding this meaning into the target language. In the framework proposed herein, the source language is Myanmar, and the target language is Wa.

A. Myanmar Language

The Myanmar text is a character string without exact boundary markup word, written from left to right sequence without constant spacing for inter-word, whereas spacing of inter-phrase may be used periodically. There are three groups of Myanmar characters: medial, vowels and consonants. The basic Myanmar consonants can be aggregated by medial. Words or Syllables of Myanmar text are organized by combination of consonants and vowels. At the same time, some syllables can be formed by only consonants without using any vowels. The special characters, signs, punctuation and numerals were included as some characters of Myanmar text. The Union of Myanmar's official language is the Myanmar language, usually referred to as Burmese, and it dates back more than a thousand years. The Burmese script is used to write the tonal and analytical language of Burma. An Indian (Brahmi) prototype served as the basis for this phonologically based script, which was developed from Mon. A Myanmar text is a collection of characters written left to right without clear word's boundary marking and without the usual inter-word space, however inter-phrase spacing may occasionally be utilized. Example sentences of Myanmar-English-Wa are shown in table 1.

B. Wa Language

Tokenization Wa, a language from the Mon-Khmer family, is spoken by around 950,000 people, predominantly in northern Burma and adjacent Thailand and China. Wa comes in three different forms: Parauk, Vax, and Avax, each of which has a variety of dialects and is occasionally thought of as a separate language. The Parauk language, also known as Phalok, Baroag, Praok, Wa, or Standard Wa, is spoken by about 400,000 people in Burma. Most of these individuals' hail from the Southeast, East, and Northeast Shan States. The Parauk language is spoken in the southwest of China.

TABLE I. EXAMPLE OF MYANMAR-ENGLISH-WA SENTENCES

Language	**Sentence**
Myanmar sentence	မင်း ဘယ် နေရာ မှာလဲ ။
English sentence	Where were you?
Wa sentence	Maix dee mawx?

IV. METHODOLOGIES OF THE PROPOSED SYSTEM

The proposed system utilizes transfer learning in machine translation. It uses a pretrained parent model (Myanmar-English) to initialize a child model (Myanmar-Wa), reducing data requirements and accelerating learning. Transfer learning in machine translation involves three stages: pre-training, fine-tuning, and inference, improving translation quality, especially in multilingual contexts. Domain adaptation is essential for specific domains.

The system employs the Huggingface NLP library's pretrained model, Mt5 (Multilingual Text-to-Text Transfer Transformer), which unifies NLP tasks into a single network using the encoder-decoder method. Each task shares the same loss function and hyperparameters in experimental results.

Mt5 inherits all capabilities of the T5 model and is trained on a modified C4 dataset with over 10,000 web page contents in 101 languages, achieving state-of-the-art performance. It outperforms other multilingual models, including multilingual BERT, XLM-R, and multilingual BERT (without Persian support), particularly in summarization tasks [5] - [6].

Figure 1 illustrates the flow of the transfer learning approach. Learning new tasks through transfer learning can be faster, more precise, and require less training data by building on previously learned tasks.

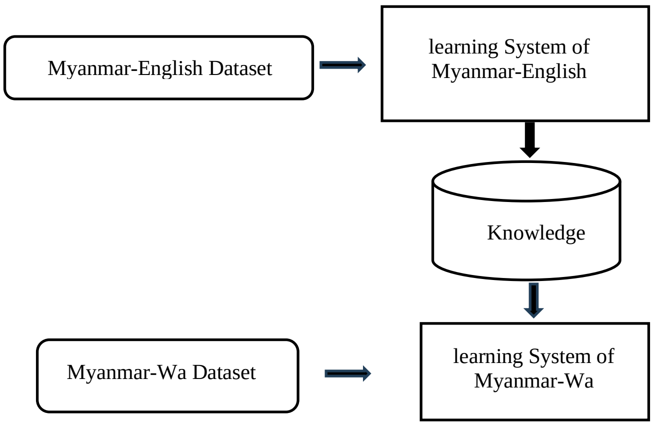

Figure 1. Flow of the Proposed Transfer Learning

V. EXPERIMENTAL RESULT

In our pursuit of enhancing Neural Machine Translation (NMT) systems, we have conducted a comprehensive comparative analysis. Our proposed NMT system, enriched with Transfer Learning capabilities, was meticulously evaluated against two baseline-traditional-NMT-models: the

Transformer-based NMT and LSTM-based NMT. This evaluation was carried out employing BLEU scores, a widely accepted metric for assessing translation quality.

Our study encompassed a substantial dataset consisting of 23,126 sentences, meticulously partitioned into training (20,813 sentences) and testing (2,313 sentences) subsets, as summarized in the table 3. For detailed insights into the experimental outcomes, please refer to table 2 and figure 2, which encapsulate the results more comprehensively. It is noteworthy to mention that this dataset holds a significant distinction as it represents the first-ever corpus for Machine Translation between the Myanmar and Wa languages, marking a pioneering step in the field of MT.

TABLE II. TABLEANALYSIS OF BLEU SCORE ON PROPOSED SYSTEM

MACHINE TRANSLATION	**BLEU SCORE**
TRADITIONAL TRANSFORMER NMT (BASELINE)	12.63
LSTM BASED NMT (BASELINE)	10.38
TRANSFER LEARNING BASED NMT	**23.36**

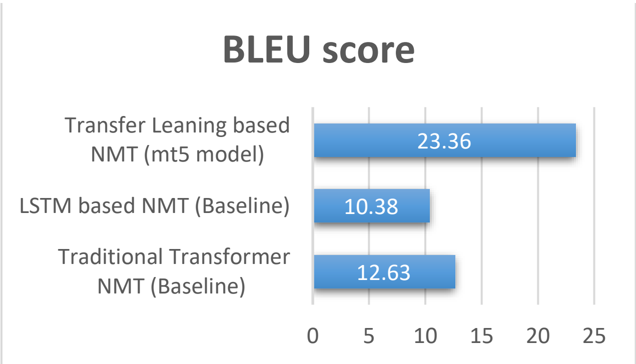

Figure 2. Experimental Results

Conforming with the experimental result, Transfer Learning based NMT using Mt5 pre-trained model achieved the best BLEU score in the other NMT applied in this proposed system.

TABLE III. DATASET DISTRIBUTION FOR MYANMAR-WA NMT EVALUATION

Dataset Subset	**Total Sentences**
Training	20,813
Testing	2,313

VI. CONCLUSION

In conclusion, The Neural Machine Translation technology is data-hungry. With traditional learning, an algorithm can only pick up new information when it has access to a sufficient amount of training data (perhaps millions of data points). It's possible that this data won't be accessible at all, or that creating and preparing the model would be too expensive. Nowadays, transfer learning based NMT are applied to overcome this problem as low-resource NMT. Transfer learning and linguistically informed data mixing are two methodologies that can aid in interlanguage communication. When utilizing Fine-tuning, using a pre-trained model for a similar job typically produces excellent results. In this proposed system, transfer learning NMT is applied the pre-trained model of huggingface called Mt5. As the conclusion of the presentation of this research, the performance BLEU score of the Transfer Learning based NMT is achieved the best result according to the experimental analysis of system.

ACKNOWLEDGMENT

This work acknowledges divine inspiration and owes its inception to the Higher Power. Gratitude to Prof. Dr. Mie Mie Khin, Rector of University of Computer Studies, Yangon, for the invaluable opportunity. Special thanks to Dr. Win Lei Lei Phyu, UCY, for supporting my career goals. Thanks to my group mates, UCSY NLP Lab members, and especially Dr. Khin Mar Soe for their guidance. Special thanks to Dr. Win Pa Pa. Grateful to Dr. Aye Mya Hlaing, Dr. Yi Mon Shwe Sin, and Mr. Ai Mhawm Kham for their contributions. My family, thank you for your love and guidance. Lastly, thanks to Pastor Ai Nap Tao and Associate Pastor Ai Khwat Pan for their unwavering support.

REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. "Attention is all you need", 2017.

[2] C. Raffel and et al., "Exploring the limits of transfer learning with a unified text-to-text transformer.", Machine Learning Research., 2020.

[3] J. Devlin and et al.: "Pre-training of deep bidirectional transformers for language understanding". Design, Analysis, and Inference. 5th Edition, John Wiley & Sons Ltd, 2019.

[4] Y. Liu and et al. "Multilingual denoising pre-training for neural machine translation", Ph.D. Thesis, Bologna University, Bologna, 2020.

[5] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and Peter J. Liu. "Exploring the limits of transfer learning with a unified text-to-text transformer". J. Machine Learning Research, 2020.

[6] L. Xue and et al. "mt5: A massively multilingual pre-trained text-to-text transformer". ArXiv, abs/2010.11934, 2020.

[7] B. Zoph and et al. "Transfer learning for low-resource neural machine translation". Proc: 2016th Conference on Empirical Methods in NLP, pp. 1568–1575, 2016.

[8] T. Kocmi and O. Bojar. "Trivial transfer learning for low-resource neural machine translation". Proc: 3rd Conference on Machine Translation, pp 244–252, Belgium, Brussels, 2018.

[9] M. Gheini and J. May. "A universal parent model for low-resource neural machine translation **transfer"**. preprint arXiv:1909.06516, 2019.

[10] M. Johnson and et al. "Google's multilingual neural machine translation system: Enabling zero-shot translation". CoRR, abs/1611.04558, 2016.